The University Of Sheffield.

Investigating the Nuclear Function of the C9orf72 Protein in Amyotrophic Lateral Sclerosis

Yolanda Barr Gibson BSc

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

The University of Sheffield
Faculty of Medicine, Dentistry and Health
Department of Neuroscience

April 2018

Abstract

Amyotrophic lateral sclerosis (ALS) is a terminal neurodegenerative disorder where death of the upper and lower motor neurons causes progressive muscle wasting and paralysis. Currently, the approved treatments for ALS provide only a small therapeutic benefit to patients. A GGGGCC repeat expansion mutation that lies within the C9orf72 gene is causal of approximately 40% of the inherited form of ALS. The mutation has been demonstrated to cause haploinsufficiency of the C9orf72 protein and this loss of function is hypothesised to contribute to pathogenesis in C9orf72-ALS. Until recently, little was known about the protein function of C9orf72. We performed a yeast two-hybrid screen to investigate the C9orf72 interactome which identified binding between C9orf72 and coilin. Coilin is the major protein component of nuclear suborganelles, Cajal bodies, which are responsible for processing snRNPs that form the spliceosome. C9orf72-ALS patients exhibit splicing defects which correlate with disease severity. We therefore hypothesised that loss of the C9orf72-coilin interaction leads to dysregulation of Cajal bodies and splicing in C9orf72-ALS.

The aims of this project were to characterise the C9orf72-coilin interaction and investigate the effect of C9orf72 on Cajal bodies and splicing. C9orf72 was found to interact with coilin directly *in vitro* and within cells, suggesting the interaction has biological importance. Depletion of C9orf72 in HEK293 cells led to an increase in the number of Cajal bodies, and a subtle defect in splicing. In contrast, C9orf72 overexpression decreased the number of Cajal bodies, but had little effect on splicing. Interestingly, investigation of Cajal bodies in C9orf72-ALS patient iAstrocytes found patients exhibited higher numbers of Cajal bodies in comparison to healthy controls.

The work included in this thesis uncovered an interaction between C9orf72 and coilin and that depletion of C9orf72 can lead to abnormal numbers of Cajal bodies. The results support evidence for a novel, uncharacterised role for C9orf72 in the nucleus which may be linked with snRNP processing, Cajal bodies and splicing. Investigation into the function of nuclear C9orf72 will lead to greater understanding of the consequences of C9orf72 haploinsufficiency in ALS and may lead to development of better drug treatments for patients.

Acknowledgements

This project would not have been possible without the help and support of a wide range of people. First, I would like to thank my supervisors, Andy and Kurt. Thank you both for your time, patience and guidance throughout the duration of the project - it has been invaluable. Most importantly however, thank you for reminding me to stay positive during the occasional PhD panic. I would also like to acknowledge The University of Sheffield for the Faculty Scholarship which provided funding for the project, and to the Medical Research Council and The Thierry Latran Foundation who also contributed funding towards aspects of the project.

Next I would like to say thank you to Dr Laura Ferraiuolo and Dr Guillaume Hautbergue for their offers of advice and lab equipment to allow me to perform the iAstrocyte and binding assay experiments. To this end, I would also like to thank my lab team member Miss Rebecca Cohen for the regular preparation of neurons for the project.

Thanks must also be said to Miss Emma Smith and Dr Chris Webster for adopting coilin and engaging with my PhD alongside your own projects. I cannot express how much help you were and how grateful I am for your time. To the rest of my SITraN friends of which there are too many to name – thanks for all your support, laboratory karaoke, Francis Newton visits and living accommodation!

I must also express my gratitude to my friends and family outside of the lab. I count myself extremely lucky to have an incredible, caring friendship group spread around the country and a loving, crazy family in Middlesbrough. To Niall, thank you for your constant encouragement and reassurance, even from the other side of the world.

Lastly, I would like to thank my mum and dad. There are no words to explain how grateful I am to you both for everything you have done for me. Your unconditional support and generosity, not just through my PhD, but through everything I have ever done, means so much.

Table of Contents

List of figures

List of tables

Abbreviations

ALS – Amyotrophic lateral sclerosis

ANOVA – Analysis of variance

APS – Ammonium persulfate

ARS2 – Arsenite resistance 2

ASO – Antisense oligonucleotide

ATG-13 – Autophagy related protein 13

ATP – Adenosine triphosphate

BAC – Bacterial artificial chromosome

Beta-gal – Beta-galactosidase

BI – Basophilic inclusions

BLAST – Basic local alignment search tool

bp – Base pair

BPS – Branch point sequence

BSA – Bovine serum albumin

C21orf2 – Chromosome 2 open reading frame 2

C9-ALS – C9orf72-related ALS

C9orf72 – Chromosome 9 open reading frame 72

C9orf72S – C9orf72 short protein isoform

C9orf72L – C9orf72 long protein isoform

CAB – Cajal body box

Cas9 – CRISPR associated protein 9

C. elegans - *Caenorhabditis elegans*

CB – Cajal body

cDNA – Complementary DNA

CFTR – Cystic fibrosis transmembrane conductance receptor

CMV - Cytomegalovirus

CNS – Central nervous system

Co-IP – Co-immunoprecipitation

CRISPR – Clustered regularly interspaced short palindromic repeats

CRM1 – Chromosome region maintenance 1

CSF – Cerebral spinal fluid

DENN – Differentially expressed in neoplastic versus normal cells

DIV – Days *in vitro*

DMEM – Dulbeccos' modified eagle's medium

DNA – Deoxyribonucleic acid

DPR – Dipeptide repeats

DSB – Double strand break

DTT – Dithiothreitol

E. coli - *Escherichia coli*

EDTA - Ethylenediaminetetraacetic acid

EGTA - ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid

ER – Endoplasmic reticulum

EV – Empty vector

FBS – Foetal bovine serum

FMRP – Fragile X mental retardation

FTD – Frontotemporal dementia

FTLD – Frontotemporal lobar degeneration

FUS – Fused in sarcoma

GEF – Guanine nucleotide exchange factor

GDP – Guanosine diphosphate

GFP – Green fluorescent protein

gRNA – Guide RNA

GSH - Glutathione

GST – Glutathione S transferase

GTP – Guanosine triphosphate

HEK293 – Human embryonic kidney cells 293

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

hnRNP – Heterogeneous ribonucleoprotein particle

INDEL – Insertion or deletion of base pairs

IP - Immunoprecipitation

iPSC – Induced pluripotent stem cells

IPTG - Isopropyl β-D-1-thiogalactopyranoside

IRES – Internal ribosome entry site

Kb - Kilobase

LB – Luria-Bertani

LSm – Like-Sm

m_3G - 2,2,7-trimethylguanosine

m^7G – 7-methylguanylate

MAPT – Microtubule associated protein tau

Mdm2 – Mouse double minute 2 homolog

miRNA – Micro RNA

mRNA – Messenger RNA

NCI – Neuronal cytoplasmic inclusion

Nek1 – NIMA-related kinase 1

NES – Nuclear export signal

NGS – Normal goat serum

NII – Neuronal intranuclear inclusion

NLS – Nuclear localisation signal

NMR – Nuclear magnetic resonance

NoLS – Nucleolar localisation signal

NPC – Nuclear pore complex

NTC – Non-targeting control

ORF – Open reading frame

P120 - Putative ribosomal RNA methyltransferase NOP2

PBS – Phosphate buffered saline

PCR – Polymerase chain reaction

PEI - Polyethylenimine

PHAX – Phosphorylated adaptor RNA export protein

PLA – Proximity ligation assay

Poly(A) - Polyadenylation

Poly-GA – Poly-GlycineAlanine

Poly-GP – Poly-GlycineProline

Poly-GR – PolyGlycineArginine

Poly-PA - PolyProlineAlanine

Poly-PR – PolyProlineArginine

PML – Promyelocytic leukemia

PRMT – Protein arginine methyltransferase

PSI-BLAST – Position-specific iterative basic local assignment search tool

qPCR – Quantitative PCR

Ran – Ras-related nuclear protein

RAN – Repeat associated non-ATG

RanGAP1 – Ran GTPase-activating protein 1

RBP – RNA-binding protein

RCC1 – Regulator of chromosome condensation 1

RIPA – Radioimmunoprecipitation assay

RISC – RNA-induced silencing complex

RNA – Ribonucleic acid

RNase – Ribonuclease

RRM – RNA recognition motif

rRNA – Ribosomal RNA

RT-PCR – Reverse transcription PCR

SART3 – Squamous cell carcinoma antigen recognised by T-cells 3

scaRNA – small Cajal body RNA

SDS – Sodium dodecyl sulfate

SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEM – Standard error of the mean

siRNA – Small interfering RNA

SOD1 – Superoxide dismutase 1

SF1 – Splicing factor 1

SF2 – Serine/arginine-rich splicing factor 1

SMA – Spinal muscular atrophy

SMN – Survival of motor neuron

snoRNA – Small nucleolar RNA

snRNA – Small nuclear RNA

snRNP – Small nuclear ribonucleoprotein

SRE – Splicing regulatory element

SRSF1 – Serine/arginine rich splicing factor 1

ss – Splice site

SSRP1 – Structure specific recognition protein 1

TBS – Tris buffered saline

TCAB1 – Telomerase Cajal body protein 1

TDP-43 – TAR DNA-binding protein 43 kDa

TERT – Telomerase reverse transcriptase

TGS1 – Trimethylguanosine synthase

ULK1 – Serine/threonine-protein kinase ULK1

UTR – Untranslated region

VRK1 – Serine/threonine-protein kinase VRK1

WDR41 – WD repeat-containing protein 41

WRAP53 – WD repeat containing antisense to TP53

Y2H – Yeast two-hybrid

1 Introduction

1.1 Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is a fatal, progressive neurodegenerative disorder categorised as a motor neuron disease (MND), and is caused by the death of upper and lower motor neurons. Upper motor neurons are located within the motor cortex and brain stem and project down to lower motor neurons found in the spinal cord. Signalling between the upper and lower motor neurons is essential for most motor functions. Loss of motor neurons in ALS leads to progressive muscle weakness, paralysis, and eventually death (Ravits and La Spada, 2009). The average life expectancy from diagnosis is 2-3 years and there is currently no effective treatment. Riluzole, the drug most commonly prescribed for treatment of ALS, has only a moderate effect and increases life expectancy by 2-3 months (Lacomblez et al., 1996). Approximately 7 in every 100,000 people are thought to suffer from the disease, with those aged between 50 and 70 at highest risk (Kurtzke, 1982). Men are also more commonly affected than women (Leblond et al., 2014).

1.2 Clinical features of ALS

Symptoms including muscle twitching, wasting and weakness, spastic dysarthria, fatigue, weight loss and dyspnoea are seen in ALS, with relatively rapid progression from symptom onset to death (Ramirez et al., 2008). Motor control decreases with disease progression and care is often required throughout the duration of the disease. Symptom onset usually occurs distally in a single limb (limb-onset ALS) or through speech and swallowing difficulties (bulbar-onset ALS), with muscle wasting and weakness gradually spreading until the respiratory muscles are unable to function (Kiernan et al., 2011). More rarely, ALS can present with initial respiratory symptoms. ALS shares clinical, neuropathological and genetic features with frontotemporal dementia (FTD) where symptoms include changes in behaviour, language, and memory, associated with degeneration of neurons in the cortex and basal ganglia. Approximately 10-15% of patients diagnosed with FTD go on to develop motor neuron degeneration, referred to as FTD-ALS (Hodges, 2001). Conversely, approximately 30% of ALS patients later develop cognitive impairment

and dementia (Rippon et al., 2006). As such, the two diseases are said to be at opposite ends of a disease spectrum within which multiple disease phenotypes exist (Figure 1.1).

Figure 1.1. ALS-FTD disease spectrum. ALS and FTD are at opposite ends of the disease spectrum with pure ALS at one end and pure FTD at the other. In between is ALS with cognitive and behavioural impairments (ALS-CBI) and FTD with motor neuron degeneration (FTD-ALS).

1.3 Pathological features of ALS

One of the defining neuropathological features of both ALS and FTD is the presence of skein-like, neuronal cytoplasmic inclusions (NCIs) of aggregated protein (Mizusawa, 1993). The NCIs are ubiquitinated, but are tau- and α-synuclein negative (Mackenzie and Feldman, 2005). One of the major proteins found in NCIs is TDP-43, which associates with the inclusions in a hyperphosphorylated and fragmented form (Arai et al., 2006; Neumann et al., 2006). The TDP-43 NCIs have been detected in glial cells as well as neurons, and are distributed across many different regions of the CNS, including the spinal cord, hippocampus, frontal cortex and temporal cortex (Arai et al., 2006; Neumann et al., 2006). The inclusions are a pathological hallmark of both sporadic and familial ALS, with the exception of SOD1 fALS, suggesting an underlying protein aggregation disease mechanism may be shared between both the sporadic and familial forms of disease (Mackenzie et al., 2007; Tan et al., 2007). The distribution pattern of TDP-43 proteinopathy has been shown to correlate with neuronal loss in ALS patient brains. In keeping with this, patients with the most widespread TDP-43 NCIs perform worse in the Mini-Mental State Examination, which requires patients to complete a questionnaire and uses patient's responses to determine cognitive ability, than patients with inclusions

localised to only a small part of the brain (Brettschneider et al., 2013). Despite this correlation, it is not yet known whether the NCIs are causal of neurodegeneration.

TDP-43-negative aggregates are also frequently observed in ALS. Bunina bodies are small, eosinophilic inclusions found in the spinal cord and brain stem of sporadic and familial ALS patients (Tomonaga et al., 1978). Bunina bodies are formed exclusively by the aggregation of the cystatin C and transferrin proteins (Mizuno et al., 2006; Okamoto et al., 1993); studies performing staining against proteins with a known propensity to aggregate fail to detect other protein components. The reason for their formation and biological significance to ALS however is not well understood.

Basophilic inclusions (BIs) appear as round, fragmented or irregular-shaped filamentous inclusions in neurons from patients diagnosed with juvenile ALS and more rarely adult-onset ALS (Kusaka et al., 1993; Matsumoto et al., 1992). BIs are defined by their pale staining appearance with haematoxylin and eosin (H&E) and their strong reaction to Nissl stain (Kusaka et al., 1990). BIs are similar in size to the nucleus and have been observed in the spinal cord, motor cortex, putamen, facial nucleus and thalamus (Aizawa et al., 2000; Kusaka et al., 1990). Several RNA-binding and/or stress granule associated proteins have been found to associate with BIs, including FUS, poly(A)-binding protein 1, T cell intracellular antigen 1 and ribosomal protein S6 (Bäumer et al., 2010; Fujita et al., 2008), suggesting the inclusions may be associated with dysfunctional RNA processing and attenuated translation (Ito, 2014).

1.4 Genetics of ALS

ALS is predominantly a sporadic disease; 90% of patients develop ALS in singular, random instances throughout a population. The remaining 10% of ALS cases are caused by inheritance of a mutation in an ALS-linked gene, referred to as familial ALS (fALS) (Rowland and Shneider, 2001). There are more than 20 known genes in which a genetic mutation can cause ALS and approximately 68% of familial ALS cases are now associated with a mutation in one of these genes (Table 1.1) (Renton et al., 2013). Mutations in the *C9orf72*, *SOD1*, *TARDBP* and *FUS* genes are collectively causative for more than half of familial ALS cases. Hence, this introduction will review the disease mechanisms of how these mutations may be pathogenic. There are large efforts currently being undertaken to discover further

ALS-risk genes. For example, Project MinE is an international collaboration between ALS research institutes, scientists and clinicians, aiming to complete whole genome sequencing of 15,000 ALS patients and 7,500 controls across 17 countries to increase the current understanding of genetic risks for ALS (Press Release Project MinE, 2013).

Table 1.1. ALS-linked genes grouped by the proposed function of the causal gene

Gene	Protein	Reference
Superoxide metabolism		
SOD1	Superoxide dismutase 1	(Rosen et al., 1993)
RNA processing		
TARDBP	TDP-43	(Sreedharan et al., 2008)
FUS	Fused in sarcoma	(Kwiatkowski Jr. et al., 2009; Vance et al., 2009)
SETX	Senataxin	(Chen et al., 2004)
ANG	Angiogenin	(Greenway et al., 2006)
MATR3	Matrin-3	(Johnson et al., 2014)
HNRNPA1/HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A1 and A2B1	(Kim et al., 2013)
Trafficking, autophagy or protein proteostasis		
C9orf72	C9orf72	(DeJesus-Hernandez et al., 2011; Renton et al., 2011)
ALS2	Alsin	(Yang et al., 2001)
FIG4	Polyphosphoinositide 5-phosphatase	(Chow et al., 2009)
ANXA11	Annexin 11	(Smith et al., 2017)
VAPB	Vesicle-associated membrane protein-associated protein B	(Nishimura et al., 2004)
OPTN	Optineurin	(Maruyama et al., 2010)
SQSTM1	Sequestosome 1	(Fecto et al., 2011)
VCP	Valosin-containing protein	(Johnson et al., 2010)
UBQLN2	Ubiquilin 2	(Deng et al., 2011)
SPG11	Spatacsin	(Orlacchio et al., 2010)
TBK1	TANK-binding kinase 1	(Cirulli et al., 2015)
CHMP2B	Charged multivesicular body protein 2B	(Parkinson et al., 2006)

Cytoskeleton		
MAPT	Microtubule-associated protein tau	(Hutton et al., 1998)
PFN1	Profilin	(Wu et al., 2012a)
TUBA4A	Tubulin, alpha 4A	(Smith et al., 2014)
DCTN1	Dynactin	(Puls et al., 2003)
NEFH	Neurofilament heavy polypeptide	(Figlewicz et al., 1994)
KIF5A	Kinesin heavy chain factor 5A	(Brenner et al., 2018)
DNA damage		
NEK1	Never in mitosis A-related kinase 1	(Brenner et al., 2016)
C21orf2	C21orf2	(van Rheenen et al., 2016)

1.4.1 Superoxide dismutase 1

Mutations in *SOD1* account for approximately 20% of familial ALS cases (Rosen et al., 1993). Approximately 140 different mutations have been identified within the *SOD1* gene, however only 20 of these have been validated as causative of ALS pathogenicity by further research (Andersen, 2006). An X-ray crystallography study found that the mutations are located in non-catalytically active SOD1 protein domains (Deng et al., 1993). The normal function of superoxide dismutase 1 (SOD1) is to 'clean up' reactive oxygen species (ROS), metabolising them into oxygen and hydrogen peroxide, thus reducing oxidative stress and damage to cells. It was originally debated whether a loss of SOD1 enzymatic activity or generation of a toxic mutant SOD1 was responsible for SOD1-ALS pathogenicity. However, there are studies which suggest SOD1 null mice do not develop ALS clinical features and that there is no correlation between SOD1 dismutase activity and ALS severity, although these results need validating (Bruijn et al., 1998; Reaume et al., 1996). On the other hand, mutant SOD1 protein has been shown to form protein aggregates within motor neurons whose appearance correlates with both disease onset and progression (Bruijn et al., 1998). Mutant SOD1 has also been shown to impair both slow and fast axonal transport of cargo proteins and mitochondria respectively (Cleveland and Williamson, 1999; De Vos et al., 2007). The latter is caused by mutant SOD1 induction of PINK1/Parkin-dependent Miro1 degradation, which detaches mitochondria from the molecular motor kinesin 1 (Moller et al., 2017). Consequently, a toxic gain-of-function from mutant SOD1 is now accepted as the major contributing factor to pathology in SOD1-related ALS. Studies in transgenic SOD1 mice have tested the use of SOD1-targeting antisense oligonucleotides

(ASO) as a means of reducing both wild-type and mutant SOD1 levels (Smith et al., 2006; Winer et al., 2013). The 2006 mouse study showed therapeutic promise, with good ASO penetration into the CNS, reduced SOD1 protein levels and an increase in mouse survival rate (Smith et al., 2006; Winer et al., 2013). Clinical trials are now underway investigating the therapeutic use of SOD1-targeting antisense oligonucleotides in humans, although these have so far been unsuccessful (Miller et al., 2013).

1.4.2 TARDBP

After *SOD1*, mutations in the *TARDBP* gene account for a further 5% of familial ALS cases (Sreedharan et al., 2008). The *TARDBP* gene encodes the Tar DNA protein of 43 kDa (TDP-43) protein, a component of the ubiquitin-positive, tau- and α-synuclein negative protein inclusions observed in the CNS of ALS patients (Arai et al., 2006). TDP-43 is a predominantly nuclear-localised, DNA- and RNA-binding protein which was originally characterised as a transcriptional repressor of an HIV gene, and a splicing factor for the cystic fibrosis-related gene *CFTR* (Buratti and Baralle, 2001; Ou et al., 1995). Although TDP-43 is mostly concentrated in the nucleus, some protein shuttles to the cytoplasm where it binds mRNA 3' UTRs, influencing their stability, transport and translation (Colombrita et al., 2012; Wang et al., 2008). In the nucleus, TDP-43 regulates RNA processing and splicing of hundreds of transcripts through an interaction with non-coding GU repeats, and has also been shown to regulate its own expression level by enhancing splicing of a TDP-43 intron (Bhardwaj et al., 2013; Polymenidou et al., 2011). TDP-43 also interacts with the Drosha complex to regulate microRNA biogenesis (Buratti et al., 2010; Kawahara and Mieda-Sato, 2012).

A number of missense mutations in the *TARDBP* gene have been identified in both sporadic and familial ALS patients, and these mutations appear mostly within the TDP-43 C-terminus which is responsible for interactions with hnRNPs and splicing regulation (Buratti et al., 2005; Kabashi et al., 2008). Mutations in the *TARDBP* gene are likely to cause ALS through a toxic gain-of-function of the mutant protein combined with loss of functional TDP-43 from the nucleus. Mutant TDP-43 protein exhibits a higher propensity to form aggregates in comparison to wild-type TDP-43 *in vitro* (Guo et al., 2011; Johnson et al., 2009). This result is also supported by *in vivo* studies which report overexpressed, mutant TDP-43 more readily forms neurotoxic inclusions in spinal cord tissue from transgenic TDP-43 mice and primary

rat cortical neurons (Barmada et al., 2010; Kabashi et al., 2010; Wils et al., 2010). As such, mutations in the *TARDBP* gene enhance the mislocalisation of TDP-43 from the nucleus into cytoplasmic inclusions.

Studies in mice depleted in TDP-43 would suggest that loss of TDP-43 function is causal of ALS; TDP-43 knockout mice exhibit typical ALS-clinical phenotypes and motor neuron degeneration (Iguchi et al., 2013; Wu et al., 2012b). Furthermore, abhorrent splicing of mRNA transcripts known to be regulated by TDP-43 has been detected in ALS spinal cord, suggesting loss of nuclear TDP-43 can disrupt RNA processing (Xiao et al., 2011). Indeed, loss of nuclear TDP-43 in patient derived fibroblasts containing a mutation in *TARDBP* can lead to changes in alternative splicing events (Highley et al., 2014). Aggregation of mutant TDP-43 into NCIs may also hinder its ability to perform cytosolic functions. Wild-type TDP-43 is involved in the axonal transport of mRNP granules to the neuronal dendrite, and this transport of mRNA is abolished in iPSC-derived motor neurons from patients exhibiting a *TARDBP* mutation (Alami et al., 2014).

1.4.3 Fused in sarcoma

Mutations in the *FUS* gene are responsible for approximately 5% of familial ALS cases and 1% of sporadic ALS cases (Kwiatkowski Jr. et al., 2009; Vance et al., 2009). Like TDP-43, the FUS protein is an RNA/DNA binding protein that shuttles between the nucleus and cytoplasm, and has roles in transcription regulation, RNA processing and mRNP transport (Wang et al., 2015; Yang et al., 2014). FUS binds nascent pre-mRNA transcripts at GGU sequences in a saw-tooth pattern, and directly facilitates intron removal by interacting with RNA polymerase II and the spliceosome component U1 snRNP (Rogelj et al., 2012; Yu and Reed, 2015). FUS can also interact with the SMN complex via direct interactions with both the U1 snRNP and the SMN protein (Yamazaki et al., 2012). As well as regulating splicing of other mRNAs, FUS also has the ability to regulate its own expression by inhibiting splicing of an intron within the FUS transcript and thus targeting the mRNA for nonsense-mediated decay (Zhou et al., 2013). Lastly, FUS has been shown to interact with mRNA and mRNPs in the cytoplasm, aiding transport to distal parts of neurons as a method of translation regulation (Fujii et al., 2005).

The pathogenic *FUS* mutations are clustered within the C-terminal domain which includes the RNA-recognition motif and nuclear localisation signal (Chiò et al., 2009; Drepper et al., 2011). As a result, mutant FUS becomes mislocalised to the cytoplasm (Dormann et al., 2010; Gal et al., 2011). The mutant FUS protein has an intrinsically disordered, prion-like structure which drives the protein into cytoplasmic aggregates (Bosco et al., 2010; Patel et al., 2015; Vance et al., 2009). It has been proposed that mislocalisation of FUS into cytoplasmic aggregates inhibits its ability to autoregulate its expression leading to increased mutant FUS levels and thus exacerbating its own protein aggregation (Zhou et al., 2013).

FUS-ALS pathogenicity is currently understood to involve both a toxic gain-of-function combined with loss of functional FUS from the nucleus. Depletion of FUS in both HeLa cells and mouse brain leads to changes in gene expression and alternative splicing events, and introduces premature stop codons in some transcripts encoding central RNA-binding proteins (Lagier-Tourenne et al., 2012; Sun et al., 2015). However, transgenic ALS-mutant FUS mice but not FUS knockout mice develop motor neuron degeneration, suggesting toxicity from the mutant protein is required to develop ALS clinical phenotypes (Scekic-Zahirovic et al., 2016; Sharma et al., 2016). Cytoplasmic aggregations of mutant FUS bind and sequester the spliceosomal U1, U11 and U12 snRNPs, preventing their import into the nucleus for catalysis of intron splicing and worsening the defective splicing caused by loss of nuclear FUS (Reber et al., 2016; Yu et al., 2015). In addition, mutant FUS binds more strongly to the SMN protein than wild-type FUS, which may cause the reduction in nuclear Gems and snRNAs observed in FUS-ALS patient fibroblasts (Sun et al., 2015). Defective transport of mRNA to neural dendrites is also a toxic consequence of the cytoplasmic FUS inclusions. mRNA which encodes proteins essential for neurite growth and maintenance, and which under normal circumstances is transported to and locally translated at the neuronal axon or dendrite, becomes sequestered by mutant FUS in cytoplasmic inclusions leading to neurite degeneration (Jun et al., 2017; Yasuda et al., 2017).

1.5 Molecular pathogenesis of ALS

There are multiple molecular pathways that appear to be disrupted in ALS, and it is likely that a complicated convergence of these eventually leads to motor neuron degeneration. Proteins in which ALS-causative mutations lie highlight RNA processing, autophagy and protein proteostasis, cytoskeleton maintenance and

DNA damage as some of the pathways involved in ALS. In addition to these, mitochondrial dysfunction, glial cell toxicity and glutamate excitotoxicity have all been shown to have some involvement in ALS. Dysfunctional RNA processing in ALS is reviewed in Section 1.7.1.

1.5.1 Autophagy and protein proteostasis

There are multiple indications that autophagy and protein proteostasis are defective in ALS. Neuronal inclusions of aggregated protein are common to almost every ALS patient, and ALS-causative mutations are commonly found within genes that regulate autophagy and protein homeostasis.

Aggregation of protein is a common theme that runs throughout ALS: (i) mutations in *TARDBP*, *FUS* and *SOD1*, all lead to the cytoplasmic aggregation of their respective proteins, (ii) ubiquitinated TDP-43-positive, tau- and α-synuclein-negative inclusions are prominent in both sporadic and familial ALS and (iii) p62-positive but TDP-43-negative cytoplasmic inclusions are an additional pathological hallmark specific to C9orf72-ALS patients (Al-Sarraj et al., 2011). The improper removal of these aggregates, which often contain the autophagy and ubiquitin proteasome system (UPS) signalling proteins p62 and ubiquitin, implies there may be underlying faults to protein clearance pathways.

Many ALS-causative mutations lie in genes which are involved in autophagy. Mutations in the *SQSTM1* gene, which encodes the autophagy adaptor p62 protein, were discovered in both familial and sporadic ALS, and patients were found to have increased levels of both p62 and TDP-43 (Fecto et al., 2011; Teyssou et al., 2013). In 2011, mutations within the *UBQLN2* gene were discovered to be causative of X-linked ALS/dementia (Deng et al., 2011). The normal function of the Ubiquilin 2 protein is to recognise protein ubiquitin chains and facilitate protein degradation by the proteasome (Seok Ko et al., 2004). ALS-mutant forms of the Ubiquilin 2 protein impair global protein degradation by the UPS (Deng et al., 2011). Lastly, a mutation in the C9orf72 protein, responsible for autophagy initiation and trafficking, is the most common cause of familial ALS known to date, and causes p62 protein accumulation and impairs autophagy (Webster et al., 2016). The complete list of mutations in ALS-linked genes associated with autophagy, protein homeostasis and trafficking are detailed in table 1.1.

Autophagy works alongside the UPS to degrade cellular components such as damaged organelles and protein aggregates. Disruption to autophagy in mice by depletion of the autophagy gene Atg7 can cause neurodegeneration and development of ALS-like ubiquitin-positive inclusions (Komatsu et al., 2006). Interestingly, induction of autophagy has been shown to alleviate TDP-43 pathology (Barmada et al., 2014). p62 is involved in the selective autophagy process and is responsible for the binding of ubiquitinated proteins targeted for autophagosome degradation and mediating their transport to the autophagosome (Lattante et al., 2015). Knockdown of p62 in zebrafish has been shown to cause motor neuron degeneration and a decrease in swimming ability which can be rescued with the autophagy inducer rapamycin, suggesting defective autophagy can be causal of problems in motor function in ALS (Lattante et al., 2015). Lastly, ageing is the best known risk factor for developing ALS (Logroscino et al., 2010; Uenal et al., 2014). Autophagic clearance of proteins has been demonstrated to naturally decrease with age, and it is thought that neurons are particularly susceptible to disruptions to autophagy because their inability to divide exacerbates the accumulation of protein (Cuervo et al., 2005; Nixon, 2013). Hence, one possibility is that age-related accumulation of protein aggregates in ALS leads to the specific degeneration of motor neurons.

1.5.2 Cytoskeleton maintenance

Mutations in several genes involved in cytoskeleton maintenance and axonal transport have been linked with ALS (Table 1.1). Profilin is an actin-binding protein that facilitates the polymerisation of monomeric actin into filamentous actin and can also regulate microtubule dynamics (Carlsson et al., 1977; Nejedla et al., 2016). ALS-linked mutations in the profilin gene lead to generation of cytoplasmic protein aggregates composed of mutant profilin, p62, ubiquitin and TDP-43, a decrease in the actin-profilin interaction, and defective axonal growth in motor neurons (Tanaka et al., 2016; Wu et al., 2012a). ALS-causative mutations in the tubulin alpha-4A gene provide further evidence that a disrupted cytoskeleton may contribute to ALS pathogenicity. The *TUBA4A* mutations interfere with the microtubule network due to the inability of tubulin alpha-4A to properly form tubulin dimers with beta-tubulin (Smith et al., 2014). In addition, an ALS-linked mutation in the axonal transport protein dynactin has been shown to hinder axonal vesicular transport and induce motor neuron degeneration (Laird et al., 2008; Puls et al., 2003). Finally, a recent study found that splice-site mutations in the C-terminal of the *KIF5A* gene, which

expresses a neuronal kinesin motor protein responsible for protein and organelle trafficking along microtubules, can be causative of ALS (Brenner et al., 2018). The mutations are suggested to lead to haploinsufficiency of the KIF5A protein, and disruption to the neuronal transport of KIF5A cargos (Brenner et al., 2018). Collectively, the mutations implicate axonal pathology and a defective neuronal cytoskeleton may be a primary cause of ALS.

Dysfunctional axonal transport is also evident in sporadic ALS and variations of familial ALS with mutations in non-cytoskeleton related genes. ALS-related mutant SOD1, TDP-43 and FUS proteins have all been demonstrated to interfere with axonal transport in primary neurons by binding to and inhibiting the function of proteins involved in this pathway (Alami et al., 2014; De Vos et al., 2007; Jun et al., 2017; Moller et al., 2017; Yasuda et al., 2017). As previously discussed, axonal transport pathology has strong links particularly with SOD1-ALS; axonal transport is slowed in mice which express an ALS mutant SOD1 protein and mutant SOD1 can directly inhibit transport of mitochondria (Cleveland and Williamson, 1999; Moller et al., 2017). The axonal accumulation of mitochondria, vesicles and neurofilaments in motor neurons from ALS patient spinal cord further implicates faults to the axonal transport pathway as a disease mechanism (Hirano et al., 1984). Moreover, neurofilament heavy polypeptide is elevated in the CSF taken from ALS patients, and an increase in neurofilament heavy polypeptide causes motor neuron degeneration in mice (Collard et al., 1995; Xu et al., 2016).

1.5.3 DNA damage

DNA damage has been widely reported in ALS, suggesting the DNA-repair machinery may be disrupted in disease. Oxidative damage to DNA introduces breaks into nuclear and mitochondrial DNA, and markers of oxidative DNA damage have been reported as elevated in urine and CSF from ALS patients, and in ALS mouse models (Aguirre et al., 2005; Bogdanov et al., 2000; Ferrante et al., 1997; Warita et al., 2001).

ALS-linked mutations are also commonly found in genes with roles in the DNA damage response. The FUS protein is recruited to sites of DNA damage and mediates the efficient repair of DNA breaks, suggesting the mislocalisation of FUS in FUS-ALS could lead to problems in DNA repair (Mastrocola et al., 2013; Wang et al., 2013b). In addition, a repeat expansion in the C9orf72 gene, which is causal for

familial ALS, has been shown to disturb ATM-mediated DNA repair, leading to an accumulation of DNA double strand breaks (Walker et al., 2017).

In 2016, two new risk genes for ALS were discovered: Nek1 and C21orf2. Nek1 mutations are proposed to cause loss-of-function of the NIMA-related protein kinase 1 (Nek1) protein (Brenner et al., 2016; Cirulli et al., 2015; Kenna et al., 2016). The Nek1 protein has functions in cilium formation and mitochondria permeability, both of which have been implicated in ALS (Chen et al., 2009; Ma et al., 2011; Shalom et al., 2008). A third function of Nek1 is in DNA repair, and depletion of Nek1 has been reported to decrease the efficiency of the DNA damage response (Pelegrini et al., 2010). Mutations in the C21orf2 gene in ALS also leads to loss-of-function of the C21orf2 protein (van Rheenen et al., 2016). The C21orf2 protein interacts with the Nek1 protein in the DNA repair pathway, and similar to Nek1, evidence suggests depletion of C21orf2 can impair homologous recombination and slow DNA repair (Fang et al., 2015).

1.5.4 Mitochondrial dysfunction

Mitochondria are dynamic organelles that have essential roles in oxidative phosphorylation and production of ATP, calcium signalling, and apoptosis. In ALS, studies have found the morphology, functionality and degradation of mitochondria can be impaired by mutant ALS proteins. The mutant SOD1 protein misfolds into a structure that mediates its localisation and aggregation within mitochondria, where it has been shown to induce mitochondrial damage (Kawamata and Manfredi, 2008; Vande Velde et al., 2008). The interaction of mutant SOD1 with mitochondria induced apoptosis and disrupted calcium homeostasis and efficient ATP production (Ferri et al., 2006; Kruman et al., 1999; Pasinelli et al., 2004). In addition, transgenic mice expressing ALS-linked mutant SOD1 have been shown to exhibit degenerating mitochondria which form vacuoles in neuronal axons and dendrites, defective electron transport chains and mitochondria respiration (Mattiazzi et al., 2002; Kong and Xu, 1998; Wong et al., 1995). Collectively, these studies are convincing for a role of mitochondrial dysfunction in SOD1-ALS.

Mitochondrial impairment has also been observed in TDP-43 and FUS models of ALS, suggesting mitochondria damage is a molecular pathway shared between the different subtypes of disease. Overexpression of TDP-43 and FUS ALS mutants in motor neurons changed mitochondria morphology and caused them to shorten in

length (Tradewell et al., 2012; Wang et al., 2013a). Expression of the pathological TDP-43 25 kDa fragment showed its localisation to mitochondria where it influenced the membrane potential, caused mitochondria clustering and induced mitophagy (Hong et al., 2012). These studies suggest that the TDP-43 pathology frequently observed in ALS patients can damage neuronal mitochondria, which may be either causal or contributory to ALS pathogenicity.

Finally, it has been reported that the activities of the respiratory chain complexes are reduced in post mortem spinal cord from ALS patients and that mitochondrial DNA is highly mutated (Wiedemann et al., 2002). In addition, ATP levels and respiratory complex I activity were likewise reduced in lymphoblast cells derived from ALS patients (Ghiasi et al., 2012). iPSC-derived motor neurons differentiated from C9orf72-ALS fibroblasts display evidence of reduced mitochondrial dysfunction, including disrupted calcium homeostasis and mitochondrial membrane potential. The decreased ATP production in ALS patients could reduce motor neurons' ability to cope with stress and disruptions to molecular pathways.

1.5.5 Glial cell toxicity

ALS involves the selective neurodegeneration of motor neurons. However, evidence is emerging for an important role of glial cells in exacerbating motor neuron degeneration, specifically microglia and astrocytes. Microglia are macrophages that mediate the immune response in the central nervous system, whereas astrocytes are supportive glial cells that 'prune' and maintain healthy synapses and neuronal circuits (Prinz and Priller, 2014; Sofroniew and Vinters, 2010).

Neuroinflammation and activation of microglia and astrocytes is frequently observed in ALS. Reactive microglia are widespread throughout the CNS of ALS patients, including the spinal cord and motor cortex (Henkel et al., 2004; Kawamata et al., 1992; Turner et al., 2004). Activation of both microglia and astrocytes is also commonly detected in SOD1-ALS mouse models (Hall et al., 1998; Kriz et al., 2002). In addition, ALS brain areas that exhibited gliosis correlated with cortical thinning and changes to the local neuronal environment (Alshikho et al., 2016). Interestingly, activation of microglia has been shown to occur prior to motor neuron degeneration, suggesting neuroinflammation can be an early marker of ALS (Henkel et al., 2006).

Interestingly, ALS but not wild type astrocytes are toxic to motor neurons. Co-culture of astrocytes expressing ALS mutant SOD1 with motor neurons can induce motor neuron degeneration (Marchetto et al., 2008; Nagai et al., 2007). In addition, astrocytes from sporadic and familial ALS patients are also toxic to motor neurons in co-culture, causing dendrite shortening and neuronal atrophy (Haidet-Phillips et al., 2011; Re et al., 2014). A recent study suggests ALS-astrocytes secrete proteins that downregulate major histocompatibility complex class I (MHCI) molecules on motor neurons, which in turn increases their susceptibility to toxicity from killer-cell immunoglobulin-like receptors on the astrocytes (Song et al., 2016). In addition, wild-type glial cells are neuroprotective to ALS-SOD1 motor neurons in a chimeric mouse model, and can increase life expectancy (Clement et al., 2003).

One of the central questions in ALS research is trying to understand how a mutation in an ALS-linked gene, or disruption to a molecular pathway, leads to the selective neurodegeneration of motor neurons whilst sparing other cells. These studies suggest that glial cells in ALS can be neurotoxic and treatments targeting neuroinflammation or activated microglia and astrocytes may extend motor neuron survival. Indeed, treatment of SOD1-ALS mice with the anti-inflammatory drug Minocycline led to a mild increase in mice survival rates (Kriz et al., 2002).

1.5.6 Glutamate excitotoxicity

The glutamate hypothesis was one of the earliest predicted mechanisms involved in ALS (Epstein et al., 1994; Shaw and Ince, 1997). Neuronal degeneration can occur from two types of excitotoxicity: classic and slow. Classic excitotoxicity occurs from the overstimulation of neurons by high levels of synaptic glutamate, which causes a disruption to calcium homeostasis from the massive influx of Na^+ and Ca^{2+} ions (Doble, 1999). Slow excitotoxicity involves an increased sensitivity to glutamate from the post-synaptic neuron, and can be caused by physical changes to the glutamate receptors (Doble, 1999; Van Den Bosch et al., 2006).

Riluzole is currently the only approved treatment for ALS in the UK, and increases ALS patients' life expectancy by approximately 2-3 months (Lacomblez et al., 1996). Riluzole exerts an anti-excitotoxic effect by inhibiting voltage-gated sodium channels on pre-synaptic neurons, thus blocking their release of glutamate (Doble, 1996). A secondary mechanism of action may also be via the competitive binding of Riluzole to post-synaptic glutamate receptors, thereby preventing glutamate binding

(Albo et al., 2004; Doble, 1996). The effectiveness of Riluzole to treat ALS, albeit moderately, suggests glutamate excitotoxicity is involved in ALS.

There is evidence of glutamate excitotoxicity in both animal models of ALS and in ALS patients. A study in SOD1-ALS mice found that glutamate levels were raised in samples of extracellular fluid taken from the cortex (Alexander et al., 2000). Increased levels of glutamate have also been detected in the CSF of a small number of ALS patients (Rothstein et al., 1990; Shaw et al., 1995). In contrast, several studies investigating glutamate in ALS patient brain and spinal cord tissue have found levels to actually be decreased (Malessa et al., 1991; Tsai et al., 1991). Interestingly, levels of the glutamate precursor *N*-acetylaspartylglutamate are also decreased in ALS patient brain and spinal cord (Rothstein et al., 1990; Tsai et al., 1991). Although contradictory, the fluctuating glutamate levels between studies is suggestive of a role for defective glutamate signalling in ALS.

1.6 Chromosome 9 open reading frame 72

The first genetic evidence for a causative link between a mutation in C9orf72 and ALS was published in 2000, and identified as a region of genetic linkage in families with ALS and FTLD on chromosome 9q21-q22 (Hosler et al., 2000), later confirmed and described as a 103-gene spanning region within 9p13.2-21.3 (Vance et al., 2006). Further genome-wide associated studies (GWAS) narrowed down the gene to chromosome 9p21.2 (Laaksovirta et al., 2010; Shatunov et al., 2010). In 2011, two papers were simultaneously published describing the discovery of a hexanucleotide (GGGGCC) repeat expansion within intron 1 of chromosome 9 open reading frame 72 (C9orf72) that is the most common genetic defect identified in familial and sporadic ALS and FTD to date, with approximately 35-40% of familial ALS and 25% of familial FTD patients estimated to possess the mutation (DeJesus-Hernandez et al., 2011; Renton et al., 2011; van Blitterswijk et al., 2012).

In a healthy individual, less than 30 GGGGCC repeats are present in the C9orf72 gene, whereas in an ALS or FTD patient, the repeat number can reach in excess of 1000 (Renton et al., 2011). Unlike other repeat expansion diseases such as Huntington's disease, where repeat number correlates with disease severity, this does not seem to be the case in C9orf72 ALS (Dols-Icardo et al., 2014). Under normal circumstances, the C9orf72 gene undergoes alternative splicing to produce three mRNA transcripts which encode two protein isoforms: C9orf72 short form (1-

222aa) derived from transcript variant 2 and C9orf72 long form (1-481aa) derived from transcript variants 1 and 3 (Figure 1.2) (DeJesus-Hernandez et al., 2011). Transcript variant 1 starts at exon 1b, whereas transcript variants 2 and 3 begin at exon 1a. The positioning of the GGGGCC repeat expansion therefore lies in the promoter for transcript variant 1 and affects its transcription in C9orf72-ALS. For transcript variants 2 and 3, the expansion lies downstream of exon 1a within the first intron, and is therefore transcribed. Both the short and long isoforms are expressed to comparable levels across most tissues and both are expressed at high levels in the brain (DeJesus-Hernandez et al., 2011; Renton et al., 2011).

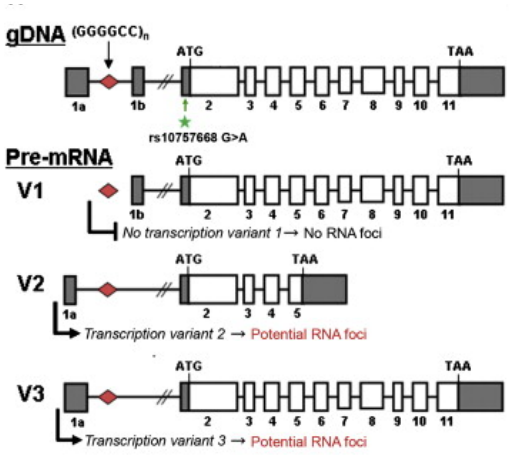


Figure 1.2. C9orf72 mRNA transcripts. The location of the repeat expansion in the C9orf72 gene is shown. White boxes represent coding exons; shaded boxes represent non-coding exons. The C9orf72 gene produces 3 pre-mRNA transcripts. The V1 pre-mRNA is transcribed from a promoter within intron 1 and is not transcribed in C9-ALS due to disruption from the repeat expansion. The V2 and V3 pre-mRNAs are transcribed upstream of the repeat expansion and produce expansion-containing pre-mRNA. The V2 transcript undergoes alternative splicing to produce an mRNA that terminates after exon 5, which encodes the C9orf72 short protein isoform. The V3 transcript encodes the C9orf72 long protein. Figure used with permission from (DeJesus-Hernandez et al., 2011).

1.6.1 C9orf72 mechanisms of disease

Three mechanisms of disease have been proposed as to how the C9orf72 mutation leads to ALS pathogenicity (Figure 1.3): (i) toxic gain-of-function from generation of RNA foci, (ii) toxic gain-of-function from generation of dipeptide repeat proteins (DPRs) and (iii) loss-of-function from haploinsufficiency of the C9orf72 protein. There is evidence in support of each pathomechanism, therefore the mechanisms are likely non-mutually exclusive.

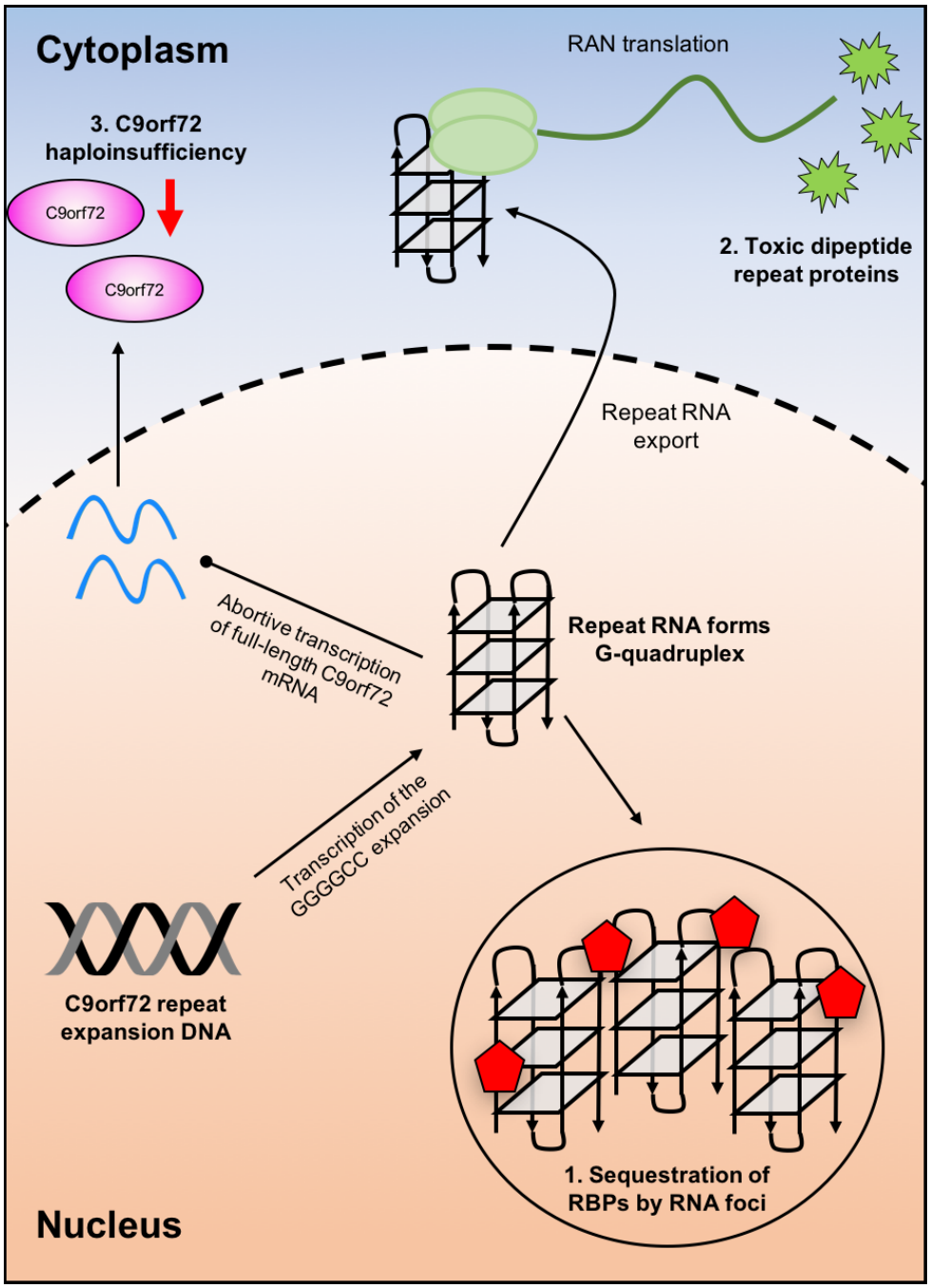


Figure 1.3. C9orf72-linked disease mechanisms. (1) Expansion containing RNA forms a stable G-quadruplex secondary structure which binds and sequesters RNA binding proteins leading to dysfunctional RNA metabolism. (2) GGGGCC repeat RNA is exported from the nucleus to the cytoplasm where it undergoes repeat associated non-ATG (RAN) translation into dipeptide repeat proteins which aggregate and cause toxicity possibly through disrupted RNA metabolism or disruption to the UPS. (3) Abortive transcription of C9orf72 caused by the repeat expansion leads to reduced C9orf72 mRNA levels and haploinsufficiency of the C9orf72 protein.

1.6.1.1 Generation of RNA foci

Bidirectional transcription of the C9orf72 gene in the sense and antisense direction generates RNA containing a GGGGCC or CCCCGG expansion respectively. The expansion-containing RNA forms RNA foci in the nucleus, and less frequently in the cytoplasm, which can sequester RNA-binding proteins (Cooper-Knock et al., 2014; Stopford et al., 2017).

The sense RNA foci were originally detected in the nuclei of post-mortem spinal cord and frontal cortex tissue from C9orf72-FTLD patients (DeJesus-Hernandez et al., 2011). The presence of both sense and antisense RNA foci, approximately 0.2-0.5 μm in size, has since been widely reported in C9orf72-expansion ALS/FTD patients (Gendron et al., 2013; Lagier-Tourenne et al., 2013; Zu et al., 2013). The distribution pattern of RNA foci is widespread; foci have been detected in cortical, hippocampal and cerebellar neurons from the brain, motor neurons from the spinal cord and astrocytes and microglia (Gendron et al., 2013; Lagier-Tourenne et al., 2013). There has also been detection of foci in peripheral cells from C9orf72 patients, such as peripheral blood leukocytes, skin-biopsy derived fibroblasts and immortalised lymphoblasts (Lagier-Tourenne et al., 2013; Zu et al., 2013). RNA foci are typically numbered 1-2 per nucleus, but as many as 40 foci per nucleus has been observed (Lagier-Tourenne et al., 2013). Sense and antisense foci can localise within the same cell, but sense foci are found in greater frequency in granule neurons from the cerebellum whereas antisense foci are more commonly detected in motor neurons and cerebellar Purkinje neurons (Cooper-Knock et al., 2015b).

Expansion containing sense RNA has been shown to form a stable G-quadruplex secondary structure which increases in stability with repeat number (Fratta et al., 2012; Reddy et al., 2013). The C9orf72-expansion RNA G-quadruplex is a four-stranded structure formed by the stacking of four guanine nucleobases that associate by Hoogsteen hydrogen bonding within the four planes (Haeusler et al., 2014). The self-interaction between the G-quadruplex structures promotes phase-separation and accelerates RNA foci formation (Fay et al., 2017; Reddy et al., 2013). RNA foci can be decreased in C9orf72 iPSC-derived cortical and motor neurons by incubation with small molecules that bind and stabilise the GGGGCC-expansion RNA in the G-quadruplex structure (Simone et al., 2017).

RNA foci are a common pathological hallmark in neurodegenerative diseases where repeat expansion mutations are the underlying cause. Myotonic dystrophy and spinocerebellar ataxia type 8 are two examples of microsatellite expansion disorders where RNA foci exert toxicity through the sequestering of RNA binding proteins (Daughters et al., 2009; Davis et al., 1997; Taneja et al., 1995). C9orf72 GGGGCC-RNA has been shown to bind the RNA-binding proteins nucleolin, hnRNP U, hnRNP F and RPL7 (Haeusler et al., 2014). C9orf72 RNA containing the antisense CCCCGG sequence in a hairpin structure has likewise been shown to interact with the hnRNP K protein (Haeusler et al., 2014). Furthermore, co-localisation between RNA foci and RNA-binding proteins such as SF2, SC35 and hnRNP H has been detected in neuronal cell lines and C9orf72-ALS Purkinje neurons (Cooper-Knock et al., 2015b; Lee et al., 2013). The GGGGCC expansion RNA in its G-quadruplex assembly has also been found to bind heme and alter its oxidative properties, although the role this may play in disease pathology is unclear (Grigg et al., 2014).

There are several studies that support RNA foci as a contributor to ALS pathogenicity. First, they have been shown to correlate with ALS clinical phenotype (Cooper-Knock et al., 2014). Second, many of the proteins that are sequestered by RNA foci have functions in RNA processing, and splicing of the mRNA targets of the sequestered proteins is altered in C9orf72-ALS patients, suggesting RNA foci can interfere with RNA metabolism (Cooper-Knock et al., 2015a; Conlon et al., 2016). Sequestration of the RNA-editing deaminase-2 (ADARB2) protein into RNA foci has also been shown to increase susceptibility of iPSC-derived neurons to glutamate excitotoxicity (Donnelly et al., 2013). Finally, GGGGCC expansion RNA has also been detected as a component of RNA granules in distal parts of iPSC-derived neurons from C9orf72 patients, where it has been shown to interact with the fragile X mental retardation protein (FMRP) and locally influence translation and neurite branching (Burguete et al., 2015; Rossi et al., 2015). These studies suggest multiple molecular pathways may be affected by sequestration of proteins by the repeat expansion RNA.

1.6.1.2 Dipeptide repeat proteins

The repeat expansion resides in a non-coding region of C9orf72 and thus would not be expected to undergo translation. However, dipeptide repeat proteins (DPRs) have been found to accumulate in both the nucleus and cytoplasm of ALS brains,

which are generated by the unconventional repeat associated non-ATG (RAN) translation of the expansion-containing RNA (Ash et al., 2013; Mori et al., 2013a). RAN translation has been found to occur in other repeat expansion disorders such as distal myopathy and spinocerebellar ataxia type 8, and is thought to require a hairpin or G-quadruplex secondary structure as is adopted by RNA containing the C9orf72 repeat expansion (Cleary and Ranum, 2013; Zu et al., 2011). The RNA structure is hypothesised to behave as an internal ribosome entry site (IRES), which recruits proteins involved in translation initiation (Stoneley and Willis, 2004). RAN translation of the C9orf72 repeat expansion can be inhibited by treatment with G-quadruplex-binding small molecules, supporting the theory that the secondary structure may be required for DPR generation (Simone et al., 2017). RNA foci have been detected in the neuronal cytoplasm, suggesting the expansion-containing RNA is exported from the nucleus, thereby becoming available for RAN translation (Donnelly et al., 2013). Exactly how the repeat expansion acts as a template for RAN translation is unclear, and there are conflicting reports in the literature. There is evidence to suggest the C9orf72 transcript is processed normally as a mature mRNA which is accurately spliced downstream of exon 2, but displays intron retention of the expansion-containing intron 1 (Niblock et al., 2016). The expansion-containing RNA has been shown to sequester the SRSF1 protein which functions in mRNA nuclear export, and this sequestration targets the RNA for export to the cytoplasm (Hautbergue et al., 2017). Inhibition of SRSF1 increased the nuclear retention of the expansion RNA, thereby decreasing DPR production and toxicity. Finally, one study found that the intron containing the GGGGCC expansion is successfully spliced in the nucleus but rather than being degraded, is exported to the cytoplasm where it can act as the RNA template for RAN translation (Cheng et al., 2018). Interestingly, the latter pathway is further upregulated by cellular stress that may arise from DPR expression, creating a positive feedback loop which increases RAN translation of the repeat expansion and DPR generation.

DPR accumulations were first described as being similar in appearance to the TDP-43 negative inclusions found in C9orf72-ALS motor neurons (Ash et al., 2013). It was later discovered that the dipeptide repeat proteins co-localise with ubiquitinated p62 positive/TDP-43 negative inclusions (Al-Sarraj et al., 2011; Mann et al., 2013). By raising antibodies against the DPRs potentially encoded by consecutive frame reads in both the sense and antisense direction, the C9orf72 repeat expansion was found to produce the following dipeptide repeat proteins via RAN translation: poly-GlycineAlanine (GA); poly-GlycineProline (GP); poly-GlycineArginine (GR); poly-

ProlineAlanine (PA); and poly-ProlineArginine (PR) (Mori et al., 2013a, 2013b; Zu et al., 2013). DPRs have been detected in multiple brain regions from C9orf72-ALS patients, but are most common in the cerebral cortex, cerebellum and hippocampus (Ash et al., 2013; Davidson et al., 2014).

Of the DPRs, expression of poly-GA, poly-GR and poly-PR appear to instigate neuronal degeneration and cell toxicity more readily than poly-PA and poly-GP, hence research efforts have focused on these proteins. The poly-GA protein is the most common DPR found in cytoplasmic inclusions in C9orf72 ALS/FTD patient tissue, and has been shown to form insoluble aggregates more readily than the other DPRs (Mori et al., 2013b; Yamakawa et al., 2014; Zhang et al., 2014). Poly-GA aggregates have also been reported to have the ability to propagate by cell-to-cell transmission in a neuroblastoma cell line (Chang et al., 2016). Expression of the poly-GA protein forms p62- and ubiquitin-positive cytoplasmic aggregates in primary cortical and hippocampal neurons, induces ER-stress via inhibition of the ubiquitin proteasome system and sequesters the Unc119 protein which is responsible for axonal maintenance (May et al., 2014; Yamakawa et al., 2014; Zhang et al., 2014). In addition, expression of a poly-GA construct in mice caused the mislocalisation of HR23A, RanGap1 and Pom121 into cytoplasmic poly-GA inclusions (Zhang et al., 2016). Finally, expression of poly-GA constructs in both cell lines and primary neurons can activate caspase-3 and the apoptosis pathway, indicative of the neurotoxicity of poly-GA aggregates (Lee et al., 2017; Zhang et al., 2014). Despite this, there is no correlation between poly-GA pathology and neurodegeneration in C9orf72-ALS/FTD patients (Mackenzie et al., 2013).

Unlike the poly-GA, poly-GP and poly-PA proteins which form large cytoplasmic inclusions, the poly-GR and poly-PR proteins migrate to the nucleus where they associate with the nucleolus (Kwon et al., 2014; Wen et al., 2014). The poly-GR and poly-PR proteins have been shown to impair numerous nuclear events including nucleocytoplasmic shuttling and RNA processing. The binding of poly-GR and poly-PR to the nucleolus induces nucleolar stress, impairs synthesis of ribosomal RNA and eventually causes cell death (Kwon et al., 2014; Tao et al., 2015). Studies have shown the poly-PR protein interacts with the nuclear pore complex preventing nuclear import and export, and nucleocytoplasmic shuttling appears to be disrupted as a consequence of DPR expression (Boeynaems et al., 2016; Freibaum et al., 2015; Jovičić et al., 2015; Shi et al., 2017). Finally, poly-GR and poly-PR proteins can interfere with RNA metabolism. Both proteins can interact with the U2 snRNP, a

component of the major splicing machinery, which shows mislocalisation to the cytoplasm in DPR-expressing iPSC-motor neurons from C9orf72-ALS patients (Yin et al., 2017). Treatment of HeLa cells with poly-GR and poly-PR peptides can inhibit spliceosome assembly and induce mislocalisation of the U2 snRNP to the cytoplasm (Yin et al., 2017). Moreover, the poly-GR and poly-PR proteins have been reported to interact with the splicing factor protein, hnRNP A2, and can cause global alteration of mRNA splicing events (Kwon et al., 2014). Indeed, proteins with a high arginine content are common regulators of RNA processing through direct RNA interactions (Bayer et al., 2005).

It is certain DPRs can be neurotoxic, however most studies have investigated their toxicity using overexpression models. In reality, the levels of DPRs in ALS/FTD patients are low and are detected in brain regions spared from neurodegeneration, with the exception of poly-GR pathology (Davidson et al., 2014; Gomez-Deza et al., 2015; Mackenzie et al., 2015). Interestingly, the poly-PR and poly-GR proteins, which demonstrate the most neurotoxicity experimentally, have the lowest abundance in the brain (Mackenzie et al., 2015). Thus, it is unlikely DPR expression is exclusively causal of C9orf72-ALS pathogenicity, but may contribute along with other disease mechanisms.

1.6.1.2.1 C9orf72 gain-of-function animal models

A number of animal models have been generated to investigate a toxic gain-of-function in C9orf72-ALS pathogenicity. The first mouse model involved expression of 66 repeats of the GGGGCC expansion using viral mediated transgenesis, which caused the mice to present with typical C9orf72-ALS clinical features such as impaired motor function, neuronal loss and pathological features such as TDP-43 inclusions, RNA foci and dipeptide repeat proteins, supporting a toxic gain-of-function mechanism (Chew et al., 2015).

Subsequently, four mouse models have been generated expressing repeat-expansion forms of C9orf72 from bacterial artificial chromosome (BAC) DNA. The studies vary in their number of GGGGCC repeats, the fragment of the C9orf72 ORF expressed, and the length of the flanking 5' and 3' UTRs. All four BAC-C9orf72 mouse models displayed RNA foci and DPRs in distribution patterns similar to post-mortem tissue from C9orf72-ALS patients (Jiang et al., 2016; Liu et al., 2016; O'Rourke et al., 2015; Peters et al., 2015). However, mice from the first two studies

did not exhibit neuronal loss or develop ALS-clinical phenotypes (O'Rourke et al., 2015; Peters et al., 2015), whereas mice from the latter studies developed motor and behavioural deficits and had motor neuron degeneration (Jiang et al., 2016; Liu et al., 2016). There is therefore conflicting evidence regarding RNA foci and DPRs and their capability to cause ALS pathogenicity in mice.

The studies discussed have investigated a gain-of-function mechanism from the repeat expansion without narrowing down the individual contribution from RNA foci versus dipeptide repeat proteins. However, studies in *Drosophila* models favour DPRs as the primary toxic species. Expression of RNA containing 36 and 103 repeats of the GGGGCC expansion in flies led to RNA foci and DPR formation, and severe degeneration of the fly eye (Mizielinska et al., 2014). However, expression of RNA containing the same amount of GGGGCC repeats but interrupted throughout the sequence with stop codons, forms RNA foci but not DPRs and does not lead to neurodegeneration, suggesting DPRs are required for toxicity (Mizielinska et al., 2014). In a second *Drosophila* model, 160 GGGGCC repeats were expressed from an intron, thus more accurately reflecting C9orf72-ALS conditions (Tran et al., 2015). Intronic expression of the repeat expansion led to generation of sense RNA foci but not DPRs, and there was no effect on locomotor abilities or signs of neurodegeneration (Tran et al., 2015). Collectively, the *Drosophila* models suggest DPRs rather than RNA foci are more important in ALS pathogenicity. This may be true only for flies however, and similar experiments in other animal models are required.

In contrast, a recent zebrafish model supports RNA foci toxicity as a causal ALS pathomechanism. Injection of both GGGGCC and CCCCGG repeat RNA impaired axon growth and caused branching defects, but did not generate DPRs, suggesting toxicity was via a DPR-independent mechanism (Swinnen et al., 2018). To confirm this possibility, injection of repeat RNA that contained stop codons regularly throughout its sequence and therefore could not be RAN translated, was sufficient to cause the same axonal defects. Lastly, the study reported that RNA foci sequestered the RNA-binding protein Pur-alpha, and overexpression of Pur-alpha rescued axonal growth and reduced RNA foci load (Swinnen et al., 2018). A second study found expression of 30 repeats of GGGGCC in a *Drosophila* model impairs locomotor function and causes cellular defects in the eye (Zhang et al., 2015). DPR proteins were not detected, and so therefore were unlikely to contribute to neurodegeneration in this model. Moreover, the defects were rescued with

RanGAP1 overexpression, a protein involved in nucleocytoplasmic shuttling which has been shown to bind C9orf72 repeat expansion RNA, suggesting RanGAP1 sequestration by RNA is the toxic mechanism at play (Zhang et al., 2015). Together, these studies suggest neurodegeneration can occur from an RNA-mediated mechanism of protein sequestration.

1.6.1.3 Haploinsufficiency of C9orf72

The original paper that described the discovery of the C9orf72 repeat expansion as an ALS-causing mutation found that, in addition to RNA foci, C9orf72 mRNA levels were decreased in expansion containing cells and CNS tissue, which may correspond with reduced C9orf72 protein expression (DeJesus-Hernandez et al., 2011). This was confirmed in post-mortem tissue from ALS/FTLD patients where a 50% reduction in C9orf72 mRNA was described (Gijselinck et al., 2012), and again in frontal cortex tissue and lymphoblasts from ALS/FTLD patients (Ciura et al., 2013). There is a lack of evidence showing decreased C9orf72 protein levels in disease tissue, mainly due to the difficulty in obtaining a specific C9orf72 antibody, but one group found levels of the C9orf72 long isoform were decreased in post-mortem frontal cortex tissue from ALS/FTLD patients who carried the C9orf72 mutation (Waite et al., 2014). The C9orf72 short isoform was not detected with the anti-C9orf72 antibody. A second study implementing C9orf72 isoform specific antibodies found the C9orf72 long protein was decreased in the frontal cortex and temporal cortex of C9orf72-ALS patients, but not in the motor cortex or cerebellum (Xiao et al., 2015). Surprisingly, the C9orf72 short protein was increased in the frontal cortex and temporal cortex of C9-ALS patients (Xiao et al., 2015).

There are several potential explanations for the decreased C9orf72 levels. First, epigenetic control of gene expression is exerted mainly through the acetylation and methylation of histones, with trimethylation of specific lysine residues on histone tails playing a major role in gene activation and repression (Barski et al., 2007). Trimethylation of lysines 9 and 27 on histone H3 are generally associated with gene silencing whereas similar methylation of lysine 36 on histone H3 can be linked with activation (Cao et al., 2002; Stewart et al., 2005; Wagner and Carpenter, 2012). Interestingly, binding was found between C9orf72 expansion DNA with trimethylated lysines 9 and 27 on histone H3, suggesting the C9orf72 gene may have decreased expression leading to reduced mRNA levels (Belzil et al., 2013). If ALS pathogenicity is caused via a toxic gain-of-function from RNA foci and DPRs

however, decreased gene expression would be expected to be beneficial. This has been supported in studies where CpG methylation of the C9orf72 promoter, which works alongside histone modifications to reduce gene expression, was found to correlate with decreased RNA foci and neuronal loss (Liu et al., 2014; McMillan et al., 2015).

Perhaps another mechanism of how reduced C9orf72 mRNA levels occur is through the G-quadruplex structure adopted by the hexanucleotide repeat expansion. Transcription of GGGGCC repeat expansion DNA was found to lead to favoured production of abortive and truncated transcripts in comparison to full length pre-mRNA in a repeat-number- and structure- dependent manner (Haeusler et al., 2014). Production of shortened 'abortive' transcripts in favour of full length transcripts in expansion carrying patients was additionally confirmed *in vivo*.

1.6.1.3.1 The C9orf72 protein

The C9orf72 protein is expressed throughout the grey matter of the CNS and shows a high level of sequence homology with non-human primates, mouse and zebrafish orthologs, indicating that its function is likely important (Suzuki et al., 2013). To support a haploinsufficiency disease mechanism, it is essential to understand the normal protein function of C9orf72. Should C9orf72 play a role in a vital cellular function, reduced levels of the protein could potentially have negative consequences. Early bioinformatics studies compared the sequence homology of C9orf72 with proteins of known function. PSI-BLAST searches found that the C9orf72 protein showed strong homology with differentially expressed in normal and neoplastic cells (DENN)-domain and DENN-like proteins, both in its primary sequence and predicted secondary structure (Levine et al., 2013; Zhang et al., 2012). DENN domain proteins, of which there are 17 in humans, act as guanine nucleotide exchange factors (GEFs) for Rab GTPases (Marat and McPherson, 2010; Zhang et al., 2012). GEFs allow the transfer of Rab-bound GDP for GTP at membrane specific locations, switching the Rab proteins from an inactive to an active state (Cherfils and Zeghouf, 2013). Once activated, Rab proteins interact with effector proteins that facilitate Rab-mediated endosomal trafficking and recycling of vesicle incorporated membrane proteins through the endosomal pathway (Stenmark, 2009). Through its shared homology with DENN domain proteins, it was hypothesised C9orf72 would function in similar pathways.

As mentioned, two C9orf72 protein isoforms are generated by alternative splicing of the C9orf72 gene. The C9orf72 long protein is a 481 amino acid protein encoded by exons 2 to 11, whereas the C9orf72 short protein is a 222 amino acid protein encoded by exons 2 to 5. From its structural similarities with DENN proteins, the C9orf72 protein is predicted to comprise three domains: an N-terminal longin domain, a central DENN domain and a C-terminal dDENN domain (Figure 1.4). The C9orf72 short protein terminates shortly into the DENN domain.

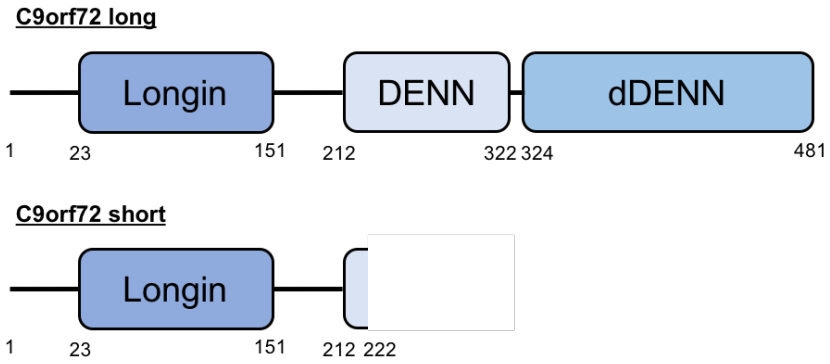


Figure 1.4. C9orf72 protein isoforms. The C9orf72 long protein is a 481 amino acid protein that contains a longin domain (23-151aa), a DENN domain (212-322aa) and a dDENN domain (324-481aa). The C9orf72 short protein is a 222 amino acid protein that contains the longin domain (23-151aa) and the first 10 amino acids of the DENN domain.

In 2014, one group reported that C9orf72 co-localised and interacted with autophagy- and endosomal trafficking- related proteins such as Ubiquilin-2 and Rabs 1, 5, 7 and 11 in neuronal cell lines, and loss of C9orf72 influenced autophagosome formation (Farg et al., 2014). The C9orf72 protein has also been shown to interact with ULK1, FIP200 and ATG13, which are involved in forming the autophagy initiation complex, suggesting C9orf72 has an early role in the autophagy pathway (Webster et al., 2016). The same study demonstrated C9orf72 influenced the trafficking of the initiation complex to the autophagophore by behaving as an effector protein for Rab1a, and iNeurons from C9orf72 patients demonstrated impairments in the autophagy pathway and p62 accumulation which could be rescued with exogenous C9orf72 expression (Webster et al., 2016). The C9orf72 protein has also been shown to interact with autophagy-related proteins SMCR8 and WDR41, and via these interactions behaves as a GEF for Rab8a and Rab39b to regulate autophagy (Sellier et al., 2016; Sullivan et al., 2016). Collectively, these

results suggest C9orf72 is a regulator of autophagy and haploinsufficiency of the C9orf72 protein in ALS could be causal of the p62 pathology.

C9orf72-regulation of trafficking may also be important in other pathways. C9orf72 has been reported to interact with and colocalise with Rab7L1 at the neuronal golgi apparatus, and depletion of C9orf72 decreased secretion of extracellular vesicles (Aoki et al., 2017). Therefore, C9orf72 may also be required for trans-golgi and extracellular vesicle trafficking.

1.6.1.3.2 C9orf72 loss-of-function animal models

There are mixed results from C9orf72 loss of function animal models regarding the contribution of C9orf72 haploinsufficiency to ALS-like neurodegeneration and motor defects. Transient knockdown of C9orf72 has been successfully completed in zebrafish using morpholino antisense oligonucleotides, leading to a number of pathological defects (Ciura et al., 2013). The knockdown had a negative effect on locomotor activity in zebrafish embryos, all of which were rescued to normal levels after reintroduction of C9orf72 mRNA, suggesting loss of the C9orf72 protein is sufficient to cause a motor phenotype. In support of this observation, loss of the *C. elegans* ortholog of C9orf72 led to worms showing increased neurodegeneration of GABAergic neurons and onset of an age-dependent motor phenotype (Therrien et al., 2013). Knockout of C9orf72 in mice can lead to changes in p62 levels, consistent with a role in autophagy, and caused mice to develop an age-dependent lethality phenotype, suggesting loss of C9orf72 can have toxic consequences (Ji et al., 2017).

In contrast, the majority of C9orf72 loss-of-function animal models exhibit lesser effects and it has been argued the motor defect observed in zebrafish may be an artefact of morpholino off-target effects (Gerety and Wilkinson, 2011). Transient knockdown of the C9orf72 mouse orthologue using antisense oligonucleotides (ASOs) targeting C9orf72 transcripts appeared to have no effect on mice. C9orf72 RNA levels were reduced to 30-40% after an intracerebroventricular injection of ASOs, with no changes in the mice seen behaviourally or neuropathologically up to 17 weeks post injection (Lagier-Tourenne et al., 2013). Since, four C9orf72 knockout mouse models have been developed. One of the studies reported that C9orf72 null mice present with a subtle motor defect and weakness of the hind-limbs (Atanasio et al., 2016), however no change was observed in motor ability,

cognitive behaviour or motor neuron degeneration in the other three studies (Jiang et al., 2016; Koppers et al., 2015; O'Rourke et al., 2016). However, whilst these studies suggest C9orf72 haploinsufficiency is not causal of ALS-like neurodegeneration, loss of the C9orf72 protein interestingly was observed to induce an immune response in mice (Atanasio et al., 2016; Jiang et al., 2016; O'Rourke et al., 2016). C9orf72 knockout mice exhibited signs of neuroinflammation including enlarged spleens, macrophage accumulation in lymph nodes, T-cell activation, autoantibody production and plasma cell activation (Atanasio et al., 2016; Jiang et al., 2016; O'Rourke et al., 2016). These studies suggest haploinsufficiency of C9orf72 may only be one of a number of converging disease mechanisms that together cause ALS and introduces a potential new role for C9orf72 in neuroinflammation. Indeed, the autophagy pathway which C9orf72 functions in is linked with the immune response. First, autophagosomes are key modes of disposal for invading microbes and pathogens in cells and is activated by immune signalling (Gutierrez et al., 2004; Nakagawa et al., 2004). Second, the autophagy pathway can influence the inflammatory response. Various autophagy protein complexes have been demonstrated to suppress pro-inflammatory signals in cells by inhibiting inflammasome activation, negative regulation of interferon signalling and suppression of pro-inflammatory IL-1α and IL-1β secretion (Castillo et al., 2012; Jounai et al., 2007; Nakahira et al., 2011; Shi et al., 2012). Disruption to the autophagy pathway through loss of the C9orf72 protein may therefore explain the hyper-immune response observed in C9orf72 null mice (Atanasio et al., 2016; O'Rourke et al., 2016).

1.7 C9orf72 in the nucleus

Investigation of the C9orf72 protein has found its distribution to be cell-wide, including some nuclear localisation (Farg et al., 2014). Immunofluorescence of the endogenous C9orf72 protein using isoform-specific antibodies suggests the C9orf72 short protein localises to the nuclear membrane whereas the C9orf72 long protein is more diffusively cytoplasmic (Xiao et al., 2015). Interactions between the C9orf72 short protein with Ran-GTPase and Importin B1 were also reported, suggesting the nuclear localisation may be linked with a role in nucleocytoplasmic shuttling (Xiao et al., 2015). Moreover, a combination of published and in house proteomic studies has identified numerous interactions between C9orf72 and nuclear proteins (Sellier et al., 2016; Sullivan et al., 2016). Interestingly, some of these proteins are involved

in RNA processing and are ALS-linked genes themselves, such as senataxin, TDP-43 and FUS (Sellier et al., 2016).

The literature therefore suggests there may be a not yet determined nuclear function of the C9orf72 protein, with evidence suggesting nucleocytoplasmic transport and RNA processing, both of which are pathways implicated in C9orf72-ALS. To this end, the work completed towards this project focused on determining a nuclear function for the C9orf72 protein and investigated whether there could be a connection between haploinsufficiency of nuclear C9orf72 and RNA misprocessing in ALS.

1.7.1 Dysfunctional RNA processing in ALS

Disruption to the RNA metabolism pathway appears to be shared between sporadic ALS and many of the genetic varieties of familial ALS, indicating that underlying problems to RNA processing may be causal of ALS pathogenicity. Motor neurons specifically have a high reliance on efficient transcription and splicing to maintain their ability to send and receive electrical signals. Any disruptions to RNA processing therefore may have a more detrimental impact on neurons in comparison to other somatic cells.

The ALS-linked TDP-43 and FUS proteins both have well known roles in RNA processing. TDP-43 can directly behave as a *trans*-acting splicing regulator of some genes, including the CFTR and MAPT genes, and can also regulate transcription via interactions with DNA (Buratti and Baralle, 2001; Gu et al., 2017). Mislocalisation of TDP-43 to the cytoplasm in ALS patient derived fibroblasts containing a mutation in *TARDBP* has been shown to lead to changes in alternative splicing events (Highley et al., 2014). Furthermore, alternative splicing of mRNA transcripts known to be regulated by TDP-43 were found to be abhorrently spliced in ALS spinal cord (Xiao et al., 2011). Taken together, these studies suggest TDP-43 pathology is one of the contributing factors to the observed defects in RNA processing in ALS. FUS likewise has roles in splicing regulation. FUS directly interacts with pre-mRNA transcripts, the U1 snRNP and the SMN complex to regulate removal of introns (Yamazaki et al., 2012; Yu et al., 2015; Zhou et al., 2013). Depletion of the FUS protein can cause alternative splicing events on transcripts encoding other RNA-binding proteins, potentially exacerbating the disruption to RNA metabolism (Sun et al., 2015). ALS-linked mutant FUS can also

sequester both major and minor snRNPs into cytoplasmic aggregates, thereby decreasing their availability for splicing catalysis and potentially downregulating global splicing (Reber et al., 2016; Yu et al., 2015).

Dysfunctional splicing is also prevalent in C9orf72-ALS, where splicing defects have been shown to correlate disease severity (Cooper-Knock et al., 2015a). RNA-sequencing of cerebellum and frontal cortex tissue from C9orf72-ALS patient brains detected hundreds of mis-splicing events (Prudencio et al., 2015). Interestingly, alternative splicing was also misregulated in brains of sporadic ALS patients, although to a much lower extent than C9orf72 patients. In both C9orf72- and sporadic-ALS patients, the alternative splicing events were approximately 8-fold higher in cerebellum versus frontal cortex. The most common mis-splicing event was the skipping of exons, followed by intron retention (Prudencio et al., 2015). This study confirms the extensive dysregulation to splicing in C9orf72-ALS.

The disruption to splicing in C9orf72-ALS is likely through convergence of several mechanisms. As previously discussed, the RNA foci formed from RNA containing the GGGGCC repeat expansion can sequester RNA binding proteins and splicing factor proteins. Indeed, splicing factors are one of the main regulatory bodies for signalling for exon inclusion/removal during alternative splicing, therefore their sequestration could lead to the exon skipping events in C9orf72-ALS. Interestingly, the arginine DPRs have been shown to bind the U2 snRNP and sequester it in the cytoplasm in C9orf72-ALS motor neurons, and can also bind the hnRNP A2 splicing factor to influence alternative splicing events (Kwon et al., 2014; Yin et al., 2017). The nuclear function of the C9orf72 protein is yet to be determined, but there are suggestions it interacts with RNA-binding proteins and may also function in nucleocytoplasmic transport (Blokhuis et al., 2016; Sellier et al., 2016; Xiao et al., 2015). Hence, the three proposed mechanisms of disease for C9orf72-ALS all may contribute to ALS pathogenicity through their interaction with, and inhibition of, RNA processing pathways.

1.7.2 Overview of RNA splicing

Introns are non-coding sequences interspaced between coding exons within a gene. Transcription of genes produces an intron-containing pre-mRNA, which undergoes spicing in the nucleus by the spliceosome, before export to the cytoplasm for translation into protein (Berget et al., 1977; Chow et al., 1977). For

some genes, all exons are constitutively expressed and only the intervening introns are removed, generating a single mRNA transcript. Alternatively, certain exons may be removed alongside introns, termed alternative splicing. Alternative splicing can therefore generate multiple mRNA transcripts, encoding different protein isoforms, from a single gene (Black, 2003).

The spliceosome can be classed as a large multimolecular complex, which is composed of five U snRNPs and associated splicing factor proteins (Will and Luhrmann, 2011). Introns contain specific sequences which recruit the spliceosome and signal for intron removal. These are the 5' and 3' splice sites (ss) which are located at the exon-intron junctions and the branch point sequence which is generally located 18-40 nucleotides upstream of the 3' ss (Burge et al., 1999). The snRNPs and splicing factors work together in a complex, step-wise pathway to catalyse the removal of introns, mediated by the RNA-RNA interaction between snRNPs and pre-mRNA. Interestingly, there are two different spliceosomes: the major spliceosome and the minor spliceosome. They differ in both their composition of U snRNPs and in the introns they remove.

1.7.2.1 Major splicing vs minor splicing

The major spliceosome is composed from the U1, U2, U4, U5 and U6 snRNPs. As discussed later, the U4, U5 and U6 snRNPs function in the splicing pathway as a tri-snRNP, which assembles in nuclear suborganelles called Cajal bodies. The major spliceosome catalyses the removal of major introns, otherwise called U2-introns, which make up more than 99% of the introns in the human genome. U2-introns contain dinucleotides GT and AG at their 5' ss and 3' ss respectively, a degenerate branch point sequence (BPS) that can be 18-40 nucleotides upstream of the 3' ss, and a polypyrimidine tract in between the BPS and 3' ss (Gao et al., 2008).

In contrast, the minor spliceosome is composed from the U11, U12, U4atac, U5 and U6atac snRNPs (Turunen et al., 2013). Thus, the U5 snRNP is shared between both the major and minor spliceosomes. The U11, U12, U4atac and U6atac snRNPs share high levels of structural similarity with their U1, U2, U4 and U6 major spliceosome snRNP counterparts, but differences in their sequences allows the minor snRNPs to recognise a different subset of introns.

The minor spliceosome removes a rare type of intron, called the minor intron or U12-type intron. U12-introns have been discovered in approximately 800 genes and make up only 0.5% of the introns in the human genome (Turunen et al., 2013). Genes typically contain only a single U12-intron, but can contain more. For example, the U12 intron database shows the spermine synthase gene contains three U12-introns. Interestingly, U12-introns are enriched in genes that have roles in voltage gated channels or in RNA processing, transcription and DNA replication, suggesting minor splicing may have some regulatory control over certain molecular pathways (Burge et al., 1998).

When U12-introns were originally discovered, they were termed ATAC-introns, due to the observation that they contained AT and AC dinucleotides at their 5' ss and 3' ss respectively, and a particularly high level of conservation at their 5' ss, unalike to U2-introns (Jackson, 1991). However, it was later discovered the majority of U12-introns also have GT-AG splice sites (Dietrich et al., 1997; Wu and Krainer, 1997). U12-introns are now defined by a highly conserved sequence downstream of the 5' ss, their unique branch point sequence, which is optimally located approximately 15 nucleotides upstream of the 3' ss, and their lack of a polypyrimidine tract (Dietrich et al., 2005; Hall and Padgett, 1996).

Another of the key differences between the major and minor spliceosomes is their efficiency at intron removal. In a *Drosophila* cell line, splicing of a reporter U2-intron was ten-fold higher than the splicing of an equivalent U12-intron (Patel et al., 2002). Within the same gene, splicing of a U12-intron was significantly slower than splicing of the adjacent U2-intron (Patel et al., 2002). Splicing occurs co-transcriptionally to ensure efficient generation of mature mRNA and protein expression. However, genes containing U12-introns generate pre-mRNA transcripts in which the U2-introns are removed but U12-introns remain, and which can therefore regulate protein levels post-transcriptionally. It has been proposed that the slower intron removal by the minor spliceosome is a consequence of the low abundance of minor snRNPs. All four of the minor-spliceosome specific snRNPs are approximately 100-fold fewer in number in comparison to the major spliceosome snRNPs (Montzka and Steitz, 1988; Tarn and Steitz, 1996). In support of this hypothesis, splicing of an exogenous reporter U12-intron was slower in a cell line with low abundance of minor snRNPs in comparison to its splicing in a cell line with high abundance of minor snRNPs (Pessa et al., 2006). However, in the same study, splicing of endogenous U12-introns in the different cell lines was unchanged, suggesting the

snRNP number does not influence the rate of endogenous minor splicing (Pessa et al., 2006). Although, the low abundance of the snRNPs can be a contributing factor under conditions of high splicing demand.

1.8 Cajal bodies and the spliceosome

In this thesis, the first steps taken towards investigating the nuclear function of C9orf72 and splicing involved characterising an interaction between the C9orf72 protein and nuclear protein coilin. Coilin is the major protein component and structural scaffold for nuclear suborganelles called Cajal bodies.

1.8.1 Cajal bodies

Cajal bodies (CBs) were first discovered in 1903 by Santiago Ramón y Cajal and were originally named nucleolar accessory bodies (Cajal, 1903). The Cajal body is a spherical, membraneless nuclear suborganelle with a diameter of approximately 0.1-0.5 μm (Cajal, 1903; Hardin et al., 1969).

The Cajal body is enriched in proteins and RNA responsible for snRNP processing and other functions (see Section 1.8.2). The two major proteins that compose Cajal bodies are coilin and SMN. Canonical Cajal bodies refer to Cajal bodies that are both coilin- and SMN-positive and are functionally involved in snRNP processing. However, non-canonical Cajal bodies lacking coilin or the SMN protein can exist, termed residual and immature Cajal bodies respectively, but these are unable to correctly recruit snRNPs (Lafarga et al., 2017; Sleeman et al., 2001; Tucker et al., 2001).

How Cajal bodies form into membraneless suborganelles from their nucleoplasmic components is not yet clear. However, evidence supports an RNA nucleation event, where other Cajal body proteins and RNA assemble in an apparently random manner (Kaiser et al., 2008; Mao et al., 2011a; Shevtsov and Dundr, 2011). Once formed, the weak interactions between the Cajal body proteins, many of which are RNA-binding proteins with low complexity domains, are thought to be important to maintaining the membraneless suborganelle in a ‘hydrogel-like’ phase within the fluid nucleoplasmic environment (Kato et al., 2012). Cajal bodies have a porous structure, where proteins may passively diffuse between Cajal bodies and the

surrounding nucleoplasm, with snRNPs shown to spend as little as 30 seconds within the Cajal body (Deryusheva and Gall, 2004; Dundr et al., 2004).

1.8.2 Cajal body function

Cajal bodies are best known for their role in snRNP processing. Many of the proteins and RNA concentrated in Cajal bodies have roles in the modification, recruitment or nucleocytoplasmic shuttling of snRNPs (Table 1.2). However, Cajal bodies are also involved in the modification of small nucleolar RNAs (snoRNA), telomere maintenance and regulating gene expression of small, non-coding RNAs, which will be briefly covered below.

Table 1.2. Cajal body proteins and RNA involved in snRNP processing

PROTEIN		
Protein Name	**Function**	**Reference**
Coilin	Cajal body scaffold	(Andrade et al., 1991)
SMN	SMN complex	(Hebert et al., 2001)
Gemins	SMN complex	(Hao et al., 2007)
Sm proteins	SMN complex	(Smoliński et al., 2011)
SART3	snRNP recycling	(Staněk et al., 2003)
TGS1	snRNA m^7G cap methylation	(Mouaikel et al., 2003)
PHAX	snRNA nuclear export	(Boulon et al., 2004)
CRM1	snRNA nuclear export	(Boulon et al., 2004)
WRAP53	SMN complex recruitment to Cajal bodies	(Mahmoudi et al., 2010)
Prp3/Prp4	snRNP recycling	(Schaffert et al., 2004)
Snurportin1	snRNP nuclear import	(Ospina et al., 2005)
SF3a60/SF3a66	snRNP nuclear import	(Nesic et al., 2004)
U2AF	Splicing factor	(Carmo-Fonseca et al., 1992)
RNA		
RNA Name	**Family**	**Reference**
U1	snRNA	(Carmo-Fonseca et al., 1992)
U2	snRNA	(Carmo-Fonseca et al., 1992)
U4	snRNA	(Carmo-Fonseca et al., 1992)
U5	snRNA	(Carmo-Fonseca et al., 1992)
U6	snRNA	(Carmo-Fonseca et al., 1992)
U11	snRNA	(Matera and Ward, 1993)

U12	snRNA	(Matera and Ward, 1993)
U4atac	snRNA	(Staněk et al., 2003)
U6atac	snRNA	(Staněk et al., 2003)
SCARNA10	scaRNA	(Jady and Kiss, 2001)
SCARNA5	scaRNA	(Darzacq et al., 2002)
SCARNA6	scaRNA	(Darzacq et al., 2002)
SCARNA12	scaRNA	(Darzacq et al., 2002)
SCARNA7	scaRNA	(Darzacq et al., 2002)
SCARNA17	scaRNA	(Darzacq et al., 2002)
SCARNA8	scaRNA	(Darzacq et al., 2002)
SCARNA13	scaRNA	(Kiss et al., 2002)
SCARNA4	scaRNA	(Kiss et al., 2004)
SCARNA1	scaRNA	(Kiss et al., 2004)

1.8.2.1 snRNP maturation

Splicing is performed by the spliceosome, which refers to five U snRNPs and their associated splicing factor proteins (Will and Luhrmann, 2011). The U snRNPs are RNA-protein complexes which contain a U snRNA which catalyses the removal of introns from pre-mRNA. This section will focus on the maturation pathway from early U snRNA transcription to fully mature spliceosomal U snRNP, and the relationship with Cajal bodies in that pathway (Figure 1.5). It is worth noting, the snRNP processing reactions can occur throughout the whole of the nucleoplasm, but by enriching the proteins and RNA in specific sites in the nucleus within Cajal bodies, the efficiency of snRNP assembly and splicing output of the cell is increased (Klingauf et al., 2006).

There are 9 U snRNAs that are involved in splicing: U1, U11, U2, U12, U4, U4atac, U5, U6 and U6atac. Similar to mRNA, the U snRNA genes are transcribed by polymerase II, except for U6 and U6atac which are transcribed from a polymerase III promoter and undergo a separate maturation pathway (Hernandez, 2001; Kunkel et al., 1986; Schramm and Hernandez, 2002). After transcription, the snRNAs are m^7G capped on their 5' end, and the 3' end undergoes snRNA specific cleavage by the integrator complex (Baillat et al., 2005). Interestingly, there is also evidence suggesting a role for the Cajal body protein coilin in the 3' end processing of nascent U snRNAs, and this might be enabled through the association of Cajal

bodies with snRNA gene loci discussed in Section 1.8.2.3 (Broome and Hebert, 2012).

snRNA transcripts are bound by the cap-binding complex, phosphorylated adaptor RNA export protein (PHAX) and arsenite resistance 2 (ARS2) which bridge an interaction between the snRNA cap and the chromosome region maintenance 1 (CRM1) protein (Fornerod et al., 1997; Hallais et al., 2013; Izaurralde et al., 1994; Ohno et al., 2000). There is evidence to suggest the export complex assembly onto snRNAs takes place in Cajal bodies. Firstly, immature snRNAs and the export complex proteins are detected in Cajal bodies, and secondly, inhibition of PHAX nuclear export causes snRNAs to accumulate in Cajal bodies (Boulon et al., 2004; Smith and Lawrence, 2000; Suzuki et al., 2010).

In the nucleus, the CRM1 protein mediates an interaction between the snRNA-export complex with the GTP-bound Ran-GTPase and facilitates nuclear export via the nuclear pore complex (Fornerod et al., 1997; Ohno et al., 2000). Once in the cytoplasm, the simultaneous GTP hydrolysis by Ran-GTPase and the dephosphorylation of PHAX releases the snRNA from the export complex (Kitao et al., 2008; Ohno et al., 2000).

The cytoplasmic snRNA is next bound by the SMN complex, a large multiprotein complex composed of the SMN protein and associated Gemins 2-7 (Massenet et al., 2002; Pellizzoni et al., 2002). The SMN complex recruits seven Sm proteins, F, E, G, D1, D2, B and D3, and acts as a platform for the Sm proteins to assemble into a heteroheptameric ring around the snRNA in an ATP-dependent manner, forming the core of the snRNP (Meister et al., 2001; Raker et al., 1996).

After assembly of the Sm protein core on the snRNP, the SMN complex recruits the trimethylguanosine synthase 1 (TGS1) protein which hypermethylates the snRNA m^7G cap to a 2,2,7-trimethylguanosine (m_3G) cap (Mouaikel et al., 2003). Together, the snRNP core and m_3G cap behave as nuclear localisation signals and target the snRNP for nuclear import (Fischer and Lührmann, 1990; Fischer et al., 1993). The nuclear import complex, comprising snurportin 1 and importin-β, assembles on the m_3G cap and facilitates nuclear import of the snRNPs (Huber et al., 1998; Palacios et al., 1997).

Once back in the nucleus, newly imported snRNPs return to Cajal bodies for further modifications (Ospina et al., 2005; Sleeman and Lamond, 1999). An interaction between the SMN protein and coilin is thought to mediate recruitment of snRNPs into Cajal bodies (Hebert et al., 2001). In Cajal bodies, the snRNPs undergo several final adjustments to become functional for pre-mRNA splicing. First, Cajal bodies are enriched in small Cajal body RNAs (scaRNAs), which catalyse the 2'-*O*-ribose-methylation and pseudouridylation of the snRNP, which increases the stability and protection of the snRNA (Jady et al., 2003). Secondly, U snRNP-specific proteins are assembled onto the complex (Nesic et al., 2004). Lastly, the mature U4, U5 and U6 snRNPs organise themselves into a U4-U6 di-snRNP and U4-U5•U6 tri-snRNP to become fully functional for pre-mRNA splicing (Novotny et al., 2011; Staněk et al., 2003; Staněk and Neugebauer, 2004). The snRNPs can then diffuse through the nucleoplasm and enter the splicing pathway (Sleeman and Lamond, 1999; Spector and Lamond, 2011).

As previously mentioned, the U6 and U6atac undergo a maturation pathway separate from the other U snRNPs in that they are transcribed by polymerase III and their biogenesis pathway takes place exclusively in the nucleus (Kunkel et al., 1986; Mroczek and Dziembowski, 2013). The 5' cap on the U6 and U6atac is a non-nucleotide gamma-monomethyl phosphate and the 3' end has a 2',3'-cyclic phosphate added (Lund and Dahlberg, 1992; Singh and Reddy, 1989). Unlike the other U snRNAs, the heteroheptameric ring on the U6 and U6atac snRNA is assembled from Like-Sm (LSm) proteins 2-8, called the LSm complex (Achsel et al., 1999; Khusial et al., 2005). The LSm complex is assembled separately in the cytoplasm and is imported into the nucleus where it associates with the 3' end of the U6/U6atac snRNA (Licht et al., 2008; Zaric et al., 2005). Final U6 snRNA modifications take place in the nucleolus, where small nucleolar RNAs (snoRNAs) guide 2'-*O*-ribose-methylation and pseudouridylation reactions (Ganot et al., 1999). The mature U6 snRNP is then recruited to Cajal bodies by the SART3 protein, where, as mentioned, it forms di-snRNPs and tri-snRNPs with the U4 and U5 snRNPs (Staněk et al., 2003; Staněk and Neugebauer, 2004).

1.8.2.1.1 U4-U5•U6 tri-snRNP assembly

The mature U4, U5 and U6 snRNPs are recruited into the splicing pathway as a U4-U5•U6 tri-snRNP. The localisation of the U4-U6 di-snRNP and U4-U5•U6 tri-snRNP within Cajal bodies has been widely reported, suggesting they are the location of

assembly (Schaffert et al., 2004; Staněk and Neugebauer, 2004). To begin, the U4 and U6 snRNPs form a di-snRNP. The U4 and U6 snRNA form RNA-RNA base pairings, mediated by the SART3 protein and LSm proteins (Achsel et al., 1999; Bell et al., 2002; Raghunathan and Guthrie, 1998). For U5 snRNP integration, the U4-U6 di-snRNP protein Prp31 interacts with the U5 snRNP protein Prp6, as well as the association of several new proteins (Liu et al., 2006; Makarova et al., 2002). If tri-snRNP assembly fails, the SART3 protein anchors immature snRNPs into Cajal bodies through an interaction with coilin, suggesting Cajal bodies also act as a quality control step, only allowing fully functional snRNPs to enter the splicing pathway (Novotný et al., 2015). After assembly, the U4-U5•U6 tri-snRNP enters the splicing pathway, where the snRNP interactions are broken. The individual snRNPs are then recycled back into Cajal bodies for reassembly.

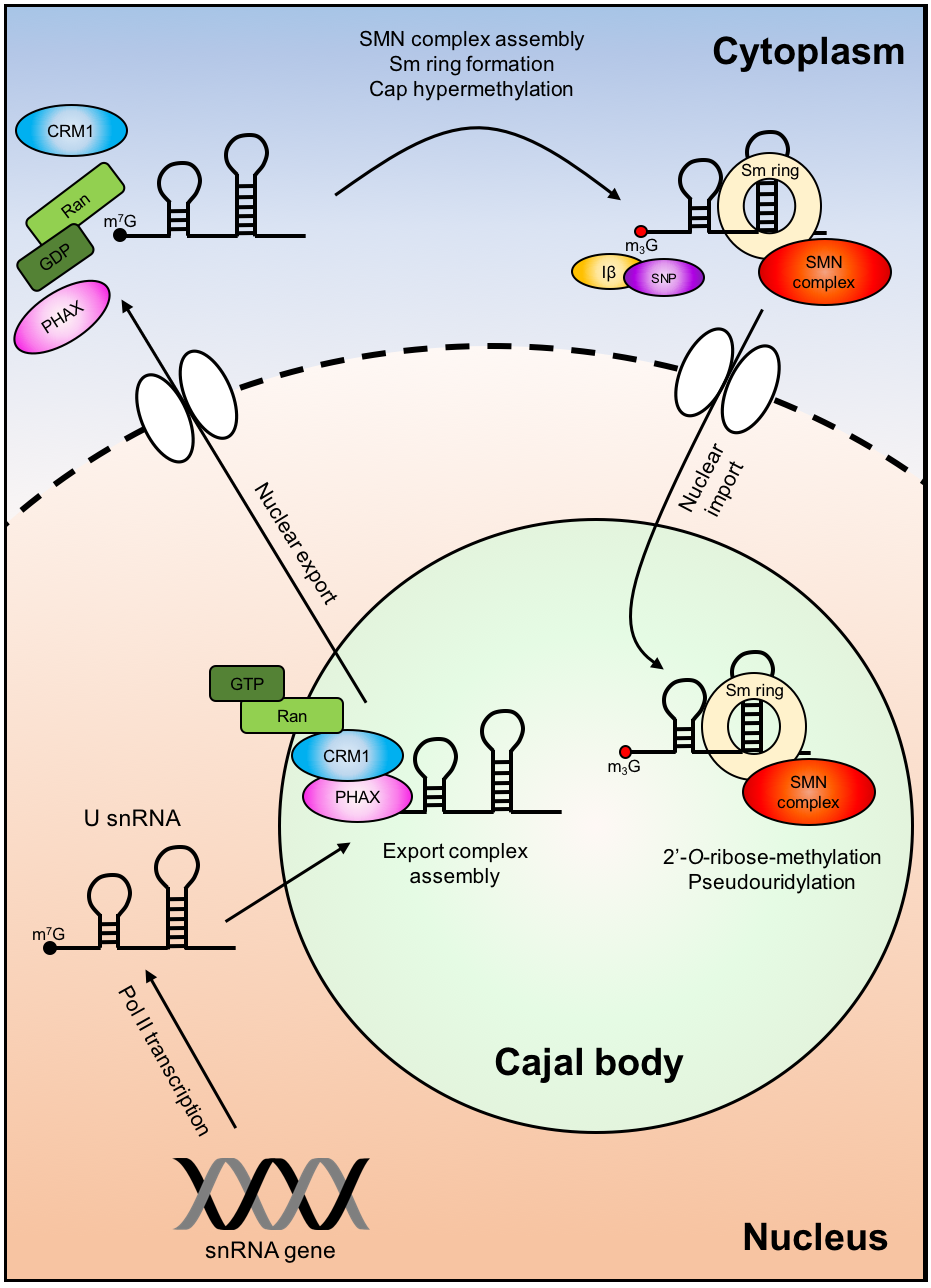


Figure 1.5. Maturation pathway of Sm-like snRNPs. The U snRNA is transcribed by Polymerase II and recruited to Cajal bodies where the export complex, composing PHAX, CRM1 and Ran-GTPase, assembles. The U snRNA is exported to the cytoplasm where the export complex disassociates. The SMN complex binds the snRNA, and mediates the formation of the Sm ring and cap hypermethylation from a m^7G cap to a m_3G cap. The m_3G cap and snRNP core act as nuclear localisation signals, and import proteins importin-β (Iβ) and snurportin 1 (SNP) bind and mediate import into the nucleus. snRNPs then localise to Cajal bodies where scaRNA catalyses the 2'-*O*-ribose methylation and pseudouridylation of the snRNAs. The snRNPs then exit Cajal bodies to join the spliceosome.

1.8.2.2 snoRNA and scaRNA modifications

In addition to snRNA, Cajal bodies are also home to another two families of RNA: scaRNA and snoRNA. scaRNA and snoRNA share similar structures, both can contain box H/ACA and/or box C/D motifs, and functionally they both mediate the 2'-*O*-ribose-methylation and pseudouridylation modifications of other RNA. However, scaRNA mediates the modification of snRNPs in Cajal bodies, whereas snoRNA mediates the modification of ribosomal RNA in the nucleolus (Darzacq et al., 2002; Decatur and Fournier, 2003; Jady et al., 2003).

Both scaRNA and snoRNA initially localise at Cajal bodies and are thought to be recruited by WRAP53 and PHAX, respectively (Boulon et al., 2004; Tycowski et al., 2009). The scaRNAs, which contain a Cajal body (CAB) box motif, are retained in Cajal bodies whereas the snoRNA, which lack a CAB box, are quickly trafficked to the nucleolus (Narayanan et al., 1999a, 1999b; Richard et al., 2003). The exact modifications to scaRNA and snoRNA that occur in Cajal bodies are unknown, but there is evidence to suggest they undergo cap hypermethylation and binding of RNA-specific proteins (Pradet-Balade et al., 2011; Verheggen et al., 2002).

1.8.2.3 Chromosome organisation and gene expression

Cajal bodies are not static structures, but migrate within the nucleoplasm where they have been shown to associate with other nuclear bodies and chromatin (Platani et al., 2000, 2002). Interestingly, Cajal bodies have been shown to associate with the genomic loci of several genes, including the U1 and U2 snRNAs and histone gene clusters (Frey et al., 1999; Frey and Matera, 1995; Smith et al., 1995). Recent evidence suggests the Cajal body can act as a 'hub' for chromosomal organisation, in which the gene loci of highly expressed snRNAs, scaRNAs, and snoRNAs form chromosomal clusters (Wang et al., 2016). Disruption of Cajal bodies led to a loss of these gene clusters and a decrease in U snRNA expression and snRNP production, suggesting Cajal bodies may influence splicing through snRNA expression in addition to snRNP processing (Wang et al., 2016).

1.8.2.4 Telomere maintenance

Cajal bodies have also been implicated in telomere maintenance. Telomeres are repeated DNA sequences at the ends of linear chromosomes and which shorten with each round of cell division (Smogorzewska and de Lange, 2004). After the shortening of telomeres to a specific length, cells signal to enter a state of senescence, or undergo programmed apoptosis (McEachern et al., 2000). To counter telomere shortening, they are maintained and synthesised by the telomerase enzyme, which is composed of a telomerase RNA and the telomerase reverse transcriptase (TERT) (Greider and Blackburn, 1989; Mitchell et al., 1999). Both the telomerase RNA and TERT have been shown to localise in Cajal bodies, which themselves can directly associate with telomeres as they're being replicated (Jady et al., 2005; Tomlinson et al., 2006; Zhu et al., 2004). In addition, one of the protein components of Cajal bodies, telomerase Cajal body protein 1 (TCAB1), recruits active telomerase to Cajal bodies and its depletion can inhibit telomere synthesis suggesting Cajal bodies may be sites of telomerase assembly (Venteicher et al., 2009). A second proposed role is that Cajal bodies are involved in the maturation of the telomerase RNA. The telomerase RNA is part of the scaRNA family, and holds similar physical properties to other scaRNAs that target it for Cajal body localisation and modification, such as the H/ACA box and CAB box motifs (Jady et al., 2004; Theimer et al., 2007).

1.8.3 Cajal body dynamics

The appearance and number of Cajal bodies varies between cell and tissue types. In neurons, the neuronal size can influence the number of Cajal bodies, but typically 1-3 Cajal bodies per nucleus are present (Cajal, 1903; Pena et al., 2001). On the other hand, cell types from various tissues, for example spleen, stomach and smooth muscle cells, lack Cajal bodies altogether (Young et al., 2000).

The number of Cajal bodies in the nucleus appears to be regulated, and can be influenced by several factors. Cajal bodies are especially prominent in cells with a high metabolic or transcription demand, which may be related to an increased need for snRNP production (Pena et al., 2001; Spector et al., 1992). Inhibition of transcription leads to disassembly of Cajal bodies and their restructuring into nucleolar caps (Shav-Tal et al., 2005), suggesting active transcription is required to

maintain canonical Cajal bodies. Cellular levels of snRNP can also influence the number of Cajal bodies; an increase in snRNPs causes Cajal bodies to increase (Sleeman et al., 2001). The effect of transcription and snRNPs on Cajal bodies are likely linked, a high transcription demand requires more snRNPs for RNA processing. In addition, within dividing cells, Cajal body numbers change in response to the cell cycle; Cajal bodies disassemble during mitosis and reassemble in daughter cells during mid-G_1 (Andrade et al., 1993; Carmo-Fonseca et al., 1993). These changes could be from the inhibition and restarting of transcription and splicing as the cell progresses through cell division, or there is also evidence to suggest Cajal bodies are regulated in the cell-cycle by post-translational modification of the coilin protein, which is discussed later. Cajal body number may also be important developmentally, as numbers of Cajal bodies are high during early development and decrease post-differentiation (Strzelecka et al., 2010). The response of Cajal bodies to factors such as serum starvation and viral infection have also led to suggested roles in pathways such as the stress response (Carvalho et al., 1999; Platani et al., 2000). Finally, overexpression or depletion of Cajal body components can also regulate the number of Cajal bodies. For example, overexpression of the SmB protein leads to an increase in Cajal bodies, whereas SMN depletion leads to a decrease in Cajal bodies (Lemm et al., 2006; Sleeman et al., 2001).

1.8.4 Cajal bodies in disease

Defects in Cajal bodies have been linked with a number of diseases, mainly through their association with the spliceosomal snRNPs. Cajal body dysregulation can influence the splicing ability of the cell and defective RNA processing has strong links with ALS (Cooper-Knock et al., 2015a; Whittom et al., 2008). Cajal bodies and splicing have been implicated in spinal muscular atrophy (SMA), a disease similar to ALS in which patients develop motor neuron degeneration as a result of SMN loss-of-function (Lefebvre et al., 1995). Depletion of the SMN protein leads to loss of canonical Cajal bodies (Lemm et al., 2006; Morse et al., 2007). Disruption to Cajal bodies is observed in the SMNΔ7 mouse model of SMA, involving relocalisation of Cajal body proteins, specifically coilin, to perinucleolar caps which disrupted ribosome biogenesis (Tapia et al., 2017). In addition, SMA patient-derived lymphoblast cells lack Cajal bodies and have reduced U4-U5•U6 tri-snRNP levels (Boulisfane et al., 2011). Hence, disruption to Cajal bodies caused by loss of the SMN protein may contribute to SMA pathogenicity.

Cajal bodies have also been linked with ALS. In the presence of mutant SOD1, Cajal bodies are decreased and lack SMN and snRNPs in neuronal cells (Kariya et al., 2012). In support of this, in a mutant human SOD1 transgenic mouse model, loss of nuclear bodies from spinal motor neurons correlates with age and ALS disease progression (Gertz et al., 2012). These studies suggest Cajal bodies may be dysfunctional in SOD1-related ALS. Moreover, ALS genes TDP-43 and FUS also have connections with Cajal bodies. TDP-43 forms intranuclear inclusions within cells and these inclusions strongly colocalise with Cajal bodies (Wang et al., 2002). Further, knockdown of TDP-43 can cause a marked decrease in the number of nuclear gem structures (Tsuiji et al., 2013) and both mutant TDP-43 and FUS expressing ALS patient fibroblasts contain lower than normal numbers of nuclear bodies, suggesting irregular Cajal body number and/or function is a theme that runs throughout a number of ALS models (Yamazaki et al., 2012).

Interestingly, Cajal bodies have also been implicated in CAG repeat diseases, such as Huntington's disease. A common characteristic of CAG repeat disorders is the formation of neuronal intranuclear inclusions (NIIs) in brains (DiFiglia et al., 1997; Paulson et al., 1997). These NIIs have been shown to closely associate with Cajal bodies in dentatorubral-pallidoluysian atrophy (DRPLA) transgenic mouse brains and it is hypothesised this may disrupt Cajal bodies and cause cell death (Yamada et al., 2001). Neurodegenerative diseases are complex disorders whose mechanisms of disease often overlap; therefore, it may be that defective Cajal bodies play a role in several diseases.

1.8.5 Coilin

At the beginning of this project, an unbiased yeast two-hybrid screen was performed with the aim of identifying novel nuclear protein interactors of the C9orf72 long protein. The Cajal body protein coilin was identified several times in the screen indicating a probable interaction between C9orf72 and coilin.

Coilin is an 80 kDa nuclear protein that is the major protein component and marker of Cajal bodies (Andrade et al., 1991; Bohmann et al., 1995). Located on chromosome 17, the coilin gene is approximately 25 kb and is comprised of 7 coding exons, which are constitutively expressed (Chan et al., 1994). Investigation of the coilin sequence has revealed two nuclear localisation signals (NLS), a single nucleolar localisation signal (NoLS) and an RG-box (Hebert and Matera, 2000).

However, the sequence does not disclose any other structural motifs from which function can be inferred and the crystal structure of coilin is yet to be determined. NMR methods have been utilised which suggest coilin is formed from three domains: an N-terminal domain, a central highly disordered domain which contains the nuclear and nucleolar localisation signals, and the C-terminal region which is thought to contain an uncharacteristic tudor domain adjacent to the RG-box (Figure 1.6) (Shanbhag et al., 2010). Although lacking a proper RNA-binding domain, the N-terminal self-association domain may contain an atypical RNA-recognition motif which allows it to interact with RNA (Bellini and Gall, 1998). The coilin RG-box is a motif of repeated arginine-glycine dipeptides which are often post-translationally symmetrically dimethylated on the arginine residues, and which commonly facilitates protein-protein interactions (Thandapani et al., 2013). The tudor domain is a barrel-like fold that preferentially binds symmetrically dimethylated arginines within RG-boxes (Sprangers et al., 2003). Interestingly, the Sm proteins and SMN protein also contain an RG-box and tudor domain, respectively. Symmetrical dimethylation of the RG-box on the SmB, SmD1 and SmD3 proteins is central to an interaction between the Sm protein RG-box and the SMN protein tudor domain (Brahms et al., 2000, 2001). Although sequence conservancy is low, the three domains in human coilin have also been described for plant coilin, suggesting the structure is shared between species (Makarov et al., 2013).

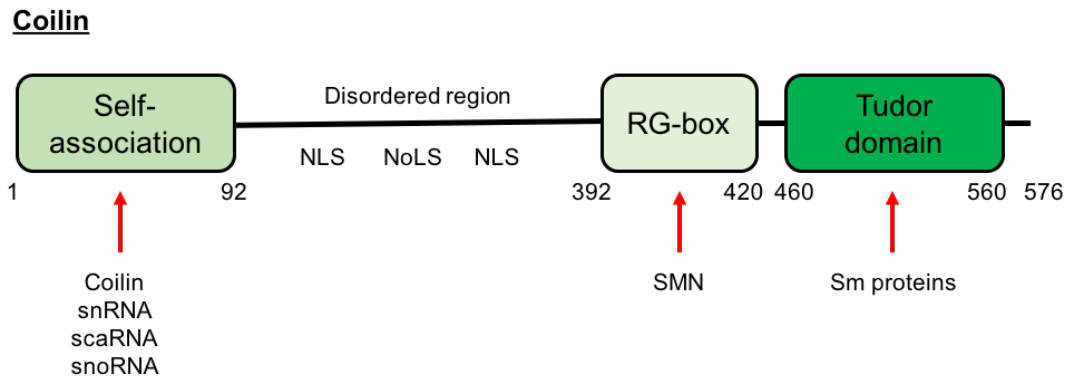


Figure 1.6. Coilin protein domains. The coilin protein contains an N-terminal domain for self-association (1-92aa), a central disordered region (92-392aa), an RG-box motif (392-420aa) and an atypical tudor domain (460-560aa). The red arrows indicate the domains with which coilin-interacting proteins or RNA associate with.

1.8.5.1 Coilin function

Coilin is the most commonly used marker for Cajal bodies. As is frequently seen for nuclear body proteins, coilin self-interacts via its conserved N-terminal domain (Hebert and Matera, 2000). This self-interacting domain, combined with the NLS, is essential for canonical Cajal body formation (Hebert and Matera, 2000). Coilin has also been shown to mediate many of the protein-protein and protein-RNA interactions in Cajal bodies, helping to scaffold the nuclear body. Studies have shown coilin is symmetrically dimethylated on the arginine residues in its RG-box which interacts with the tudor domain on the SMN protein to recruit snRNPs into Cajal bodies (Boisvert et al., 2002; Hebert et al., 2001; Tucker et al., 2001). In addition, coilin associates with Sm proteins and snRNPs via an interaction between the Sm-fold on Sm proteins and the tudor domain on coilin (Xu et al., 2005). Coilin has also been widely documented to interact with multiple snoRNA, scaRNA and telomerase RNA and may be directly involved in the processing of these RNAs within Cajal bodies (Enwerem et al., 2015; Machyna et al., 2014; Poole et al., 2016). The literature therefore suggests that the main function of coilin is to form the 'structural scaffold' of Cajal bodies and enrich and recruit the proteins required for snRNP assembly.

Indeed, coilin is essential for Cajal body formation. Numerous studies have investigated the essential nature of coilin through its knockdown and knockout in both cell and animal models. HeLa cells depleted of coilin have decreased cell proliferation and splicing efficiency but still show viability (Lemm et al., 2006; Whittom et al., 2008). Similarly, *Drosophila* coilin null mutants show full viability but lose the ability to form Cajal bodies (Liu et al., 2009). In zebrafish, coilin is essential for embryogenesis; coilin knockdown led to decreased cell proliferation and embryo lethality (Strzelecka et al., 2010). In addition, Cajal body integrity was lost (although some residual Cajal bodies remained) and the spliceosomal snRNPs dispersed from foci, leading to decreased snRNP assembly and defective splicing (Strzelecka et al., 2010). Interestingly, embryo viability was rescued with injections of mature snRNPs, suggesting coilin knockdown caused a downstream effect where snRNP assembly was reduced due to Cajal body loss, which in turn impacted the embryos ability to efficiently perform splicing (Strzelecka et al., 2010). Finally, in a mouse study, coilin was shown to be essential for canonical Cajal body formation. Coilin knockout mice lacked Cajal bodies containing SMN and snRNPs, but residual Cajal bodies containing fibrallarin did remain (Tucker et al., 2001). The homozygous coilin

knockout mutants additionally showed decreased viability (Tucker et al., 2001). From the loss-of-function coilin studies it appears coilin is essential for correct Cajal body formation and loss of Cajal bodies can affect viability in varying degrees of severity, from no effect in *Drosophila* to embryo lethality in zebrafish. The strength of the phenotype for Cajal body disruption may be dependent on the splicing demand of the organism, explaining the severe phenotype observed in zebrafish during embryo development.

Manipulating the coilin protein has been shown to influence splicing. Both expression levels and post-translational modifications of coilin have been shown to have an effect on the splicing efficiency of cells (Whittom et al., 2008). Coilin depletion caused decreased splicing of an intron from an artificial reporter whereas hypomethylation of coilin increased splicing (Whittom et al., 2008). However, as Cajal bodies are sites of snRNP maturation, and coilin is the major protein component of Cajal bodies, the effect on splicing is likely explained by Cajal body regulation. In contrast, a study in *Arabidopsis thaliana* reported coilin loss-of-function mutants led to increased splicing of an artificial reporter, purportedly by influencing the gene expression of a small number of stress-genes (Kanno et al., 2016). More work is required to understand if there is a direct relationship between coilin and splicing.

Although coilin concentrates in Cajal bodies, it is found predominantly dispersed throughout the nucleoplasm of cells and, as not all cells contain Cajal bodies but do have nucleoplasmic coilin, there are suggestions of additional functions for coilin (Lam et al., 2002; Young et al., 2000). Coilin has been shown to bind double stranded DNA and contains DNA-dependent RNase activity with which it can process snRNA 3' ends (Broome and Hebert, 2012). In the same study, coilin depletion led to accumulation of immature U snRNAs, suggesting coilin may directly be involved in snRNA 3' end processing. Another interpretation however, is that loss of coilin disrupts Cajal bodies and their association at snRNA gene loci, and the loss of the interaction may influence snRNA processing. There is also evidence that coilin plays a part in the DNA damage response. UV-irradiation causes Cajal bodies to disassemble and interestingly, coilin is one of the first proteins recruited to the DNA breaks (Bártová et al., 2014; Cioce et al., 2006). In addition, coilin redistributes to the nucleolus in response to cisplatin-induced DNA damage, where it has been shown to inhibit rRNA biogenesis by interacting with the RPA-194 subunit of

Polymerase I (Gilder et al., 2011). Collectively, these studies indicate nucleoplasmic coilin has Cajal body-independent functions.

1.8.5.2 Post-translational modifications of coilin

Coilin is a phosphoprotein that undergoes many post-translational modifications which can regulate its localisation and formation of Cajal bodies. Coilin contains segments of serine rich sequences and undergoes phosphorylation, often in a cell cycle dependent manner (Carmo-Fonseca et al., 1993; Chan et al., 1994). Proteomics approaches have shown coilin to contain a minimum of 11 potential phosphorylation sites in its sequence (Beausoleil et al., 2006; Olsen et al., 2006). Phosphorylation of coilin has been shown specifically to increase during the mitosis phase of the cell cycle and the phosphorylation state of coilin is able to regulate Cajal bodies and protein interactions (Carmo-Fonseca et al., 1993). SMN, for example, shows preferential binding for dephosphorylated coilin, whereas the SmB protein shows a higher affinity of binding to a phosphorylated coilin (Toyota et al., 2010). Overexpression of a dephosphorylated coilin mimic in cells caused a loss of canonical Cajal body formation and uncharacteristic localisation, suggesting coilin phosphorylation is a prerequisite for Cajal body formation (Hearst et al., 2009). However, in the same study it was shown that hyperphosphorylation of all 11 phosphorylation sites also caused Cajal body disassembly, thus there may be an 'in the middle' phosphorylation state that favours Cajal body formation. The hyperphosphorylation state of coilin when entering the mitosis stage of the cell cycle may then be responsible for the disassembly of Cajal bodies.

Methylation of coilin can also regulate its function and localisation. As mentioned, coilin can be symmetrically dimethylated on its arginine residues in the RG-box motif, which enhances the interaction with the SMN protein (Boisvert et al., 2002; Hebert et al., 2001). Inhibition of coilin methylation weakens the coilin-SMN interaction, and causes SMN to localise into nuclear gems instead of Cajal bodies, suggesting the methylation state of coilin may be the reason some cell lines display nuclear gems but not Cajal bodies (Hebert et al., 2002). An MCF7 cell line lacking the 5' methylthioadenosine phosphorylase (MTAP) enzyme has been used to study the effect of hypomethylation on coilin localisation. The cell line has increased levels of MTAP which inhibits protein arginine methyltransferases (PRMT), thus decreasing coilin methylation. In this cell line, hypomethylated coilin accumulates in

the nucleolus whereas methylated coilin remains in canonical Cajal bodies (Tapia et al., 2010). Of course, the PRMTs will catalyse the methylation of target proteins other than coilin, and you cannot rule out an indirect effect of this on coilin mislocalisation. Lastly, a HeLa cell line that expresses a hypomethylated version of coilin, was found to more efficient at splicing a reporter intron than normal HeLa cells expressing methylated coilin, suggesting coilin methylation and its effects on Cajal bodies can influence splicing (Whittom et al., 2008).

1.9 Summary, hypothesis and project aims

The GGGGCC hexanucleotide repeat expansion in the C9orf72 gene is the leading cause of amyotrophic lateral sclerosis. Three non-mutually exclusive disease mechanisms have been proposed, including haploinsufficiency of the C9orf72 protein. The C9orf72 protein has roles in autophagy and endosomal trafficking, but little is currently known about the nuclear function of C9orf72. The C9orf72 protein exhibits nuclear localisation and interacts with nuclear proteins, indicating there may be an essential nuclear function that is not yet understood. Characterisation of this nuclear function could lead to better understanding of C9orf72-linked disease mechanisms and development of more effective treatments.

Unto this end, the interaction between C9orf72 and coilin was investigated. Coilin is the major protein component of Cajal bodies (CBs), nuclear suborganelles which act as sites of splicing machinery maturation and assembly. C9orf72 ALS patients have splicing defects that correlate with disease severity and disruption to Cajal bodies has been implicated in several neurodegenerative disorders. We hypothesised that haploinsufficiency of the C9orf72 protein in ALS leads to dysregulation of Cajal bodies and downstream splicing.

The project aims were therefore:

1. To characterise the C9orf72-coilin interaction
2. To investigate how C9orf72 regulates Cajal bodies and splicing
3. To confirm the relevance of the C9orf72/Cajal body/Splicing relationship in ALS

2 Materials and Methods

2.1 Materials

Materials were obtained from Sigma unless otherwise stated.

2.1.1 Stock solutions

2.1.1.1 Purchased stock solutions

Acrylamide-bis-acrylamide (30% w/v)

B-27 supplement

β-mercaptoethanol

Chloroform

Ethanol (100%)

Ethylenediaminetetraacetic acid (EDTA) (0.5 M)

Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA) (0.25 M)

Ethidium bromide (10 mg/ml)

Foetal bovine serum (FBS)

Formaldehyde (37-40%)

Glacial acetic acid (100%)

Glycerol (100%)

HEPES, pH 7.9 (1 M)

Methanol (100%)

Nuclease free water

Normal goat serum (NGS)

NP-40 (100%)

Pen/Strep (5000 U/ml penicillin, 5000 U/ml streptomycin)

Propan-2-ol

Poly-L-Lysine (0.1% w/v)

Protein assay dye (Bradford) reagent (5X)

Tetramethylethylenediamine (TEMED)

Triton-X100 (100%)

Trypsin/EDTA

Tween-20 (100%)

2.1.1.2 Stock solutions prepared in house

The following were prepared in dH_2O unless otherwise stated:

Ammonium persulfate (APS) (10% w/v)
Ampicillin (100 mg/ml)
Bovine Serum Albumin (BSA) (100 mg/ml)
Chloramphenicol (34 mg/ml) dissolved in ethanol
Dithiothreitol (DTT) (0.1 M)
Isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 M)
Kanamycin (100 mg/ml)
L-glutamine (200 mM)
Magnesium chloride (1 M)
Phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4)
Polyethylenimine (1 mM)
Potassium chloride (4 M)
Sodium dodecyl sulfate (SDS) (10% w/v)
Sodium chloride (5 M)
Sodium pyruvate (100 mM)
Tris buffered saline (TBS) (50 mM Tris-HCl, 150 mM NaCl)
Tris-HCl (1.5 M, pH 8.8)
Tris-HCl (0.5 M, pH 6.8)

2.1.2 Molecular biology reagents

2.1.2.1 Vectors and expression plasmids

The vectors and expression plasmids used throughout the project are detailed in Table 2.1 and Table 2.2.

Table 2.1. Backbone vectors

Vector	Expression	Supplier
pCI-neo	Mammalian	Promega
pCI-neo-myc	Mammalian	Promega – modified in house by Dr Kurt De Vos
pCI-neo-FLAG	Mammalian	Promega – modified in house by Dr Kurt De Vos
pCI-neo-mCherry	Mammalian	Promega – modified in house by Dr Kurt De Vos
pRK5	Mammalian	BD Biosciences
pGEX6p1	Bacterial	GE Healthcare Life Sciences
PL-SIN-PGK-cPPT-WHV (pLenti-VOS)	Viral	In house: Prof. Mimoun Azzouz

Table 2.2. Plasmids used for protein expression

Plasmid	Protein Expression	Vector Backbone	Supplier
pRK5-myc-C9orf72S	myc-C9orf72S	pRK5	In house: Dr Matthew Walsh and Dr Adrian Higginbottom
pRK5-myc-C9orf72L	myc-C9orf72L	pRK5	In house: Dr Matthew Walsh and Dr Adrian Higginbottom
pCB6-P120	P120 minigene	pCB6	Prof Richard Padgett, Cleveland Clinic
pLenti-VOS-mVenus-C9orf72S	mVenus-C9orf72S	pLenti-VOS	In house: Dr Annekathrin Moeller
pLenti-VOS-mVenus-C9orf72L	mVenus-C9orf72L	pLenti-VOS	In house: Dr Annekathrin Moeller
pLenti-VOS-EGFP	EGFP	pLenti-VOS	In house: Dr Annekathrin Moeller
pLenti-VOS-emGFP-miR negative control	emGFP-miRNA negative control	pLenti-VOS	In house: Dr Annekathrin Moeller. See Webster et al. 2016 for details
pLenti-VOS-emGFP-miR C9orf72	emGFP-miRNA C9orf72	pLenti-VOS	In house: Dr Annekathrin Moeller. See Webster et al. 2016 for details
pRSV-β-Galactosidase	β-Galactosidase	pRSV	Addgene (#24058)
pCI-neo-myc-coilin	myc-coilin	pCI-neo	This project (Section 2.2.1.12)
pCI-neo-FLAG-coilin	FLAG-coilin	pCI-neo	This project (Section 2.2.1.12)
pCI-neo-mCherry-coilin	mCherry-coilin	pCI-neo	This project (Section 2.2.1.12)
pEGFP-coilin	EGFP-coilin	pEGFP-C3	Addgene (#36906)
pGEX6p1-GST	GST	pGEX6p1	GE Healthcare Life Sciences (28-9546-48)
pGEX6p1-GST-C9orf72S	GST-C9orf72S	pGEX6p1	In house: Dr Christopher Webster
pGEX6p1-GST-C9orf72L	GST-C9orf72L	pGEX6p1	In house: Dr Christopher Webster

2.1.2.2 Reagents for *E.Coli* bacterial cultures

LB broth and LB agar were purchased in powder form.

LB broth (Miller) powder components were:
10 g/L Tryptone
10 g/L NaCl
5 g/L Yeast extract

For LB broth reconstitution, 25 g LB broth was dissolved in 1 L of dH_2O and autoclaved. LB broth was supplemented with either 100 µg/ml ampicillin or 100 µg/ml kanamycin.

LB agar (Miller) powder components were:
15 g/L Agar
10 g/L Tryptone
10 g/L NaCl
5 g/L Yeast extract

For LB agar reconstitution, 40 g LB agar was dissolved in 1 L of dH_2O and autoclaved. LB agar was supplemented with either 100 µg/ml ampicillin or 100 µg/ml kanamycin.

Antibiotics were added upon cooling of the LB agar/broth to approximately 60 °C.

2.1.2.3 Reagents for small-scale preparation of plasmid DNA

For plasmid purification from *E.Coli*, the NucleoSpin® Plasmid Kit (Macherey-Nagel) was used.

The following reagents were included with the kit:

NucleoSpin® plasmid columns
Collection tubes
Resuspension buffer A with RNase A supplementation
Lysis buffer A2
Neutralisation buffer A3

Wash buffer AW
Wash buffer A4
Elution buffer AE

2.1.2.4 Reagents for genomic DNA extraction

QuickExtract™ DNA extraction solution (Lucigen)

2.1.2.5 Reagents for C9orf72 exon 2 and coilin cloning

2.1.2.5.1 Primers for generation of C9orf72 exon 2 and coilin fragments

C9orf72 exon 2 was amplified from genomic DNA using the primers below. A XhoI-coilin-NotI product was generated from the pEGFP-coilin plasmid using a forward and reverse primer which contained the XhoI and NotI cut site sequence on their 5' and 3' ends respectively.

C9orf72 exon 2 forward: 5'-AAATCATTTGGGGTTTTGATGG-3'

C9orf72 exon 2 reverse: 5'-TGTAAAACGACGGCCAGTACCA-3'

XhoI-coilin forward: 5'-CTCGAGATGGCAGCTTCCGAGACG-3'

NotI-coilin reverse: 5'-GCGGCCGCTCAGGCAGGTTCTGTACTTGA-3'

2.1.2.5.2 Reagents for Phusion® High-Fidelity PCR

PCR amplification of C9orf72 exon 2 and coilin was performed using the Phusion® High-Fidelity DNA polymerase (New England BioLabs):

5X Phusion® HF buffer
10 mM dNTPs
Phusion® DNA polymerase

2.1.2.6 Reagents for agarose gel electrophoresis

1x Tris-Acetate (TAE) buffer:
40 mM Tris
20 mM Glacial acetic acid
1 mM EDTA, pH 8

Agarose (Bioline)

DNA loading buffer (Bioline)

2.1.2.7 Reagents for extraction of DNA bands from agarose gels

DNA was excised and purified from agarose gels using the GenElute™ Gel Extraction Kit (Sigma).

The following reagents were included with the kit:
GenElute™ binding column G
Collection tubes
Gel solubilisation buffer
Column preparation buffer
Wash solution
Elution solution

In house reagents included:
100% Propan-2-ol

2.1.2.8 Reagents for subcloning C9orf72 exon 2 and coilin PCR products into the pCR™-Blunt II-TOPO® vector

The C9orf72 exon 2 and coilin DNA products were ligated into the pCR™-Blunt II-TOPO® vector using the Zero Blunt® TOPO® PCR cloning kit (Thermo Fisher Scientific).

The ligation was performed using the reagents provided in the manufacturers kit:

10 ng/µl linearised pCR™-Blunt II-TOPO® vector in:
50% (v/v) Glycerol

50 mM Tris-HCl, pH 7.4

1 mM EDTA

2 mM DTT

0.1% (v/v) Triton-X100

100 µg/ml BSA

30 µM Bromophenol blue

Salt solution:

1.2 M NaCl

0.06 M $MgCl_2$

2.1.2.9 Reagents for transformation of XL10-Gold® Ultracompetent cells

XL10-Gold® Ultracompetent cells

XL10-Gold® β-Mercaptoethanol

2.1.2.10 Reagents for colony PCR

Coilin was amplified from pCR™-Blunt II-TOPO®-coilin using the following primers:

XhoI-coilin forward: 5'-CTCGAGATGGCAGCTTCCGAGACG-3'

NotI-coilin reverse: 5'-GCGGCCGCTCAGGCAGGTTCTGTACTTGA-3'

PCR amplification of the coilin insert used the following reagents:

5X FIREPol® Master Mix Ready to Load (Solis Biodyne) which contained:

FIREPol® DNA polymerase

5X reaction buffer B (0.5 M Tris-HCl, 0.1 M $(NH_4)_2SO_4$, 0.1% Tween-20 (v/v))

7.5 mM $MgCl_2$

2 mM dNTPs of each

Blue dye

Yellow dye

2.1.2.11 Reagents for restriction enzyme digestion of DNA

The coilin insert was excised from the pCR™-Blunt II-TOPO® vector using the following FastDigest reagents from Thermo Fisher Scientific:

FastDigest Green Buffer (10X)
FastDigest XhoI enzyme
FastDigest NotI enzyme

2.1.2.12 Reagents for coilin ligation into pCI-neo-tag vectors

The Quick Ligation™ Kit (New England BioLabs) was used to perform the ligation of XhoI-coilin-NotI into linearised pCI-neo-tag vectors which had been previously cut with XhoI and NotI.

The kit reagents included:

Quick Ligase Reaction Buffer (2X)
Quick T4 DNA ligase

2.1.2.13 Reagents for site-directed mutagenesis

2.1.2.13.1 Primers for generation of siRNA resistant pRK5-myc-C9orf72

Primers used in the first round of mutagenesis introduced two T>C base pair changes at T439 and T445 in the C9orf72 cDNA sequence using missense mutations in the primers. The primers were used to mutate the C9orf72 siRNA #2 binding site within the pRK5-myc-C9orf72S and pRK5-myc-C9orf72L plasmids. Primers used in the second round of mutagenesis introduced four T>C base pair changes at T703, T709, T715 and T718 in the C9orf72 cDNA sequence using missense mutations in the primers. The primers were used to mutate the C9orf72 siRNA #D binding site within the C9orf72 siRNA #2 resistant pRK5-myc-C9orf72S and pRK5-myc-C9orf72L plasmids. Primers were designed using the online QuikChange Primer Design Program.

C9orf72 T439C, T445C forward:
5'-GACAAGACAAAAAACTTTACGTCTATGGCACCACTCTCTGCATTTCG-3'

C9orf72 T439C, T445C reverse:
5'-CGAAATGCAGAGAGTGGTGCCATAGACGTAAAGTTTTTTGTCTTGTC-3'

C9orf72 T703C, T709C, T715C, T718C forward:
5'-
CTCCAGTAAGCATTGGAATGATGCTCTGGCCCTGGTCTTCCATTCTCTCTGTGC
-3'

C9orf72 T703C, T709C, T715C, T718C reverse:
5'-
GCACAGAGAGAATGGAAGACCAGGGCCAGAGCATCATTCCAATGCTTACTGGA
G-3'

2.1.2.13.2 PCR reagents for site-directed mutagenesis

PCR was performed using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent):

10X QuikChange Lightning Reaction Buffer
dNTP mix
QuikSolution Reagent
QuikChange Lightning Enzyme
Dpn I enzyme

2.1.2.14 Sequencing primers

The SP6 primer was used to sequence C9orf72 exon 2 in the pCR™-Blunt II-TOPO® vector and pRK5-myc-C9orf72S and pRK5-myc-C9orf72L after site-directed mutagenesis.

SP6: 5'-ATTTAGGTGACACTATAG-3'

Coilin was sequenced in the pCR™-Blunt II-TOPO® vector using the SP6 primer and four primers within the coilin sequence spaced approximately 400 base pairs apart to ensure full coverage.

SP6: 5'-ATTTAGGTGACACTATAG-3'
Coilin sequencing 1: 5'-GGAGGGTGAAGAAACTGAACC-3'
Coilin sequencing 2: 5'-GATGAATCTATCAGTGATGGTCCC-3'

Coilin sequencing 3: 5'-TCCCTGCTAGTTTAGGAAGAGG-3'
Coilin sequencing 4: 5'-ACCTGCCTTGAGAGAACCTG-3'

2.1.2.15 Reagents for expression of GST-proteins in Rosetta pLysS cells

GST-protein expression was induced in Rosetta pLysS bacteria cells (Novagen) transformed with the pGEX6p1 expression plasmids detailed in table 2.2 using the following reagents:

Terrific Broth (TB):
24 g/L Yeast extract
20 g/L Tryptone
0.4% (v/v) Glycerol
17 mM KH_2PO_4
72 mM K_2HPO_4

TB supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol
TB supplemented with 100 µg/ml ampicillin
0.5 mM IPTG

2.1.3 Cell culture and transfection

2.1.3.1 Reagents for subculturing of HEK293 cells

Complete DMEM media:
DMEM (with 4.5 g/L glucose)
10% Foetal bovine serum (Life Science Production)
1 mM Sodium pyruvate

PBS
Trypsin/EDTA (Lonza)

2.1.3.2 Reagents for culturing rat primary neurons

Poly-L-Lysine (0.1% w/v in H_2O)

Complete Neurobasal Medium:
Neurobasal medium (Thermo Fisher Scientific)
2% B-27 supplement (Thermo Fisher Scientific)
100 IU/ml penicillin
100 mg/ml streptomycin
2 mM L-glutamine (Lonza)

2.1.3.3 Reagents for subculturing iAstrocytes

Human plasma Fibronectin purified protein (Merck)
Accutase®

iAstrocyte media:
DMEM (with 4.5 g/L glucose)
100 IU/ml penicillin
100 mg/ml streptomycin
10% Foetal bovine serum (Life Science Production)
0.2 % N-2 Supplement (Thermo Fisher Scientific)

2.1.3.4 Reagents for plasmid DNA transfection

Opti-MEM Reduced Serum Media (Thermo Fisher Scientific)
Polyethylenimine (PEI)
Lipofectamine® 2000 Reagent (Thermo Fisher Scientific)

2.1.3.5 Reagents for siRNA transfection

Opti-MEM Reduced Serum Media (Thermo Fisher Scientific)
Lipofectamine® RNAiMAX Reagent (Thermo Fisher Scientific)

MISSION siRNA Universal Negative Control #1
C9orf72 #2 siRNA: GUGCUAUAGAUGUAAAGUU
C9orf72 #D siRNA: GAUCAGGGUCAGAGUAUUA
Coilin #1 siRNA: GGAACAACAUCUUCCAGUU
Coilin #2 siRNA: CAUCUACUAUUAUCCAGAA
Coilin #3 siRNA: CAGAGAAAUUACCAACUGA

2.1.4 Reagents for biochemical methods

2.1.4.1 Reagents for whole cell lysis

Trypsin/EDTA
PBS

Radioimmunoprecipitation (RIPA) buffer:
100 mM Tris-HCl, pH 6.8
150 mM NaCl
1 mM EDTA
1 mM EGTA
1 mM DTT
0.1% (w/v) SDS
0.5% (w/v) sodium deoxycholate
1% (v/v) NP40
Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific)

2.1.4.2 Reagents for nuclear fractionation

Hypotonic lysis buffer:
10 mM HEPES, pH 7.9
1.5 mM $MgCl_2$
10 mM KCl
0.5 mM DTT

RIPA buffer

2.1.4.3 Reagents for Bradford protein assay

Bovine Serum Albumin (100 mg/ml)
Protein assay dye reagent (Biorad)

2.1.4.4 Reagents for preparation of samples for SDS-PAGE electrophoresis

5X Laemmli buffer:

250 mM Tris-HCl, pH 6.8

10% (w/v) SDS

0.5% (w/v) Bromophenol blue

50% (v/v) Glycerol

25% (v/v) β-mercaptoethanol

2.1.4.5 Reagents for immunoprecipitation

1 µg of antibody (Table 2.3)

Protein G Sepharose 4 Fast Flow (GE Healthcare)

RIPA buffer

2X Laemmli buffer

Table 2.3. Immunoprecipitation antibodies

IP antibody	Protein pull down	Antibody species	Supplier
Anti-coilin (ab87913)	Coilin	Mouse	Abcam
Anti-myc (9B11)	myc-C9orf72	Mouse	Cell Signalling

2.1.4.6 Reagents for *in vitro* GST-binding assay

In vitro production of ^{35}S-coilin:

Reticulocyte from the T7 TnT® Quick Coupled Transcription/Translation System (Promega)

^{35}S-methionine (Perkin Elmer)

RB100 buffer:

25 mM HEPES, pH 7.5

100 mM KOAc

10 mM $MgCl_2$

1 mM DTT

0.05% (v/v) Triton-X100
10% (v/v) Glycerol

Pull down:
Glutathione Sepharose High Performance (GE Healthcare)

GSH elution buffer:
50 mM Tris-HCl, pH 7.5
100 mM NaCl
40 mM reduced glutathione

Radioactivity detection:
Whatman filter paper
BAS Storage Phosphor Screen (GE Healthcare)

2.1.4.7 Reagents for SDS-PAGE electrophoresis

10% resolving gel:
10% (w/v) acrylamide
375 mM Tris-HCl, pH 8.8
0.1% (w/v) SDS
0.1% (w/v) APS
0.01% (v/v) TEMED

6% stacking gel:
6% (w/v) acrylamide
125 mM Tris-HCl, pH 6.8
0.1% (w/v) SDS
0.06% (w/v) APS
0.3% (v/v) TEMED

SDS-PAGE running buffer:
25 mM Tris-HCl
190 mM glycine
0.1% (w/v) SDS

2.1.4.8 Reagents for coomassie staining of polyacrylamide gels

Coomassie Brilliant Blue G250 stain:
1 g/L Coomassie Brilliant Blue
50% (v/v) Methanol
10% (v/v) Glacial acetic acid

Destain solution:
30% (v/v) Methanol
10% (v/v) Glacial acetic acid

2.1.4.9 Reagents for electrophoretic transfer of proteins

Transfer buffer:
25 mM Tris-HCl
190 mM Glycine
20% (v/v) Methanol

Whatman filter paper
Nitrocellulose blotting membrane 0.45 μm pore (Amersham)

Ponceau S stain:
0.1% (w/v) Ponceau S
5% (v/v) Glacial acetic acid

2.1.4.10 Reagents for immunoblot of nitrocellulose membrane

TBST:
0.1% (v/v) Tween-20 in TBS

Blocking:
5% milk (MARVEL) in TBST

Antibodies used for immunoblot are detailed in Table 2.4.

Table 2.4. Immunoblot antibodies

Antibody	Isotype	Supplier	Catalogue Number	Working Dilution
Primary Antibody				
Anti-myc	Rabbit	Abcam	ab9106	1:1000
Anti-myc	Mouse	Cell Signalling	9B11	1:2000
Anti-coilin	Rabbit	ProteinTech	10976-1-AP	1:1000
Anti-SMN	Mouse	BD Bioscience	610646	1:300
Anti-β-Gal	Rabbit	Abcam	ab9619	1:500
Anti-C9orf72	Rabbit	Santa Cruz	sc-138763	1:500
Anti-C9orf72	Rabbit	ProteinTech	25757-1-AP	1:1000
Actin C4	Mouse	Millipore	mab1501	1:5000
Anti-GAPDH	Rabbit	Cell Signalling	2118	1:2500
Anti-FLAG	Mouse	Sigma	F3165	1:2000
Anti-mCherry	Rabbit	Abcam	ab167453	1:1000
Anti-GFP	Mouse	Clontech	JL8	1:5000
Anti-SSRP1	Mouse	Biolegend	609702	1:500
Secondary Antibody				
Anti-mouse-HRP	Goat	Dako	P0447	1:5000
Anti-rabbit-HRP	Goat	Dako	P0448	1:5000

2.1.4.11 Reagents for chemiluminescent protein detection

SuperSignal™ West Pico PLUS Chemiluminescent Substrate components:
SuperSignal™ West Pico PLUS Luminol/Enhancer Solution
SuperSignal™ West Pico PLUS Stable Peroxide Solution

Hyperfilm ECL (Amersham)

Developing solution (Ilford Hypam)
Fixing solution (Ilford Hypam)

2.1.4.12 Reagents for RNA extraction

TRIzol™ reagent (Ambion)
Chloroform
Propan-2-ol

75% Ethanol
Nuclease free water

2.1.4.13 Reagents for RT-PCR generation of cDNA

10X DNase I Buffer (New England BioLabs)
DNase I enzyme (New England BioLabs)
25 mM EDTA
Oligo dT(18) primer (Thermo Fisher Scientific)
dNTPs solution mix (New England BioLabs)

The following are supplied with the SuperScript® III Reverse Transcriptase (Thermo Fisher Scientific):
DTT
5X First Strand Buffer
SuperScript® III Reverse Transcriptase (200 U/µl)

2.1.4.14 Reagents for qPCR

5X HOT FIREPol EvaGreen® dye (Solis Biodyne)
Nuclease free water

The following primers (100 µM stock) were used in qPCR to measure relative mRNA levels:

Human C9orf72 (working dilution 1:40)
Forward: 5'-GTTGATAGATTAACACATATAATCCGG-3'
Reverse: 5'-CAGTAAGCATTGGAATAATACTCTGA-3'

Human CRISPR edited C9orf72 (working dilution 1:40)
Forward: 5'-TCCAATGCTTACTGGAGAAGTG-3'
Reverse: 5'-ACTACAACGGAACAGCCACA-3'

Rat C9orf72 (working dilution 1:40)
Forward: 5'-GTTGACAGGCTAACGCACATCATTCGAA-3'
Reverse: 5'-CAGTAAGCATAGGGATGATACTCTGA-3'

Human Coilin (working dilution 1:10)
Forward: 5'-AGGGCATCCTGTTTCCTGTG-3'
Reverse: 5'-CCTTCTTGGGTGTCTCTACTGG-3'

Human GAPDH (working dilution 1:33)
Forward: 5'-GGGTGGGGCTCATTTGCAGGG-3'
Reverse: 5'-TGGGGGCATCAGCAGAGGGG-3'

Rat RPL19 (working dilution 1:40)
Forward: 5'-CTCGATGCCGGAAGAACACC-3'
Reverse: 5'-GAGCGTTGGCAGTACCCTT-3'

2.1.4.15 Reagents for PCR amplification of the pCI-neo and P120 introns

FIREPol® Master Mix Ready To Load
Nuclease free water

Primers for amplification of the pCI-neo intron:
pCI-neo forward: 5'-TCACTAGAAGCTTTATTGCGGTAG-3'
pCI-neo reverse: 5'-TGTATCTTATCATGTCTGCTCGAAG-3'

Primers for amplification of the P120 intron:
P120 forward: 5'-TGGATGAGGAACCATTTGTG-3'
P120 reverse: 5'-AGAACGAGACCGCCCTTC-3'

2.1.5 Reagents for microscopy

2.1.5.1 Reagents for immunofluorescence

PBS
4% (w/v) Formaldehyde in PBS
0.2% (v/v) Triton-X100 in PBS
5% (v/v) Normal goat serum in PBS
Hoechst 33342
Fluorescence mounting media (Dako)

Antibodies used in immunofluorescence are detailed in table 2.5.

Table 2.5. Immunofluorescence antibodies

Antibody	Isotype	Supplier	Catalogue Number	Working Dilution
Primary Antibody				
Anti-myc	Mouse	Cell Signalling	9B11	1:2000
Anti-myc	Rabbit	Abcam	ab9106	1:1000
Anti-coilin	Rabbit	ProteinTech	10976-1-AP	1:1000
Anti-coilin	Mouse	Abcam	ab87913	1:1000
Anti-SMN	Mouse	BD Bioscience	610646	1:300
Anti-FLAG	Mouse	Sigma	F3165	1:2000
Anti-mCherry	Rabbit	Abcam	ab167453	1:1000
Anti-GFP	Mouse	Clontech	JL8	1:5000
Anti-β-Gal	Rabbit	Abcam	ab9619	1:500
Secondary Antibody				
Anti-mouse-AF488	Donkey	Thermo Fisher Scientific	A21202	1:500
Anti-mouse-AF568	Donkey	Thermo Fisher Scientific	A10037	1:500
Anti-rabbit-AF488	Goat	Thermo Fisher Scientific	A11008	1:500
Anti-rabbit-AF568	Goat	Thermo Fisher Scientific	A11011	1:500

2.1.5.2 Reagents for the proximity ligation assay

In house PLA reagents:

PBS

4% Formaldehyde in PBS

0.2% Triton-X100 in PBS

5% Normal goat serum in PBS

Hoechst 33342

Fluorescence mounting media (Dako)

The following Duolink® PLA technology reagents were used (Sigma):

Duolink® PLA Probe Anti-Mouse MINUS:

5X PLA Probe Anti-Mouse MINUS

1X Blocking Solution

1X Antibody Diluent

Duolink® PLA Probe Anti-Rabbit PLUS:

5X PLA Probe Anti-Rabbit PLUS

1X Blocking Solution

1X Antibody Diluent

Duolink® In Situ Detection Reagents Orange:

5X Ligation Solution

1X Ligase

5X Amplification Solution

1X Polymerase

Duolink® In Situ Wash Buffers, Fluorescence:

Wash Buffer A

Wash Buffer B

Antibodies used in PLA are detailed in table 2.6.

Table 2.6. PLA antibodies

Antibody	Species	Supplier	Dilution
Anti-coilin (10976-1-AP)	Rabbit	ProteinTech	1:1000
Anti-myc (9B11)	Mouse	Cell Signalling	1:2000

2.2 Methods

2.2.1 Molecular biology methods

2.2.1.1 Plasmid growth from bacterial glycerol stocks

Bacterial glycerol stocks were used for long-term plasmid storage. 500 µl of bacterial culture was combined with 500 µl of sterile 50% glycerol and stored at -80 °C.

To recover bacteria from glycerol stocks, a sterile loop was used to remove frozen bacteria, which were then streaked onto LB agar plates containing the appropriate antibiotic. The plates were incubated at 37 °C for 16 hours. A sterile loop was used to pick a single colony for inoculation into 1 ml LB broth containing the appropriate antibiotic. The bacterial culture was shaken at 220 rpm at 37 °C for 8 hours. 50 µl of the 1 ml culture was inoculated into 5 ml LB broth containing the appropriate antibiotic and shaken at 220 rpm at 37 °C for 16 hours.

2.2.1.2 Small-scale preparation of plasmid DNA

For plasmid purification, the NucleoSpin® Plasmid Kit from Macherey-Nagel was used following the manufacturers protocol. *E.Coli* were pelleted from the 5 ml bacterial culture by centrifugation at 4,000 rpm for 10 minutes at RT in a Sigma 1-15 PK centrifuge using a 12124 rotor. The LB broth was discarded and the bacterial pellet resuspended in 250 µl resuspension buffer A1. The bacteria were lysed by addition of 250 µl of lysis buffer A2 and incubation at RT for 5 minutes. 300 µl of neutralisation buffer A3 was added and the lysate centrifuged at 11,000 g for 10 minutes to pellet protein, genomic DNA and cell debris. The plasmid-containing supernatant was applied to the NucleoSpin® Plasmid column to bind DNA to the silica membrane and subjected to centrifugation at 11,000 g for 1 minute. The run-through was discarded and the membrane washed with 500 µl wash buffer AW by centrifugation at 11,000 g for 1 minute. The run-through was discarded and the membrane washed with 600 µl wash buffer A4 by centrifugation at 11,000 g for 1 minute. The run-through was discarded and the membrane dried by centrifugation at 11,000 g for 2 minutes. The plasmid DNA was eluted by addition of 50 µl of elution buffer AE and centrifugation at 11,000 g for 1 minute.

2.2.1.3 Genomic DNA extraction

Genomic DNA was extracted from the C9orf72 CRISPR-edited cell lines B2, B17 and B12 for sequencing of C9orf72 exon 2. Cells were pelleted by centrifugation at 400 g for 4 minutes at 4 °C. Cells were washed by resuspension in PBS and pelleted by centrifugation at 14,000 rpm for 30 seconds at 4 °C in a Sigma 1-15 PK centrifuge using a 12124 rotor. 100 µl of QuickExtract™ DNA extraction Solution (Lucigen) was added to cells and vortexed for 15 seconds. The solution was incubated at 65 °C for 2 hours then 99 °C for 2 minutes. DNA was stored at 4 °C.

2.2.1.4 DNA quantification

The concentration and quality of DNA samples were determined using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) which measured the absorbance of the samples at 260 nm and 280 nm. The concentration of the DNA in ng/µl is based on the A_{260}; a DNA sample with concentration 1 ng/µl and a 1 cm path length has A_{260} = 50. The ratio of A_{260}/A_{280} was used as an assessment of DNA purity where a ratio of ~1.8 was taken as pure DNA.

2.2.1.5 PCR amplification of C9orf72 exon 2 and coilin to generate blunt-ended products

The C9orf72 exon 2 fragment to be sequenced was generated using Phusion® High-Fidelity DNA polymerase (New England BioLabs), the primers described in Section 2.1.2.5.1 and genomic DNA extracted from B2, B17 and B12 cells as the DNA template. A XhoI-Coilin-NotI product for cloning into pCI-neo-tag expression vectors was generated using Phusion® High-Fidelity DNA polymerase, the primers described in Section 2.1.2.5.1 and the pEGFP-coilin plasmid as the template.

A 20 µl reaction mix was prepared as below:

12.4 µl Nuclease-free water
4 µl 5X Phusion® HF buffer
0.4 µl 10 mM dNTPs
1 µl 10 µM Forward primer
1 µl 10 µM Reverse primer
1 µl Template DNA (5 ng/µl)

0.2 µl Phusion® DNA polymerase

The reaction was subjected to the following thermocycling conditions:

Table 2.7. Thermocycling conditions for Phusion® High-Fidelity PCR

Step	Cycles	Temp	Time
Initial Denaturation	1	98 °C	30 s
Denaturing	35	98 °C	10 s
Annealing		60 °C	20 s
Extension		72 °C	30 s C9orf72 exon 2 / 60 s coilin
Final extension	1	72 °C	10 mins

2.2.1.6 Agarose gel electrophoresis of DNA

A 1-2% (w/v) agarose gel was prepared by dissolving agarose in boiling TAE buffer. Upon cooling, 1 µl of ethidium bromide was added per 100 ml before pouring into casting apparatus containing a horizontal comb. The gel was placed into a Geneflow Horizontal Gel Tank, immersed in TAE buffer and samples loaded. If required, DNA loading buffer (Bioline) was diluted 1 in 5 with DNA sample. A nucleic acid molecular weight marker (Hyperladder, Bioline) was loaded for determination of DNA bands in sample wells. The gel was subjected to electrophoresis at 120 V for approximately 40 mins and bands visualised using the Syngene G:Box Chemi XRQ.

2.2.1.7 Gel extraction of DNA

DNA extraction from agarose gels was performed using the GenElute Gel Extraction Kit (Sigma). DNA bands were visualised on a UV transilluminator and excised from the gel using a scalpel. The gel slice was weighed and solubilised in 3 gel volumes of the Gel Solubilisation Buffer by incubation at 60 °C for 10 minutes, with vortexing every 2 minutes. A binding column containing a silica membrane was placed into a collection tube and prepared for DNA binding by addition of 500 µl of the Column Preparation Solution and centrifugation at 16,000 g for 1 minute. The run-through was discarded. 1 gel volume of 100% propan-2-ol was added to the solubilised gel solution and mixed until homogenous. The solubilised gel mixture was added to the binding column and centrifuged at 16,000 g for 1 minute. The run-

through was discarded. The silica membrane was washed by addition of 700 µl of Wash Solution to the binding column and centrifuged at 16,000 g for 1 minute. The run-through was discarded and the membrane was dried by centrifugation at 16,000 g for 1 minute. DNA was eluted by addition of 20 µl of Elution Solution and centrifugation at 16,000 g for 1 minute. DNA concentration was measured as detailed in Section 2.2.1.4.

2.2.1.8 Subcloning into the pCR™-Blunt II-TOPO® vector

The blunt-ended C9orf72 exon 2 and XhoI-Coilin-NotI products were subcloned into the pCR™-Blunt II-TOPO® vector using the Zero Blunt® TOPO® PCR cloning kit (Thermo Fisher Scientific) following the manufacturers protocol. The pCR™-Blunt II-TOPO® plasmid is provided linearised and bound by virus DNA topoisomerase I enzyme at the 3' end. Briefly, the system functions by site-specific cleavage of the DNA product by topoisomerase I followed by formation of a covalent bond between the 3' phosphate of the DNA and a tyrosyl residue in topoisomerase I. The bond is then broken by attacking of the cleaved 5' hydroxyl, removing the enzyme and ligating the DNA into the plasmid (Shuman, 1994).

The 6 µl ligation reaction was prepared as follows:
4 µl gel extracted/purified blunt-ended DNA product
1 µl salt solution
0.5 µl sterile water
0.5 µl linearised pCR™-Blunt II-TOPO® vector

The reaction was incubated at RT for 15 minutes before transformation into XL10-Gold® Ultracompetent Cells.

2.2.1.9 Transformation of XL10-Gold® Ultracompetent Cells

XL10-Gold® Ultracompetent Cells (Stratagene) exhibit the Hte phenotype which lends to increased transformation efficiency of ligated plasmids. An empty eppendorf and an aliquot of cells were chilled on ice until the cells had thawed. A 20 µl aliquot of cells was removed into the chilled eppendorf and 1 µl of β-mercaptoethanol was added. The β-mercaptoethanol was mixed into cells by gentle swirling of the pipette tip in the cells and incubated on ice for 1 minute. 2 µl of ligated plasmid was added to cells and mixed by gentle swirling of the pipette tip in

the cells and incubated on ice for 30 minutes. The cells were heat shocked at 42 °C for exactly 30 s then placed on ice for 2 minutes. 180 µl of LB broth was added to cells and incubated at 37 °C for 1 hour with shaking at 220 rpm. The 200 µl bacterial culture was split into a 50 µl and 150 µl aliquot and spread onto two LB agar plates that incorporated the antibiotic for which the bacteria expressed resistance. The plates were incubated at 37 °C for 16 hours. Colonies containing the pCR™-Blunt II-TOPO® vectors with C9orf72 exon 2 insert were selected for plasmid purification as described in Section 2.2.1.2 for sequencing as described in Section 2.2.1.14. Colonies containing pCR™-Blunt II-TOPO® vectors with coilin insert were selected for colony PCR screening.

2.2.1.10 Colony PCR screening of bacterial colonies

Colonies were selected for the coilin insert by colony PCR. The following 10 µl reaction was prepared in a 0.2 ml PCR tube using primers described in Section 2.1.2.10:

2 µl 5X FIREPol® Master Mix Ready to Load (Solis Biodyne)
0.3 µl Forward primer
0.3 µl Reverse primer
7.4 µl Sterile water

A bacterial colony was picked with a sterile pipette tip and streaked onto a replica plate which was incubated at 37 °C for 8 hours. The residual bacteria on the pipette tip was transferred to the reaction tubes by swirling the pipette tip in the PCR mix. The reactions were subjected to the following thermocycling conditions:

Table 2.8. Thermocycling conditions for colony PCR

Step	Cycles	Temp	Time
Bacteria lysis/Initial denaturation	1	95 °C	5 mins
Denaturation	35	95 °C	30 s
Annealing		58 °C	30 s
Extension		72 °C	2 mins
Final extension	1	72 °C	10 mins

The PCR mix was subjected to DNA agarose gel electrophoresis as described in Section 2.2.1.6 for analysis of DNA bands. Colonies which contained the coilin insert were picked from the replica plate for plasmid purification, sequencing and restriction enzyme digest.

2.2.1.11 Restriction enzyme digest of DNA

pCR™-Blunt II-TOPO®-coilin plasmids with the correct coilin sequence was digested with XhoI and NotI restriction enzymes (Thermo Fisher Scientific) for cloning into the required expression vectors. The following 20 µl restriction digest reaction was prepared:

3.4 µl pCR™-Blunt II-TOPO® with coilin insert (0.584 µg/µl)
2 µl FastDigest Green Buffer
13.6 µl Sterile water
0.5 µl FastDigest XhoI enzyme
0.5 µl FastDigest NotI enzyme

The reactions were incubated at 37 °C for 10 minutes. The total 20 µl reaction mix was loaded onto an agarose gel and subjected to electrophoresis for separation of the vector and coilin insert. The coilin insert was gel extracted as described in Section 2.2.1.7 for ligation into expression vectors.

2.2.1.12 Ligation of coilin insert into pCI-neo-tag expression vectors

The purified coilin insert was ligated into pCI-neo-myc, pCI-neo-FLAG and pCI-neo-mCherry expression vectors. The vectors had previously been digested with XhoI and NotI and were stored linearised and dephosphorylated at -20 °C. The Quick T4 DNA Ligase (New England BioLabs) was used to perform the following ligation reaction which required a 3:1 molar ratio of insert to vector:

50 ng pCI-neo-tag vector (5.5 kb)
50 ng coilin insert (1.8 kb)
10 µl Quick Ligase Reaction Buffer
Up to 19 µl with sterile water
1 µl Quick T4 DNA Ligase

The reactions were incubated at RT for 10 minutes before transformation into XL10-Gold® Ultracompetent Cells as described in Section 2.2.1.9 and purified.

2.2.1.13 Site-directed mutagenesis of pRK5-myc-C9orf72

Site-directed mutagenesis was performed using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent). Two rounds of mutagenesis were performed; the first round of mutagenesis was performed on the pRK5-myc-C9orf72S and pRK5-myc-C9orf72L plasmids using the primers described in Section 2.1.2.13.1. The second round of mutagenesis was performed on the mutant pRK5-myc-C9orf72S and pRK5-myc-C9orf72L plasmids using the primers described in Section 2.1.2.13.1.

The mutagenesis reactions were prepared as follows:

39 µl Sterile water
5 µl 10X QuikChange Lightning Reaction Buffer
1 µl pRK5-myc-C9orf72 vector (25 ng/µl)
1.25 µl Forward primer (100 ng/µl)
1.25 µl Reverse primer (100 ng/µl)
1 µl dNTP mix
1.5 µl QuikSolution Reagent
1 µl QuikChange Lightning Enzyme

The reactions were subjected to the following thermocycling conditions:

Table 2.9. Thermocycling conditions for QuikLightning Site-Directed Mutagenesis

Step	Cycles	Temp	Time
Initial Denaturation	1	95 °C	2 mins
Denaturation	18	95 °C	20 s
Annealing		60 °C	10 s
Extension		68 °C	3.5 mins
Final Extension	1	68 °C	5 mins

After mutagenesis, 2 µl of *DpnI* restriction enzyme was added to the reaction mix and incubated at 37 °C for 5 minutes to digest the parent plasmid DNA. The mutated plasmid was transformed into XL10-Gold® Ultracompetent cells, purified and sequenced.

2.2.1.14 DNA Sequencing

DNA sequencing was outsourced to The University of Sheffield Core Genomic Facility. 10 µl of 100 ng/µl DNA and 10 µl of 100 pmol/µl sequencing primers were supplied to the facility per sequencing reaction. The primers used for sequencing are described in Section 2.1.2.14. The sequencing files were analysed using 4Peaks (Nucleobytes) and Serial Cloner (Serialbasics) software.

2.2.1.15 Expression of GST-tagged proteins in Rosetta pLysS cells

GST-control and GST-tagged C9orf72 proteins were expressed in Rosetta pLysS bacterial cells transformed with the pGEX6p1 expression plasmids detailed in table 2.2. A sample of the cells was removed from glycerol stocks and inoculated into 20 ml Terrific Broth containing ampicillin and chloramphenicol antibiotics. The pGEX6p1 vector carries resistance against ampicillin and the Rosetta pLysS cells contain a plasmid which carries chloramphenicol resistance. The bacterial cultures were grown at 37 °C for 16 hours with shaking at 220 rpm. 10 ml of the bacterial culture was added to 750 ml Terrific Broth containing ampicillin antibiotic and mixed well until homogenous. 1 ml of this culture was removed into a cuvette for measurement of the A_{600} using the WPA Spectrawave S1200 Visible Spectrophotometer (Biochrom Ltd). The 750 ml culture was incubated at 37 °C with shaking at 220 rpm and 1 ml samples removed at 30 minute intervals to measure the A_{600}. When the A_{600} reached 0.7, 750 µl of 0.5 M IPTG (final concentration 0.5 mM) was added to induce GST-protein expression and the culture incubated 37 °C with shaking at 220 rpm for 3 hours. The bacteria were pelleted by centrifugation at 4,000 g at RT for 20 mins and stored at -20 °C before use in the *in vitro* GST-binding assay described in Section 2.2.3.5.

2.2.2 Cell culture and transfection

2.2.2.1 Subculturing of HEK293 cells

HEK293 cells were grown in 20 ml of complete DMEM at 37°C/5% CO_2 in T75 cell culture flasks and sub-cultured twice per week when cells reached approximately 90% confluency. Media was removed and cells washed with PBS. Cells were detached from the flask by addition of 2 ml of Trypsin/EDTA solution and incubation at 37°C/5% CO_2 for 2 minutes. The trypsin was neutralised by addition of 8 ml of complete DMEM and triturated to create a single cell suspension. A volume of the 10 ml cell suspension was seeded into new flasks as stock cells or into the appropriate cell culture vessel for experimental purposes as described in Table 2.10. For immunofluorescence experiments, cells were seeded onto glass coverslips in a 24-well plate.

Table 2.10. Seeding volumes for HEK293 cells

Experimental purpose	Seeding culture vessel	Volume cell suspension (ml)	Volume complete DMEM (ml)	Final volume (ml)
Stock flasks	T75 flask	1	19	20
Plasmid transfection	24-well plate	1.2	10.8	12
	6-well plate	1.5	10.5	12
	10 cm dish	1.5	8.5	10
siRNA transfection	24-well plate	0.3	11.7	12

2.2.2.2 Preparation of rat primary neurons

Rat primary neurons were prepared by Miss Rebecca Cohen. Rat cortical and hippocampal neurons were isolated from E18 Sprague Dawley rat embryos (Charles River) and seeded onto poly-L-lysine coated coverslips in a 24-well plate at approximately 190,000 cells per well. Neurons were cultured in complete neurobasal media at 37°C/5% CO_2, and maintained by replacement of half of the media with fresh media at 5, 8 and 12 days *in vitro* (DIV).

2.2.2.3 Seeding of iAstrocyte cell lines

iAstrocyte cells were differentiated from neural progenitor cells (NPCs) converted from healthy or C9orf72-ALS patient skin fibroblasts by Dr Laura Ferraiuolo (Meyer et al., 2014). The human skin fibroblasts were obtained by Prof Dame Pamela J

Shaw and informed consent was acquired from all subjects prior to collection of samples (Study number STH16573, Research Ethics Committee reference 12/YH/0330). The demographics of the cell lines is detailed in table 2.11.

Table 2.11. iAstrocyte cell line patient demographics

Cell line	ALS-type	Sex	Age at biopsy collection (years)	Ethnicity	Onset to death (months)
155	Non-ALS control	M	40	Caucasian	-
209	Non-ALS control	F	69	Caucasian	-
3050	Non-ALS control	M	55	Caucasian	-
78	C9-ALS	M	66	Caucasian	31.7
183	C9-ALS	M	50	Caucasian	27
201	C9-ALS	F	66	Caucasian	19.4

The iAstrocytes were grown in iAstrocyte media at 37°C/5% CO_2 in cell culture flasks. Media was removed and for seeding, cells were washed with PBS and detached from the flask by addition of 1 ml of Accutase® and incubation at 37°C/5% CO_2 for 2 minutes. The Accutase® was quenched by addition of 3 ml of iAstrocyte media and the cell suspension centrifuged at 200 g for 4 minutes at RT. The media was removed and the cells resuspended in 3ml of fresh iAstrocyte media by gently pipetting up and down. 10 µl of cell suspension was removed and the cell count determined using a hemocytometer. The iAstrocyte cell suspension was diluted and seeded onto fibronectin-coated coverslips in a 24-well plate at 50,000 cells per well.

2.2.2.4 Plasmid DNA transfection

Cells were seeded such that they were at 70% confluency for next-day plasmid transfection. HEK293 and iAstrocyte cells were transfected using polyethylenimine (PEI) and Lipofectamine 2000 (Invitrogen) transfection reagents respectively. HEK293 cells were transfected using a PEI to DNA ratio of 3:1 and iAstrocytes were transfected using a Lipofectamine 2000 to DNA ratio of 2:1.

Table 2.12. Plasmid transfection of HEK293 cells with PEI

Culture Vessel	PEI per well (µl)	PEI Opti-MEM volume per well (µl)	DNA per well (µg)	DNA Opti-MEM volume per well (µl)
24-well plate	1.5	25	0.5	25
6-well plate	6	100	2	100
10 cm dish	60	500	10	500

Table 2.13. Plasmid transfection of iAstrocyte cells with Lipofectamine 2000

Culture Vessel	Lipofectamine 2000 per well (µl)	Lipofectamine 2000 Opti-MEM volume per well (µl)	DNA per well (µg)	DNA Opti-MEM volume per well (µl)
24-well plate	1	25	0.5	25

The stated volume of transfection reagent was added to Opti-MEM, vortexed and incubated at RT for 5 minutes. Separately, the plasmid DNA was diluted in Opti-MEM. The diluted transfection reagent was added to the diluted DNA and incubated at RT for 20 minutes. The transfection mix was added drop-wise onto cells and incubated at 37°C/5% CO_2 for 6 hours. The transfection media was removed from cells and replaced with fresh complete DMEM or iAstrocyte media. Cells were incubated at 37°C/5% CO_2 for 24 hours to allow for protein expression before fixing or harvesting.

2.2.2.5 siRNA transfection

HEK293 cells were seeded such that they were at approximately 40% confluency for next-day siRNA transfection. Dry siRNA was dissolved in sterile water to give a 100 µM stock siRNA. The stock siRNAs were diluted 1 in 10 in sterile water to generate 10 µM siRNAs which were used for transfections. siRNA was transfected into cells using Lipofectamine RNAiMax reagent (Invitrogen) following the manufacturers protocol (Table 2.14).

Table 2.14. siRNA transfection of HEK293 cells with Lipofectamine RNAiMax

Culture Vessel	Lipofectamine RNAiMax per well (µl)	Lipofectamine RNAiMax Opti-MEM volume (µl)	siRNA per well (µl)	siRNA Opti-MEM volume (µl)
24-well plate	0.6	25	0.6	25

The stated volume of Lipofectamine RNAiMax was added to Opti-MEM, vortexed and incubated at RT for 5 minutes. Separately, the siRNA was diluted in Opti-MEM. The diluted Lipofectamine RNAiMax was added to the diluted siRNA and incubated at RT for 20 minutes. The transfection mix was added drop-wise onto cells and incubated at 37°C/5% CO_2 for 6 hours. The transfection media was removed from cells and replaced with fresh complete DMEM. After 96 hours, cells were fixed for immunofluorescence or harvested for protein or RNA. If the experiment required it, plasmid transfection was performed 72 hours post siRNA transfection. Cells were fixed for immunofluorescence or harvested for protein or RNA 24 hours post plasmid transfection.

2.2.2.6 Viral transduction of rat primary neurons

Rat primary neurons were transduced by Miss Rebecca Cohen at 5 DIV. Neurobasal media was removed and cells received 4 transducing units per cell of lentivirus construct diluted in 250 µl complete neurobasal media. Cells were incubated at 37°C/5% CO_2 for 16 hours before the media was replaced with a 1:1 mix of fresh complete neurobasal media and conditioned complete neurobasal media. Neurons were incubated at 37°C/5% CO_2 for 7 days before fixing for immunofluorescence or harvested for protein or RNA.

2.2.3 Biochemical Methods

2.2.3.1 Whole cell lysis

Media was removed and cells washed with PBS. Cells were detached from the culture vessel by addition of Trypsin/EDTA solution for full coverage of the cells and incubation at 37°C/5% CO_2 for 2 minutes. The cells were harvested and trypsin quenched by addition of DMEM. The cell suspension was centrifuged at 400 g for 4 minutes at 4 °C to pellet cells. Cells were washed by removal of media and gentle

resuspension in PBS. Cells were pelleted by centrifugation at 400 g for 4 minutes at 4 °C. Cells were lysed by resuspension in RIPA buffer and incubation on ice for 30 minutes. Lysates were clarified by centrifugation at 14,000 rpm for 30 mins at 4 °C in a Sigma 1-15 PK centrifuge using a 12124 rotor and the supernatant transferred to a fresh tube.

2.2.3.2 Nuclear fractionation

Media was removed and cells washed with PBS. Cells were detached from the culture vessel by addition of Trypsin/EDTA solution for full coverage of the cells and incubation at 37°C/5% CO_2 for 2 minutes. The cells were harvested and trypsin quenched by addition of DMEM. Cells were pelleted by centrifugation at 400 g for 4 minutes at 4 °C. The media was removed and cells washed by gentle resuspension in PBS. If a total lysate sample was required, one third of the PBS cell suspension was removed and whole cell lysis performed. The remaining PBS cell suspension was centrifuged at 400 g for 4 minutes at 4 °C to pellet cells. The PBS was removed and the pellet washed quickly with hypotonic lysis buffer to remove residual salt. The hypotonic lysis buffer lyses the cell membrane but leaves the nuclear membrane intact. The cell pellet was gently resuspended in hypotonic lysis buffer and incubated on ice for 30 minutes to lyse the cell membrane. Nuclei were pelleted by centrifugation at 4,000 rpm for 3 minutes at 4 °C in a Sigma 1-15 PK centrifuge using a 12124 rotor. The supernatant 'cytoplasmic fraction' was transferred to a fresh eppendorf. The nuclear pellet was washed twice with hypotonic lysis buffer by gentle resuspension and centrifugation at 4,000 rpm for 3 minutes at 4 °C in a Sigma 1-15 PK centrifuge using a 12124 rotor to remove any residual cytoplasmic fraction. The nuclear pellet was lysed by resuspension in RIPA buffer and incubation on ice for 30 minutes. The nuclear lysate was clarified by centrifugation at 14,000 rpm for 30 mins at 4 °C in a Sigma 1-15 PK centrifuge using a 12124 rotor and the supernatant transferred to a fresh eppendorf.

2.2.3.3 Bradford protein assay

The Bradford protein assay was used to measure the protein concentration of cell lysates following whole cell lysis or nuclear fractionation (Bradford, 1976). Bradford reagent was diluted 1 in 5 with water and used to make a 1 in 100 dilution of 1 mg/ml BSA. Two-fold serial dilutions of BSA were completed on a 96-well flat-bottom plate to produce a standard curve ranging from 1000 ng/ml to 0.5 ng/ml.

Lysates were diluted 1 in 100 in diluted Bradford reagent and two-fold serial dilutions completed on the 96-well plate. The final volume in each well was 100 µl and diluted Bradford reagent alone was used as a blank. The A_{595} was measured using the PHERAstar FS plate reader (BMG Labtech) and the A_{595} of the blank was subtracted from the A_{595} of standards and samples. The standard curve was generated by plotting the A_{595} against protein concentration and the gradient of the curve measured. The protein concentration of lysates was determined using the equation: concentration = (A_{595} x curve gradient x dilution factor).

2.2.3.4 Immunoprecipitation

For input samples, lysates were diluted in water and 5X Laemmli buffer to a 1mg/ml protein concentration in a final volume of 50 µl (e.g. 10 µl lysate at 5 mg/ml + 10 µl 5X Laemmli + 30 µl water). The remaining lysate was used for immunoprecipitation (IP). IP samples were diluted in RIPA buffer to have a minimum of 1 mg total protein in a final volume of 750 µl (e.g. 200 µl lysate at 5 mg/ml + 550 µl RIPA buffer). 1 µg of antibody for protein immunoprecipitation was added per IP sample and incubated for 16 hours at 4 °C with agitation. A 50% slurry of Protein G Sepharose beads in RIPA buffer was prepared. 30 µl of the bead slurry was added per IP sample and incubated for 3 hours at 4 °C with agitation. The beads were pelleted by centrifugation at 3,000 g for 3 mins at 4 °C. The supernatant was discarded and the beads washed five times by gentle resuspension in RIPA buffer and centrifugation at 3,000 g for 3 mins at 4 °C. After the final wash, care was taken to remove all residual supernatant from the beads using a gel-loading tip. Protein was eluted from the beads by addition of 25 µl of 2X Laemmli buffer and boiling for 5 minutes at 100 °C. The complete 25 µl elution was analysed by immunoblot.

2.2.3.5 *In vitro* GST-binding assay

^{35}S-methionine radiolabeled coilin was generated using the T7 TnT® Quick Coupled Transcription/Translation System (Promega) from the pCI-neo-myc-coilin construct which contained a T7 promoter. The 10 µl reaction was prepared as follows behind a plastic shield to protect against radiation:

1.5 µl pCI-neo-myc-coilin plasmid (333 ng/µl)
0.5 µl ^{35}S-Methionine (PerkinElmer)
8 µl Reticulocyte lysate

The reaction was incubated at 37 °C for 90 minutes. Frozen Rosetta pLysS bacterial cells expressing the GST-tagged proteins were weighed into eppendorfs. GST, GST-C9orf72S and GST-C9orf72L cells were weighed out at 0.1 g, 0.2 g and 0.3 g respectively due to differences in protein expression. 1 ml of ice-cold RB100 buffer was added and cells lysed by sonication until no visible clumps remained. The lysates were centrifuged at 17,000 g for 5 minutes at 4 °C to pellet cell debris. 1 ml of the clarified lysate was transferred to separate tubes containing 30 µl of 50% glutathione-agarose (GSH) bead slurry in RB100 buffer. The lysates were incubated with the beads for 30 minutes at 4 °C with agitation. Beads were pelleted by centrifugation at 500 g for 1 minute at 4 °C and the bacteria lysate discarded. The beads were washed twice by resuspension in RB100 buffer and centrifugation at 500 g for 1 minute at 4 °C. Beads were resuspended in 500 µl RB100 buffer and 8 µl of the ^{35}S-coilin reticulocyte reaction was added and incubated for 1 hour at 4 °C with agitation. Beads were pelleted and washed twice by resuspension in RB100 buffer and centrifugation at 500 g for 1 minute at 4 °C. Care was taken to remove any residual liquid and proteins eluted by adding 40 µl of GSH elution buffer and incubating at RT for 10 minutes with frequent agitation. The GST control was prepared for immunoblot by taking 1 µl of the GST elution and adding 15 µl of GSH elution buffer and 4 µl of 5X Laemmli buffer. For GST-C9orf72S and GST-C9orf72L, immunoblot samples were prepared by diluting 16 µl of elution with 4 µl of 5X Laemmli buffer. Input samples were prepared by diluting 0.5 µl of the ^{35}S-coilin reticulocyte reaction in 15.5 µl of GSH elution buffer and 4 µl of 5X Laemmli buffer. The whole of the 20 µl samples were subjected to SDS-PAGE electrophoresis followed by coomassie staining of the gel for GST-protein visualisation. The gel was dried and backed onto Whatman filter paper and exposed to a Phosphor Screen for 14 days. A Typhoon Phosphor Imager was used to detect radioactivity and digitise results.

2.2.3.6 SDS-PAGE electrophoresis

Protein samples were boiled at 100 °C for 5 minutes in Laemmli buffer to denature proteins for SDS-PAGE electrophoresis. A 10% polyacrylamide resolving gel was prepared. After the resolving gel had set, a 6% polyacrylamide stacking gel was prepared on top of it with a 10- or 15-well comb depending on the number of samples to be run. The gel was placed into a Mini-PROTEAN Tetra Cell tank (Biorad) and immersed in running buffer. Protein samples were loaded into wells

and subjected to electrophoresis at 100 V for approximately 2 hours. 3 µl of Precision Plus Protein All Blue Standard was loaded as a molecular weight marker.

2.2.3.7 Coomassie staining of polyacrylamide gels

After SDS-PAGE electrophoresis of the samples from the *in vitro* GST-binding assay, the polyacrylamide gel was stained with Coomassie G-250 by heating in a microwave for 15 s and incubating at RT for 20 minutes with agitation. The Coomassie was replaced with destain solution and microwaved for 15 s. The gel was incubated in destain at RT for 20 minutes and the destain replaced with fresh destain solution every 20 minutes until the GST-proteins were clearly visible and the gel clear. The gel was dried and backed onto Whatman paper and exposed to a Phosphor Screen for 14 days. A Typhoon Phosphor Imager was used to detect radioactivity and digitise results.

2.2.3.8 Electrophoretic transfer of proteins

For immunoblotting, proteins were electrotransferred from the polyacrylamide gel onto nitrocellulose membrane. The gel and nitrocellulose membrane were sandwiched in a gel holder cassette as follows: sponge, 2 filter papers, polyacrylamide gel, nitrocellulose membrane, 2 filers papers, sponge. The cassette was orientated into a Mini Trans-Blot Cell (Biorad) such that the protein transfer from gel to nitrocellulose was aligned with the electric current from the cathode towards the anode. The cassette was submerged in transfer buffer and subjected to 100 V for a 1 hour transfer or 30 V for a 16 hour transfer. The membrane was stained with Ponceau S for protein detection and to allow for more accurate cutting of the membrane if required.

2.2.3.9 Immunoblot of nitrocellulose membrane

Membranes were blocked in 5% milk in TBST for 1 hour at RT. Primary antibody was diluted in 5% milk in TBST and added to the membrane for 1 hour at RT or for 16 hours at 4 °C. The membrane was washed three times with 10 minute TBST incubations with rocking. HRP-conjugated secondary antibody was diluted in TBST and added to the membrane for 1 hour at RT. The membrane was washed three times with 10 minute TBST incubations with rocking. Protein detection was performed by incubation of the blot with SuperSignal West Pico Chemiluminescent

Substrate for five minutes before placement into a film cassette. The substrate oxidises luminol in the presence of HRP and peroxide thus generating a chemiluminescent signal proportionate to the amount of HRP present which can be detected by autoradiograph film. The following steps were completed in a dark room. Autoradiograph Hyperfilm (Amersham) was positioned above the blot in the cassette. Exposure time varied from 1 s to 1 hour depending on the band intensity. The film was placed into developing solution for 1 minute, rinsed quickly with water and placed into fixing solution for 2 minutes. The film was air-dried and scanned onto a computer for analysis.

2.2.3.10 Quantification of protein and DNA bands using densitometry

Densitometry was performed on protein and DNA bands using ImageJ software (Abràmoff et al., 2004). A box was drawn around the band to be analysed using the rectangular selection. For measuring band densitometry in the splicing assays, the box was drawn around both the upper unspliced band and the lower spliced band. Histograms for each band were generated by selecting analyse > gels > plot lanes. A straight line was drawn level to the background signal and the area under the histogram was measured by selection with the wand tool. This area was taken as representative of protein level. Protein bands were normalised to the housekeeping protein GAPDH which was included as a loading control.

2.2.3.11 RNA extraction

Cells were washed once with PBS and lysed in 200 µl TRIzol™ reagent (Thermo Fisher Scientific) per well of a 24-well plate following the manufacturers protocol. 40 µl of chloroform was added and lysates shaken immediately for 15 s. Samples were incubated at RT for 10 minutes before centrifugation at 12,000 g for 15 minutes at 4 °C. Centrifugation separates the sample into three distinct liquid phases: a lower phenol-chloroform mix that contains protein, a middle interphase that contains DNA and an upper aqueous phase containing RNA. The upper aqueous phase was removed into a new RNA-free tube. 100 µl of 100% propan-2-ol was added to the aqueous phase and mixed by gentle flicking to precipitate RNA. The mix was incubated at RT for ten minutes before centrifugation at 12,000 g for 10 minutes at 4 °C to pellet the RNA. The supernatant was discarded and the RNA pellet washed by addition of 200 µl of 75% ethanol and brief vortexing. The RNA was pelleted by

centrifugation at 7,500 g for 5 minutes at 4 °C. The ethanol was removed and the pellet air dried for 10 minutes. The RNA was resuspended in 9 µl of nuclease free water by pipetting up and down and incubating at 60 °C for 15 minutes. The RNA concentration and quality was determined using the NanoDrop 1000 spectrophotometer by measuring the absorbance at 260 and 280 nm. An A_{260}/A_{280} of ~2.0 was taken as pure RNA.

2.2.3.12 RT-PCR generation of cDNA

The RNA was reverse transcribed to produce cDNA which could be quantified by qPCR. 2 µg of RNA was prepared in 8 µl of nuclease free water. 1 µl of DNaseI buffer was added to the RNA and mixed. 1 µl of DNAase I enzyme was added and incubated at 37 °C for 10 minutes to remove DNA contamination. The DNase I enzyme was inactivated by addition of 1 µl of 25 mM EDTA and incubating at 75 °C for 10 minutes. 2 µl of 25 mM dNTP mix and 1 µl of Oligo(dT)18 primer was added to the RNA and incubated at 65 °C for 5 minutes to prime the RNA. 4 µl of 5X reverse transcriptase buffer and 2 µl of 0.1 M DTT was added and incubated at 25 °C for 5 minutes. cDNA was transcribed by addition of 0.5 µl SuperScript III enzyme and incubation at 50 °C for 1 hour. The reaction was inactivated by incubation at 75 °C for 15 minutes. cDNA was stored at 4 °C.

2.2.3.13 qPCR

The cDNA was quantified by qPCR using HOT FIREPol EvaGreen® dye and gene-specific primers at an optimised dilution (detailed in Section 2.1.4.14). The following 20 µl reactions were prepared:

11 µl Nuclease free water
4 µl HOT FIREPol EvaGreen® qPCR mix
2 µl Forward primer at optimised dilution
2 µl Reverse primer at optimised dilution
1 µl of 1 in 10 dilution of cDNA

Each cDNA sample was run in triplicate on the CFX96 Real-Time PCR Detection System (Biorad). The qPCR thermocycling conditions were as follows:

Table 2.15. Thermocycling conditions for qPCR

Step	Cycles	Temp	Time
Initial denaturation	1	95 °C	10 mins
Denaturation	35	95 °C	30 s
Annealing/Extension		60 °C	1 min

The Ct values from the cDNA triplicates were averaged and normalised to GAPDH or RPL19 mRNA levels for human or rat mRNA respectively.

2.2.3.14 PCR amplification of the pCI-neo intron

The pCI-neo intron was PCR amplified using FIREPol® DNA polymerase, primers described in Section 2.1.4.15 and either pCI-neo plasmid or cDNA from cells as the DNA template. The following 10 µl reactions were prepared:

6.8 µl Nuclease free water
2 µl FIREPol® Master Mix Ready To Load
0.1 µl Forward primer (100 µM)
0.1 µl Reverse primer (100 µM)
1 µl pCI-neo plasmid (100 ng/µl) OR 1 in 10 dilution of cDNA

cDNA was run in triplicate and subjected to the following thermocycling conditions:

Table 2.16. Thermocycling conditions for pCI-neo intron amplification

Step	Cycles	Temp	Time
Initial Denaturation	1	95 °C	3 mins
Denaturation	15	95 °C	30 s
Annealing (Touchdown)		65 °C to 50 °C	45 s
Extension		72 °C	1 min
Denaturation	25	95 °C	30 s
Annealing		58 °C	45 s
Extension		72 °C	1 min
Final Extension	1	72 °C	10 mins

The whole 10 µl reaction was loaded onto an agarose gel and subjected to electrophoresis. Unspliced and spliced DNA bands were quantified as described in Section 2.2.3.10.

2.2.3.15 PCR amplification of the P120 intron

The P120 intron was PCR amplified using FIREPol® DNA polymerase, primers described in Section 2.1.4.15 and either P120 minigene plasmid or cDNA from cells as the DNA template. The following 10 µl reactions were prepared:

6.8 µl Nuclease free water
2 µl FIREPol® Master Mix Ready To Load
0.1 µl Forward primer (100 µM)
0.1 µl Reverse primer (100 µM)
1 µl P120 minigene plasmid (100 ng/µl) OR 1 in 10 dilution of cDNA

cDNA was run in triplicate and subjected to the following thermocycling conditions:

Table 2.17. Thermocycling conditions for P120 intron amplification

Step	Cycles	Temp	Time
Initial Denaturation	1	95 °C	3 mins
Denaturation	15	95 °C	30 s
Annealing (Touchdown)		65 °C to 50 °C	45 s
Extension		72 °C	1 min
Denaturation	10 exogenous P120/ 25 endogenous P120	95 °C	30 s
Annealing		58 °C	45 s
Extension		72 °C	1 min
Final Extension	1	72 °C	10 mins

The whole 10 µl reaction was loaded onto an agarose gel and subjected to electrophoresis. Unspliced and spliced DNA bands were quantified as described in Section 2.2.3.10.

2.2.4 Microscopy

2.2.4.1 Immunofluorescence

Cells were plated onto 13 mm glass coverslips for immunofluorescence experiments. Media was removed and cells washed gently with PBS. Cells were fixed with 4% formaldehyde in PBS for 20 minutes. Cells were washed twice with PBS to remove formaldehyde. Cells were permeabilised with 0.2% Triton-X100 in PBS for 3 minutes. Cells were washed twice with PBS to remove the Triton and blocked with 5% Normal Goat Serum (NGS) in PBS for 1 hour. Primary antibody was diluted in 5% NGS in PBS (Table 2.5) and incubated with cells for 1 hour at RT. Cells were washed three times with PBS to remove primary antibody. Fluorophore-conjugated secondary antibody and DNA stain Hoechst 33342 were diluted in 5% NGS in PBS and incubated with cells for 1 hour at RT. Cells were washed three times with PBS to remove secondary antibody. Coverslips were mounted onto glass microscope slides using fluorescence mounting medium. Slides were stored at 4 °C before visualisation with a 63x Plan Apochromat 1.4NA objective on a Zeiss Axioplan 2 microscope installed with a Zyla 4.2 sCMOS camera. Images were collected using MicroManager 1.4 software and analysed using ImageJ software (Edelstein et al., 2014).

2.2.4.2 Proximity Ligation Assay

The proximity ligation assay was performed using the DuoLink® PLA Technology reagents (Sigma) following the manufacturers protocol. Cells were plated onto 13 mm glass coverslips for PLA experiments. Media was removed and cells washed gently with PBS. Cells were fixed with 4% Formaldehyde in PBS for 20 minutes then washed twice with PBS to remove formaldehyde. Cells were permeabilised with 0.2% Triton-X100 in PBS for 3 minutes then washed twice with PBS to remove Triton. Cells were blocked using the Blocking Solution for 30 minutes at 37 °C. Primary antibody was diluted in 5% NGS in PBS (Table 2.6) and incubated with cells for 1 hour at RT. Cells were washed twice with Wash Buffer A to remove primary antibody. Secondary PLA probes were diluted in Antibody Diluent and incubated with cells for 30 minutes at 37 °C. Cells were washed twice with Wash Buffer A to remove PLA probes. The PLA probes were ligated by incubation of cells with the PLA ligation reagents for 30 minutes at 37 °C. Cells were washed twice with Wash Buffer A and rolling circle amplification initiated by incubation of cells with

the PLA amplification reagents for 100 minutes at 37 °C. Cells were washed once with 1X Wash Buffer B and incubated with Hoechst 33342 diluted in 0.01X Wash Buffer B for 15 minutes to stain nuclei. Cells were washed once with 0.01X Wash Buffer B and mounted onto glass microscope slides using fluorescence mounting medium. Images were recorded as described in Section 2.2.4.1. Cells were selected for image capture if they were GFP-positive by searching on the 488 nm channel.

2.2.4.3 Cajal body quantification

Images were opened on ImageJ software for analysis. Cajal bodies were manually quantified by counting the number of coilin foci per nucleus. For non-transfected cells, 10 cells were selected randomly per field of view for Cajal body quantification. For transfected cells, Cajal bodies were counted in cells which had appropriate expression shown by immunofluorescence using an antibody against the transfected protein.

2.2.4.4 PLA signal quantification

Images were opened on ImageJ software for analysis. A region of interest (ROI) was drawn around either the whole cell on the GFP channel or the nucleus on the DAPI channel. On the PLA signal channel, the brightness and contrast of images was edited by setting the minimum and maximum display range to a determined value such that PLA signals appeared as bright visible puncta with no background signal. The ROI was transferred onto the PLA signal channel and the number of PLA signals contained within the ROI were manually counted.

2.2.4.5 Performing statistical tests

All statistical tests were performed using Prism 7 software (GraphPad Software Inc., San Diego, CA). Specific details of tests are given in the appropriate sections of the results chapters.

3 Optimisation and development of tools for the project

3.1 Introduction

A GGGGCC hexanucleotide repeat expansion in the C9orf72 gene is the leading cause of familial ALS. The mutation is thought to cause disease via three non-mutually exclusive mechanisms: haploinsufficiency of the C9orf72 protein (loss of function), generation of toxic RNA foci and dipeptide repeat proteins (gain of function). Decreased levels of C9orf72 mRNA and protein in ALS patient brains support haploinsufficiency of the C9orf72 protein as a disease contributor. There have been recent advancements in understanding the cellular functions and localisation of the C9orf72 protein, but a role for C9orf72 in the nucleus is yet to be discovered.

The aim of this project was to increase understanding of the nuclear function of C9orf72. To do this, C9orf72 protein overexpression and knockdown experiments were frequently used to look at the effect of C9orf72 on Cajal bodies and splicing. In addition to transient C9orf72 knockdowns, a stable C9orf72 knockout cell line was used. This chapter describes the characterisation of these CRISPR/Cas9 generated C9orf72 mutant cell lines, and the development and optimisation of tools used throughout the project.

3.2 Results

3.2.1 Characterisation of stable C9orf72 knockout cell lines

Haploinsufficiency of the C9orf72 protein is one of three mechanisms proposed to contribute to ALS pathogenesis. In trying to understand what the function of C9orf72 might be, it can be useful to knockdown the C9orf72 protein and probe what effect this has on different cellular pathways. This helps inform not only the pathway which C9orf72 may be involved in, but also pinpoints what might be occurring in disease. Transient knockdown of C9orf72 can be achieved with some efficiency by transfection of cell lines with C9orf72-targeting siRNAs. However, siRNAs can have off-target effects and the acute nature of transfections is not as representative of

physiological conditions as a stable knockout. For these reasons, we used CRISPR/Cas 9 genome editing to try and establish a C9orf72 knockout cell line.

Dr Christopher Webster co-transfected HEK293 cells with two Cas9 nickase plasmids. Each plasmid expressed the Cas9 nickase enzyme, puromycin resistance and different C9orf72 exon 2 complimentary guide RNAs (gRNA). The gRNAs associate with the Cas9 nickase and target the enzyme to the C9orf72 exon 2 DNA. After DNA binding, each nickase cuts a single DNA strand creating a 'nick'; together, these two nicks create a double strand break (DSB). The DSB is repaired by inefficient non-homologous end joining which introduces errors into the genome via insertions or deletions (INDELs) of base pairs (Mali et al., 2013; Ran et al., 2013). If the inserted or deleted base pairs lead to a frame shift in the open reading frame (ORF), a premature stop codon is often introduced. Premature stop codons signal to the cells nonsense-mediated decay surveillance pathway to target the mRNA for degradation, thereby preventing translation and creating a loss-of-function mutant (Hug et al., 2016). Having two gRNAs targeting C9orf72 exon 2 increased the specificity of the gene editing technology, and was hypothesised to lead to loss of both the short and long C9orf72 protein isoforms. Cells were selected for successful transfection by growth in puromycin containing media. Living cells were seeded at increasing dilutions until isogenic cell colonies were apparent. Cell cultures were scaled up from these colonies and were preliminarily screened by Dr Christopher Webster for C9orf72 mRNA levels as measured by RT-qPCR. From this, three clones were selected for further characterisation: B2, B17 and B12.

To determine if the C9orf72 gene in the B2, B17 and B12 cell lines was successfully edited, I performed DNA sequencing of C9orf72 exon 2. HEK293 cells are pseudotriploid and therefore contain three copies of the C9orf72 gene (Bylund et al., 2004; Lin et al., 2014). Genomic DNA was extracted from the C9orf72 edited CRISPR lines. A PCR amplifying C9orf72 exon 2 was performed on the genomic DNA using the Phusion High-Fidelity DNA polymerase to create a C9orf72 exon 2 blunt-ended product. The PCR product was resolved on an agarose gel and gel extracted and purified. The blunt-ended C9orf72 exon 2 product was ligated into the pCR-Blunt II-TOPO vector and transformed into XL10-Gold Ultracompetent cells. Since HEK293 cells are triploid, each pCR-Blunt II-TOPO-C9orf72 exon 2 construct was hypothesised to contain one of three possible C9orf72 exon 2 sequences. Plasmid DNA was purified from bacterial colonies and sequenced using the SP6 forward primer located in the pCR-Blunt II-TOPO vector which would give full

sequencing coverage of the C9orf72 exon 2 (Figure 3.1). We sequenced 9 colonies per stable cell line. If we did not detect an appropriate ratio of alleles representing a triploid genome, we sequenced additional colonies.

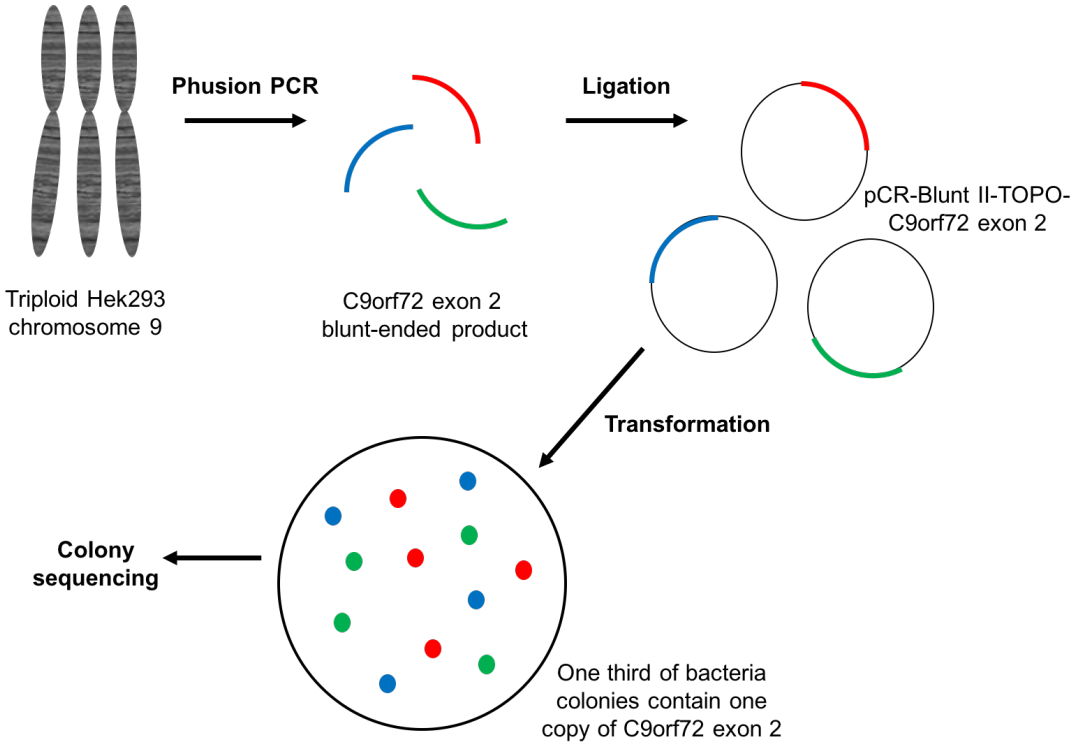
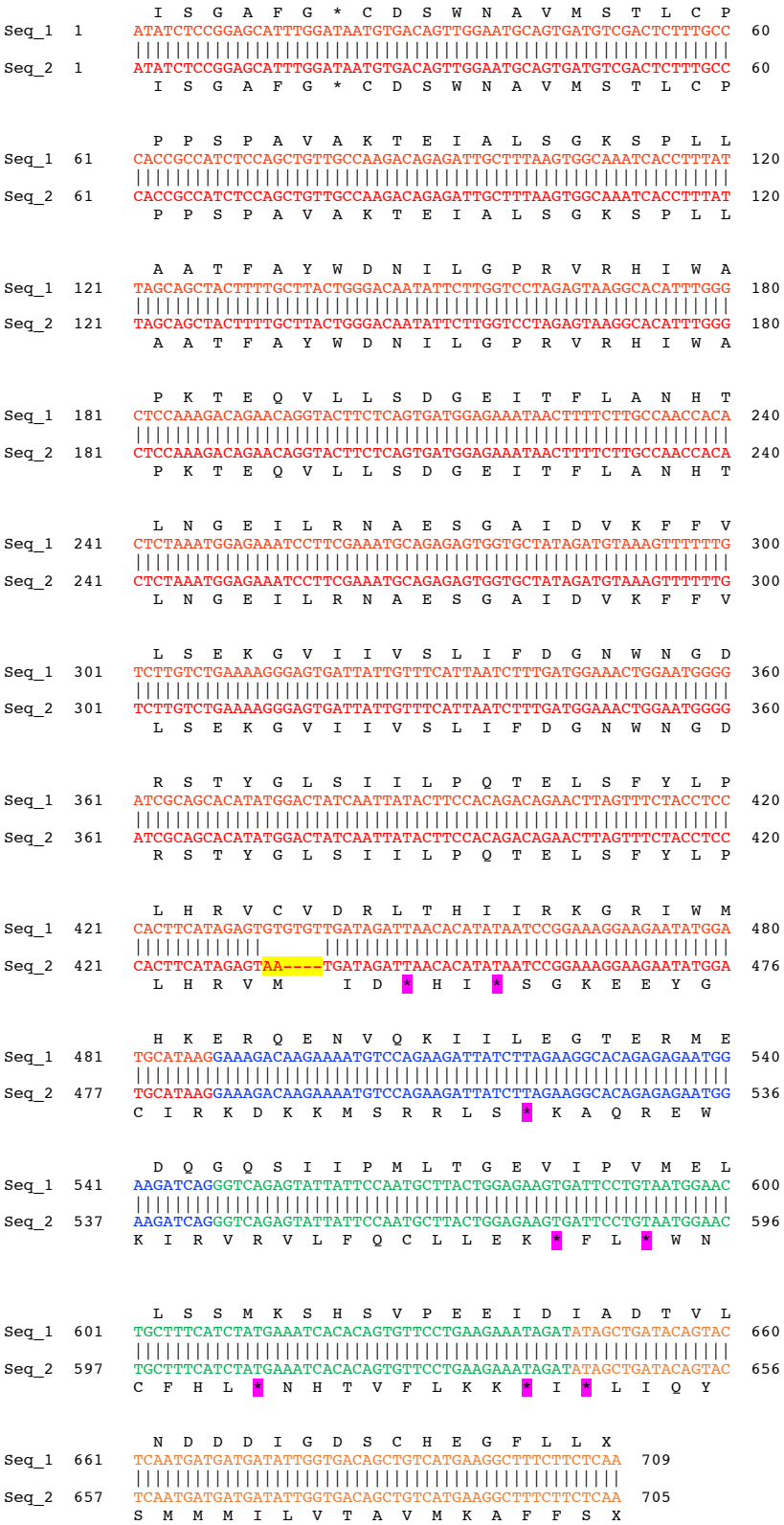


Figure 3.1. Cloning of CRISPR edited C9orf72 exon 2 for sequencing. The C9orf72 exon 2 was amplified from genomic DNA extracted from the C9orf72 knockout CRISPR cell lines and ligated into the pCR-Blunt II-TOPO vector. Constructs were transformed into XL10-Gold Ultracompetent cells before minipreparation of plasmid DNA for sequencing of the C9orf72 exon 2 insert.

For the B2 CRISPR line, nine colonies were successfully sequenced (Table 3.1). Out of the nine colonies sequenced, 5 colonies exhibited a +2/-4 bp mutation (Figure 3.2), 2 colonies a -10 bp mutation (Figure 3.3) and 2 colonies a -35 bp mutation (Figure 3.4). This suggests all three copies of C9orf72 in the B2 clone were successfully edited and each had acquired a different INDEL mutation. The three mutant C9orf72 sequences were aligned with the wild-type C9orf72 sequence and all three mutations were found to lead to a frameshift in the C9orf72 ORF and coded for premature stop codons (Figure 3.2-3.4). Therefore, it was hypothesised based on sequence that the B2 clone should have no translated C9orf72 protein and was a stable C9orf72 knockout cell line.

B2 mutation	Number of colonies with mutation	% of colonies with mutation	Premature stop codon
+2/-4 bp	5	55.5	Yes
-10 bp	2	22.2	Yes
-35 bp	2	22.2	Yes

Table 3.1. C9orf72 exon 2 mutations in the B2 CRISPR line. Plasmids from nine bacteria colonies were successfully sequenced for their C9orf72 exon 2 insert. In the nine C9orf72 sequences returned, three INDEL mutations recurred: +2/-4 bp, -10 bp and -35 bp. All three mutants were found to produce a premature stop codon in their sequence.



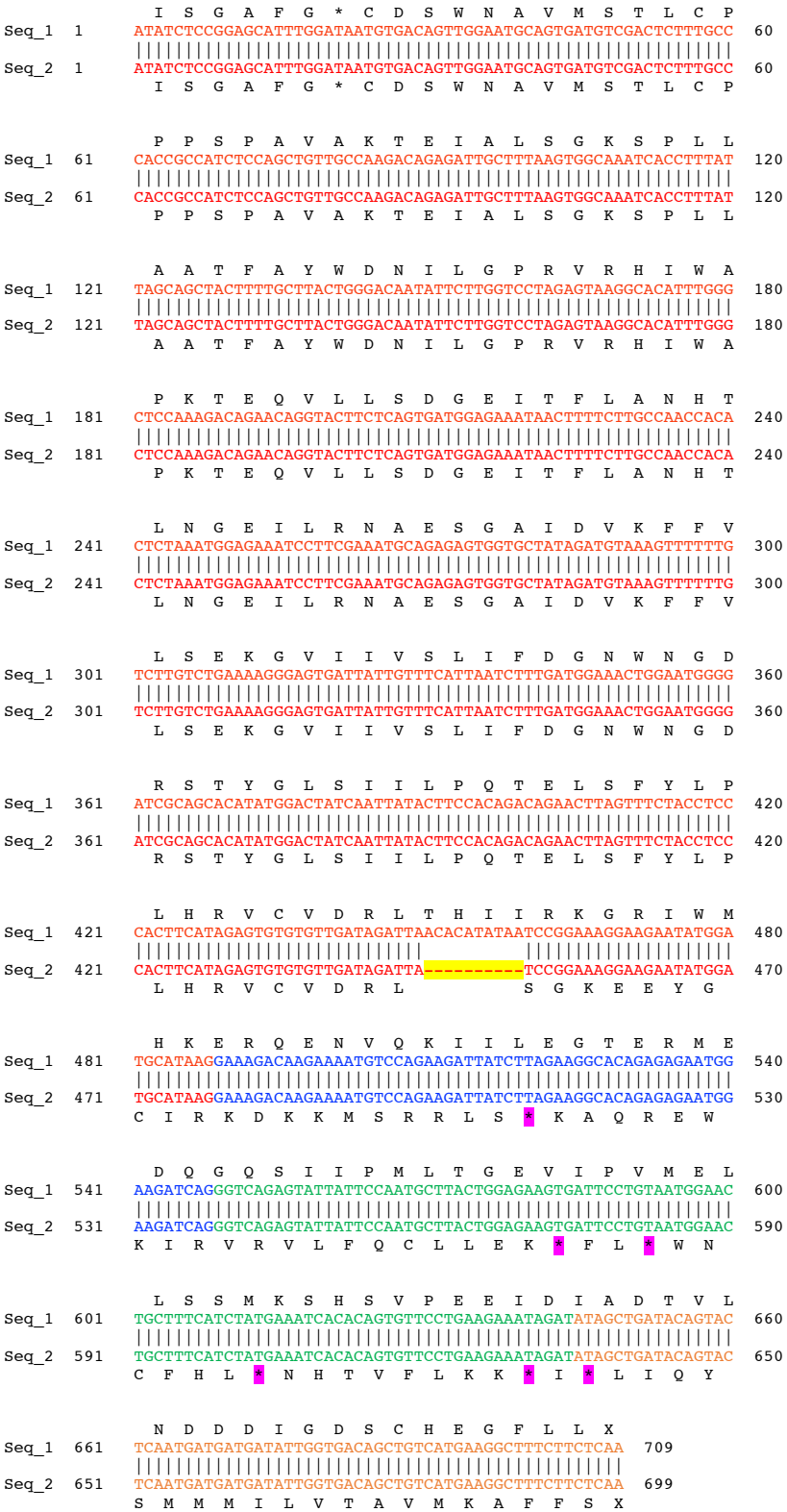
Key:

Exon 2
Exon 3
Exon 4
Exon 5

INDEL

* premature stop codon

Figure 3.2. Sequence alignment of B2 +2/-4 bp mutant with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the B2 +2/-4 bp mutant (Seq_2). Exon 2 (red), exon 3 (blue), exon 4 (green) and exon 5 (orange) are shown. The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and B2 C9orf72 respectively. Premature stop codons are labelled with an asterisk (pink). The B2 INDEL is highlighted in yellow.



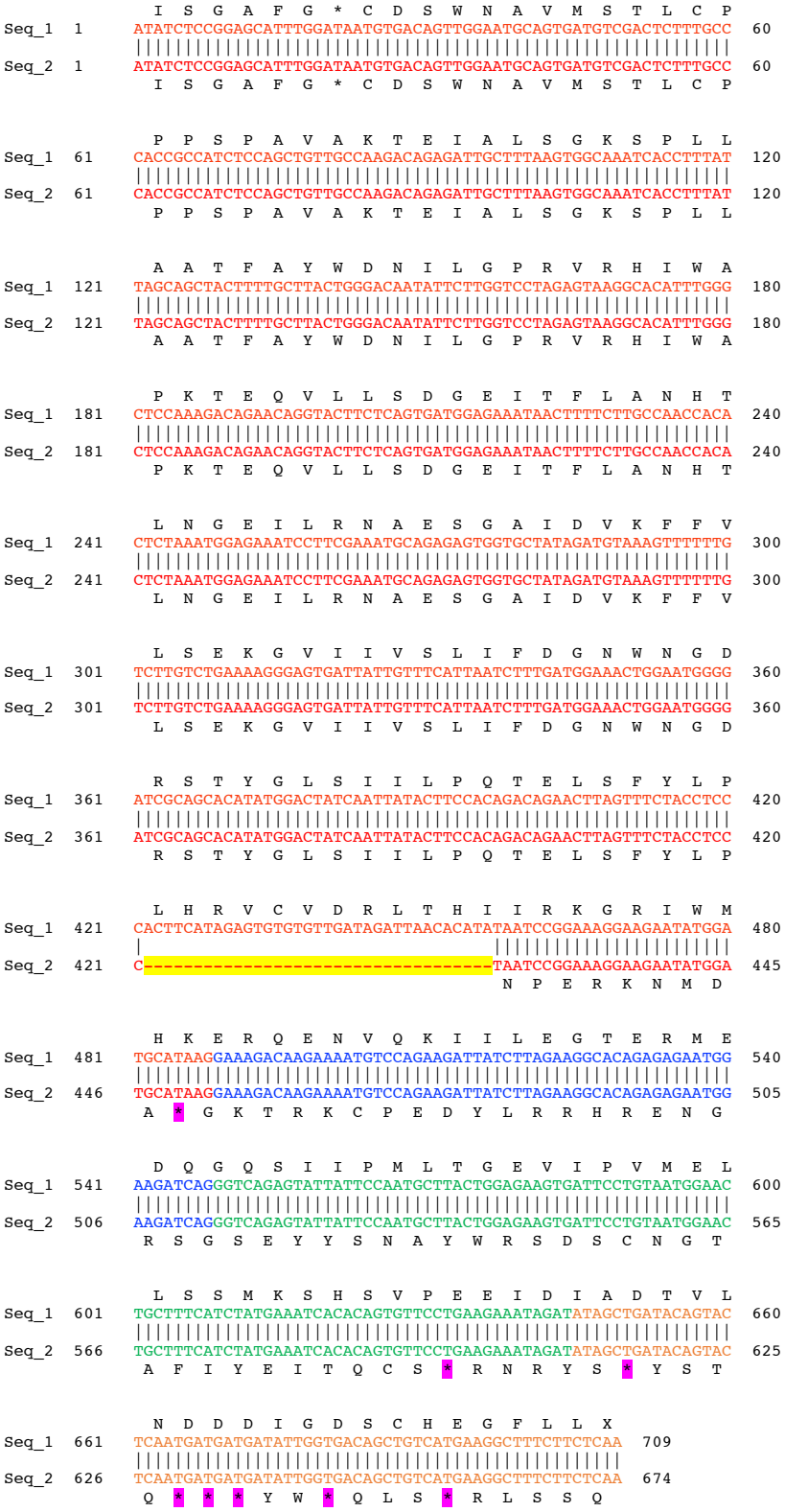
Key:

Exon 2
Exon 3
Exon 4
Exon 5

INDEL

* premature stop codon

Figure 3.3. Sequence alignment of B2 -10 bp mutant with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the B2 -10 bp mutant (Seq_2). Exon 2 (red), exon 3 (blue), exon 4 (green) and exon 5 (orange) are shown. The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and B2 C9orf72 respectively. Premature stop codons are labelled with an asterisk (pink). The B2 INDEL is highlighted in yellow.



Key:

Exon 2
Exon 3
Exon 4
Exon 5

INDEL

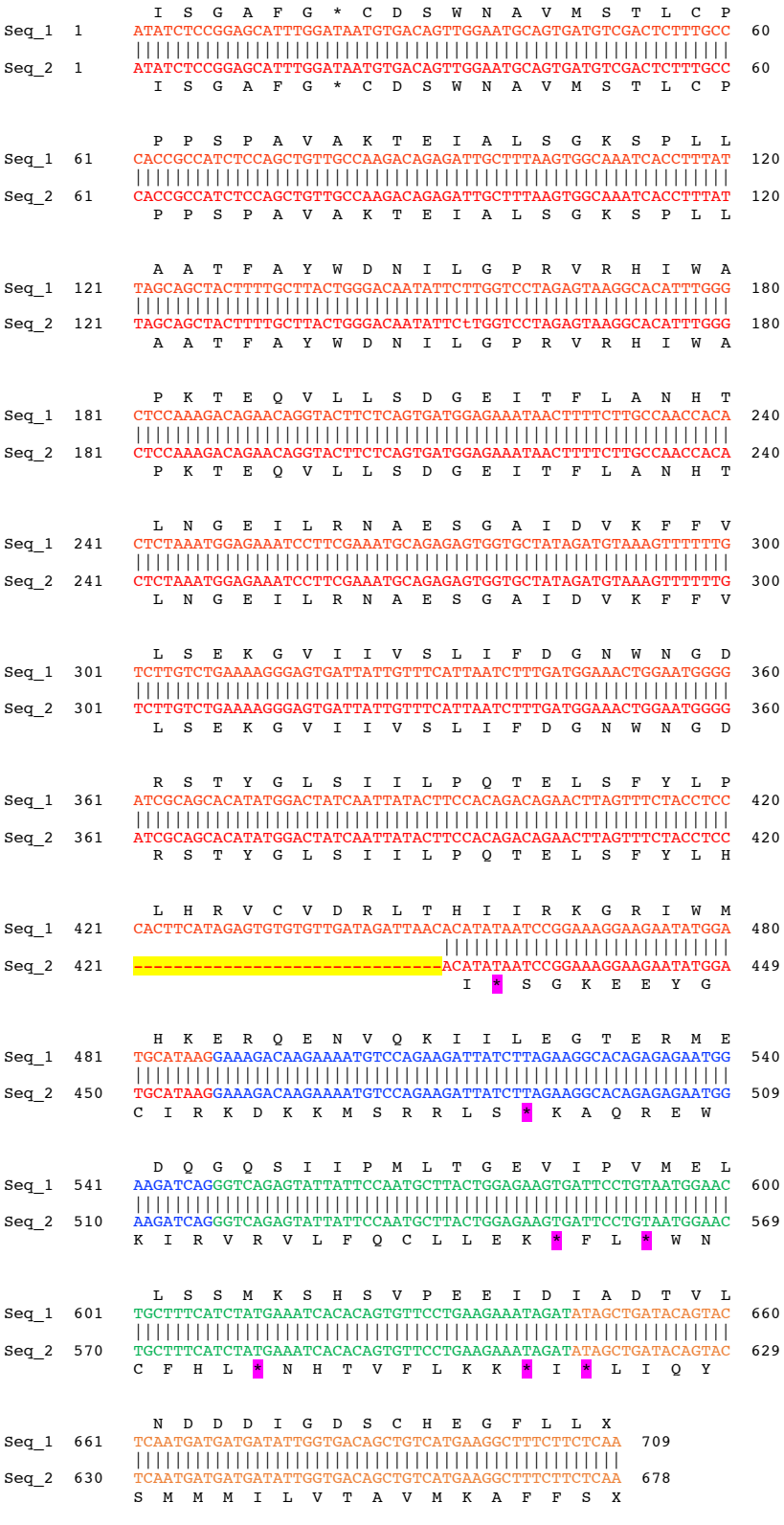
* premature stop codon

Figure 3.4. Sequence alignment of B2 -35 bp mutant with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the B2 -35 bp mutant (Seq_2). Exon 2 (red), exon 3 (blue), exon 4 (green) and exon 5 (orange) are shown. The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and B2 C9orf72 respectively. Premature stop codons are labelled with an asterisk (pink). The B2 INDEL is highlighted in yellow.

For the B12 CRISPR line, fifteen colonies were successfully sequenced (Table 3.2). Out of the fifteen colonies sequenced, 5 colonies exhibited a -31 bp mutation (Figure 3.5) and 10 colonies a +48 bp mutation (Figure 3.6), suggesting two copies of the C9orf72 gene had acquired the same 48 bp insertion mutation. The two mutant C9orf72 sequences were aligned with the wild-type C9orf72 sequence to determine whether the mutations created premature stop codons (Figure 3.5-3.6). The 48 bp insertion mutant introduced novel amino acids into the C9orf72 sequence but did not create a frameshift or premature stop codon. The -31 bp mutant did create a premature stop codon. Hence, it was assumed two copies of the C9orf72 gene could produce potentially functional C9orf72 protein and therefore the B12 cell line could not be used as a C9orf72 loss-of-function cell model.

B12 mutation	Number of colonies with mutation	% of colonies with mutation	Premature stop codon
-31 bp	5	33.3	Yes
+48 bp	10	66.6	No

Table 3.2. C9orf72 exon 2 mutations in the B12 CRISPR line. Plasmids from fifteen bacteria colonies were successfully sequenced for their C9orf72 exon 2 insert. In the fifteen C9orf72 sequences returned, two INDEL mutations recurred: -31 bp and +48 bp. Only the -31 bp mutant was predicted to generate a premature stop codon.



Key:

Exon 2
Exon 3
Exon 4
Exon 5

INDEL

* premature stop codon

Figure 3.5. Sequence alignment of B12 -31 bp mutant with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the B12 -31 bp mutant (Seq_2). Exon 2 (red), exon 3 (blue), exon 4 (green) and exon 5 (orange) are shown. The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and B12 C9orf72 respectively. Premature stop codons are labelled with an asterisk (pink). The B12 INDEL is highlighted in yellow.

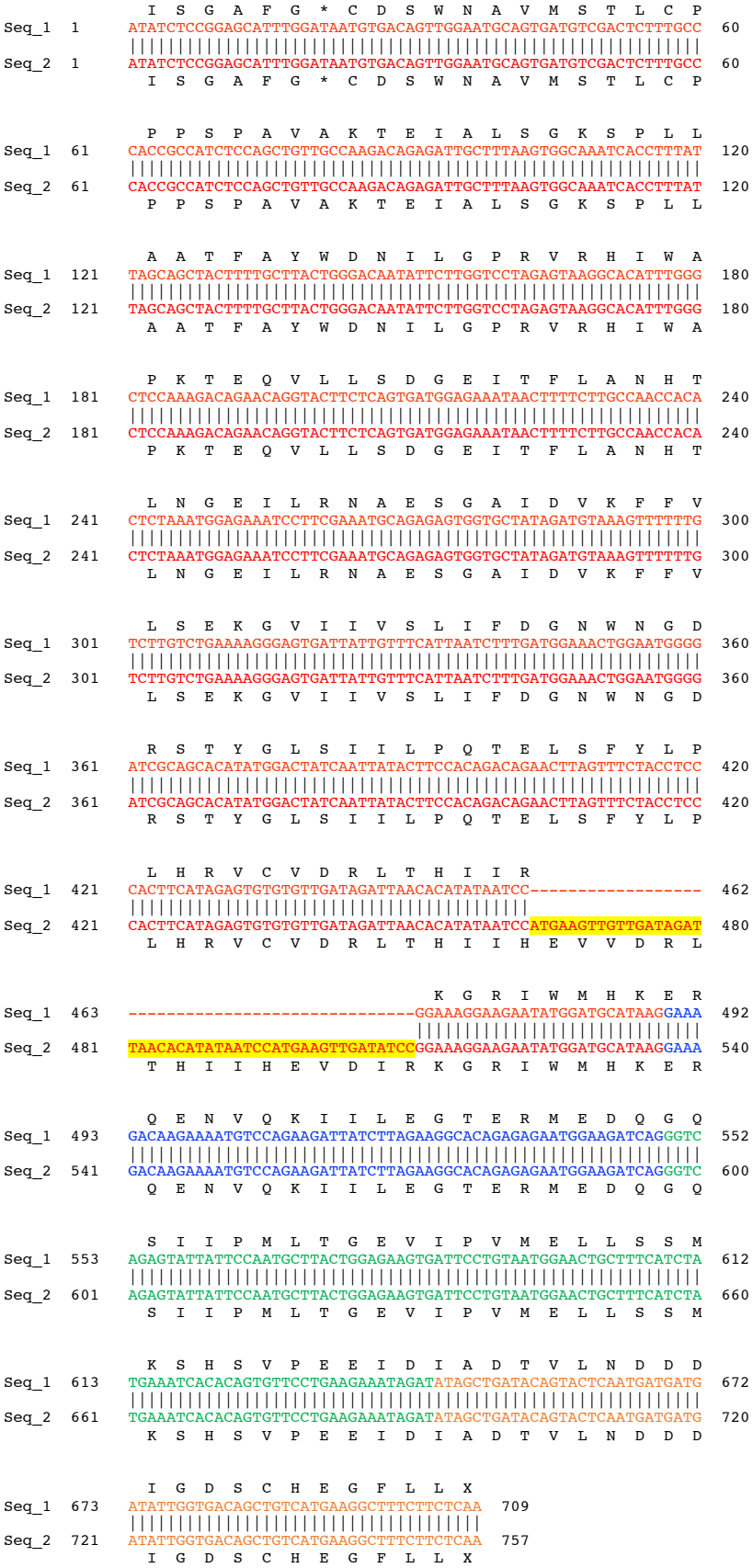
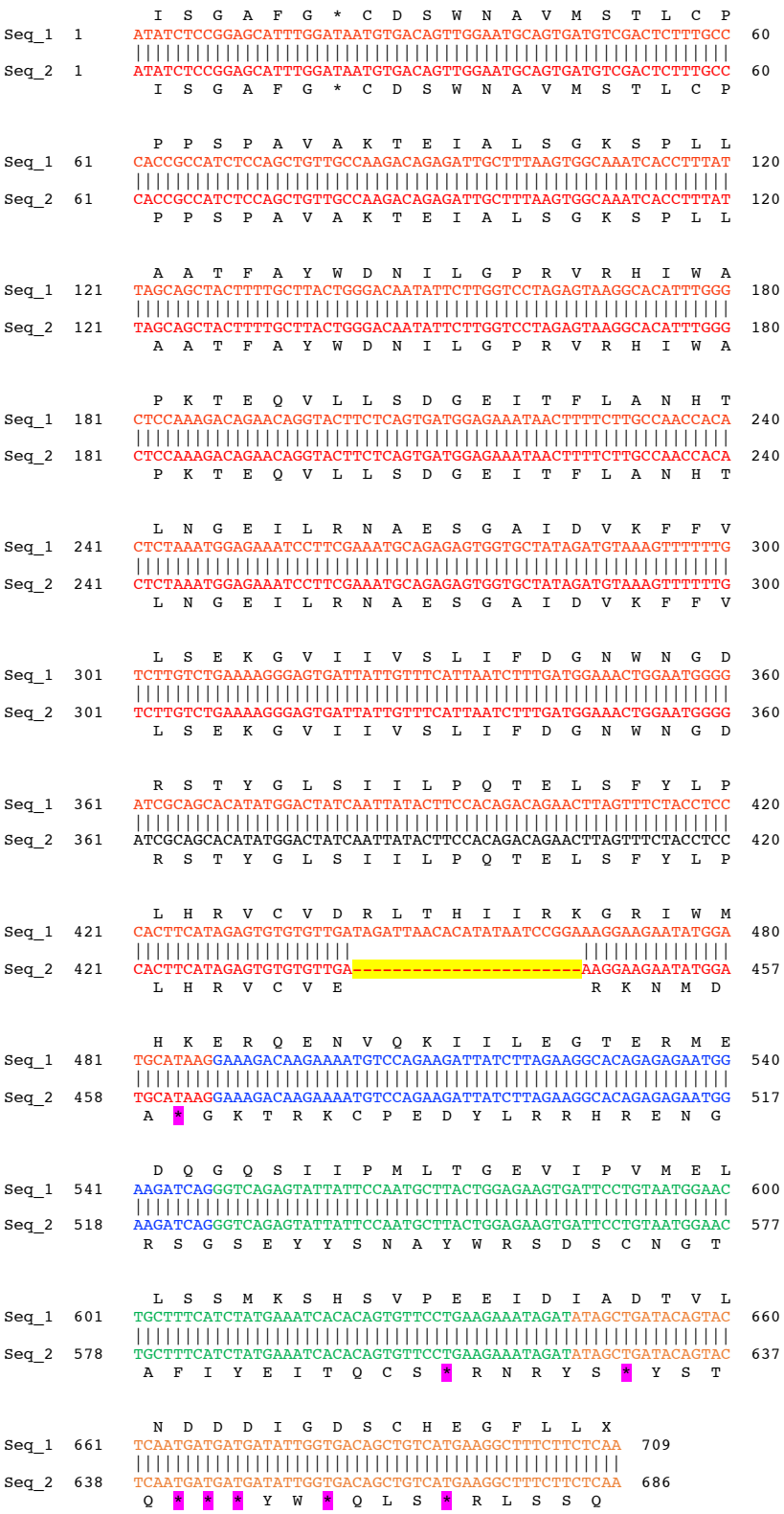

I S G A F G * C D S W N A V M S T L C P
Seq_1 1 ATATCTCCGGAGCATTTGGATAATGTGACAGTTGGAATGCAGTGATGTCGACTCTTTGCC 60
Seq_2 1 ATATCTCCGGAGCATTTGGATAATGTGACAGTTGGAATGCAGTGATGTCGACTCTTTGCC 60
I S G A F G * C D S W N A V M S T L C P
P P S P A V A K T E I A L S G K S P L L
Seq_1 61 CACCGCCATCTCCAGCTGTTGCCAAGACAGAGATTGCTTTAAGTGGCAAATCACCTTTAT 120
Seq_2 61 CACCGCCATCTCCAGCTGTTGCCAAGACAGAGATTGCTTTAAGTGGCAAATCACCTTTAT 120
P P S P A V A K T E I A L S G K S P L L
A A T F A Y W D N I L G P R V R H I W A
Seq_1 121 TAGCAGCTACTTTTGCTTACTGGGACAATATTCTTGGTCCTAGAGTAAGGCACATTTGGG 180
Seq_2 121 TAGCAGCTACTTTTGCTTACTGGGACAATATTCTTGGTCCTAGAGTAAGGCACATTTGGG 180
A A T F A Y W D N I L G P R V R H I W A
P K T E Q V L L S D G E I T F L A N H T
Seq_1 181 CTCCAAAGACAGAACAGGTACTTCTCAGTGATGGAGAAATAACTTTTCTTGCCAACCACA 240
Seq_2 181 CTCCAAAGACAGAACAGGTACTTCTCAGTGATGGAGAAATAACTTTTCTTGCCAACCACA 240
P K T E Q V L L S D G E I T F L A N H T
L N G E I L R N A E S G A I D V K F F V
Seq_1 241 CTCTAAATGGAGAAATCCTTCGAAATGCAGAGAGTGGTGCTATAGATGTAAAGTTTTTTG 300
Seq_2 241 CTCTAAATGGAGAAATCCTTCGAAATGCAGAGAGTGGTGCTATAGATGTAAAGTTTTTTG 300
L N G E I L R N A E S G A I D V K F F V
L S E K G V I I V S L I F D G N W N G D
Seq_1 301 TCTTGTCTGAAAAGGGAGTGATTATTGTTTCATTAATCTTTGATGGAAACTGGAATGGGG 360
Seq_2 301 TCTTGTCTGAAAAGGGAGTGATTATTGTTTCATTAATCTTTGATGGAAACTGGAATGGGG 360
L S E K G V I I V S L I F D G N W N G D
R S T Y G L S I I L P Q T E L S F Y L P
Seq_1 361 ATCGCAGCACATATGGACTATCAATTATACTTCCACAGACAGAACTTAGTTTCTACCTCC 420
Seq_2 361 ATCGCAGCACATATGGACTATCAATTATACTTCCACAGACAGAACTTAGTTTCTACCTCC 420
R S T Y G L S I I L P Q T E L S F Y L P
L H R V C V D R L T H I I R
Seq_1 421 CACTTCATAGAGTGTGTGTTGATAGATTAACACATATAATCC---------- 462
Seq_2 421 CACTTCATAGAGTGTGTGTTGATAGATTAACACATATAATCCATGAAGTTGTTGATAGAT 480
L H R V C V D R L T H I I H E V V D R L
K G R I W M H K E R
Seq_1 463 ----------GGAAAGGAAGAATATGGATGCATAAGGAAA 492
Seq_2 481 TAACACATATAATCCATGAAGTTGATATCCGGAAAGGAAGAATATGGATGCATAAGGAAA 540
T H I I H E V D I R K G R I W M H K E R
Q E N V Q K I I L E G T E R M E D Q G Q
Seq_1 493 GACAAGAAAATGTCCAGAAGATTATCTTAGAAGGCACAGAGAGAATGGAAGATCAGGGTC 552
Seq_2 541 GACAAGAAAATGTCCAGAAGATTATCTTAGAAGGCACAGAGAGAATGGAAGATCAGGGTC 600
Q E N V Q K I I L E G T E R M E D Q G Q
S I I P M L T G E V I P V M E L L S S M
Seq_1 553 AGAGTATTATTCCAATGCTTACTGGAGAAGTGATTCCTGTAATGGAACTGCTTTCATCTA 612
Seq_2 601 AGAGTATTATTCCAATGCTTACTGGAGAAGTGATTCCTGTAATGGAACTGCTTTCATCTA 660
S I I P M L T G E V I P V M E L L S S M
K S H S V P E E I D I A D T V L N D D D
Seq_1 613 TGAAATCACACAGTGTTCCTGAAGAAATAGATATAGCTGATACAGTACTCAATGATGATG 672
Seq_2 661 TGAAATCACACAGTGTTCCTGAAGAAATAGATATAGCTGATACAGTACTCAATGATGATG 720
K S H S V P E E I D I A D T V L N D D D
I G D S C H E G F L L X
Seq_1 673 ATATTGGTGACAGCTGTCATGAAGGCTTTCTTCTCAA 709
Seq_2 721 ATATTGGTGACAGCTGTCATGAAGGCTTTCTTCTCAA 757
I G D S C H E G F L L X
Key:
Exon 2
Exon 3
Exon 4
Exon 5
INDEL
* premature stop codon

Figure 3.6. Sequence alignment of B12 +48 bp mutant with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the B12 +48 bp mutant (Seq_2). Exon 2 (red), exon 3 (blue), exon 4 (green) and exon 5 (orange) are shown. The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and B12 C9orf72 respectively. Premature stop codons are labelled with an asterisk (pink). The B12 INDEL is highlighted in yellow.

For the B17 CRISPR line, nine colonies were successfully sequenced (Table 3.3). Out of the nine colonies sequenced, 3 colonies exhibited a -23 bp mutation (Figure 3.7) and 6 colonies a -27 bp mutation (Figure 3.8), suggesting two copies of the C9orf72 gene had acquired the same 27 bp deletion mutation. The two mutant C9orf72 sequences were aligned with the wild-type C9orf72 sequence to determine whether the mutations created premature stop codons (Figure 3.7-3.8). The 27 bp deletion mutant removed amino acids from the C9orf72 sequence but did not create a frameshift or premature stop codon. The -23 bp mutant did create a premature stop codon. Two copies of the C9orf72 gene could produce potentially functional C9orf72 protein and therefore the B17 cell line could not be used as a C9orf72 loss-of-function cell model.

B17 mutation	Number of colonies with mutation	% of colonies with mutation	Premature stop codon
-23 bp	3	33.3	Yes
-27 bp	6	66.6	No

Table 3.3. C9orf72 exon 2 mutations in the B17 CRISPR line. Plasmids from nine bacteria colonies were successfully sequenced for their C9orf72 exon 2 insert. In the nine C9orf72 sequences returned, two INDEL mutations recurred: -23 bp and -27 bp. Only the -23 bp mutant was predicted to generate a premature stop codon.



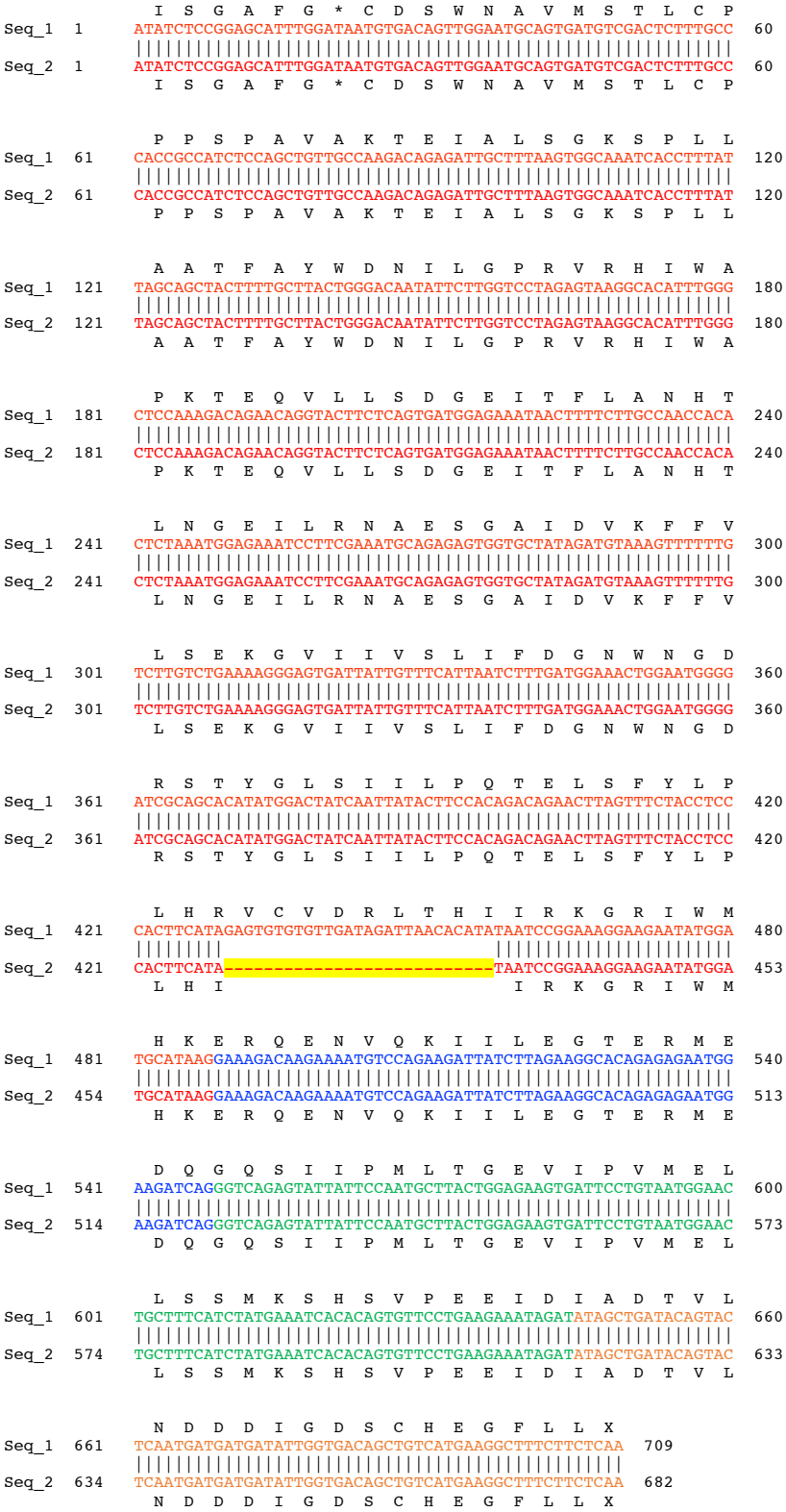
Key:

Exon 2
Exon 3
Exon 4
Exon 5

INDEL

* premature stop codon

Figure 3.7. Sequence alignment of B17 -23 bp mutant with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the B17 -23 bp mutant (Seq_2). Exon 2 (red), exon 3 (blue), exon 4 (green) and exon 5 (orange) are shown. The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and B17 C9orf72 respectively. Premature stop codons are labelled with an asterisk (pink). The B17 INDEL is highlighted in yellow.



Key:

Exon 2
Exon 3
Exon 4
Exon 5

INDEL

* premature stop codon

Figure 3.8. Sequence alignment of B17 -27 bp mutant with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the B17 -27 bp mutant (Seq_2). Exon 2 (red), exon 3 (blue), exon 4 (green) and exon 5 (orange) are shown. The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and B17 C9orf72 respectively. Premature stop codons are labelled with an asterisk (pink). The B17 INDEL is highlighted in yellow.

The sequencing results suggested clone B2 was the only CRISPR line to contain a mutation in all three C9orf72 alleles predicted to generate a premature stop codon. Both B12 and B17 were predicted to have two copies of the C9orf72 gene from which a mutant C9orf72 protein could be translated. To investigate this, C9orf72 mRNA and protein levels were measured in the CRISPR cell lines. To quantify C9orf72 mRNA levels, RNA was extracted from cells and an RT-qPCR performed, data shown in figure 3.9A. Primers used in the qPCR targeted C9orf72 exon 4 and exon 6 and therefore were C9orf72 long isoform specific. B17 C9orf72 mRNA levels were comparable to those seen in the HEK293 control. B2 was the only clone that appeared to have reduced C9orf72 mRNA, with a 50% reduction. Interestingly, C9orf72 mRNA in the B12 CRISPR line was measured at approximately 10-fold higher than in control HEK293 cells (Figure 3.9A). To investigate the effect on C9orf72 protein level in the CRISPR lines, whole cell lysate was prepared and subjected to SDS-PAGE electrophoresis. Protein was transferred onto nitrocellulose membrane and blotted for C9orf72 using an anti-C9orf72 antibody (Proteintech, 25757-1-AP). The C9orf72 protein was detected with an observed size of 50 kDa in the HEK293 cell control, but was absent in the B17, B2 and B12 CRISPR lines (Figure 3.9B). A smaller, unknown 37 kDa protein was detected in the B12 clone, labelled with an asterisk. The anti-C9orf72 antibody, which reportedly recognises an epitope within the C9orf72 N-terminal, did not detect a C9orf72 short protein isoform in any of the clones, which has a predicted molecular weight of 25 kDa. The immunoblot suggests all three clones lack the full length C9orf72 long protein. The B2 clone contained three predicted loss-of-function mutations and had both an observed knockdown in both C9orf72 mRNA and protein. Taken together, these results strongly suggest this cell line was a loss-of-function C9orf72 mutant and was the only clone taken forward for future use in experiments as a stable C9orf72 knockout cell model.

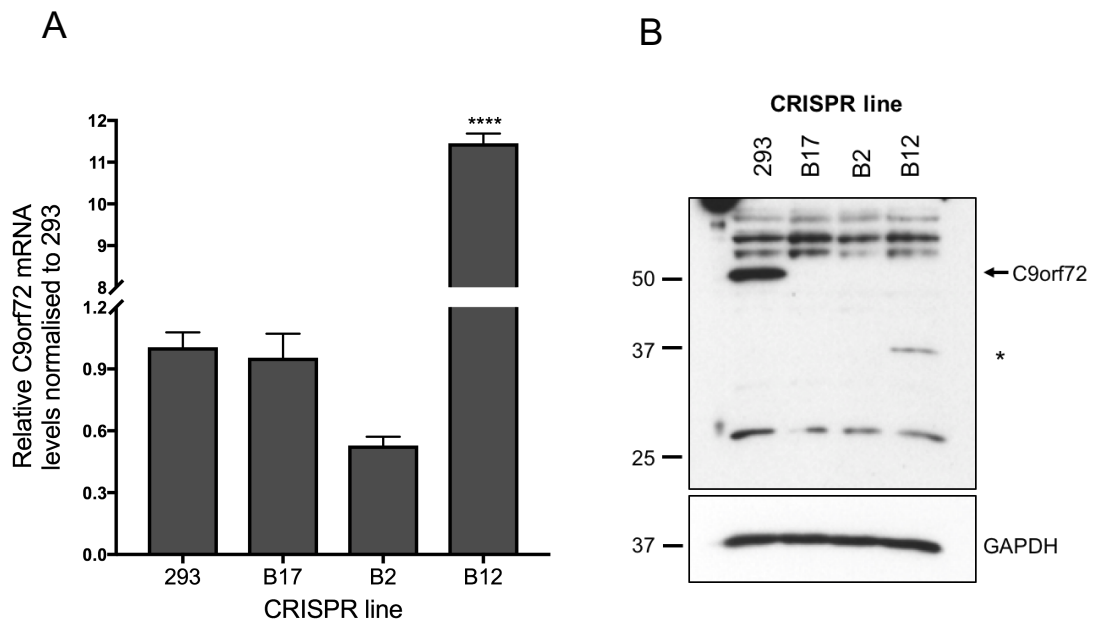


Figure 3.9. C9orf72 mRNA and protein levels in C9orf72 CRISPR lines. (A) Relative C9orf72 mRNA levels as measured by RT-qPCR. Data represent mean ± SEM; One-way ANOVA with Dunnett's test, **** $p \leq 0.0001$. n=3 triplicates from one experiment. (B) C9orf72 immunoblot of whole cell lysate from C9orf72 CRISPR lines. Blots were probed with anti-C9orf72 (25757-1-AP) and anti-GAPDH. Asterisk notes smaller, unknown protein of interest at 37 kDa present in the B12 cell line.

3.2.2 Characterisation of C9orf72 targeting siRNAs

The stability of the C9orf72 knockout in the B2 CRISPR line is useful for probing long-term effects of a loss of the C9orf72 protein. However, there is the potential that the cells compensate for the loss of the C9orf72 protein by upregulation of other proteins and/or pathways. Therefore, throughout the project transient knockdown of C9orf72 in cell lines was also performed to look at more acute effects. For these knockdowns, two C9orf72 siRNAs were used: C9orf72 siRNA #2 and C9orf72 siRNA #D (Webster et al., 2016). These siRNA were used in a pooled 1:1 mix for C9orf72 knockdown, unless otherwise stated as used individually. Until recently, commercially available anti-C9orf72 antibodies were nonspecific and unable to detect the low levels of endogenous C9orf72 in cells. Hence, after C9orf72 siRNA transfections, C9orf72 knockdown was confirmed by measuring C9orf72 mRNA levels by RT-qPCR, and is the technique included throughout this project. However, a new rabbit anti-C9orf72 antibody was recently made available which can detect endogenous C9orf72 long protein. mRNA can be a good indicator of knockdown but it does not directly infer the protein level (Schwanhäusser et al.,

2011). Therefore, the new anti-C9orf72 antibody was used to detect endogenous C9orf72 levels following C9orf72 siRNA transfection for comparison with measured C9orf72 mRNA levels by RT-qPCR. HEK293 cells were transfected with either non-targeting control (NTC), C9orf72 #2, C9orf72 #D, or C9orf72 pool siRNA. To look at C9orf72 mRNA levels, RNA was extracted and an RT-qPCR performed. C9orf72 mRNA was reduced relative to the NTC by approximately 40% for both C9orf72 siRNAs when used individually, and approximately 50% for C9orf72 siRNA pooled (Figure 3.10A). For C9orf72 protein levels, whole cell lysates were prepared and subjected to SDS-PAGE and immunoblot using the anti-C9orf72 antibody. The endogenous C9orf72 long protein was detected in the NTC and was significantly decreased in all C9orf72 siRNA transfected samples (Figure 3.10B). The level of reduction of the C9orf72 protein after C9orf72 siRNA transfection appeared significantly greater than the apparent 40-50% reduction in C9orf72 mRNA. Consequently, the C9orf72 mRNA levels measured throughout the project as confirmation of C9orf72 knockdown may be underestimating real protein loss.

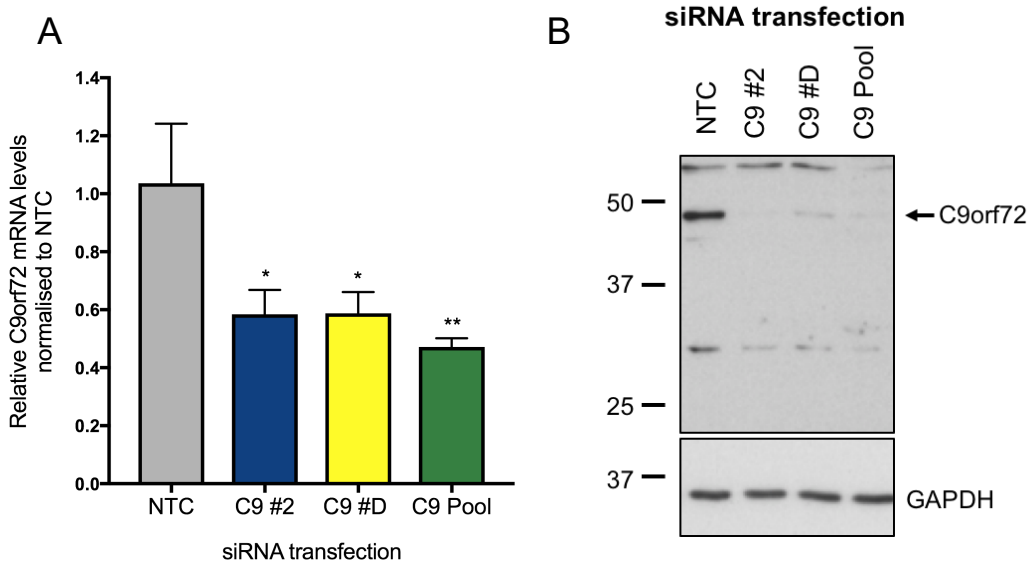
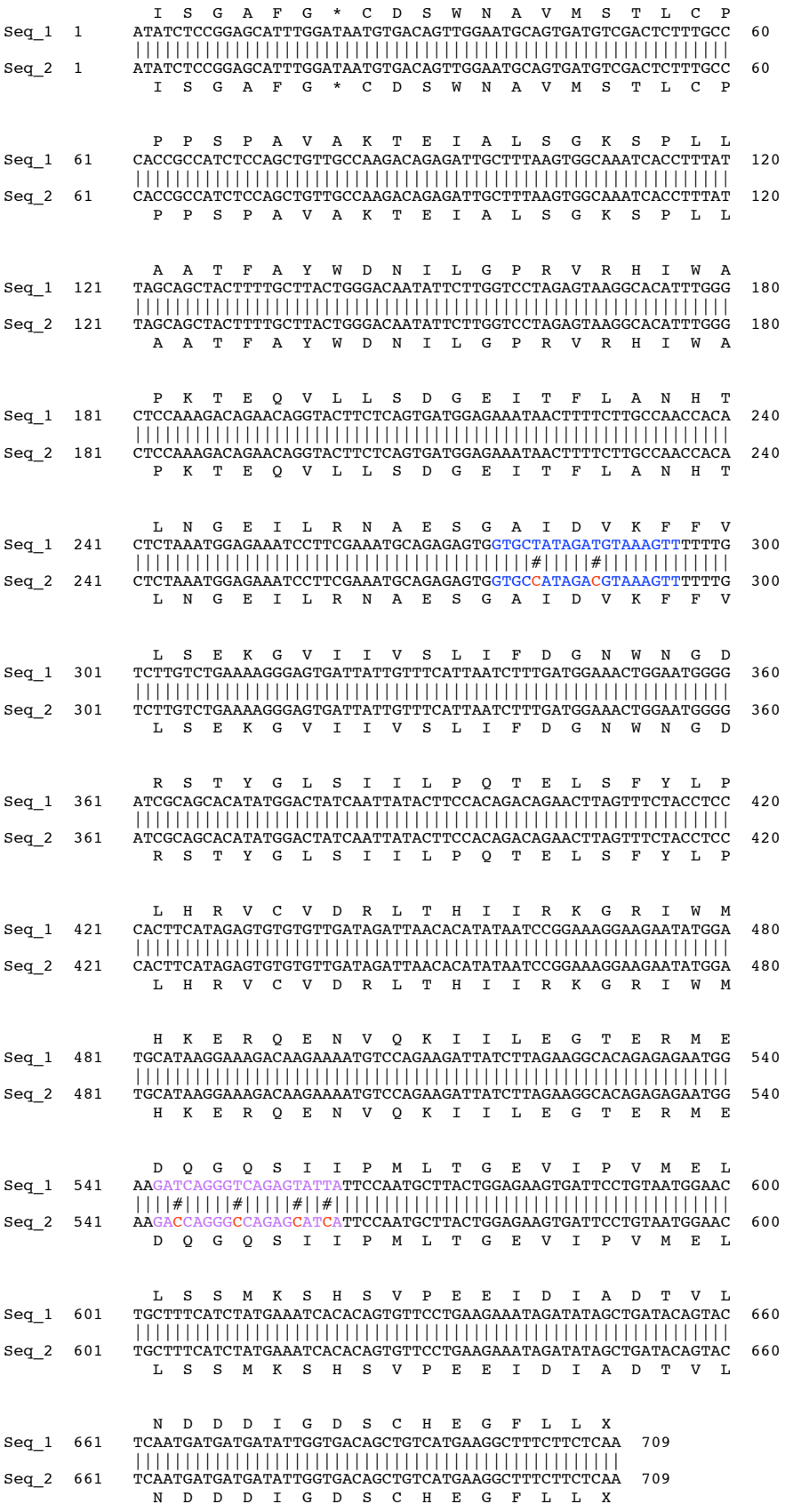


Figure 3.10. C9orf72 mRNA and protein levels following C9orf72 siRNA transfection. (A) HEK293 cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pooled siRNA. Relative C9orf72 mRNA levels were measured by RT-qPCR. Data represent mean ± SEM; One-way ANOVA with Dunnett's test, * $p \leq 0.05$, ** $p \leq 0.01$. n=3 triplicates from one experiment. (B) C9orf72 immunoblot of HEK293 whole cell lysate following C9orf72 siRNA transfection. Blots were probed with anti-C9orf72 (25757-1-AP) and anti-GAPDH.

3.2.3 Generation of siRNA resistant myc-C9orf72 constructs

One method of reducing C9orf72 protein levels in cell lines was through C9orf72 siRNA transfection. After knockdown, rescue experiments were performed to confirm specificity. The aim was to reintroduce C9orf72 into cells. The constructs frequently used for C9orf72 overexpression were pRK5-myc-C9orf72S and pRK5-myc-C9orf72L which expressed N-terminal myc-tagged C9orf72 short and long protein under a CMV promoter respectively. It was hypothesised that the C9orf72 siRNA in cells would target the transfected myc-C9orf72 and prevent overexpression. Thus, site-directed mutagenesis was performed on the myc-C9orf72 constructs to mutate the C9orf72 #2 and #D siRNA binding sites to provide siRNA resistance and allow for use in rescue experiments. The C9orf72 #2 siRNA binding site was located within the C9orf72 exon 2 and the C9orf72 #D siRNA binding site spanned the junction between exon 3 and exon 4. Thymine nucleotides were selected for mutagenesis to cytosines, with care taken to ensure the amino acid sequence was unchanged. Both myc-C9orf72 short and long constructs underwent a first round of site-directed mutagenesis to generate two T>C mutations within the C9orf72 #2 siRNA binding site. After mutagenesis, the constructs were transformed into XL10-Gold Ultracompetent cells and the plasmid purified. The plasmid was sequenced using the SP6 primer to confirm successful mutation. Successfully mutated plasmids underwent a second round of site-directed mutagenesis to generate four T>C mutations within the C9orf72 #D siRNA binding site. As before, constructs were transformed into XL10-Gold Ultracompetent cells and plasmids miniprepped. The constructs were sequenced with the SP6 primer to confirm both siRNA binding sites were mutated. The final sequence of the siRNA resistant constructs is shown in Figure 3.11.



Key

C9orf72 #2 siRNA binding site

C9orf72 #D siRNA binding site

Mutated T>C base

Figure 3.11. Sequence alignment of siRNA resistant C9orf72 with wild-type C9orf72. The sequence of wild-type C9orf72 exons 2-5 (Seq_1) is aligned with the sequence of C9orf72 exons 2-5 from the siRNA resistant mutant (Seq_2). The translated amino acid sequence is shown above and below the sequence alignment for wild-type C9orf72 and siRNA resistant C9orf72 respectively. The C9orf72 #2 binding site is shown in blue. The C9orf72 #D binding site is shown in purple. Site-directed mutagenesis was performed within these binding sites to introduce a T>C base pair change (red). Base pair mismatches are shown with a #.

The sequencing results confirmed both siRNA binding sites in the myc-C9orf72 constructs were successfully mutated. To test their functionality for rescue experiments, the constructs were expressed in cells after C9orf72 siRNA transfection and compared against non-siRNA resistant myc-C9orf72 expression. HEK293 cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pool siRNA. 72 hours post knockdown, cells were transfected with pCI-neo (empty vector control), myc-C9orf72S, myc-C9orf72S siRNA resistant, myc-C9orf72L or myc-C9orf72L siRNA resistant constructs. 24 hours post overexpression, whole cell lysates were prepared and subjected to SDS-PAGE and immunoblot with anti-myc antibody to probe myc-C9orf72 expression levels. Interestingly, myc-C9orf72S and myc-C9orf72L were detected in C9orf72 siRNA treated samples, but to slightly lower levels than in the NTC sample, suggesting there was only a small effect of the siRNA on plasmid expression (Figure 3.12A). This was likely due to the strength of the expression from the CMV promoter. Both myc-C9orf72S and myc-C9orf72L siRNA resistant constructs appeared to express myc-C9orf72 to a higher level than the non-siRNA resistant constructs. Importantly, the levels of the myc-C9orf72 siRNA resistant proteins were not reduced in C9orf72 siRNA treated samples versus NTC, suggesting they were resistant to siRNA (Figure 3.12A). C9orf72 knockdown in the experiment was confirmed to be between 40-50% by RNA extractions followed by RT-qPCR (Figure 3.12B). The siRNA resistant myc-C9orf72S and myc-C9orf72L constructs were therefore taken forward for use in rescue experiments where C9orf72 knockdown had been achieved with C9orf72 siRNA.

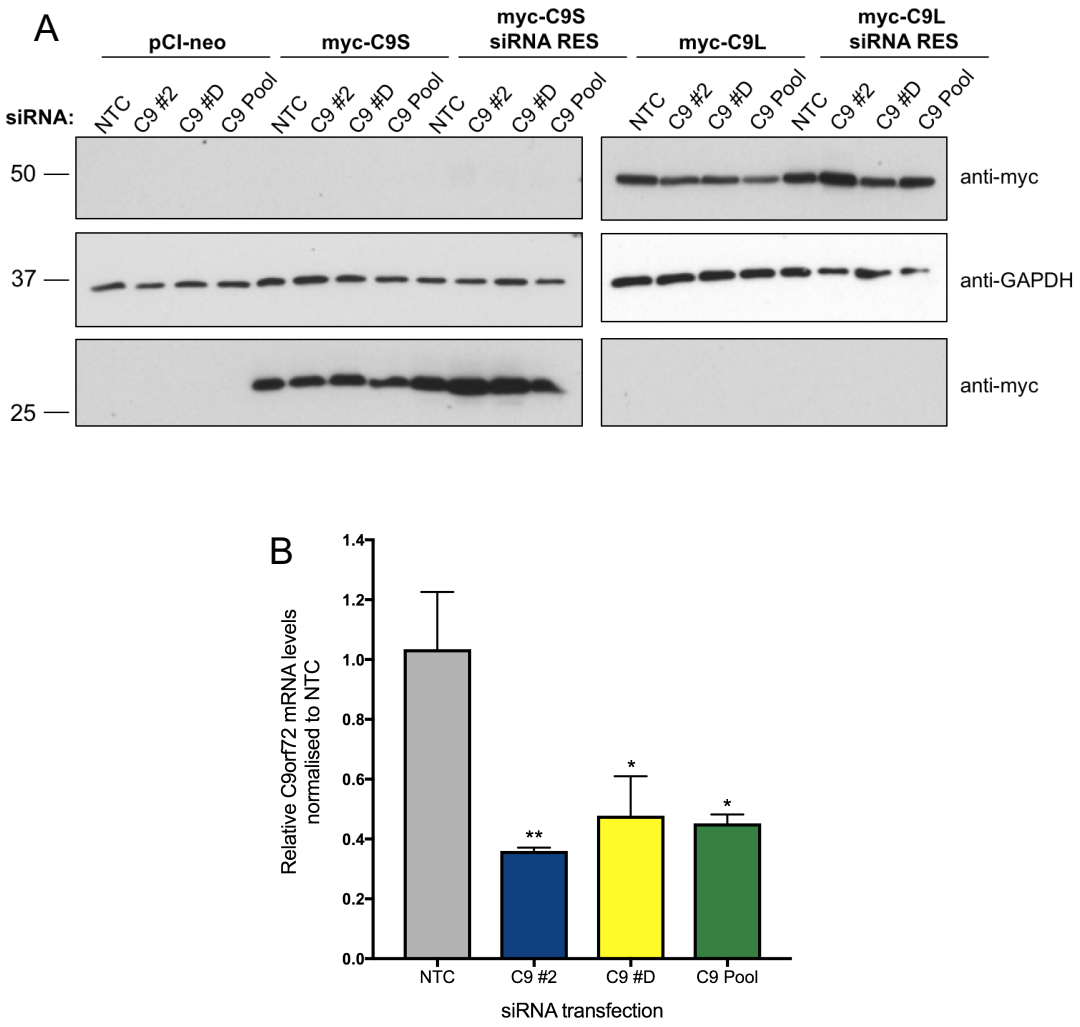


Figure 3.12. Expression of myc-C9orf72 siRNA resistant constructs. (A) myc-C9orf72 and myc-C9orf72 siRNA resistant constructs were expressed in HEK293 cells following NTC or C9orf72 siRNA transfection. Anti-myc immunoblot of whole cell lysate shows myc-C9orf72 protein levels. Blots were probed with anti-myc (9B11) and anti-GAPDH. (B) C9orf72 knockdown was confirmed by RT-qPCR. Data represent mean ± SEM; One-way ANOVA with Dunnett's test, * $p \leq 0.05$, ** $p \leq 0.01$. n=3 triplicates from one experiment.

3.2.4 Expression of tagged-coilin proteins

To establish a function for C9orf72, it's relationship with coilin was examined. Coilin is a nuclear protein that localises to Cajal bodies, which are involved in the assembly of the cells splicing machinery, where it acts as a scaffolding protein. The interaction between C9orf72 and coilin is described in Chapter 4 and the involvement of C9orf72 in Cajal bodies and splicing is investigated in Chapters 5 and 6 respectively. To investigate the behaviour and relationship between coilin and C9orf72 in cells, a GFP-coilin construct was obtained for coilin overexpression. Coilin was subcloned from the pEGFP-C3 backbone vector into pCI-neo-myc, pCI-neo-FLAG and pCI-neo-mCherry vectors to allow for a selection of different tagged-coilin proteins for expression experiments and sequenced to confirm the coilin protein sequence had not mutated during the subcloning process. To test for correct expression, the constructs were transfected into HEK293 cells and analysed by western blot and immunofluorescence. For immunoblot, HEK293 cells were either non-transfected or transfected with GFP-, myc-, FLAG-, or mCherry-coilin. Whole cell lysates were prepared and subjected to SDS-PAGE electrophoresis and proteins transferred to nitrocellulose membrane. The blot was cut such that GFP-coilin, myc-coilin, FLAG-coilin and mCherry-coilin samples were separate and received anti-GFP, anti-myc, anti-FLAG and anti-mCherry antibodies respectively. All antibodies detected their respective tagged-coilin protein at correct molecular weights which were absent in non-transfected samples, suggesting each coilin protein had a functional tag. A second immunoblot was performed on all samples using an anti-coilin antibody. The antibody detected endogenous coilin in all non-transfected samples and the overexpression of GFP-, myc-, FLAG- and mCherry coilin (Figure 3.13). Together these blots suggest the tagged-coilin constructs were expressed correctly and were detectable in HEK293 cells.

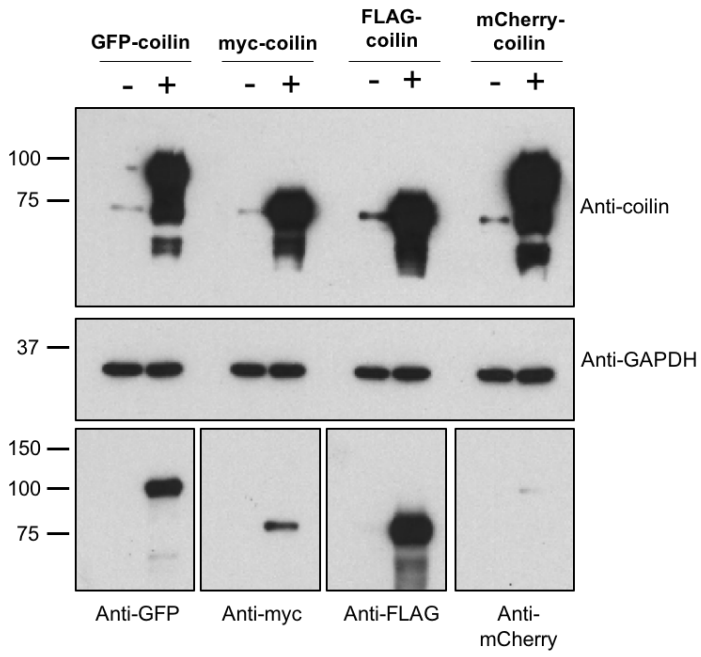


Figure 3.13. Tagged-coilin constructs expressed in HEK293 cells. GFP-, myc-, FLAG-, and mCherry-coilin constructs were transfected into HEK293 cells and analysed for correct tagged-protein expression by immunoblot with anti-GFP, anti-myc (9B11), anti-FLAG and anti-mCherry antibodies respectively. Immunoblot with anti-coilin was included

The immunoblots confirmed the tagged-coilin proteins of the expected molecular weight were expressed. As mentioned, coilin localises strongly to Cajal bodies which appear as bright nuclear spots (Cioce and Lamond, 2005). To investigate correct localisation of the tagged-coilin, an immunocytochemistry experiment was performed. HEK293 cells were transfected with GFP-coilin, myc-coilin, FLAG-coilin or mCherry-coilin. Cells were fixed and immunostained using anti-GFP, anti-myc, anti-FLAG and anti-mCherry antibodies respectively. All tagged-coilin proteins were successfully detected with their respective antibodies (Figure 3.14). However, all were expressed diffusively throughout the nucleus and distinct Cajal bodies were not detected, suggesting these proteins were unable to localise to Cajal bodies. Therefore, these constructs were unsuitable for experiments where Cajal bodies were the focus.

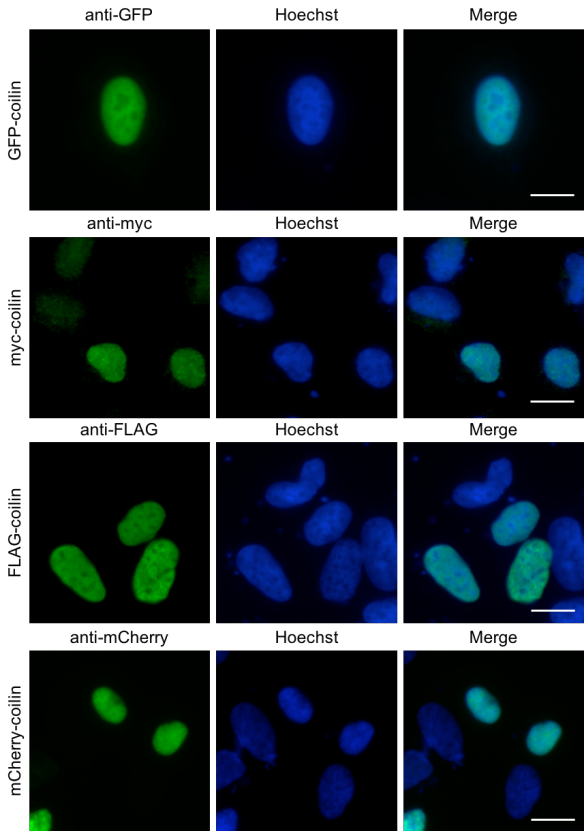


Figure 3.14. Tagged-coilin localises in the nucleus but not Cajal bodies. GFP-, myc-, FLAG-, and mCherry-coilin constructs were transfected into HEK293 cells and analysed for correct tagged-protein expression by immunostaining with anti-GFP, anti-myc (9B11), anti-FLAG and anti-mCherry respectively (green). Scale bar = 10 µm.

3.2.5 Optimisation of an anti-coilin antibody for Cajal body immunostaining

Overexpressed tagged-coilin localised to the nucleus but not Cajal bodies, suggesting endogenous coilin immunostaining should be used as a marker of CBs. Consequently, a rabbit anti-coilin antibody was acquired and tested in HEK293 cells for Cajal body detection. Cells were fixed and immunostained using the antibody at dilutions ranging from 1 in 50 to 1 in 1000. All dilutions immunostained Cajal bodies, imaged as bright green foci (Figure 3.15). At lower dilutions, there was cytoplasmic signal, nucleoplasmic coilin staining and Cajal body staining. The 1 in 1000 dilution effectively detected coilin localised in CBs and therefore this dilution was selected for use in future experiments.

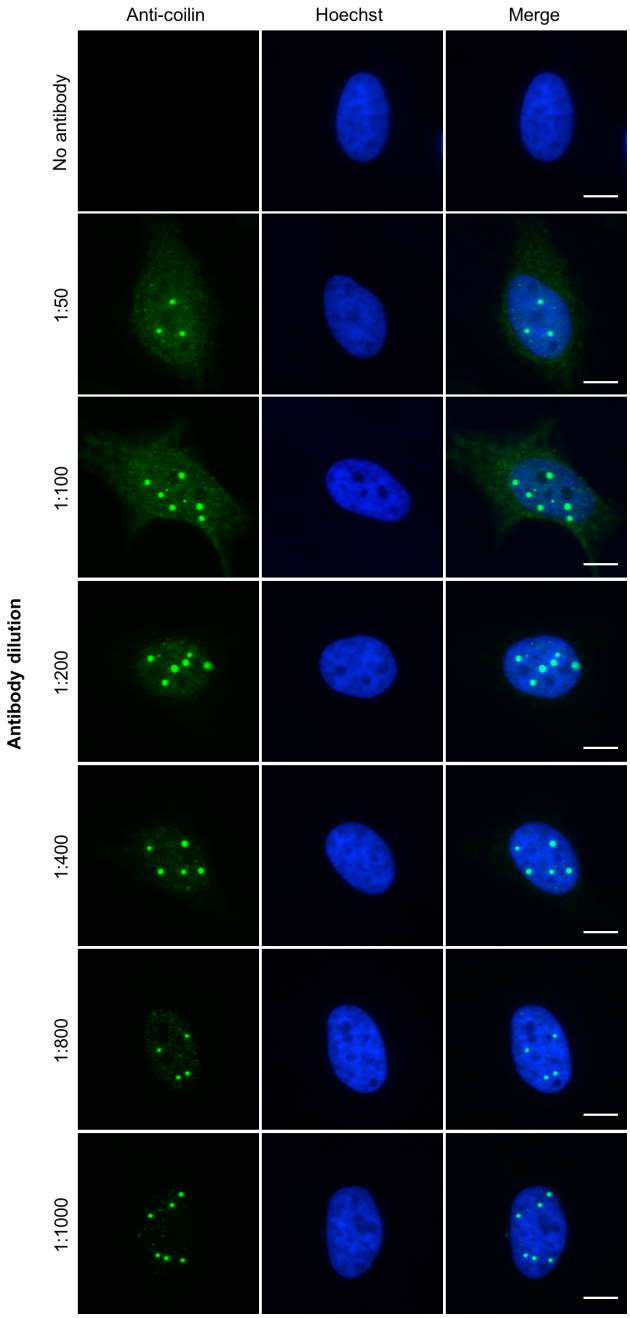


Figure 3.15. Optimisation of a rabbit anti-coilin antibody for Cajal body immunostaining. HEK293 cells were fixed and immunostained using a rabbit anti-coilin antibody at varying dilutions for Cajal body (green foci) visualisation. Scale bar = 10 μm.

3.2.6 Optimisation of coilin siRNAs

For coilin knockdown, siRNA transfections were used. Three commercially-available coilin-targeting siRNAs were purchased and optimised to determine which combination of siRNAs produced the most effective knockdown of both coilin protein and mRNA. HEK293 cells were transfected with NTC siRNA as a control. Cells were also transfected with individual coilin 1, 2 and 3 siRNA and pooled combinations thereof. To probe coilin protein knockdown, cell lysates were subjected to SDS-PAGE and immunoblotted with anti-coilin antibody. All coilin siRNAs reduced coilin protein levels (Figure 3.16A). Relative coilin protein levels were quantified by normalisation to the GAPDH loading control and are presented relative to the NTC (Figure 3.16B). Individually, the coilin 1 and 3 siRNAs were most effective for coilin protein knockdown, and siRNAs 1+3 and 2+3 were the best pooled combinations. An RT-qPCR was also performed on RNA extracted from siRNA transfected cells to look at coilin mRNA levels. In agreement with the immunoblot, all siRNAs decreased coilin mRNA levels, and pooled siRNAs 1+3 and 2+3 were best (Figure 3.16C). The pooled combination of coilin siRNAs 1 and 3 effectively knocked down both coilin mRNA and protein and therefore was selected as the combination for future coilin knockdowns. To determine if the pooled combination of coilin siRNAs 1+3 caused Cajal body loss, HEK293 cells were transfected with either NTC or coilin 1+3 siRNA. Cells were fixed and immunostained using an anti-coilin antibody to mark Cajal bodies. Cells transfected with the pooled coilin siRNA had no apparent nuclear CBs, which were present in all NTC transfected cells (Figure 3.16D).

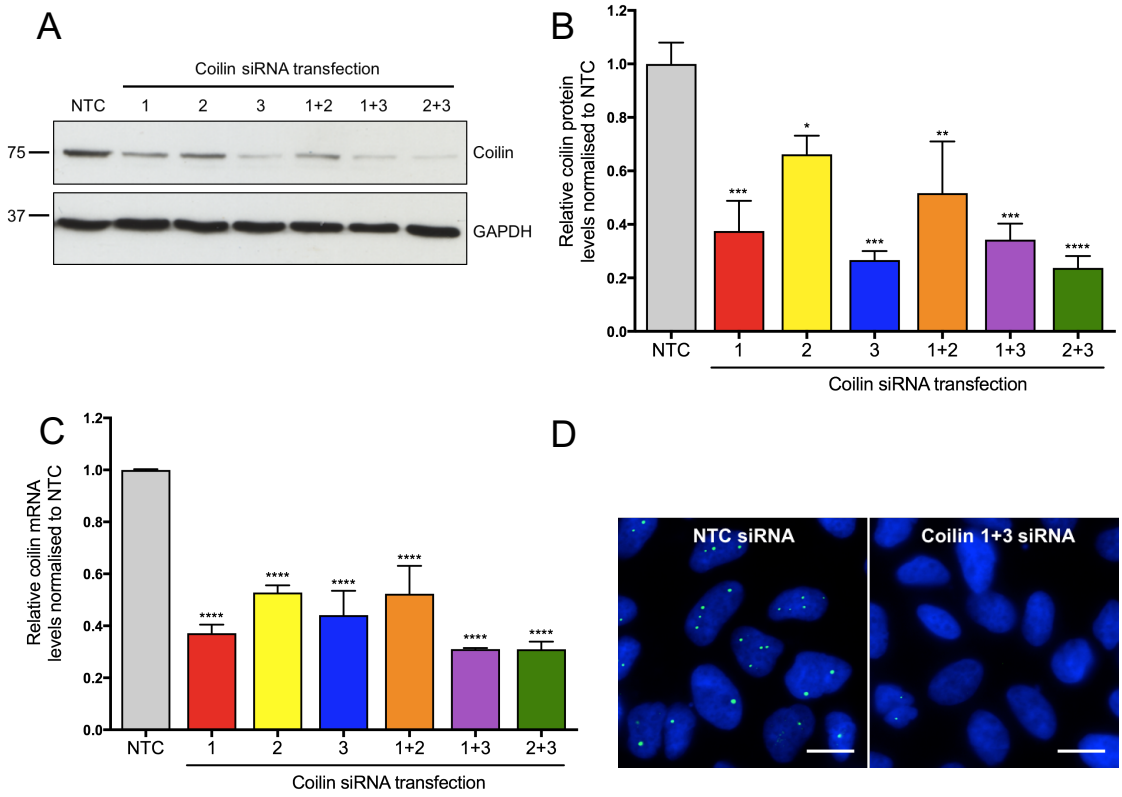


Figure 3.16. Optimisation of coilin siRNA. (A) Anti-coilin immunoblot of HEK293 whole cell lysates following transfection of individual and pooled combinations of coilin #1, #2 and #3 siRNAs. Blots were probed with anti-coilin and anti-GAPDH. (B) Relative coilin protein levels quantified from (A). Coilin protein was normalised to GAPDH and is shown relative to NTC. (C) Relative coilin mRNA levels as measured by RT-qPCR. All data represent mean ± SEM; One-way ANOVA with Dunnett's test, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. n=3 experiments. (D) HEK293 cells transfected with NTC or Coilin 1+3 pooled siRNA and immunostained with anti-coilin (green). Scale bar = 10 µm.

3.3 Discussion

The aim of this chapter was to describe the process of generating and optimising tools utilised throughout the project. This included characterising a C9orf72 knockout CRISPR cell line, optimising new siRNAs and antibodies and generating siRNA resistant protein constructs. The C9orf72 knockout CRISPR line was generated to investigate the effect of C9orf72 loss of protein in a stable model and to compliment transient knockdown experiments. Exon 2 of the C9orf72 gene was targeted for mutation using CRISPR/Cas9 nickase technology, and three cell lines generated from this were characterised for loss of C9orf72 in three ways: sequencing of the mutation, measuring C9orf72 mRNA levels and immunoblot of the C9orf72 protein. The B2 clone had INDEL mutations in each copy of the C9orf72 gene which were predicted to lead to premature stop codons, the C9orf72 mRNA showed a 50% reduction and an anti-C9orf72 antibody did not detect C9orf72 protein. Newly transcribed mRNA may not have been targeted for nonsense-mediated decay at the time of RNA extraction, explaining the C9orf72 mRNA levels. It is unlikely, with the number of premature stop codons found downstream of the INDELs, to translate into full-length protein. We were therefore confident in implementing B2 as a new C9orf72 knockout cell line.

The B17 and B12 clones however, showed inconsistencies across their results. B17 was found to have two mutations in the C9orf72 gene, a 23 bp and 27 bp deletion. The 27 bp deletion was found twice as often than the 23 bp deletion, hence it was probable that two copies of the C9orf72 gene had both acquired the same INDEL. The -27 bp mutant caused the deletion of nine amino acids from the C9orf72 protein, but did not lead to a frameshift in the ORF. Therefore, it was assumed the B17 line had two copies of the C9orf72 gene that could be transcribed and translated to produce a C9orf72 mutant protein lacking amino acids 129-138, which are located within the N-terminal longin domain. C9orf72 mRNA levels in the B17 clone were comparable with the HEK293 control, as in-frame mRNA transcripts could still be generated from two copies of the C9orf72 gene. Interestingly, the immunoblot of the B17 clone did not detect C9orf72 protein, even though we would expect a 9 amino acid deletion. The epitope of the anti-C9orf72 antibody is described as the N-terminal of C9orf72. Therefore, if the deleted amino acids in the B17 -27 bp INDEL mutant, specifically the VDRLTHI residues, form all or part of the anti-C9orf72 epitope, the protein would be undetectable by immunoblot.

Similarly, the B12 clone had a +48 bp mutation that occurred in two copies of the C9orf72 gene which did not produce a premature stop codon mutation but led to the insertion of 16 novel amino acids into the C9orf72 sequence. It may be possible that the INDEL mutation stabilised, or inhibited destabilisation, of the C9orf72 mRNA, leading to the ten-fold higher levels of B12 C9orf72 mRNA versus HEK293 control. The higher mRNA levels could also be clone specific as opposed to a direct effect of the INDEL mutation. The immunoblot contradicted the qPCR data with no detectable C9orf72 protein. As before, this result could occur if the +48 bp INDEL had disrupted the epitope recognised by the anti-C9orf72 antibody. Interestingly, the B12 clone had an additional protein band at 37 kDa, which was absent in HEK293 cells and the other CRISPR lines, suggesting the 48 bp insertion was leading to generation of a smaller C9orf72 isoform that was still detectable with the anti-C9orf72 antibody. The inserted base pairs could insert a novel splice site from which a new C9orf72 isoform can be alternatively spliced. If alternative splicing was occurring from this novel splice site, it would produce a C9orf72 protein lacking the first N-terminal 139 amino acids and the predicted molecular weight of this new isoform would be 37.6 kDa, fitting with the detected protein on the immunoblot. Interestingly, the 48 bp insertion codes for residues VDRLTHI, which would allow the anti-C9orf72 antibody to recognise and bind if this is indeed the antibody epitope.

The aim was to take forward a stable C9orf72 knockout cell model and as such, the B17 and B12 clones were not investigated further. However, the B12 clone could be an interesting cell model for future experiments if it is indeed expressing a stable C9orf72 ∆longin mutant. To test these hypotheses, mutagenesis could be undertaken on a C9orf72 construct to substitute the residues predicted to be involved in the anti-C9orf72 antibody epitope, to confirm the lack of protein detected on the immunoblot was for this reason. Unfortunately, other C9orf72 antibodies which may target other epitopes are not specific or sensitive enough yet to perform another immunoblot with. Redesign of qPCR primers could inform as to whether alternative splicing is occurring at the insertion mutation. If alternatively spliced, a forward primer in the sequence upstream of the insertion should not amplify product.

The discrepancy between C9orf72 mRNA and C9orf72 protein levels seen in the CRISPR lines was also seen for cells transfected with C9orf72 siRNA. Due to the lack of suitable C9orf72 antibodies for most of the duration of the project,

knockdown of C9orf72 was confirmed by measuring mRNA levels. On average, knockdown of C9orf72 mRNA was between 40-70%. At the end of the project, a new anti-C9orf72 antibody became available which could detect C9orf72 and this antibody was tested on HEK293 cells transfected with C9orf72 siRNAs for comparison with mRNA levels. In all knockdown conditions, the reduction in C9orf72 mRNA was approximately 50% whereas the C9orf72 immunoblot suggested there was a much greater loss of protein. At the point of cell harvest, the RT-qPCR will detect C9orf72 out-of-frame mRNA which has not yet been targeted by the RNA-Induced Silencing Complex (RISC) for degradation thus mRNA levels may not correlate with protein levels. In addition, transcription of C9orf72 may be highly active but followed by slow translation and/or high protein turnover. This would explain the low C9orf72 protein levels described in the literature (Waite et al., 2014; Xiao et al., 2015). The non-correlation between mRNA and protein levels has been examined further in other studies (Schwanhäusser et al., 2011; Vogel et al., 2010). In the context of the project, it is possible that the C9orf72 knockdowns measured as reduction in C9orf72 mRNA are underestimating the actual effect on C9orf72 protein levels.

Of note, the anti-C9orf72 antibody did not detect an endogenous C9orf72 'short' isoform thought to be translated from the alternatively spliced V2 transcript (DeJesus-Hernandez et al., 2011). The short isoform is terminated at exon 5, therefore contains the same N-terminus as the long isoform for recognition by the antibody. There has been debate in the C9orf72 research community as to the existence of the C9orf72 short isoform, and the immunoblot with this anti-C9orf72 antibody would support there only being one C9orf72 protein in the HEK293 cell line. Other groups have shown detection of the short isoform in patient tissue but not cell lines (Waite et al., 2014; Xiao et al., 2015), suggesting it's expression may be cell-specific.

This chapter also detailed the generation of tagged-coilin protein constructs for expression in cells. Coilin was subcloned into four different expression vectors such that it was tagged with GFP, myc, FLAG or mCherry on the N-terminal. Immunoblot of these proteins confirmed successful expression of the proteins which had the expected molecular weights and were detected by antibodies recognising their tags (Figure 3.13). Endogenous coilin localises to the nucleus, where approximately 70-80% of it is dispersed throughout the nucleoplasm and the remaining is concentrated in Cajal bodies (Lam et al., 2002). The tagged-coilin constructs

correctly localised to the nucleus but did not appear to form Cajal bodies (Figure 3.14). One reason for this could be because of the location of tag. The N-terminal of coilin is defined as the 'self-association domain', without which it cannot interact with itself or form canonical Cajal bodies (Hebert and Matera, 2000). If the N-terminal tag is affecting the proper folding or interactions between the self-association domains, then the tagged-coilin would be unable to localise to Cajal bodies. To determine whether Cajal bodies still exist in these cells, immunostaining with another Cajal body marker such as SMN would be required. The C-terminal of coilin is responsible for mediating protein-protein interactions with other Cajal body proteins, such as the SMN and Sm proteins, suggesting this is also not an ideal location for a tag (Hebert et al., 2001). This is supported by the observance of an unusual staining pattern of a coilin-GFP protein in HeLa cells (Hebert and Matera, 2000). On the other hand, it may be the overexpression of coilin itself that leads to the disassembly of Cajal bodies. It was described by Hebert & Matera in 2000 that GFP-coilin could be seen in Cajal bodies in low-expressing cells. Supporting this, a stable GFP-coilin HeLa cell line, where the GFP-coilin was expressed to a similar level as endogenous coilin, was generated in which Cajal bodies were successfully formed. The GFP-coilin was noted as having the same biochemical and localisation patterns as endogenous coilin (Platani et al., 2000). These data suggest coilin expression can affect Cajal bodies. Accordingly, for all experiments where there was an interest in Cajal bodies, they were investigated using endogenous coilin.

4 C9orf72 interacts with nuclear protein coilin

4.1 Introduction

In this chapter, the nuclear localisation of C9orf72 was investigated. Often, a protein's function is closely linked with its localisation in a cell. The exact localisation of endogenous C9orf72 has thus far been difficult to determine, with studies reporting different localisation patterns that vary between cell and tissue types. Contributing to the problem is a lack of antibodies recognising C9orf72 with high specificity or the ability to distinguish between the short and long isoforms, which may have different functions. However, in 2015, C9orf72 isoform-specific antibodies were generated which suggested the endogenous short isoform localised to the nuclear envelope whereas the long isoform appeared cytoplasmic (Xiao et al., 2015). Both isoforms showed nuclear localisation when overexpressed in cell lines.

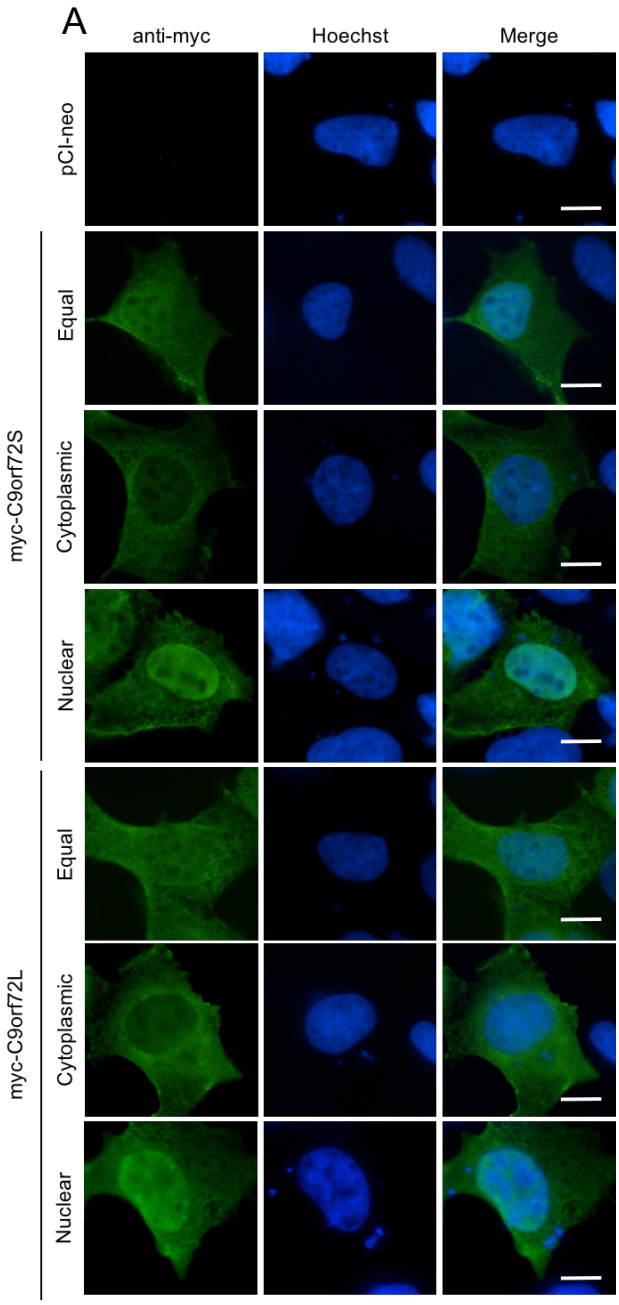
Moreover, proteomic analysis studies undertaken by various groups have identified potential protein interactors of C9orf72 (Blokhuis et al., 2016; Sellier et al., 2016; Sullivan et al., 2016). Nuclear proteins appear frequently throughout the list of proteins included in the C9orf72 interaction network (See appendix). Taken together, these studies suggest C9orf72 has some nuclear involvement.

In addition to probing C9orf72 localisation, a yeast two-hybrid (Y2H) screen was conducted to identify novel C9orf72 protein interactors which produced multiple hits for an interaction between C9orf72 and nuclear protein coilin. Two Y2H screens were outsourced and performed by staff at the proteomics facility at the German Cancer Research Centre: 1) A complete human brain random library screened with both short and long C9orf72 isoforms and 2) A human full-length cDNA library screened with C9orf72 long isoform only. The latter screen identified two protein interactions only, one of which was coilin and the results of which are included in this chapter. Hits from the other screen are referenced in the appendix. The interaction between C9orf72 and coilin was confirmed using an *in vitro* GST-binding assay, co-immunoprecipitation experiments and a proximity ligation assay (PLA).

4.2 Results

4.2.1 Exogenous C9orf72 short and long isoforms localise to the nucleus in HEK293 cells

To investigate C9orf72 localisation after protein overexpression, myc-C9orf72 short (S) or long (L) plasmid constructs were transfected into HEK293 cells. pCI-neo was transfected as an empty vector control. After transfection, cells were fixed and immunostained using an anti-myc antibody to investigate the localisation of myc-C9orf72. Both myc-C9orf72S and myc-C9orf72L showed similar staining and localised to both the nucleus and cytoplasm in all transfected cells (Figure 4.1A). However, the strength of nuclear localisation varied from cell to cell for both C9orf72 isoforms. Thus, cells were categorised as either 'Equal', 'Nuclear' or 'Cytoplasmic' accordingly and quantified (Figure 4.1B). Myc-C9orf72S showed equal, nuclear and cytoplasmic localisation in approximately 40%, 30% and 30% of cells analysed respectively. Myc-C9orf72L showed a similar distribution with equal, nuclear and cytoplasmic localisation in approximately 60%, 20% and 20% of cells analysed respectively. Myc-C9orf72S appeared to show more preference towards distinct cytoplasmic or nuclear localisation in comparison to myc-C9orf72L which exhibited diffusive localisation in a higher proportion of cells. Overall, the subcellular localisation of myc-C9orf72 was varied within the population of transfected HEK293 cells.


A
anti-myc
Hoechst
Merge
pCI-neo
myc-C9orf72S
Equal
Cytoplasmic
Nuclear
myc-C9orf72L
Equal
Cytoplasmic
Nuclear

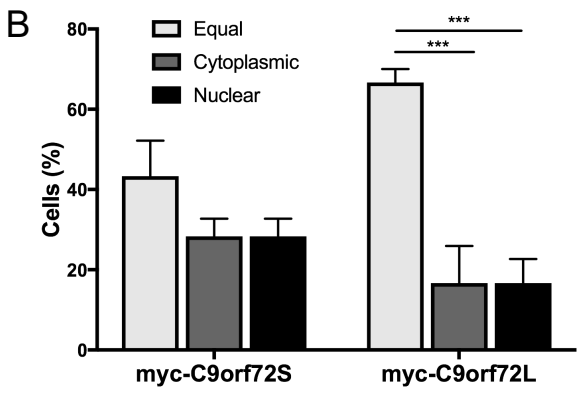

B
Equal
Cytoplasmic
Nuclear
***
***
Cells (%)
80
60
40
20
0
myc-C9orf72S
myc-C9orf72L

Figure 4.1. myc-C9orf72 localisation is varied within the HEK293 cell population. (A) HEK293 cells were transfected with pCI-neo (control), myc-C9orf72S or myc-C9orf72L and immunostained using an anti-myc antibody (9B11) (green). myc-C9orf72 localisation was categorised as 'Equal', 'Cytoplasmic' or 'Nuclear'; Representative images are shown. Scale bar = 10 μm. (B) Percentage of cells exhibiting equal, cytoplasmic or nuclear localisation of myc-C9orf72S and myc-C9orf72L. Data represent mean ± SEM; Two-way ANOVA with Sidak's test, *** $p \leq 0.001$. n=3 experiments (20 cells per experiment).

Immunostaining against myc-C9orf72 in HEK293 cells showed both the C9orf72 short and long protein localised in the nucleus and cytoplasm when overexpressed. To validate the immunofluorescence results, a nuclear fractionation was undertaken on HEK293 cells to investigate the localisation of myc-C9orf72 in total, cytoplasmic or nuclear fractions. As before, cells were transfected with pCI-neo (empty vector control), myc-C9of72S or myc-C9orf72L. Cells were harvested and a sample removed and lysed in RIPA buffer for the 'total' cell fraction. The remaining cells were lysed in hypotonic lysis buffer which breaks down the cell membrane but leaves the nuclear membrane intact. The nuclei were pelleted and the supernatant collected as the 'cytoplasm' fraction. The nuclear pellet was lysed in RIPA buffer to produce the 'nuclear' fraction. Myc-C9orf72 localisation across the fractions was examined by immunoblot. To confirm successful nuclear fractionation, SSRP1 and actin were used as nuclear and cytoplasm markers respectively. SSRP1 is a protein component of the Facilitates Chromatin Transcription (FACT) complex and binds DNA at nucleosomes (Röttgers et al., 2000). Actin is a microfilament protein involved in cytoskeleton shaping (Stricker et al., 2010). In all transfected samples, SSRP1 was absent in the cytoplasm fraction and enriched in the nuclear fraction, suggesting the cytoplasm fraction was free of nuclear contamination (Figure 4.2). Conversely, actin was detected in both the nuclear and cytoplasmic fractions, suggesting unlysed cells remained or there was some cytoplasmic contamination of the nuclear fraction. To look at myc-C9orf72 localisation, an anti-myc antibody was used. Myc-C9orf72L was detected in total, cytoplasmic and nuclear fractions. Myc-C9orf72S was likewise detected in all fractions, however the protein level appeared highest in the nuclear fraction (Figure 4.2). The results support the immunofluorescence data, suggesting C9orf72 localises to both the cytoplasm and nucleus.

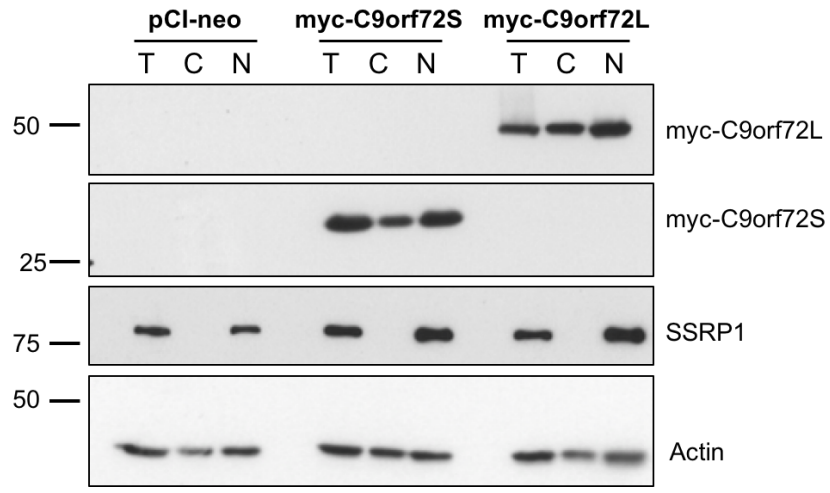


Figure 4.2. C9orf72 localises to both the cytoplasm and nucleus in HEK293 cells. HEK293 cells were transfected with pCI-neo (control), myc-C9orf72S or myc-C9orf72L and subjected to nuclear fractionation. Total (T), cytoplasmic (C) and nuclear (N) fractions were immunoblotted using anti-myc antibody (9B11) to investigate myc-C9orf72 subcellular localisation. Anti-SSRP1 and anti-actin antibodies were used as nuclear and cytoplasmic markers respectively.

4.2.2 The C9orf72 short isoform localises to nuclear speckles in rat cortical neurons

Myc-C9orf72 was shown to localise to both the nucleus and cytoplasm when overexpressed in HEK293 cells. However, while useful to understand the location of C9orf72 in this cell line to gain information about function, the location of C9orf72 in primary neurons holds more importance in the context of ALS. For this reason, the localisation of C9orf72 was also investigated in primary rat cortical neurons. Neurons were prepared and transduced by Miss Rebecca Cohen. At 5 DIV, neurons were transduced with pLenti virus expressing GFP (control), mVenus-C9orf72S or mVenus-C9orf72L and fixed at 12 DIV. Both mVenus-C9orf72S and mVenus-C9orf72L localised to the nucleus, cytoplasm and dendrites of neurons, suggesting C9orf72 localises evenly throughout the whole cell (Figure 4.3A). Interestingly, mVenus-C9orf72S but not mVenus-C9orf72L localised to nuclear speckles in many transduced neurons. Confocal microscope imaging showed clearly the localisation of mVenus-C9orf72S in nuclear speckles in a subset of cortical neurons (Figure 4.3B). The images show that in neurons, the two C9orf72 isoforms have different localisation patterns.

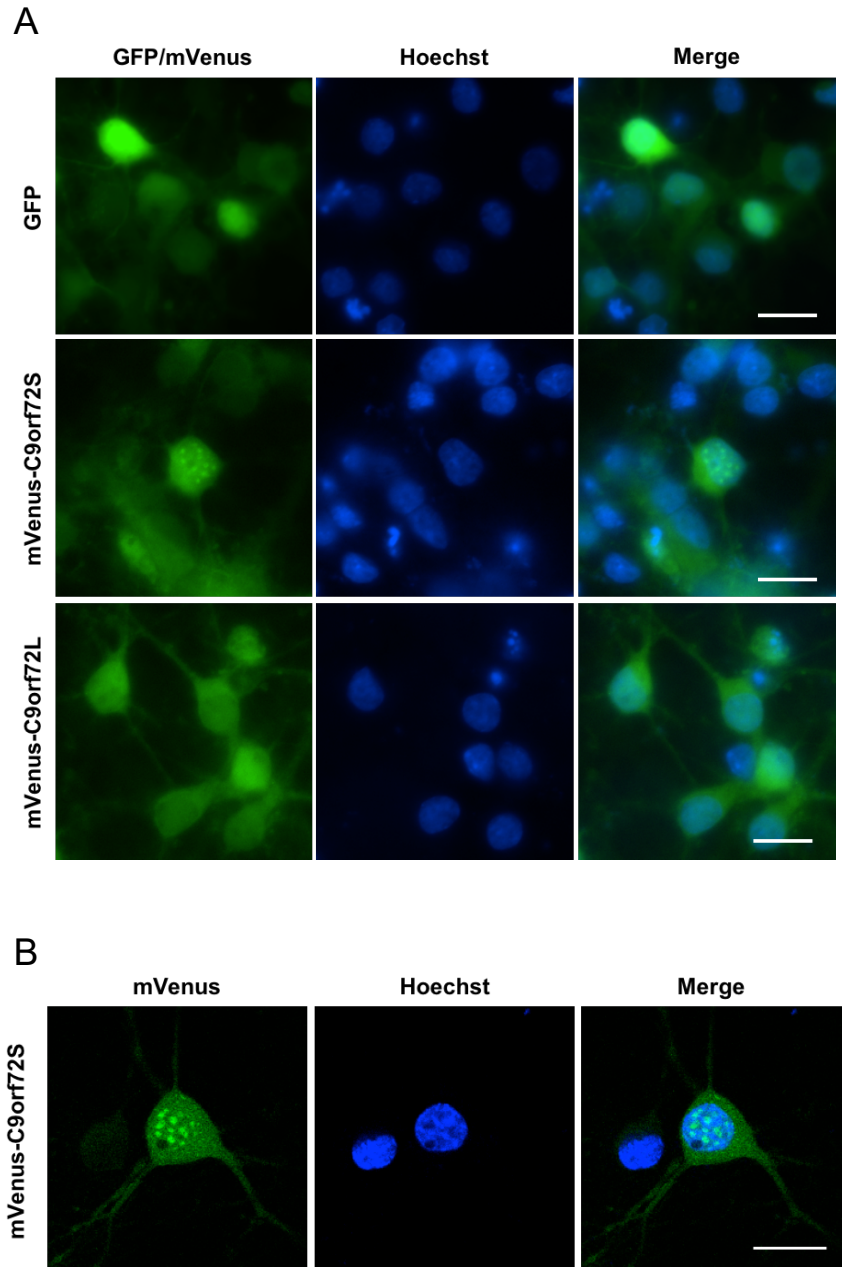


Figure 4.3. mVenus-C9orf72S forms nuclear speckles in rat cortical neurons. (A) 5 DIV rat cortical neurons were transduced with GFP (control), mVenus-C9orf72S or mVenus-C9orf72L. Neurons were fixed at 12 DIV and imaged for mVenus-C9orf72 localisation (green). Scale bar = 10 μm (B) Confocal microscope image of mVenus-C9orf72S localised in nuclear 'speckles'. Scale bar = 10 μm. Neurons were prepared and transduced by Miss Rebecca Cohen.

4.2.3 A yeast two-hybrid screen identified coilin as a C9orf72 interacting protein

The function of C9orf72 in the nucleus is unknown. It has been hypothesised that the short C9orf72 protein may be involved in nucleocytoplasmic shuttling due to its localisation at the nuclear envelope in certain cell types (Xiao et al., 2015). Identifying protein interactors can increase understanding of a protein's potential function. Thus, a yeast two-hybrid (Y2H) screen was conducted, with special attention paid to novel nuclear protein interactors of C9orf72. The screen was outsourced to the Protein Interaction Screening of the Genomics and Proteomics Core Facilities, German Cancer Research Centre, Heidelberg, Germany and involved screening a human full-length cDNA library using C9orf72L as the bait protein. Briefly, C9orf72L was cloned into the bait vector pGBKT7, for yeast expression of a C9orf72L protein tagged at its N-terminal end with the transcription factor GAL4 DNA binding-domain. The cDNA library which was screened against contained full-length ORFs expressing 'prey' proteins fused with the corresponding GAL4 activation domain. If the bait and prey proteins interact, GAL4 becomes functional and facilitates transcription from a reporter gene. The screen selected for high-confidence interactions only by rejecting prey proteins that: 1) only had a single hit against C9orf72L and therefore lacked reproducibility or 2) were 'promiscuous' prey proteins that are commonly pulled down in Y2H screens due to their tendency to bind to numerous different proteins. The Y2H identified two proteins as C9orf72L interactors: coilin and synapsin III (Table 4.1). Coilin had 5 hits and synapsin III had 37 hits; both were predicted to be biologically significant. The interaction with coilin, which is a nuclear protein, was taken forward for further investigation to gain a better understanding of the role of C9orf72 in the nucleus. The C9orf72 interaction with synapsin III is currently being investigated by Miss Rebecca Cohen within the lab.

Bait	Prey gene	Number of times prey isolated	Number of times the fragment starts in the 5' UTR	Number of times the fragment starts in the coding sequence	Number of times the fragment starts in the 3' UTR	Average number of bases of the 5'UTR that are part of the hybrid prey protein	Prey promiscuity	Gene description
C9orf72L	SYN3	37	0	37	0	0	1	synapsin III
C9orf72L	COIL	5	0	5	0	0	1	coilin

Table 4.1 Coilin and synapsin III interact with C9orf72L. A Y2H screen used C9orf72L as bait protein and screened a human full-length cDNA library. Proteins isolated more than once and a prey promiscuity less than 4 were identified as positive hits. Prey promiscuity is defined as prey that are repeatedly identified in several screens due to their non-specific binding of bait proteins.

4.2.4 C9orf72 and coilin directly interact *in vitro*

Coilin is a nuclear protein and is the major protein component of nuclear suborganelles called Cajal bodies. It is thought to mediate protein-protein interactions within Cajal bodies, thus acting as a scaffolding protein. The Y2H screen suggested coilin is a protein interactor of C9orf72L. To confirm the interaction and to determine if the interaction is direct, an *in vitro* GST-binding assay was performed. GST, GST-C9orf72S or GST-C9orf72L protein was expressed in Rosetta™ pLysS competent cells as the C9orf72 does not express well in bacterial systems and these cells are designed to enhance expression of eukaryotic proteins that contain rare codons. Glutathione sepharose added to bacterial lysates for protein immobilisation. ^{35}S radiolabelled coilin was made *in vitro* using a coupled transcription/translation reticulocyte system by incorporating ^{35}S-methionine residues, although it is worth noting coilin has only 2 methionines in its protein sequence and hence may only have minimal labelling. ^{35}S-coilin reticulocyte was incubated with the GSH before elution with excess glutathione. Samples were subjected to SDS-PAGE electrophoresis and the pull-down was confirmed by coomassie staining. ^{35}S-coilin was detected using a phosphoimager. The GST pull down was successful, with GST, GST-C9orf72S and GTS-C9orf72L observed by coomassie staining at the correct molecular weights (Figure 4.4). The additional protein bands were previously characterised by Dr Chris Webster using mass spectrometry and were identified as bacterial chaperone proteins CH60 and DnaK.

^{35}S-coilin was pulled down with both C9orf72 short and long protein but not with the GST control, suggesting coilin directly interacts with both C9orf72 protein isoforms.

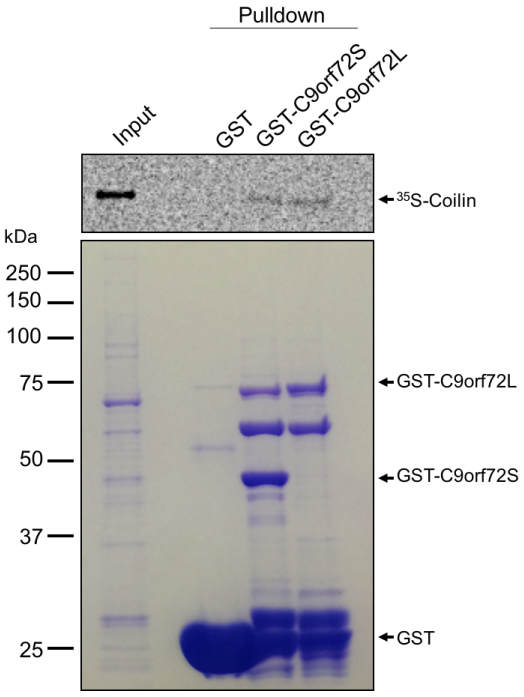


Figure 4.4. Coilin directly binds to C9orf72 *in vitro*. GST-, GST-C9orf72S or GST-C9orf72L was expressed and purified from Rosetta pLysS competent cells and a pull down performed using GSH beads. GST-C9orf72 protein expression was confirmed by Coomassie staining (lower panel). Recombinant ^{35}S-coilin was produced *in vitro* (input) and incubated with the GST-C9orf72 pull down. Eluted proteins were subjected to SDS-PAGE electrophoresis and ^{35}S-coilin detected by phosphor imaging (top panel).

4.2.5 C9orf72 interacts with coilin in HEK293 cells

The *in vitro* GST-binding assay showed C9orf72 directly interacts with coilin. However, while confirming the proteins can bind *in vitro,* the assay is limited in that it cannot be concluded from the results if the interaction is physiologically relevant. For this reason, a co-immunoprecipitation experiment was performed to investigate the interaction between the C9orf72 proteins and coilin in cells. Myc-C9orf72S/L was transfected into HEK293 cells. pCI-neo was included as an empty vector control. Whole-cell lysates were prepared and a sample removed to confirm transfections were successful (Inputs). Anti-coilin antibody was added to the remaining lysate before addition of Protein G sepharose beads for coilin pull-down.

Beads were washed and proteins were eluted by addition of Laemmli buffer and boiling. Samples were subjected to SDS-PAGE electrophoresis for protein separation and transferred to a nitrocellulose membrane for immunoblotting. Anti-coilin and anti-C9orf72 antibodies were used for detection of endogenous coilin and myc-C9orf72S/L respectively. Endogenous coilin was detected in the pCI-neo, myc-C9orf7S and myc-C9orf72L transfected 'Input' samples. Myc-C9orf72S and myc-C9orf72L were likewise detected in their input samples confirming successful overexpression (Figure 4.5A). Coilin protein levels were considerably increased in the IP samples, showing successful immunoprecipitation. In these samples, myc-C9orf72S and myc-C9orf72L were both co-immunoprecipitated with coilin, suggesting the interaction between C9orf72 and coilin does occur in the cellular environment. It appeared myc-C9orf72S was pulled down more than myc-C9orf72L, however there also appeared to be greater expression of myc-C9orf72S vs myc-C9orf72L in the input samples thus there was more protein available for interaction. To support this result, the experiment was performed in reverse. HEK293 cells were transfected with pCI-neo or myc-tagged C9orf72 short or long protein and subjected to nuclear fractionation, to enrich for nuclear proteins. As before, a sample of lysate was removed for inputs. To remaining lysate, a pull-down of myc-C9orf72 was performed by incubation with an anti-myc antibody followed by addition of Protein G sepharose beads. Proteins were eluted and subjected to SDS-PAGE and immunoblot with anti-coilin and anti-myc antibodies. Endogenous coilin and myc-C9orf72S and myc-C9orf72S were detected as expected in the input samples (Figure 4.5B). Both myc-C9orf72S and myc-C9orf72L were enriched in the IP samples, confirming successful immunoprecipitation. Endogenous coilin was pulled down with both C9orf72 isoforms, but as seen in Figure 4.5A, this interaction appeared to be stronger with the C9orf72 short isoform. Together, the two co-immunoprecipitation experiments indicate C9orf72 and coilin interact *in vivo.*

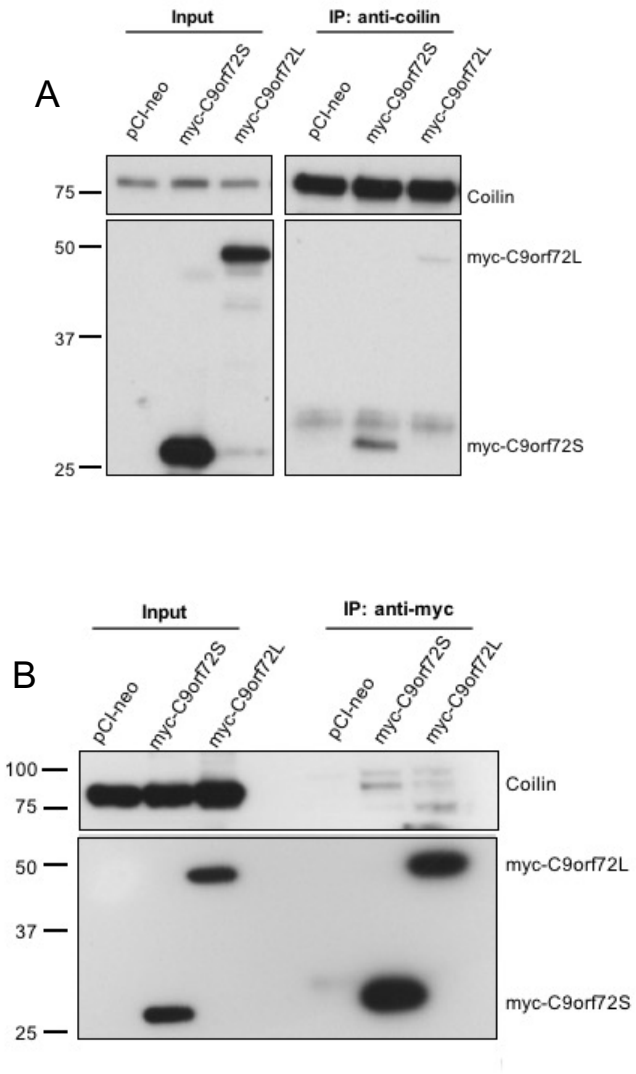


Figure 4.5. Endogenous coilin and myc-C9orf72 interact in cells. HEK293 cells were transfected with pCI-neo (control), myc-C9orf72S or myc-C9orf72L. Immunoprecipitation of endogenous coilin was performed on whole cell lysate using an anti-coilin antibody (A) or immunoprecipitation of myc-C9orf72 was performed on HEK293 nuclear fractions using an anti-myc antibody (9B11) (B). For coilin detection, the top blots were probed using an anti-coilin antibody. For myc-C9orf72 detection, the bottom blots were probed using anti-C9orf72 (sc-138763) (A) or anti-myc (9106) (B).

4.2.6 C9orf72 and coilin are at close proximity in both the nucleus and the cytoplasm

The previous sections in this chapter have shown that C9orf72 localises to both the nucleus and cytoplasm, and interacts with coilin. Next, the localisation of where the C9orf72 and coilin proteins interact was investigated. To do this, a proximity ligation assay was implemented, which can visualise protein-protein interactions *in situ*

(Figure 4.6). Briefly, primary antibodies recognising the two proteins of interest, one raised in mouse and the other in rabbit, are incubated on fixed and permeabilised cells. Secondary antibodies are then added, called PLA probes, which are conjugated to DNA oligonucleotides. Next, hybridising oligonucleotides, which are complimentary to the PLA probe oligonucleotides, are added and ligated to form a closed, circularised DNA product. This circular DNA then acts as a template for DNA polymerase to perform rolling circle amplification, generating a large DNA concatemer at the site of the protein-protein interaction. Fluorescent hybridisation probes, complimentary to the DNA amplicon, are added which bind to produce a single fluorescent spot (PLA signal), allowing the location of the protein interaction to visualised by microscopy. For the proximity ligation assay to produce a PLA signal, the two proteins of interest must be within 40 nm, therefore the assay is only a measure of proximity rather than interaction.

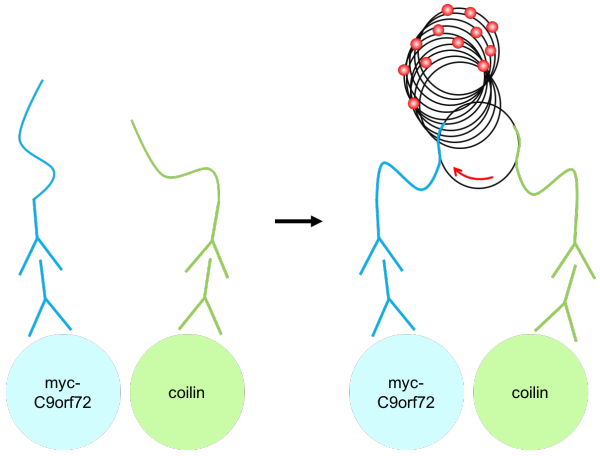
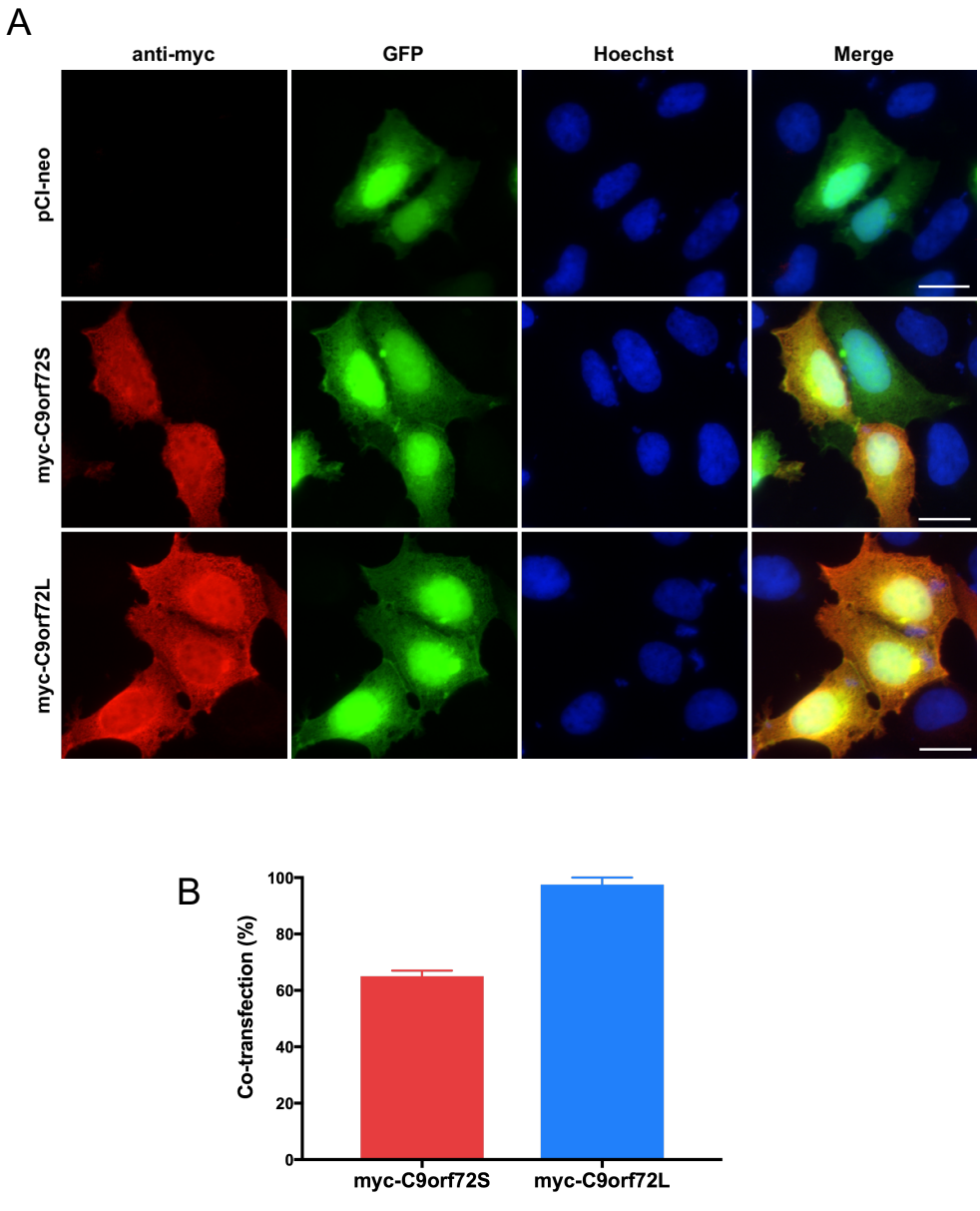


Figure 4.6. The proximity ligation assay. Primary antibodies raised in different species are added against target proteins. Secondary antibodies conjugated to oligonucleotides (PLA probes) bind to the primary antibodies. Hybridisation oligonucleotides bind to the PLA probes and form a circular DNA template which undergoes rolling circle amplification. The RCA generates a DNA concatemer with up to 1000 repeats of the sequence. Fluorescent probes bind the DNA product and concentrate at the site of the protein-protein interaction, producing a single fluorescent PLA signal which can be visualised using fluorescence microscopy.

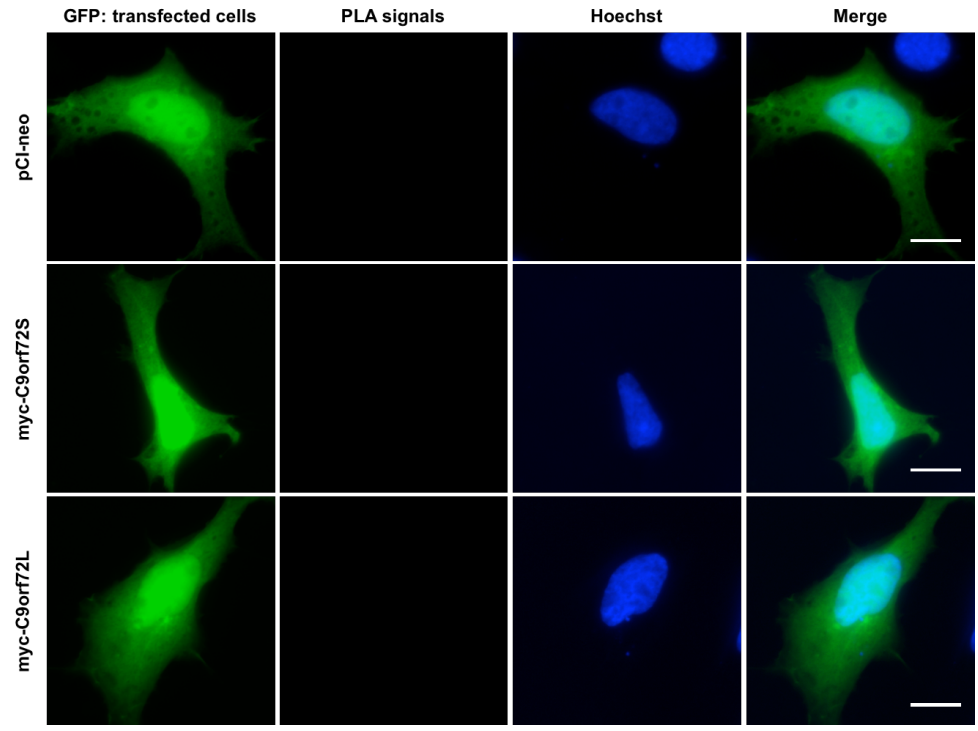
The proximity ligation assay was used to investigate where in the cell myc-C9orf72 and endogenous coilin are in close proximity. HEK293 cells were transfected with pCI-neo or myc-tagged C9orf72 short or long protein. Immunostaining of myc-C9orf72 in the PLA protocol was not possible, therefore in addition to these

transfections, cells received 10% GFP of total DNA transfected. It was hypothesised that GFP transfected cells would be co-transfected with myc-C9orf72, therefore GFP positive cells could be selected for analysis of their PLA signals, ensuring we were only analysing cells overexpressing C9orf72. Cells were fixed, permeabilised and used for either immunofluorescence or PLA. Immunofluorescence cells were immunostained using an anti-myc antibody to confirm successful myc-C9orf72 transfection. Cells successfully expressed myc-C9orf72S, myc-C9orf72L and GFP (Figure 4.7A). The co-transfection efficiency of GFP and the myc-C9orf72 proteins was quantified as the percentage of GFP-positive cells which also expressed myc-C9orf72. GFP and myc-C9orf72L had 100% co-transfection efficiency, whereas myc-C9orf72S was only co-transfected in 60% of GFP-transfected cells (Figure 4.7B).

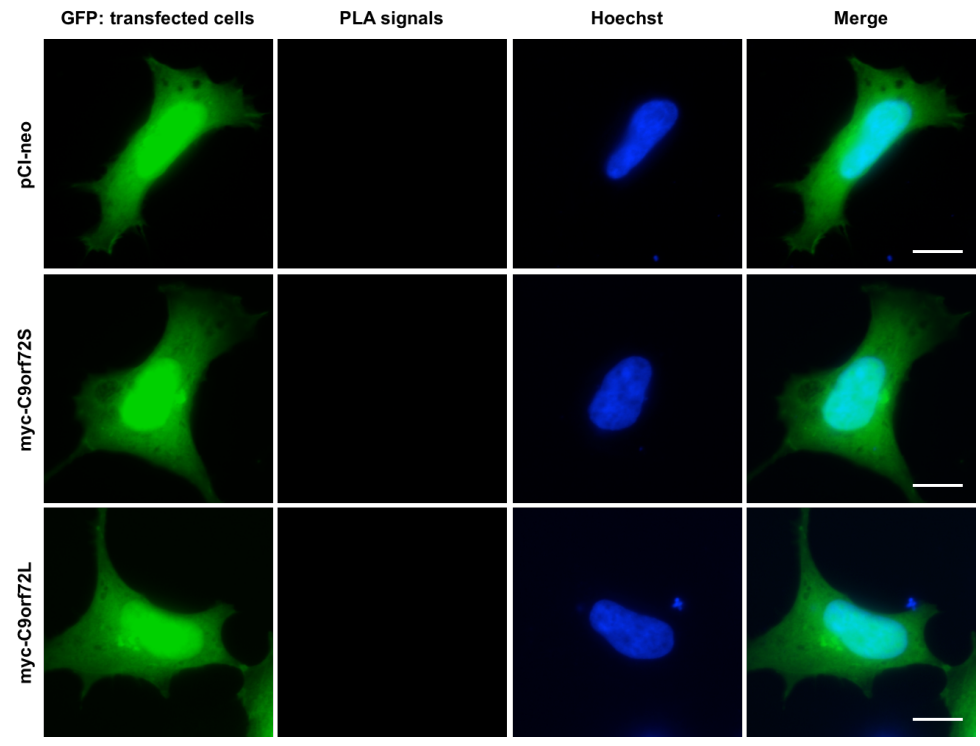
For PLA, cells received either no antibody, anti-coilin antibody, anti-myc antibody, or both anti-coilin and anti-myc antibody. The PLA protocol was completed as described above and PLA signals viewed by fluorescence microscopy. All transfected cells which received no primary antibody, had no detected PLA signals (Figure 4.7C). Likewise, transfected cells which received anti-coilin antibody alone did not generate PLA signals (Figure 4.7D). Cells which were myc-C9orf72S or myc-C9orf72L but not pCI-neo transfected, and which received anti-myc antibody had some PLA signals (Figure 4.7E). Cells which were myc-C9orf72S or myc-C9orf72L but not pCI-neo transfected, and received both anti-coilin and anti-myc antibodies had increased PLA signals located in both the nucleus and cytoplasm of the cells (Figure 4.7F). The PLA signals were quantified per cell (Figure 4.7G), per nucleus (Figure 4.7H) or per cytoplasm (Figure 4.7I). Although some background PLA signals were generated in myc-C9orf72 cells with only the anti-myc antibody, there were significantly more PLA signals in myc-C9orf72 cells which received both antibodies, suggesting the PLA detected close proximity between C9orf72 and coilin (Figure 4.7G/H/I). Both the images and quantification showed approximately half of the total PLA signals were in the nucleus, suggesting the interaction occurs equally between the cytoplasm and nucleus.


A
anti-myc
GFP
Hoechst
Merge
pCI-neo
myc-C9orf72S
myc-C9orf72L
B
Co-transfection (%)
100
80
60
40
20
0
myc-C9orf72S
myc-C9orf72L

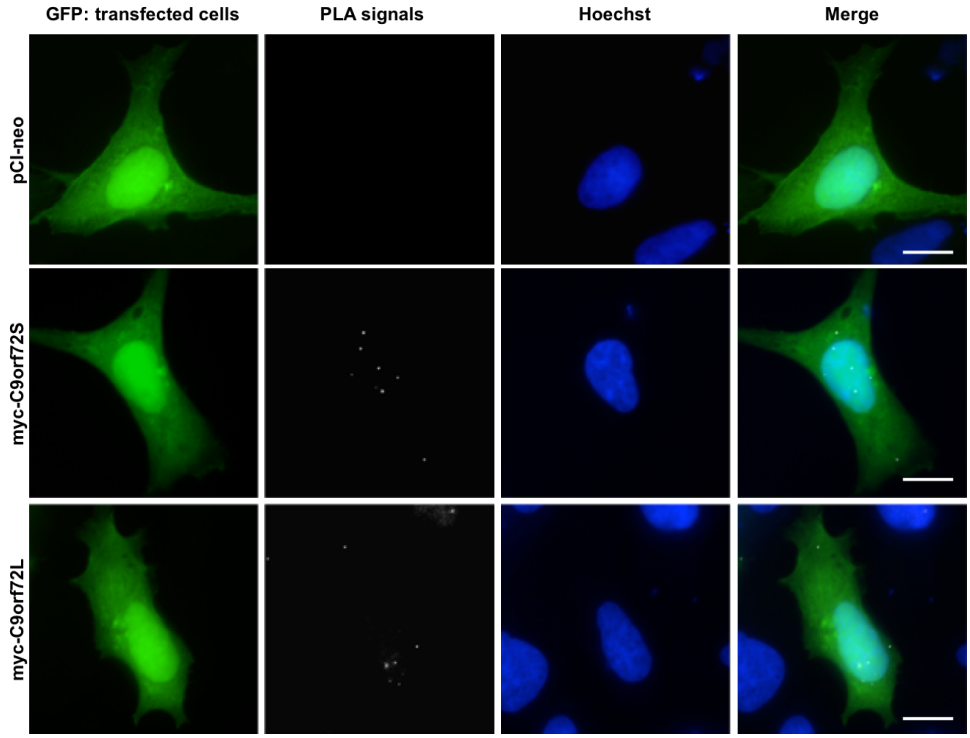
C


GFP: transfected cells
PLA signals
Hoechst
Merge
pCI-neo
myc-C9orf72S
myc-C9orf72L

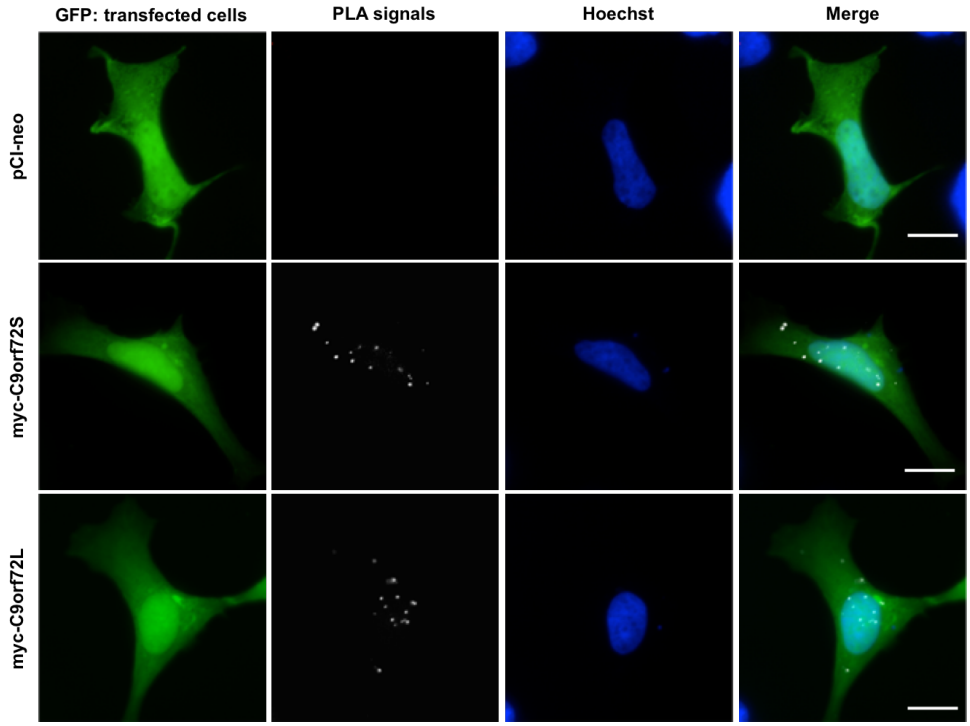
D


GFP: transfected cells
PLA signals
Hoechst
Merge
pCI-neo
myc-C9orf72S
myc-C9orf72L

E


GFP: transfected cells
PLA signals
Hoechst
Merge
pCI-neo
myc-C9orf72S
myc-C9orf72L

F


GFP: transfected cells
PLA signals
Hoechst
Merge
pCI-neo
myc-C9orf72S
myc-C9orf72L

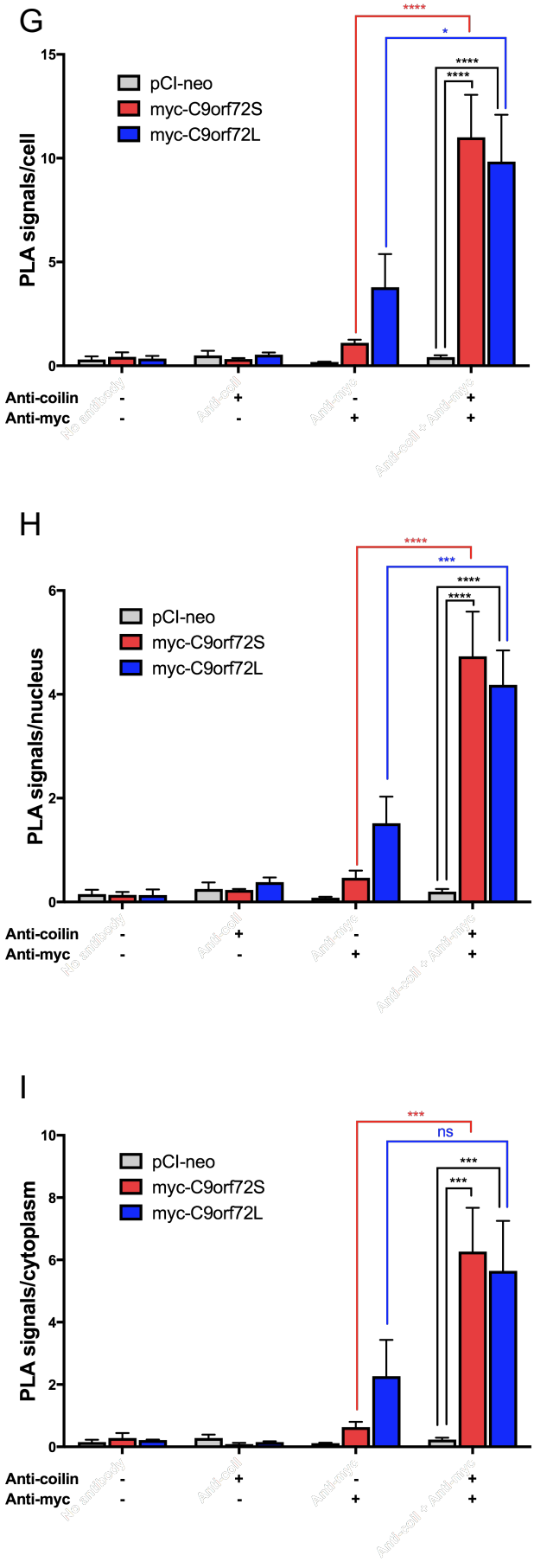

G
PLA signals/cell
pCI-neo
myc-C9orf72S
myc-C9orf72L
****
*
****
****
Anti-coilin
Anti-myc
H
PLA signals/nucleus
pCI-neo
myc-C9orf72S
myc-C9orf72L
****
***
****
****
Anti-coilin
Anti-myc
I
PLA signals/cytoplasm
pCI-neo
myc-C9orf72S
myc-C9orf72L
***
ns
***
***
Anti-coilin
Anti-myc

Figure 4.7. Coilin and C9orf72 are in close proximity in both the nucleus and the cytoplasm. HEK293 cells were co-transfected with pCI-neo (control), myc-C9orf72S or myc-C9orf72L + GFP. (A) Cells were fixed and immunostained using an anti-myc antibody to confirm successful co-transfection of myc-C9orf72 (red) with GFP (green). (B) Co-transfection efficiency quantified as percentage of GFP-positive cells co-transfected with myc-C9orf72S and myc-C9orf72L. Data represent mean ± SEM. n=4 experiments (20 cells per experiment). (C-F) Cells were fixed and a proximity ligation assay performed using no primary antibody (C), anti-coilin antibody (D), anti-myc antibody (E), or anti-coilin and anti-myc antibodies (F). GFP transfected cells (green) were selected for visualisation of PLA signals (white) representative of an in situ protein-protein interaction. Scale bar = 10 μm. (G-I) Quantification of the number of PLA signals per cell (G), per nucleus (H) or per cytoplasm (I). Data represent mean ± SEM; Two-way ANOVA with Tukey's test, * $p \leq 0.05$, *** $p \leq 0.001$, **** $p \leq 0.0001$. n=3 experiments (20 cells per experiment).

4.3 Discussion

In this chapter, C9orf72 was found to localise to the nucleus and interact with the nuclear protein coilin. The nuclear localisation of C9orf72 was shown in both the HEK293 cell line and primary rat cortical neurons. At the time, antibodies that can successfully detect endogenous C9orf72 by immunofluorescence were unavailable and so overexpression of the C9orf72 protein was used. The discussion of the nuclear localisation of C9orf72 in this chapter refers only to the overexpressed protein. High levels of exogenous C9orf72 expression may lead to abnormal distribution of the protein in cells. Both short and long C9orf72 isoforms localised to both the nucleus and cytoplasm, and analysis of their protein sequence reveals a lack of a nuclear localisation signal (NLS), suggesting the proteins do not enter the nucleus by active nuclear transport through the nuclear pore complex. The threshold for passive diffusion through the nuclear pore complex is predicted to be between 30-60 kDa, allowing the short isoform (25 kDa) and possibly the long isoform (50 kDa) to cross the nuclear membrane in this way (Timney et al., 2016). Based on antibody staining, the endogenous short isoform appears more nuclear, whereas the long isoform is cytoplasmic (Xiao et al., 2015). Further studies using additional antibodies are needed to confirm this observation. We suggest entry could be via passive diffusion and the overexpression of the C9orf72 proteins in our studies drives the long isoform into the nucleus.

The short C9orf72 isoform showed more of a distinct nuclear or cytoplasmic localisation in comparison to the long isoform which was more commonly found diffusively through the whole cell (Figure 4.1B). In addition, the short isoform formed nuclear speckles in rat cortical neurons. The nucleus has many nuclear bodies, such as Cajal bodies, gemins, PML bodies and paraspeckles which all have their own distinct protein make-up (Mao et al., 2011b). The C9orf72S nuclear speckles did not exhibit the staining pattern that would be expected for Cajal bodies, which are usually found numbered 1-2 in primary neurons. To determine the identity of the speckles, immunostaining with antibodies recognising protein markers of different nuclear bodies could be used. The localisation and formation of nuclear speckles seen only for the short C9orf72 isoform is consistent with a specific nuclear role for this isoform of C9orf72.

Studies into the biological function of C9orf72 have so far focused on its role in the cytoplasm and involvement in trafficking and autophagy (Sellier et al., 2016; Webster et al., 2016). However, many of the protein interactors of C9orf72 identified in proteomic analysis studies were found to be nuclear proteins, and we and others have shown there is likely a nucleus-associated function. The proteomic screens of C9orf72 completed to date have mostly identified C9orf72 protein interactors by immunoprecipitation of C9orf72 followed by mass spectrometry analysis of its binding partners. The proteins identified using this method may not directly interact with C9orf72 however, but are pulled down as part of the same protein complex. In this project, an unbiased yeast two-hybrid screen was performed to look for novel C9orf72 protein interactors using a different approach. The yeast two-hybrid assay suggested there was an interaction between C9orf72L and coilin. Coilin is the major protein component of Cajal bodies, where it scaffolds and interacts with proteins involved in processing of the splicing machinery, such as the SMN protein, small nuclear ribonucleoproteins (snRNPs) and Sm proteins. Loss of the SMN protein caused by a mutation in the *SMN1* gene leads to the development of Spinal Muscular Atrophy (SMA). SMA has similarities to ALS, characterised by the death of the lower motor neurons and muscular atrophy, and is the leading genetic cause of death in infants (Burghes and Beattie, 2009). Loss of nuclear bodies has also been observed in TDP-43 and FUS-related ALS (Yamazaki et al., 2012). This suggests that there could be a link between Cajal body proteins and motor neuron degeneration, hence the interaction between C9orf72 and coilin was investigated further.

The interaction with coilin was investigated using three different approaches. An *in vitro* GST-binding assay confirmed there was a direct interaction between C9orf72 and coilin, and the co-immunoprecipitation experiments and proximity ligation assay suggested the interaction was biologically relevant. The ^{35}S-coilin bands detected in the GST-binding assay pull downs were faint relative to the input, suggesting the interaction may be weak. The weakness of the signal could have also been due to the low level of ^{35}S radiolabelling of the recombinant coilin protein. To radiolabel coilin, ^{35}S-methionine was added to the *in vitro* coupled transcription/translation system for incorporation into coilin. Analysis of the coilin amino acid sequence shows there are only three methionine residues spread throughout the 576-amino acid sequence (Figure 4.8). As such, even with 100% efficient labelling of coilin, the amount of ^{35}S incorporated would be low, and therefore difficult to detect. A way to increase radiolabelling of coilin would be to use ^{35}S-cysteine in addition to ^{35}S-

methionine (Chen and Casadevall, 1999). Coilin contains 12 cysteine residues (Figure 4.8) therefore labelling with ^{35}S-cysteine would increase the strength of radioactive signal for detection. Unfortunately, commercially available ^{35}S-cysteine products are unsuitable for *in vitro* translation, and in house attempts at its use in such experiments have found its radiolabelling efficiency to be weak, therefore we were unable to take this approach in the GST-binding assay.

MAASETVRLRLQFDYPPPATPHCTAFWLLVDLNRCRVVTDLISLIRQRFGFSSGAFLGLY
LEGGLLPPAESARLVRDNDCLRVKLEERGVAENSVVISNGDINLSLRKAKKRAFQLEEGE
ETEPDCKYSKKHWKSRENNNNNEKVLDLEPKAVTDQTVSKKNKRKNKATCGTVGDDNEEA
KRKSPKKKEKCEYKKKAKNPKSPKVQAVKDWANQRCSSPKGSARNSLVKAKRKGSVSVCS
KESPSSSSESESCDESISDGPSKVTLEARNSSEKLPTELSKEEPSTKNTTADKLAIKLGF
SLTPSKGKTSGTTSSSSDSSAESDDQCLMSSSTPECAAGFLKTVGLFAGRGRPGPGLSSQ
TAGAAGWRRSGSNGGGQAPGASPSVSLPASLGRGWGREENLFSWKGAKGRGMRGRGRGRG
HPVSCVVNRSTDNQRQQQLNDVVKNSSTIIQNPVETPKKDYSLLPLLAAAPQVGEKIAFK
LLELTSSYSPDVSDYKEGRILSHNPETQQVDIEILSSLPALREPGKFDLVYHNENGAEVV
EYAVTQESKITVFWKELIDPRLIIESPSNTSSTEPA

Figure 4.8. Coilin protein sequence. The coilin amino acid sequence is shown. Methionine residues that were S^{35} radiolabelled in the GST-binding assay are highlighted in yellow. Cysteine residues that can additionally be ^{35}S radiolabelled are highlighted in turquoise.

Co-immunoprecipitation experiments between overexpressed myc-C9orf72 and endogenous coilin confirmed the interaction occurred in HEK293 cells. The Co-IPs showed coilin was pulled down with myc-C9orf72 and vice versa. Both proteins were pulled down after immunoprecipitation of the other, suggesting the interaction is genuine and not an artefact of the experiment. The amount of co-immunoprecipitated protein was similar for both Co-IPs also, confirming the strength of the interaction. Unfortunately, an IgG only control was omitted from the anti-coilin IP. Inclusion of the IgG control would have given more confidence that myc-C9orf72 was not binding to the beads.

The proximity ligation assay was used to investigate where in the cell C9orf72 and coilin were in close proximity. There were similar numbers of PLA signals generated with transfection of both myc-C9orf72S and myc-C9orf72L, suggesting coilin interacts with both C9orf72 isoforms equally. However, the co-transfection efficiency of myc-C9orf72S with GFP was found to be only 60%. Therefore, 40% of the cells analysed were unable to generate PLA signals as they were not co-transfected. This will have decreased the mean PLA signals per cell and thus underestimate the interaction between coilin and C9orf72S. Because coilin is a nuclear protein, one might think that coilin would interact more strongly with the C9orf72 short isoform, which has a stronger nuclear presence than C9orf72 long. Interestingly, the PLA showed that coilin and C9orf72 were in close proximity in both the nucleus and cytoplasm. This result was unexpected, as coilin has been shown to have two nuclear localisation signals and has not been detected in the cytoplasm in any of our or others' experiments by western blotting or immunofluorescence (Carmo-Fonseca et al., 1993). One study has described the shuttling of coilin between the nucleus and cytoplasm in *Xenopus* oocytes, suggesting there could be some low levels of coilin in the cytoplasm, but there is a lack of evidence for a shuttling coilin protein in a human cell model (Bellini and Gall, 1999). Coilin interacts with multiple proteins and RNAs in the SMN complex which shuttles between the nucleus, cytoplasm and Cajal bodies (Hebert et al., 2001; Toyota et al., 2010; Xu et al., 2005). C9orf72S has also been suggested to have a role in nucleocytoplasmic shuttling due to its localisation pattern and interactions with nuclear pore complex proteins Importin β1 and Ran-GTPase (Xiao et al., 2015). Therefore, the interaction between C9orf72 and coilin may be facilitating trafficking of proteins and/or RNA complexes between the nucleus and cytoplasm. Coilin has also been shown to undergo degradation by the ubiquitin proteasome system in the cytoplasm. Coilin is ubiquitinated by the Mdm2 ubiquitin ligase in the nucleus and exported to the cytoplasm where it is targeted to the proteasome for degradation (Cantarero et al., 2015). The close proximity between C9orf72 and coilin in the cytoplasm may then be linked with coilin protein degradation. Additionally, coilin will also be available for binding by C9orf72 after it's translation in the cytoplasm.

Of note, PLA signals were also generated in cells which were transfected with myc-C9orf72 but only received anti-myc antibody, suggesting there was some non-specific binding of the rabbit secondary PLA probe to myc-C9orf72. These non-specific PLA signals were significantly less in number versus the PLA signals generated in cells which received both anti-myc and anti-coilin antibody, suggesting

we could detect true C9orf72-coilin interactions in cells which received both anti-myc and anti-coilin antibodies. It should be taken into consideration that some of these PLA signals would be generated from the non-specific binding of the rabbit PLA probe to myc-C9orf72 and do not represent a genuine C9orf72-coilin interaction.

When these experiments were conducted, antibodies recognising endogenous C9orf72 were not available and so the interaction was confirmed using overexpressed myc-C9orf72. Now, an anti-C9orf72 antibody is available which was shown in the previous chapter to successfully detect C9orf72, and these experiments could be repeated to investigate the interaction between endogenous C9orf72 and coilin. Apart from the *in vitro* GST-binding assay, the interaction was investigated in the HEK293 cell line, which was beneficial for confirming that the interaction does take place *in vivo*. However, the C9orf72 mutation causes death of motor neurons, and to determine if the C9orf72-coilin interaction is relevant in ALS, the Co-IP and PLA experiments could also be performed on a primary neuron culture. It would be especially interesting to perform the PLA on the rat cortical neurons in which mVenus-C9orf72S forms nuclear speckles, to see if an interaction with coilin is localised to these speckles.

By using different methods to study the interaction, it imparts confidence the interaction is real and likely has a biological function. Coilin has an N-terminal self-association domain, a central highly disordered region and a C-terminal tudor domain which facilitates protein-protein interactions. It would therefore be predicted that the interaction with C9orf72 is taking place via the C-terminal domain. Conversely, both C9orf72 isoforms were shown to interact with coilin, suggesting coilin binding is on the C9orf72 N-terminal longin domain. A ∆longin-C9orf72 construct could be generated and used in the Co-IP experiments to confirm this. The PLA results suggest this is both a nuclear and cytoplasmic interaction. However, coilin is best known for its role in scaffolding Cajal bodies. Spliceosomal snRNPs undergo a complicated processing pathway that moves between the nucleus, Cajal bodies and cytoplasm, described in detail in section 1.8.2.1 and reviewed in (Matera and Wang, 2014). The role of Cajal bodies in this pathway is thought to involve snRNP remodelling and maturation. Therefore, it could be that C9orf72 is involved in the processing or trafficking of splicing machinery components, and encounters coilin in this pathway. The immunofluorescence experiments suggest C9orf72 is not enriched in Cajal bodies but the consensus

view is that Cajal bodies concentrate spliceosome components to increase efficiency of the processing reactions, but that these also occur throughout the nucleoplasm at a slower rate (Klingauf et al., 2006; Pena et al., 2001). This is supported by the fact CBs are non-essential and are absent in certain cell types (Tucker et al., 2001; Young et al., 2001). Also, the protein components of Cajal bodies are predominantly dispersed throughout the nucleus, and not all coilin-interacting proteins exhibit concentrated localisation in Cajal bodies (Cantarero et al., 2015; Ogg and Lamond, 2002). The next chapter explores what effect C9orf72 has on the appearance and number of Cajal bodies.

5 C9orf72 influences Cajal body numbers

5.1 Introduction

Coilin is the major protein component of nuclear, membrane-less suborganelles called Cajal bodies. They measure approximately 0.1-0.5 μm in diameter and the number of Cajal bodies per nucleus varies by cell type, but typically is between 0-10 (Ogg and Lamond, 2002). Cajal bodies are dynamic structures and the number and size of Cajal bodies has been shown to be affected by many factors such as cell cycle and the transcriptional activity of cells. For example, Cajal bodies are more numerous in cells with a higher transcription and splicing demand, such as neurons and transformed cell lines (Pena et al., 2001; Spector et al., 1992; Young et al., 2001).

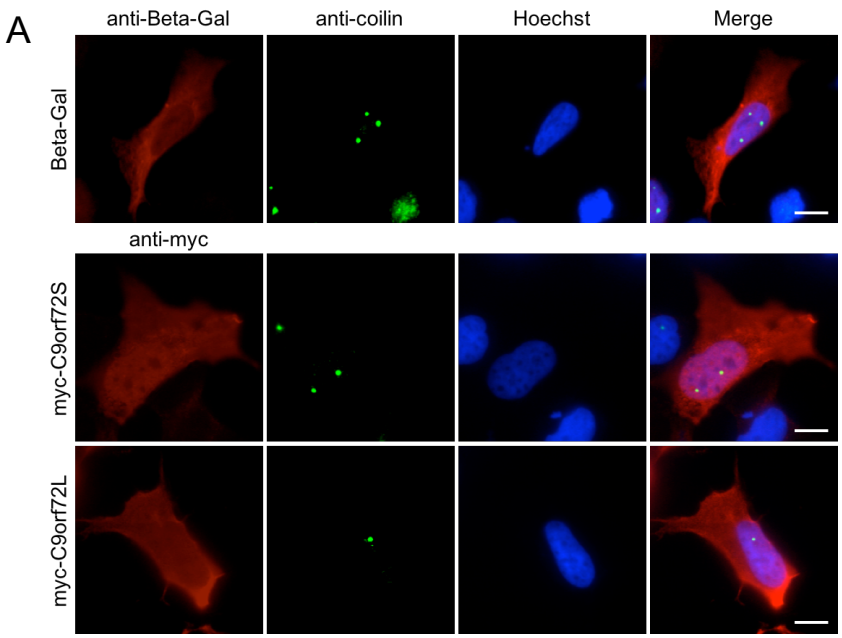
Cajal bodies are involved in the processing and maturation of the small nuclear ribonucleoproteins (snRNPs) that make up the splicing machinery. snRNPs go through a complicated pathway of modifications in the nucleus and cytoplasm before they are fully functional; it has been proposed that snRNPs enter Cajal bodies twice during this pathway where RNP-remodelling and RNA-processing takes place, discussed in detail in Section 1.8.2.1. Cajal bodies can be visualised by immunofluorescence using antibodies against coilin and SMN, the two major protein markers of Cajal bodies. Coilin acts as a scaffolding protein which mediates protein-protein and protein-RNA interactions between the Cajal body components, allowing snRNP processing to occur efficiently (Boisvert et al., 2002; Hebert et al., 2001; Xu et al., 2005). Coilin localisation within Cajal bodies is essential for canonical Cajal body formation and function (Hebert and Matera, 2000; Tucker et al., 2001).

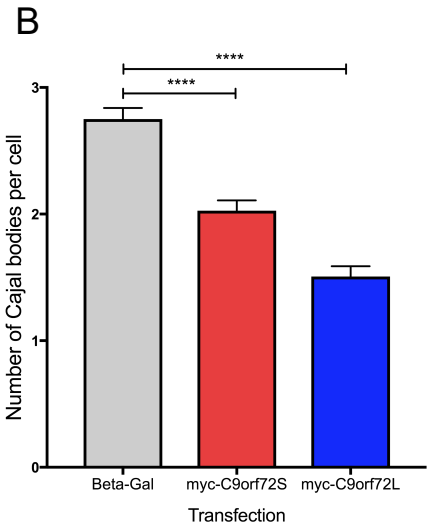
Loss of Cajal bodies in Spinal Muscular Atrophy and the association between several ALS-linked proteins and Cajal bodies suggests they could be defective and contribute to pathology in neurodegenerative disorders (Boulisfane et al., 2011; Wang et al., 2002). In the previous chapter, we showed that C9orf72 interacts with coilin. Haploinsufficiency of C9orf72 is hypothesised to contribute to ALS pathogenesis, and therefore this chapter investigates the effect of C9orf72 on Cajal bodies. C9orf72 protein was overexpressed or depleted in HEK293 cells, primary neurons and iAstrocyte patient cell lines and Cajal bodies investigated by immunofluorescence using an anti-coilin antibody.

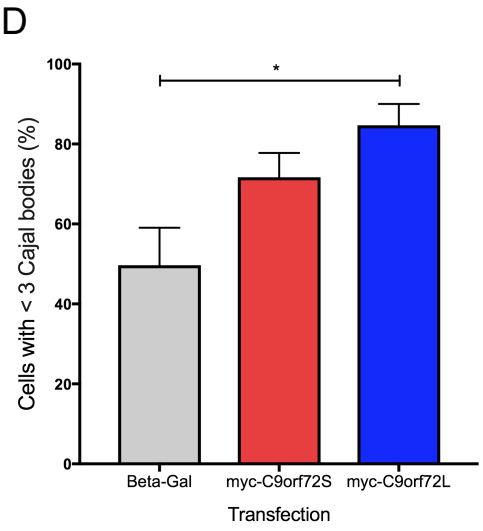
5.2 Results

5.2.1 C9orf72 overexpression reduces the number of Cajal bodies in HEK293 cells

To investigate the effect of C9orf72 on Cajal bodies, the short and long C9orf72 isoforms were overexpressed in HEK293 cells. Cells were transfected with beta-galactosidase (control), myc-C9orf72 short or myc-C9orf72 long. Beta-gal was used as a control for protein expression and was not expected to repeat the effect of transfection and overexpression in HEK293 cells, so we can observe any specific effect of C9orf72. Twenty-four hours post-transfection, cells were fixed and immunostained using an anti-beta-galactosidase antibody or anti-myc antibody to confirm transfections were successful. Co-immunostaining with an anti-coilin antibody was used to visualise Cajal bodies in cells, which appeared as bright spherical foci in the nucleus (Figure 5.1A). To determine if C9orf72 affected Cajal body numbers, the mean number of Cajal bodies per cell was quantified. Cells transfected with either myc-C9orf72S or myc-C9orf72L had an approximate 30% or 50% decrease in mean Cajal bodies in comparison to cells transfected with beta-galactosidase, respectively (Figure 5.1B). To visualise the effects of C9orf72 expression on individual cells we plotted the frequency distribution of cells containing between 0-7 Cajal bodies. The frequency distributions were similar but shifted to the left with C9orf72S or C9orf72L transfection, indicating cells had fewer Cajal bodies after C9orf72 overexpression (Figure 5.1C). The frequency distributions suggested the myc-C9orf72 transfected cell population had more cells with 0 and 1 Cajal bodies in comparison to the beta-gal control. To confirm this, the percentage of cells with < 3 Cajal bodies per cell was quantified for statistical analysis. Both myc-C9orf72S and myc-C9orf72L had more cells with < 3 Cajal bodies per cell (Figure 5.1D). Collectively, these results suggest that overexpression of the C9orf72 protein decreases the number of Cajal bodies in HEK293 cells.


A
anti-Beta-Gal
anti-coilin
Hoechst
Merge
Beta-Gal
anti-myc
myc-C9orf72S
myc-C9orf72L


B
****
****
Number of Cajal bodies per cell
0
1
2
3
Beta-Gal
myc-C9orf72S
myc-C9orf72L
Transfection


D
*
Cells with < 3 Cajal bodies (%)
0
20
40
60
80
100
Beta-Gal
myc-C9orf72S
myc-C9orf72L
Transfection

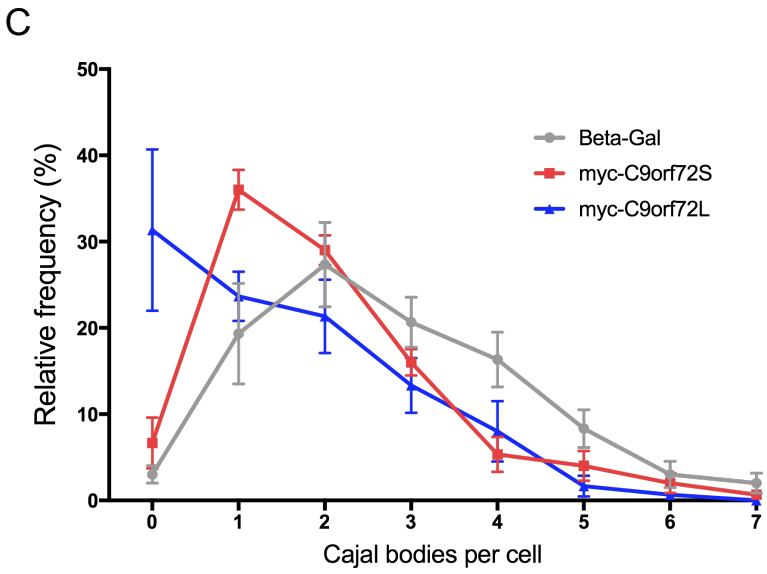

C
Relative frequency (%)
0
10
20
30
40
50
0
1
2
3
4
5
6
7
Cajal bodies per cell
Beta-Gal
myc-C9orf72S
myc-C9orf72L

Figure 5.1. C9orf72 overexpression decreases Cajal bodies in HEK293 cells. (A) HEK293 cells were transfected with Beta-Galactosidase (control), myc-C9orf72S or myc-C9orf72L. Cells were fixed and immunostained using anti-Beta-Gal antibody or anti-myc antibody (9B11) (red) and anti-coilin antibody (green) for Cajal body staining. Scale bar = 10 μm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Dunnett's test, **** $p \leq 0.0001$. n=300 cells from 3 replicate experiments. (C) Frequency distribution of Cajal bodies. Data represent mean ± SEM; n=3 experiments (100 cells per experiment). (D) Cells were quantified for percentage of cells with < 3 Cajal bodies. Data represent mean ± SEM; One-way ANOVA with Dunnett's test, * $p \leq 0.05$. n=3 experiments (100 cells per experiment).

5.2.2 C9orf72 knockdown increases the number of Cajal bodies in HEK293 cells

Overexpression of C9orf72 was shown to decrease Cajal bodies in HEK293 cells. Next, the effect of C9orf72 knockdown on Cajal bodies in HEK293 cells was investigated. Cells were transfected with either a non-targeting control (NTC) siRNA or pooled C9orf72 siRNAs. After 72 hours, cells were fixed and immunostained using anti-coilin and anti-SMN antibodies as markers of Cajal bodies. For both the NTC and C9orf72 siRNA, coilin and SMN showed perfect co-localisation in Cajal bodies (Figure 5.2A). The SMN protein, which shuttles between the nucleus and cytoplasm as part of the SMN complex, was also dispersed throughout the cytoplasm. Cajal bodies were quantified as mean number of Cajal bodies per cell from 300 cells (Figure 5.2B). Cells which were transfected with C9orf72 siRNA had increased Cajal bodies in comparison to the NTC. The frequency distribution of Cajal bodies in the NTC and C9orf72 siRNA cells was plotted which showed C9orf72 knockdown cells had fewer cells with 0, 1 and 2 Cajal bodies per cell in comparison to the NTC but more cells with > 4 Cajal bodies per cell (Figure 5.2C). The significance of the increase in cells containing > 4 Cajal bodies per cell was statistically confirmed in Figure 5.2D, with 30% of C9orf72 siRNA cells containing > 4 Cajal bodies per cell versus 10% of NTC siRNA cells, suggesting a loss of the C9orf72 protein in HEK293 cells can lead to an increase in the number of Cajal bodies in cells. C9orf72 knockdown was confirmed by measuring C9orf72 mRNA levels by RT-qPCR which showed an average 50% reduction across the three experiments (Figure 5.2E).

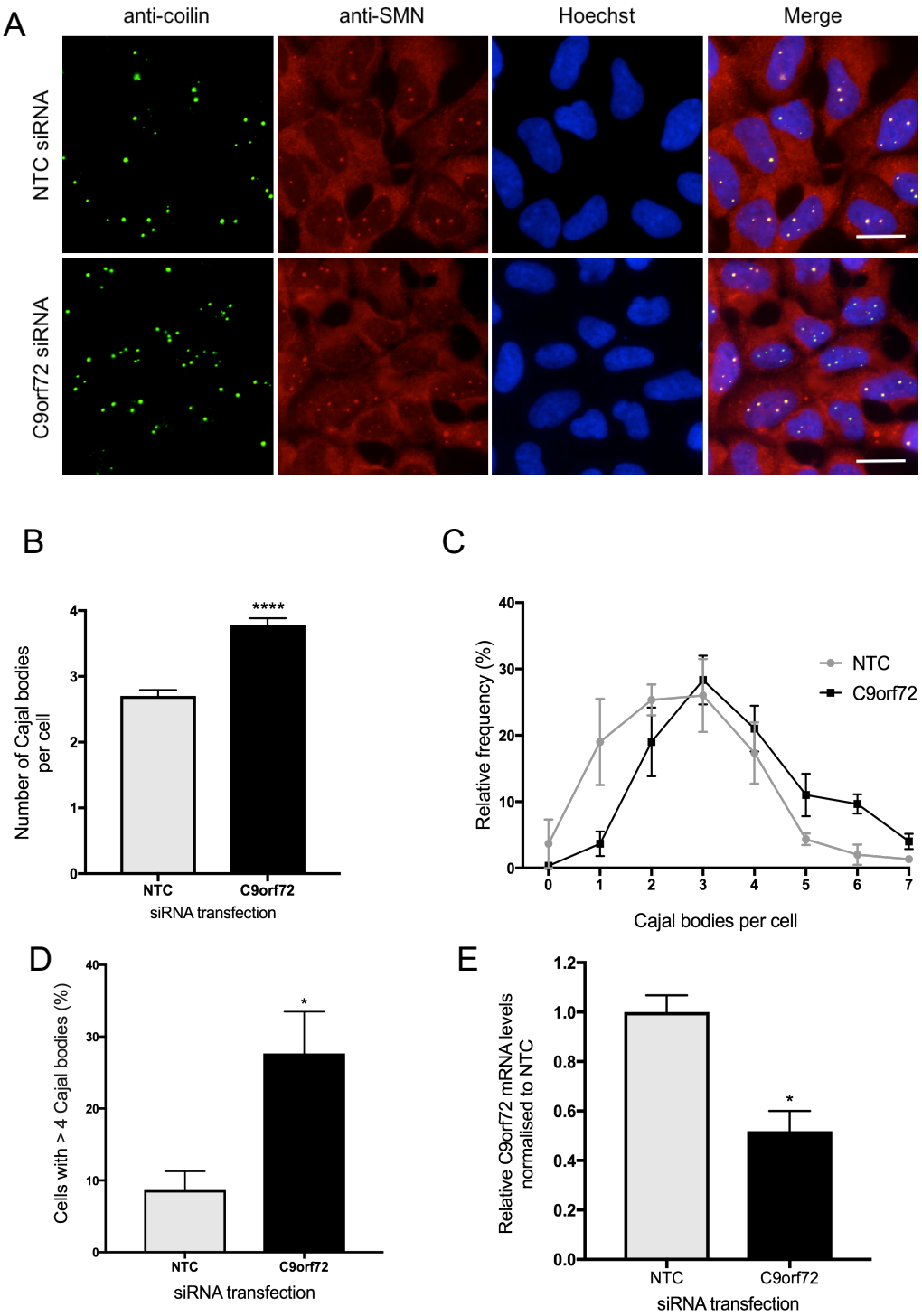


Figure 5.2. C9orf72 knockdown increases Cajal bodies in HEK293 cells. (A) HEK293 cells were transfected with NTC or C9orf72 siRNA. Cells were fixed and immunostained using anti-coilin antibody (green) for Cajal body staining. Scale bar = 10 μm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; Unpaired t-test, **** $p \leq 0.0001$. n=300 cells from 3 replicate experiments. (C) Frequency distribution of Cajal bodies. Mean ± SEM; n=3 experiments (100 cells per experiment). (D) Cells were quantified for percentage of cells with > 4 Cajal bodies. Mean ± SEM; Unpaired t-test, * $p \leq 0.05$. n=3 experiments (100 cells per experiment). (E) C9orf72 knockdown was confirmed by RT-qPCR. Mean ± SEM; Unpaired t-test, * $p \leq 0.05$. n=3 experiments.

To confirm the increase in the number of Cajal bodies observed after C9orf72 knockdown was specific to a loss of C9orf72 protein as opposed to an siRNA off-target effect, the two C9orf72 siRNAs were transfected separately and Cajal body numbers analysed. The C9orf72 siRNAs have different sequences and target different exons of C9orf72, hence the likelihood of them yielding the same off-target effect is improbable. HEK293 cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pooled siRNA. Cells were fixed and immunostained using an anti-coilin antibody as a marker of Cajal bodies (Figure 5.3A). The individual C9orf72 siRNAs as well as the pooled transfection led to an increase in the mean number of Cajal bodies per cell (Figure 5.3B). Analysis of the frequency distribution of Cajal bodies in the different conditions showed that cells that received C9orf72 siRNA had more cells with > 2 Cajal bodies per cell in comparison to the NTC. The percentage of cells with 1 or 2 Cajal bodies per cell was also dramatically decreased in C9orf72 siRNA cells versus the NTC. Successful knockdown of C9orf72 was determined by measuring the C9orf72 mRNA levels by RT-qPCR which showed a 40-50% decrease in C9orf72 mRNA levels with all C9orf72 siRNAs. Both C9orf72 siRNAs caused an increase in Cajal body number when used individually, suggesting a loss of the C9orf72 protein can directly influence Cajal bodies in HEK293 cells.

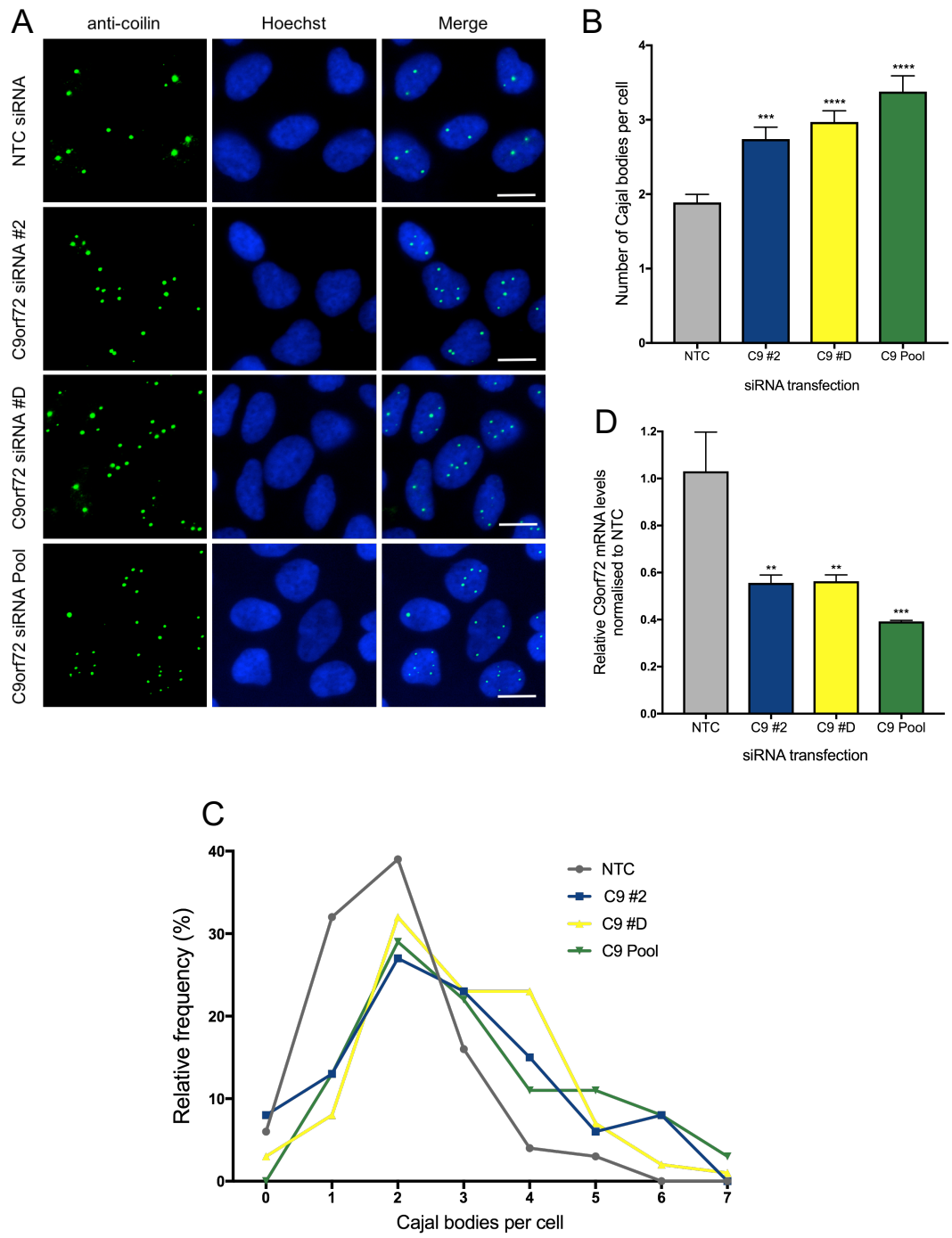
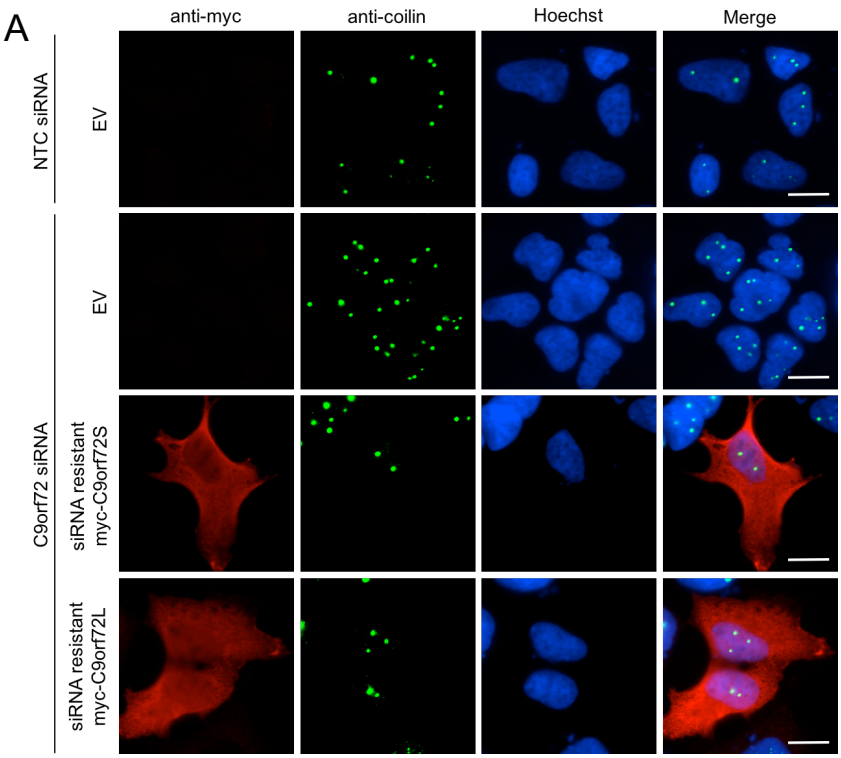
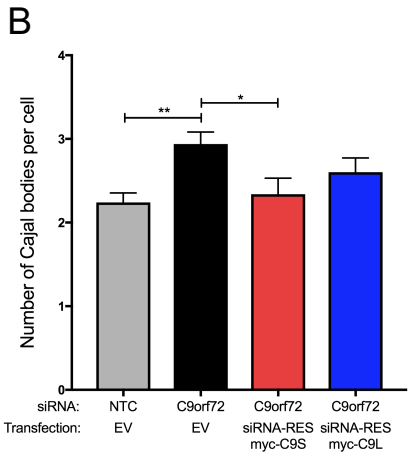


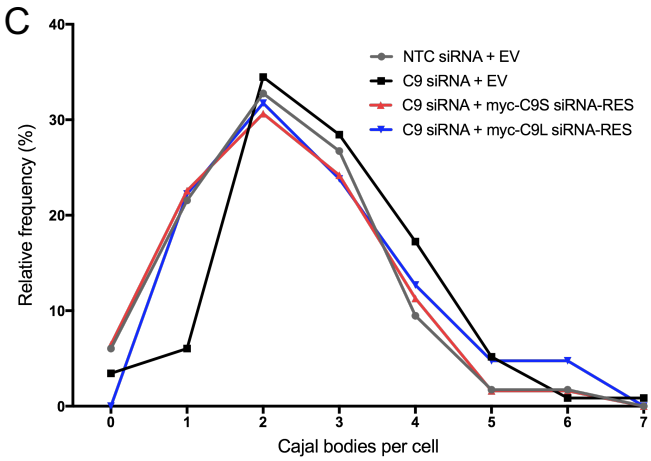
Figure 5.3. Individual C9orf72 siRNA transfection increases Cajal bodies. (A) HEK293 cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pool siRNA. Cells were fixed and immunostained using anti-coilin antibody (green) for Cajal body staining. Scale bar = 10 µm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Dunnett's test, *** $p \leq 0.001$, **** $p \leq 0.0001$. n=100 cells. (C) Frequency distribution of Cajal bodies, n=1. (D) C9orf72 knockdown was confirmed by RT-qPCR. Mean ± SEM; One-way ANOVA with Dunnett's test, ** $p \leq 0.01$, *** $p \leq 0.001$. n=3 triplicates from 1 experiment.

5.2.3 C9orf72 overexpression partially rescues Cajal bodies after C9orf72 knockdown

If the increase in Cajal bodies was specific to a loss of C9orf72, it was hypothesised that Cajal body numbers would be restored with rescue transfections of the C9orf72 protein. HEK293 cells were transfected with NTC or C9orf72 pooled siRNAs. After 72 hours, cells were transfected with pCI-neo (empty vector control), siRNA resistant myc-C9orf72S or siRNA resistant myc-C9orf72L. The generation of the siRNA resistant myc-C9orf72 constructs is described in Chapter 3, Section 3.2.3. Cells were fixed and co-stained using anti-myc and anti-coilin antibodies, for detection of myc-C9orf72 and Cajal bodies respectively (Figure 5.4A). As seen before, the number of Cajal bodies was increased in C9orf72 siRNA cells versus the NTC (Figure 5.4B). There was a partial rescue with myc-C9orf72S but not myc-C9orf72L transfection; expression of myc-C9orf72S significantly decreased the number of Cajal bodies per cell in comparison to C9orf72 siRNA cells transfected with empty vector (Figure 5.4B). The frequency distributions showed C9orf72 siRNA/EV cells had considerably fewer cells that contained 1 Cajal body versus the NTC/EV, consistent with previous results (Figure 5.4C). The frequency distributions of the myc-C9orf72 rescue transfections were more similar in appearance to the NTC distribution than the C9orf72 siRNA distribution, suggesting that overexpression of the C9orf72 proteins could rescue Cajal bodies to normal numbers. The exception to this was a higher percentage of cells with 6 Cajal bodies per cell in the myc-C9orf72L rescue transfection in comparison to all other conditions. C9orf72 knockdown was confirmed as a 60% reduction in C9orf72 mRNA by RT-qPCR (Figure 5.4D). The results suggest loss of the C9orf72 protein can lead to an increase in Cajal bodies in HEK293 cells, and reintroduction of the C9orf72 short or long proteins may be able to rescue Cajal body numbers.


A
anti-myc
anti-coilin
Hoechst
Merge
NTC siRNA
EV
C9orf72 siRNA
EV
siRNA resistant myc-C9orf72S
siRNA resistant myc-C9orf72L


B
Number of Cajal bodies per cell
**
*
siRNA:
Transfection:
NTC
EV
C9orf72
EV
C9orf72
siRNA-RES myc-C9S
C9orf72
siRNA-RES myc-C9L


C
Relative frequency (%)
Cajal bodies per cell
NTC siRNA + EV
C9 siRNA + EV
C9 siRNA + myc-C9S siRNA-RES
C9 siRNA + myc-C9L siRNA-RES

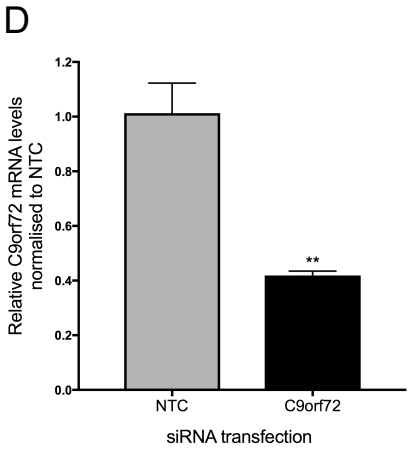
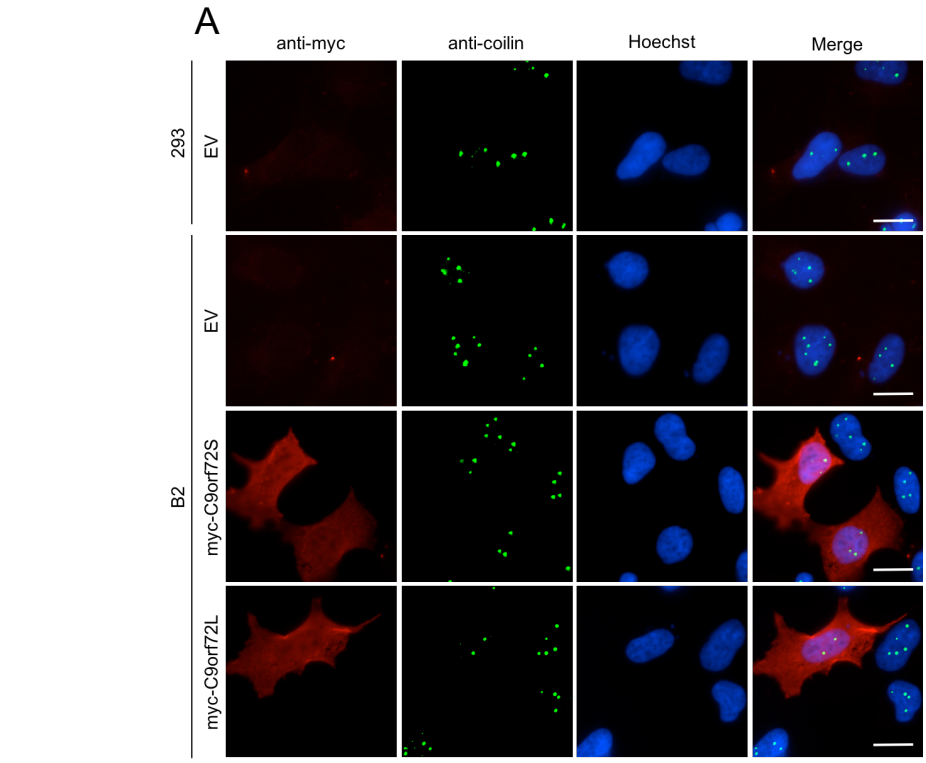

D
Relative C9orf72 mRNA levels normalised to NTC
**
NTC
C9orf72
siRNA transfection

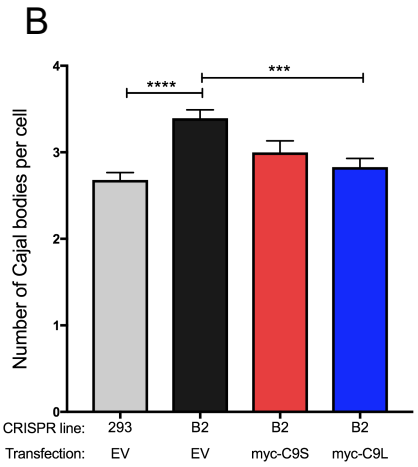
Figure 5.4. C9orf72 expression rescues Cajal body numbers after C9orf72 knockdown in HEK293 cells. (A) HEK293 cells were transfected with NTC or C9orf72 siRNA. Post-knockdown, cells were transfected with pCI-neo (empty vector control; EV), myc-C9orf72S siRNA resistant or myc-C9orf72L siRNA resistant for re-introduction of C9orf72 expression. Cells were fixed and immunostained using anti-myc antibody (red) and anti-coilin antibody (green) for Cajal body staining. Scale bar = 10 μm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Tukey's test, * $p \leq 0.05$, ** $p \leq 0.01$. n (cells) = NTC/EV: 116, C9 siRNA/EV: 116, C9 siRNA/C9S: 62, C9 siRNA/C9L: 63. (C) Frequency distribution of Cajal bodies, n=1. (D) C9orf72 knockdown was confirmed by RT-qPCR. Mean ± SEM; Unpaired t-test, ** $p \leq 0.01$. n=3 triplicates from 1 experiment.

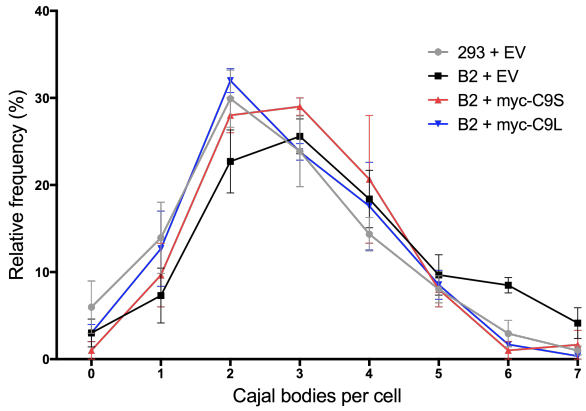
5.2.4 The number of Cajal bodies are increased in a C9orf72 knockout cell line

Knockdown of C9orf72 using two individual siRNAs was shown to increase Cajal bodies in HEK293 cells, and overexpression of the C9orf72 proteins showed a minor rescue of the number of Cajal bodies. siRNA is an acute method of reducing protein expression but in ALS patients we would expect to see a long-term reduction in C9orf72. To investigate the chronic effects of C9orf72 depletion, Cajal bodies were investigated in B2 cells. The B2 cell line was generated from HEK293 cells that underwent CRISPR/Cas9 gene editing to stably knockout the C9orf72 gene and is described in Chapter 3, Section 3.2.1. HEK293 cells (control) or B2 cells were transfected with pCI-neo (empty vector control), myc-C9orf72S or myc-C9orf72L. Cells were fixed and co-stained with anti-myc and anti-coilin antibodies for myc-C9orf72 and Cajal body detection respectively (Figure 5.5A). As expected, the B2 cell line had an increased number of Cajal bodies in comparison to HEK293 cells (Figure 5.5B). The number of Cajal bodies was rescued with C9orf72 protein expression, but the rescue was only significant with C9orf72L (Figure 5.5B). The frequency distribution of Cajal bodies in the B2 cell line shifted to the right in comparison to the HEK293 distribution, suggesting more cells contained higher numbers of Cajal bodies (Figure 5.5C). Transfection with both myc-C9orf72S and myc-C9orf72L in the B2 cell line caused the frequency distribution to shift to the left suggesting expression of the C9orf72 isoforms can rescue Cajal body numbers (Figure 5.5C). Observation of the images and frequency distributions suggested the biggest change in Cajal bodies between B2 and HEK293 cells appeared to be in the number of cells that contained > 4 Cajal bodies per cell. This was confirmed by

quantification of the percentage of cells that had > 4 Cajal bodies per cell (Figure 5.5D). The percentage of cells with > 4 Cajal bodies was two-fold higher in the B2 cell line versus HEK293 cells. Interestingly, expression of both the C9orf72 short and long isoform rescued the percentage of cells with > 4 Cajal bodies in B2 cells to normal levels (Figure 5.5D). However, there was no statistical significance across these differences. The increase in Cajal bodies after both transient and stable C9orf72 knockdown strongly suggests a role of C9orf72 in Cajal body regulation in the HEK293 cell line.


A
anti-myc
anti-coilin
Hoechst
Merge
293
EV
B2
EV
myc-C9orf72S
myc-C9orf72L


B
Number of Cajal bodies per cell
****
***
CRISPR line: 293 B2 B2 B2
Transfection: EV EV myc-C9S myc-C9L


C
Relative frequency (%)
Cajal bodies per cell
293 + EV
B2 + EV
B2 + myc-C9S
B2 + myc-C9L

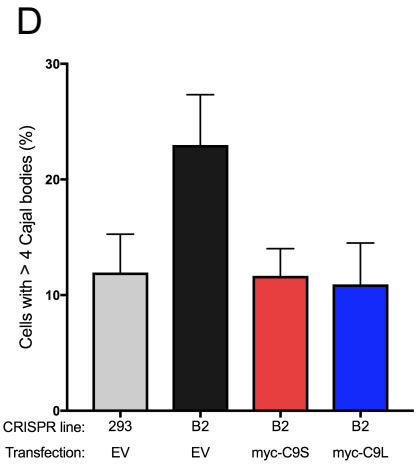

D
Cells with > 4 Cajal bodies (%)
CRISPR line: 293 B2 B2 B2
Transfection: EV EV myc-C9S myc-C9L

Figure 5.5. C9orf72 expression rescues Cajal body numbers in the B2 CRISPR line. (A) Cells were transfected with pCI-neo (empty vector control; EV), myc-C9orf72S or myc-C9orf72L. Cells were fixed and immunostained using anti-myc antibody (red) and anti-coilin antibody (green) for Cajal body staining. Scale bar = 10 μm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Tukey's test, * $p \leq 0.05$, ** $p \leq 0.01$. n (cells from 3 replicate experiments) = 293/EV: 291, B2/EV: 314, B2/C9S: 110, B2/C9L: 194. (C) Frequency distribution of Cajal bodies. Mean ± SEM; n=3 experiments (D) Cells were quantified for percentage of cells with > 4 Cajal bodies. Mean ± SEM; One-way ANOVA with Tukey's test. n=3 experiments.

5.2.5 C9orf72 has no effect on the number of Cajal bodies in rat primary neurons

The results in this chapter have so far suggested C9orf72 can influence Cajal bodies in the HEK293 cell line. Of most interest, loss of C9orf72 led to an increase in Cajal bodies in the nucleus. Haploinsufficiency of the C9orf72 protein has been proposed to contribute to motor neuron degeneration in ALS. We therefore next investigated the effect of C9orf72 on Cajal bodies in a primary neuron cell model. Rat cortical neurons were prepared and transduced at 5 DIV with 4 transducing units of lentivirus expressing GFP (control), mVenus-C9orf72S or mVenus-C9orf72L by Miss Rebecca Cohen. At 12 DIV, neurons were fixed and immunostained using an anti-coilin antibody to mark Cajal bodies. mVenus-C9orf72S formed nuclear speckles but these did not show co-localisation with Cajal bodies (Figure 5.6A). GFP and mVenus-C9orf72L appeared dispersed throughout the neurons. Quantification of the number of Cajal bodies per cell showed there was no difference between mVenus-C9orf72L expressing neurons and GFP, but suggested there was a slight increase in the number of Cajal bodies in mVenus-C9orf72S expressing neurons (Figure 5.6B). However, the frequency distributions of Cajal bodies in GFP, mVenus-C9orf72S and mVenus-C9orf72L neurons were very similar (Figure 5.6C), suggesting overexpression of C9orf72 in neurons has no effect on the number of Cajal bodies.

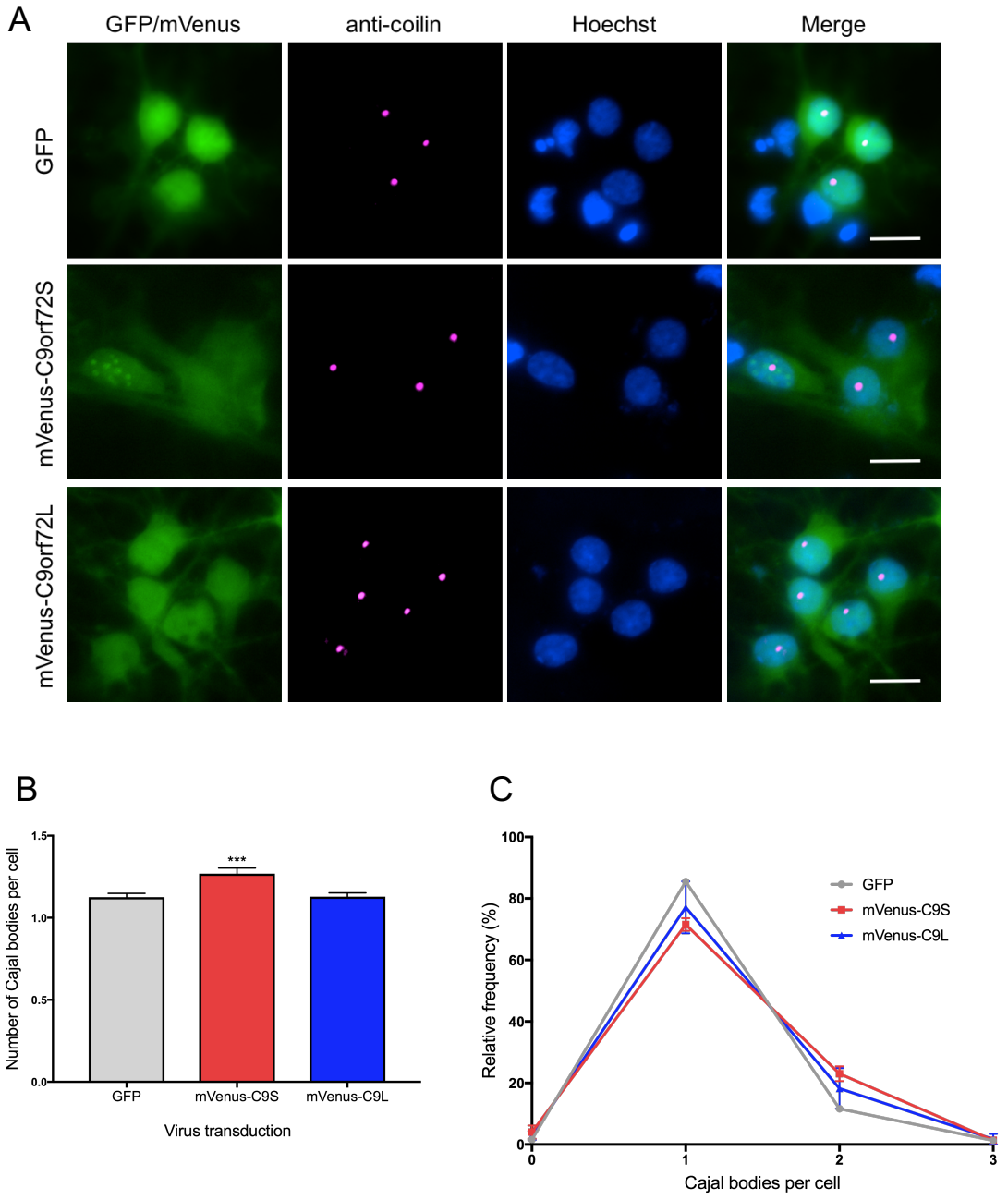


Figure 5.6. Cajal bodies in rat cortical neurons after C9orf72 overexpression. (A) 5 DIV rat cortical neurons were transduced with GFP (control), mVenus-C9orf72S or mVenus-C9orf72L. Neurons were fixed at 12 DIV and immunostained using anti-coilin antibody (magenta) for Cajal body staining. Scale bar = 10 µm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Dunnett's test, *** $p \leq 0.001$. n = approximately 300 cells from 3 replicate experiments (C) Frequency distribution of Cajal bodies. Mean ± SEM; n=3 experiments. Neurons were prepared and transduced by Miss Rebecca Cohen.

In HEK293 cells, knockdown of C9orf72 caused an increase in Cajal bodies. Here, C9orf72 was knocked down in rat hippocampal neurons to investigate if the result is repeatable in a primary cell model. Rat hippocampal neurons were prepared and transduced at 5 DIV with 4 transducing units of lentivirus expressing either GFP-NTC-miRNA (control) or GFP-C9orf72-miRNA by Miss Rebecca Cohen. At 12 DIV, neurons were fixed and immunostained using an anti-coilin antibody for Cajal body detection (Figure 5.7A). Quantification of the mean number of Cajal bodies per cell showed there was no difference in Cajal bodies in neurons transduced with NTC or C9orf72 miRNA (Figure 5.7B). In addition, the frequency distribution of Cajal bodies in NTC and C9orf72 miRNA neurons were the same, confirming knockdown of C9orf72 in rat hippocampal neurons did not change the number of Cajal bodies (Figure 5.7C). C9orf72 knockdown was confirmed by measuring C9orf72 mRNA levels by RT-qPCR and on average was reduced by approximately 50% (Figure 5.7D). Interestingly, overexpression and knockdown of C9orf72 in primary rat neurons did not appear to influence Cajal bodies as was observed in the HEK293 cell line.

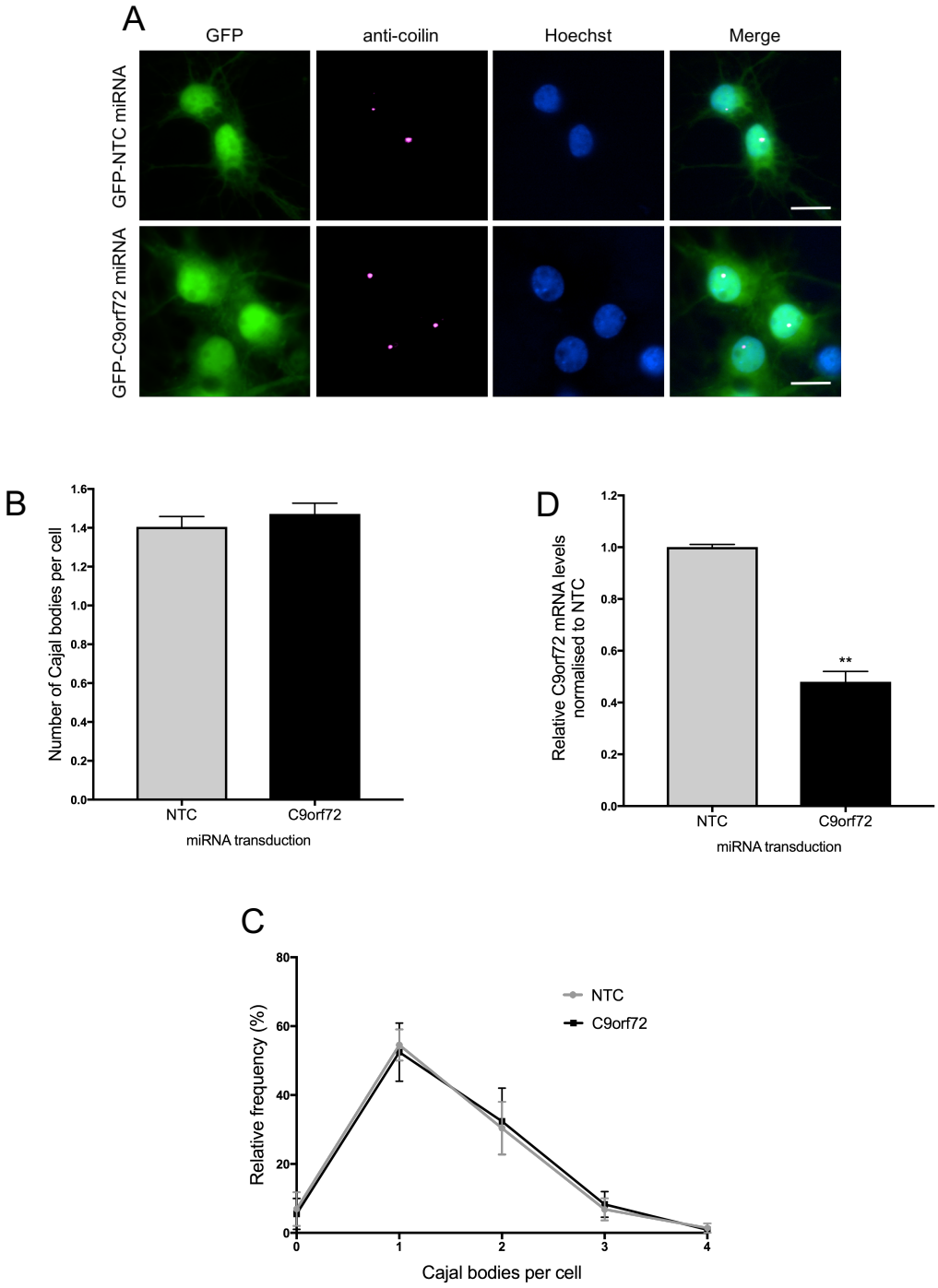


Figure 5.7. C9orf72 knockdown in rat hippocampal neurons does not change Cajal bodies. (A) 5 DIV rat hippocampal neurons were transduced with GFP-NTC or GFP-C9orf72 miRNA. Neurons were fixed at 12 DIV and immunostained using anti-coilin antibody (magenta) for Cajal body staining. Scale bar = 10 µm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; Unpaired t-test. n = 210 cells from 2 replicate experiments (C) Frequency distribution of Cajal bodies. Mean ± SEM; n=2 experiments. (D) C9orf72 knockdown was confirmed by RT-qPCR. Mean ± SEM; Unpaired t-test, ** $p \leq 0.01$. n=2 experiments. Neurons were prepared and transduced by Miss Rebecca Cohen.

5.2.6 Investigating Cajal bodies in C9orf72-ALS iAstrocyte cell lines

Manipulation of C9orf72 protein levels has been shown to influence Cajal bodies in HEK293 cells but not primary rat neurons. To consider the relevance of C9orf72 and Cajal bodies in ALS, we implemented C9orf72-ALS iAstrocyte cell lines, generously developed and provided by Dr Laura Ferraiuolo. The iAstrocytes were differentiated from induced neural progenitor cells from three C9-ALS patients and three non-ALS healthy controls (Meyer et al., 2014). The iAstrocyte cell model was selected because their differentiation protocol is fully optimised and can be generated with ease. In addition, they have been well characterised and the C9orf72-ALS patient cells have been shown to recapitulate C9-ALS neuropathological hallmarks e.g. RNA foci and defective autophagy (Webster et al., 2016). The C9-ALS iAstrocyte lines were investigated by Dr Laura Ferraiuolo for their C9orf72 mRNA levels by performing RT-qPCR on RNA extracted from the iAstrocyte cells and normalising C9orf72 mRNA to GAPDH mRNA. All were found to have decreased C9orf72 mRNA levels in comparison to controls (Table 5.1). RNA foci have been confirmed in the C9-ALS iAstrocyte lines by Dr Adrian Higginbottom and the presence of DPR proteins is yet to be confirmed (Hautbergue et al., 2017).

iAstrocyte line	Sex	Age at biopsy collection (years)	Onset to death (months)	Relative C9orf72 mRNA levels
Non-ALS control 209	F	69	-	0.99
Non-ALS control 155	M	40	-	0.99
Non-ALS control 3050	M	55	-	1.02
C9-ALS patient 201	F	66	19.4	0.43
C9-ALS patient 183	M	50	27	0.35
C9-ALS patient 78	M	66	31.7	0.29

Table 5.1. Age, sex and C9orf72 mRNA levels of iAstrocyte lines. Three non-ALS control and three C9-ALS patient iAstrocyte lines were used. Sex, age at biopsy collection, onset to death and relative C9orf72 mRNA levels are documented. Relative C9orf72 mRNA levels were determined by RT-qPCR by Dr Laura Ferraiuolo.

Non-ALS control 209 and C9-ALS patient 201 were age and sex matched as pair 1. iAstrocytes were transfected with pCI-neo (empty vector control), myc-C9orf72S or myc-C9orf72L. Cells were fixed and immunostained with anti-myc and anti-coilin antibodies for myc-C9orf72 and Cajal body visualisation respectively. Interestingly, control 209 iAstrocytes contained very few or no Cajal bodies (Figure 5.8A). Conversely, all patient 201 iAstrocytes, including myc-C9orf72 transfected cells, had detectable Cajal bodies (Figure 5.8A). Quantification of the number of Cajal bodies per cell showed patient iAstrocytes had significantly more Cajal bodies in comparison to control cells (Figure 5.8B). Expression of myc-C9orf72S had no effect on the number of Cajal bodies in these cells, but expression of myc-C9orf72L interestingly appeared to increase Cajal body numbers (Figure 5.8B). The frequency distribution of Cajal bodies in the control 209 iAstrocytes showed that most cells did not contain Cajal bodies, but approximately 10% had 1 (Figure 5.8C). There was no apparent difference in the frequency distributions between the empty vector, myc-C9orf72S or myc-C9orf72L transfected patient 209 iAstrocytes showing C9orf72 protein expression in these cells did not rescue Cajal body numbers. The results suggest in the iAstrocyte pair 1, Cajal bodies are increased in ALS cells but this may not be due to haploinsufficiency of the C9orf72 protein.

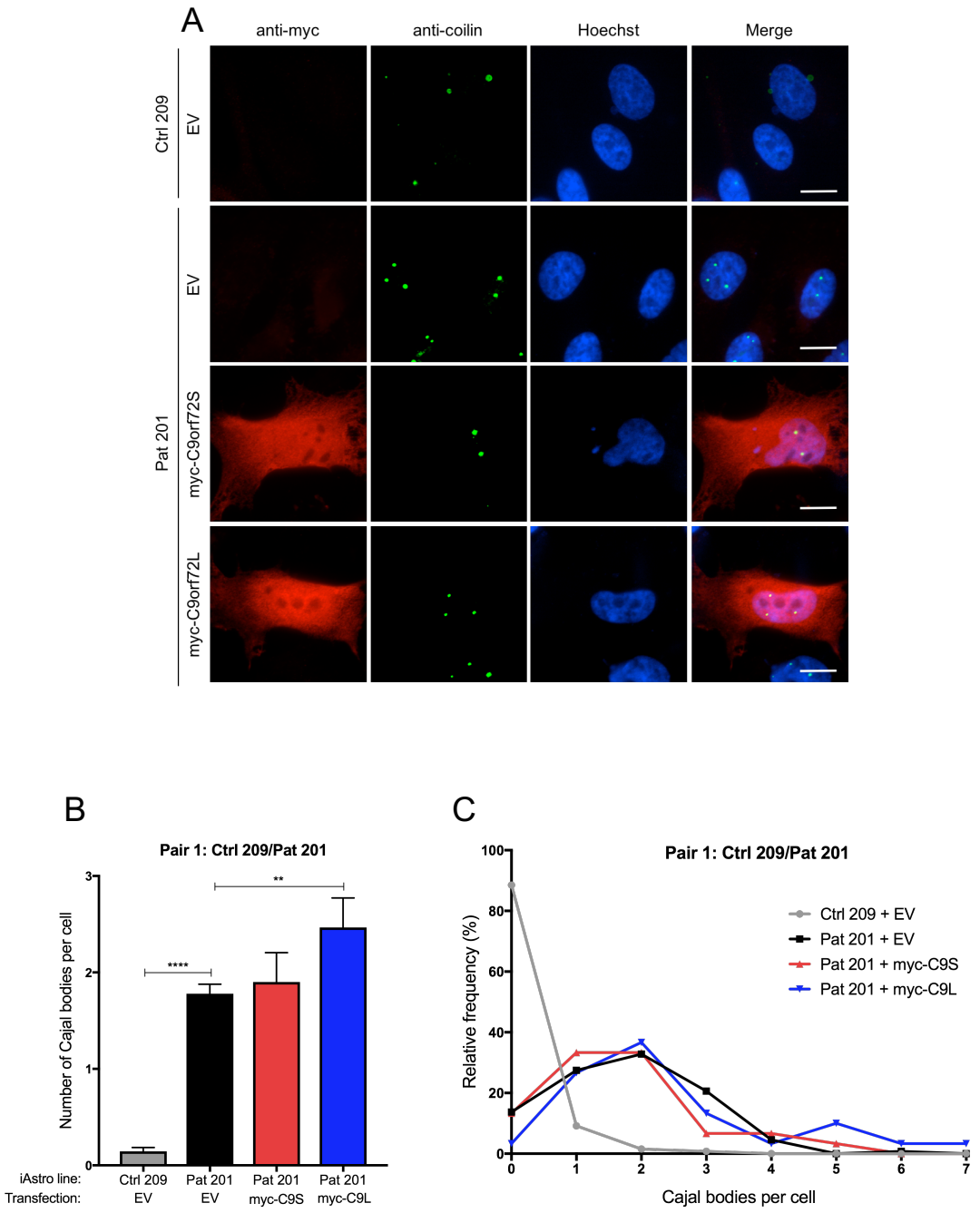


Figure 5.8. Cajal bodies are increased in C9ALS iAstrocyte cells from control/patient pair 1. (A) iAstrocyte cells derived from control (Ctrl 209) or C9ALS patient (Pat 201) iPSC's were transfected with pCI-neo (empty vector control; EV), myc-C9orf72S or myc-C9orf72L. Cells were fixed and immunostained using an anti-myc antibody (red) or anti-coilin antibody (green) as a Cajal body marker. Scale bar = 10 µm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Tukey's test. ** $p \leq 0.01$, **** $p \leq 0.0001$. n (cells) = 209/EV: 131, 201/EV: 131, 201/C9S: 30, 201/C9L: 30. (C) Frequency distribution of Cajal bodies, n=1. iAstrocytes were generated by Dr Laura Ferraiuolo.

We next looked at Cajal bodies in age and sex matched iAstrocyte pair 2: non-ALS control 155 and C9-ALS patient 183. Control 155 iAstrocytes did not contain Cajal bodies (Figure 5.9A). Patient 183 iAstrocytes had Cajal bodies in most cells, and the Cajal bodies remained after myc-C9orf72 transfection (Figure 5.9A). Quantification of mean Cajal bodies per cell confirmed patient 183 had significantly more Cajal bodies versus control 155 (Figure 5.9B). Of interest, overexpression of myc-C9orf72 short and long proteins in the patient 183 iAstrocytes led to an increase in the number of Cajal bodies per cell (Figure 5.9B). The frequency distribution of Cajal bodies in control iAstrocytes showed 80% of cells had no Cajal bodies. Conversely, approximately 70% of patient iAstrocytes contained 1 to 6 Cajal bodies per cell. The frequency distributions for patient 183 iAstrocytes expressing myc-C9orf72 appeared similar to patient 183 cells transfected with empty vector, but there were fewer cells with 0 Cajal bodies and more cells with 4+ Cajal bodies per cell. Again, these results suggest Cajal bodies are misregulated in C9-ALS patients, and expression of C9orf72 protein in these cells may cause Cajal body numbers to increase further.

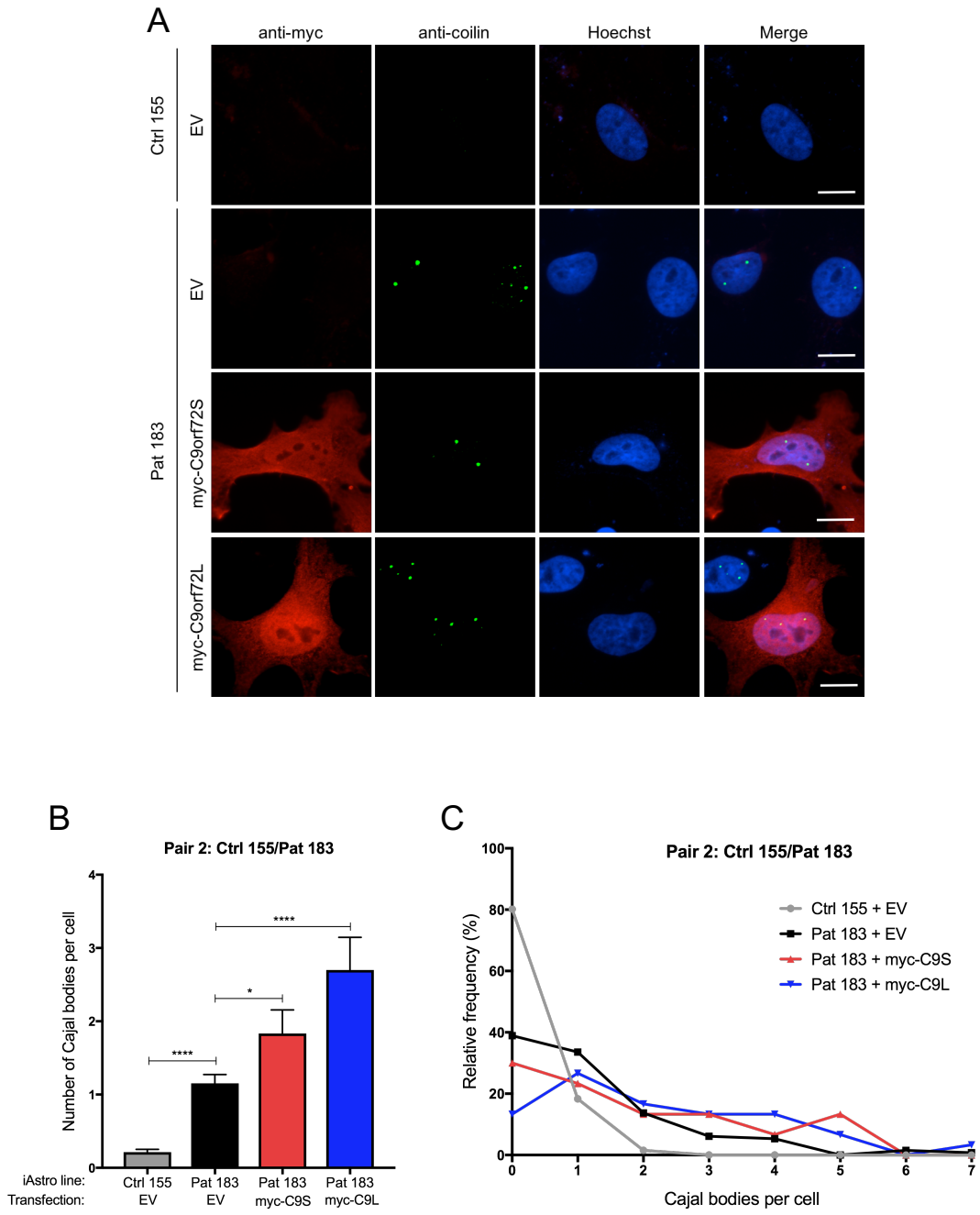


Figure 5.9. Cajal bodies are increased in C9ALS iAstrocyte cells from control/patient pair 2. (A) iAstrocyte cells derived from control (Ctrl 155) or C9ALS patient (Pat 183) iPSC's were transfected with pCI-neo (empty vector control; EV), myc-C9orf72S or myc-C9orf72L. Cells were fixed and immunostained using an anti-myc antibody (red) or anti-coilin antibody (green) as a Cajal body marker. Scale bar = 10 μm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Tukey's test. * $p \leq 0.05$, **** $p \leq 0.0001$. n (cells) = 209/EV: 131, 201/EV: 131, 201/C9S: 111, 201/C9L: 111. (C) Frequency distribution of Cajal bodies, n=1. iAstrocytes were generated by Dr Laura Ferraiuolo.

The final iAstrocyte pair included non-ALS control 3050 and C9-ALS patient 78. As before, the cells were transfected with pCI-neo (empty vector control), myc-C9orf72S or myc-C9orf72L before fixing and immunostaining with anti-myc and anti-coilin antibodies. Control 3050 iAstrocytes had no detectable Cajal bodies (Figure 5.10A). Most patient 78 iAstrocytes had no Cajal bodies, but they could be seen in a small number of cells (Figure 5.10A). Quantification of the mean number of Cajal bodies showed there was no difference in Cajal bodies between control and patient iAstrocytes, or after C9orf72 expression, although there was a notable non-significant increase in Cajal bodies in patient 78 iAstrocytes expressing myc-C9orf72L. The frequency distribution of Cajal bodies appeared similar across all conditions however patient cells with C9orf72L expression had a higher percentage of cells containing > 2 Cajal bodies per cell (Figure 5.10C). Here, it appeared C9-ALS patient 78 did not have increased Cajal bodies in comparison to the non-ALS healthy controls but expression of the C9orf72 long isoform may cause an increase in Cajal bodies.

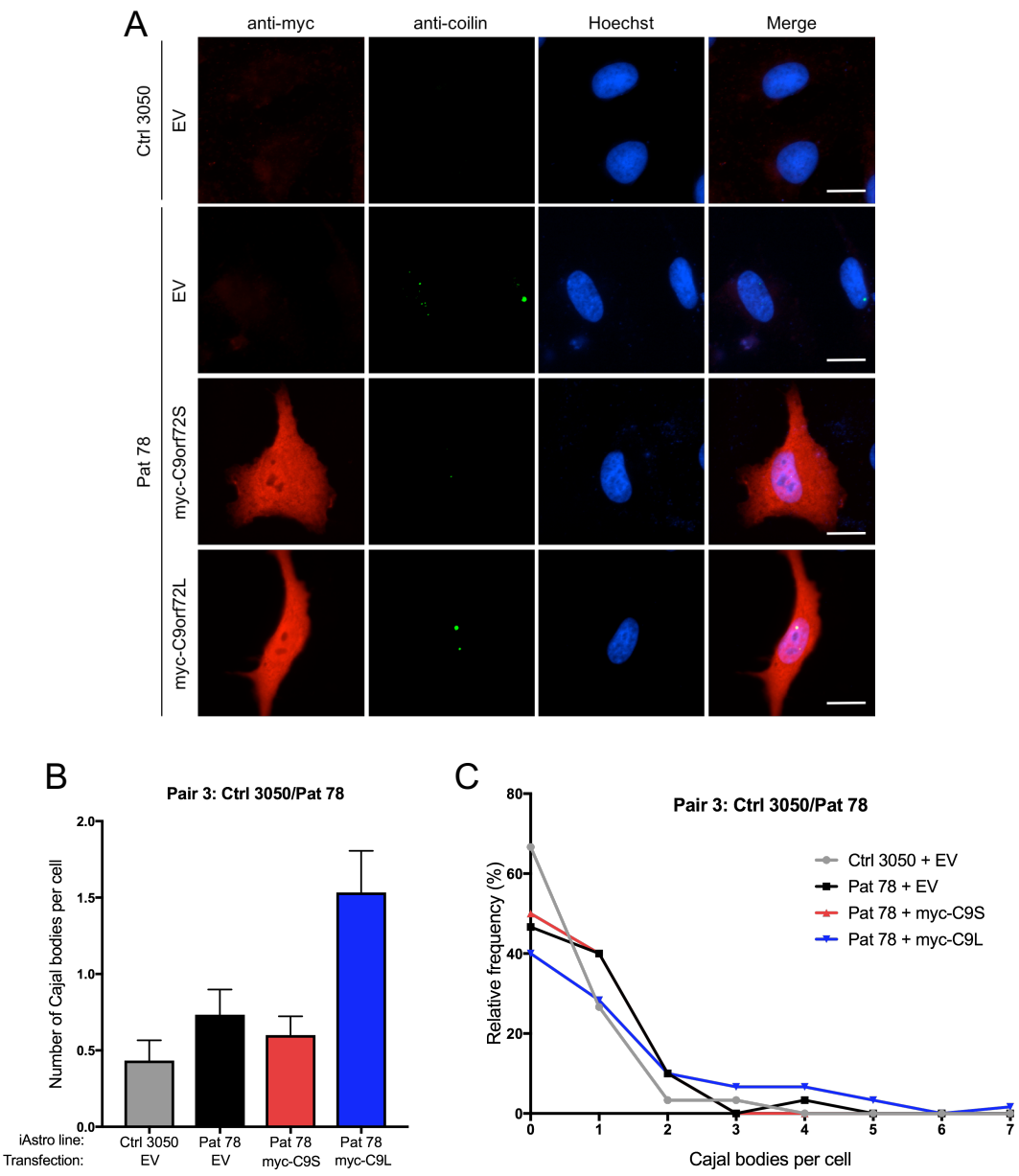


Figure 5.10. Cajal bodies are unchanged in C9ALS iAstrocyte cells from control/patient pair 3. (A) iAstrocyte cells derived from control (Ctrl 3050) or C9ALS patient (Pat 78) iPSC's were transfected with pCI-neo (empty vector control; EV), myc-C9orf72S or myc-C9orf72L. Cells were fixed and immunostained using an anti-myc antibody (red) or anti-coilin antibody (green) as a Cajal body marker. Scale bar = 10 µm. (B) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Tukey's test. n = 30 cells. (C) Frequency distribution of Cajal bodies, n=1. iAstrocytes were generated by Dr Laura Ferraiuolo.

Figures 5.8-5.10 represent the number of Cajal bodies in the individual patient and control iAstrocyte cells and compared these within age and sex matched pairs. An alternative representation of the data can be to group the three non-ALS controls together and compare these against the combined C9-ALS cell lines to show the difference between healthy vs disease. The same data presented in this format is shown in Figure 5.11. As was seen in the pairs, C9-ALS patient cells contained a significantly higher number of Cajal bodies in comparison to non-ALS controls (Figure 5.11A). Overexpression of the C9orf72 short isoform had no effect on the number of Cajal bodies in C9-ALS patients whereas expression of the long isoform led to an increase in Cajal bodies (Figure 5.11A). The frequency distribution of Cajal bodies in non-ALS controls showed most cells had none or only 1 Cajal body, whereas cells derived from C9-ALS patients had a wide range of Cajal body numbers, which was unchanged after C9orf72 expression (Figure 5.11B).

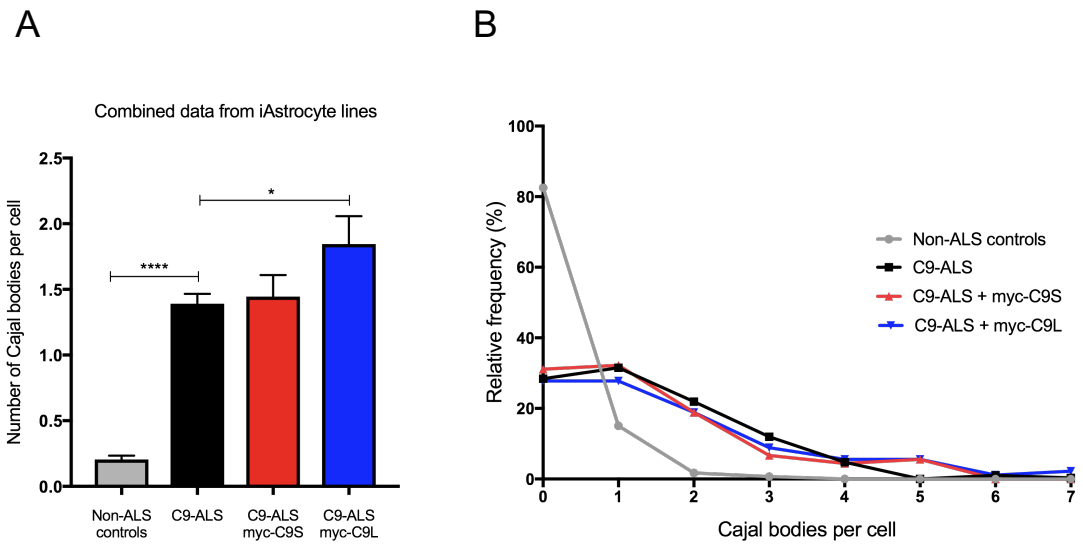


Figure 5.11. Combined results from the iAstrocyte lines. iAstrocyte cells derived from non-ALS control or C9-ALS iPSC's were transfected with pCI-neo (empty vector control; EV), myc-C9orf72S or myc-C9orf72L. Cells were fixed and immunostained using an anti-myc antibody (red) or anti-coilin antibody (green) as a Cajal body marker. (A) Quantification of number of Cajal bodies per cell. Data represent mean ± SEM; One-way ANOVA with Tukey's test. * $p \leq 0.05$, **** $p \leq 0.0001$. n (cells) = 209/EV: 292, 201/EV: 292, 201/C9S: 171, 201/C9L: 171. (B) Frequency distribution of Cajal bodies, n=1. iAstrocytes were generated by Dr Laura Ferraiuolo.

5.3 Discussion

This chapter investigated the relationship between the C9orf72 protein and Cajal bodies. C9orf72 has been shown to interact with coilin and the best-known function of coilin is as the scaffolding protein in Cajal bodies. Hence, it was logical to investigate whether there is an effect on Cajal bodies by C9orf72. We looked at this relationship in three different cell models: HEK293 cells, primary rat neurons and C9-ALS patient iAstrocytes. Interestingly, the effect of C9orf72 on Cajal bodies in the different cell models varied, suggesting the nuclear function of C9orf72 may be cell-type specific.

In the HEK293 cell line, C9orf72 was shown to influence Cajal body numbers. Overexpression of C9orf72 caused a decrease in Cajal bodies whereas knockdown of C9orf72 led to an increase in Cajal bodies. Cajal bodies were increased in both the B2 stable C9orf72 knockout cell line and in response to transient knockdown of C9orf72 using siRNAs. Expression of both C9orf72 isoforms in cells following C9orf72 knockdown appeared to rescue Cajal body numbers, suggesting the change in Cajal body numbers was a result of a direct loss of the C9orf72 protein. To determine if the Cajal body phenotype was relevant to ALS, anti-coilin immunostaining was performed on C9orf72-ALS iAstrocytes. Fibroblasts were collected from three non-ALS controls or three C9orf72-ALS patients and differentiated into induced neural progenitor cells. From that stage, they were further differentiated into iAstrocytes which were shown to have defining C9-ALS characteristics such as reduced C9orf72 mRNA levels and RNA foci which was personally communicated in house. Experiments looking at whether the iAstrocytes produce DPRs are still to be performed. All non-ALS control iAstrocytes had very few or no Cajal bodies. Interestingly, two of the three C9-ALS patients had increased Cajal bodies; patient 78 had Cajal bodies comparable to non-ALS controls. However, Cajal body numbers were not rescued with expression of the C9orf72 protein, suggesting the increased Cajal bodies were not a response to C9orf72 haploinsufficiency. Whilst changes to the number of Cajal bodies were observed in HEK293 and iAstrocyte cells, there was no effect seen in primary rat neurons. Overexpression and knockdown of C9orf72 in neurons had no effect on the number of Cajal bodies. This is likely due to differences in the cells. HEK293 and iAstrocyte cells are mitotically active whereas the primary cells are post-mitotic neurons from rat cortex or hippocampus.

Cajal bodies are dynamic structures and their number and appearance in cells has been shown to change in response to a number of different factors (Table 5.2). The wide range of Cajal body numbers in HEK293 cells suggests they are more dynamic structures in transformed cells. The changes to Cajal bodies listed in Table 5.2 were all observed in transformed cell lines. Table 5.2 shows that manipulation of proteins that localise in Cajal bodies and/or are involved in the snRNP processing pathway can cause an increase or decrease in Cajal bodies. The integrator complex is involved in the processing of the 3' end of newly transcribed snRNAs (Baillat et al., 2005). PHAX and CRM1 are responsible for the nuclear export of snRNAs (Fornerod et al., 1997; Ohno et al., 2000). The SMN protein alongside gemin proteins associates with snRNAs to form the SMN complex and regulate snRNP assembly in the cytoplasm (Massenet et al., 2002; Meister et al., 2001). Finally, the SmB protein is one of the Sm proteins which forms the core heteroheptameric ring of snRNPs (Raker et al., 1996). Changes in these proteins influences Cajal bodies, therefore it could be that C9orf72 also plays a part in the snRNP processing pathway. Fluorescence imaging of C9orf72 in the different cell lines does not show a concentrated localisation in Cajal bodies, therefore it is unlikely to be directly involved in the processing reactions that take place within them.

Table 5.2. Regulating factors of Cajal bodies.

Stimulus	Effect on Cajal bodies	Reference
UV-C DNA damage	Fragmentation	(Cioce et al., 2006)
Adenovirus infection	Fragmentation	(Rebelo et al., 1996)
Serum starvation	Decrease	(Andrade et al., 1993; Cantarero et al., 2015)
SMN overexpression	Increase	(Hao et al., 2007; Sleeman et al., 2003)
SMN depletion	Decrease	(Lemm et al., 2006)
SmB overexpression	Increase	(Sleeman et al., 2001)
Mitosis	Disassemble	(Carmo-Fonseca et al., 1993)
Increased transcription demand	Increase	(Lafarga et al., 1998)
Integrator complex depletion	Coilin mislocalisation into perinucleolar caps	(Takata et al., 2012)
PHAX depletion	Coilin mislocalisation into perinucleolar caps	(Lemm et al., 2006; Takata et al., 2012)
CRM1 inhibition	Coilin mislocalisation into perinucleolar caps	(Takata et al., 2012)
VRK1 depletion	Decrease	(Cantarero et al., 2015)

C9orf72 could be involved in the nuclear export of snRNAs. A study in *Xenopus* oocytes showed that U snRNAs localise to Cajal bodies before export to the cytoplasm, and inhibition of the RNA exporter PHAX caused an accumulation of the U1 snRNA into Cajal bodies (Suzuki et al., 2010). The co-localisation of snRNAs with their nuclear export factors in Cajal bodies has led to the wide-held idea that Cajal bodies act as quality control sites where export proteins assemble onto snRNAs before trafficking to the cytoplasm. Failures in the RNA export machinery therefore leads to an accumulation of immature snRNAs which are tethered within Cajal bodies. A loss of the C9orf72 protein could hence cause a build-up of snRNAs in the nucleus which then get 'stored' in Cajal bodes, the number of which increases in response. In trying to understand Cajal body differences between cell lines, one study found overexpression of the SMN protein, which increases Cajal bodies in transformed cells, did not induce Cajal body formation in primary DFSF1 cells (Sleeman et al., 2001). Interestingly however, overexpression of the SmB protein did lead to an increase in Cajal bodies in both primary and transformed cells. The conclusion drawn from these studies was there must be a signal requiring increased snRNP biogenesis to upregulate Cajal bodies in primary cells. This supports the idea that the presence of Cajal bodies in neurons is due to their increased transcription demand and hence a greater need for snRNP production. This is observed in figures 5.6-5.10 where wild-type neurons contain approximately 1 Cajal body per cell whereas wild-type iAstrocyte cells contain < 0.5 Cajal bodies per cell. Leading from this, the C9orf72 protein in primary neurons could function in the pathways discussed above, for example in the trafficking of the snRNA export complex, but changes in this pathway do not generate the same Cajal body phenotype.

There is evidence the C9orf72 protein may function in nucleocytoplasmic transport. An interaction between C9orf72 and Ran-GTPase has previously been confirmed by co-immunoprecipitation (Xiao et al., 2015). Interestingly, Ran-GTPase interacts with CRM1 and PHAX to form the snRNA export complex and facilitate transport of snRNAs from the nucleus to the cytoplasm (Ohno et al., 2000). C9orf72 has been shown to interact with Rab-GTPases and several of the proteins in the autophagy initiation complex to facilitate their trafficking in the cytoplasm (Sullivan et al., 2016; Webster et al., 2016). Ran-GTPase and Rab-GTPases are similar proteins and are both part of the Ras superfamily of small GTPases (Wennerberg et al., 2005). Hence, it is feasible C9orf72 could form part of the snRNA export complex via an interaction with Ran-GTPase. Expression of the PHAX or CRM1 proteins tagged

with a photoactivatable fluorescent protein would allow us to visualise the snRNA trafficking from Cajal bodies to the cytoplasm in real time, and the effect loss of C9orf72 has on this pathway. It would also be valuable to do co-immunoprecipitation experiments between C9orf72 with CRM1 and PHAX to determine if C9orf72 is a protein component of the snRNA export complex. Finally, it would also be interesting to knockdown Ran-GTPase and observe Cajal bodies numbers in response. The increased Cajal body phenotype we observe could be expected if Ran-GTPase and C9orf72 are interacting in the same protein complex in the same pathway. Interestingly, the arginine-containing C9orf72 DPR proteins have been demonstrated to induce nuclear Ran defects, suggesting a possible convergence between C9orf72 loss of function and toxic DPR generation (Boeynaems et al., 2016; Jovičić et al., 2015; Shi et al., 2017).

Another common regulator of Cajal bodies is the cell cycle. Cajal bodies have been shown to disassemble during mitosis and reassemble in daughter cells during early-$G_{1,}$ with the number of Cajal bodies peaking during the G_1/S phase (Andrade et al., 1993; Carmo-Fonseca et al., 1993; Dundr et al., 2004). The disassembly and reassembly of Cajal bodies throughout the cell cycle is thought to be regulated by coilin phosphorylation. Coilin is a phosphoprotein, with 11 serine residues identified so far as available for phosphorylation (Toyota et al., 2010). Mitotic coilin is hyperphosphorylated which has been shown to decrease the ability of coilin to self-interact and cause Cajal body disassembly (Carmo-Fonseca et al., 1993; Hearst et al., 2009; Hebert and Matera, 2000). Dephosphorylation of some, but not all, coilin serine residues in the G_1 phase of the cell cycle causes reassembly of Cajal bodies, and interestingly, the SMN protein shows preferential binding to a hypophosphorylated coilin which may facilitate de novo Cajal body formation post-mitosis (Carrero et al., 2011). Complete dephosphorylation of coilin however, ablates the ability for coilin to form Cajal bodies, suggesting some coilin phosphorylation is required for Cajal body assembly (Hearst et al., 2009). The phosphorylation state of coilin is controlled by the nuclear kinase VRK1 which itself is cell cycle regulated and can phosphorylate coilin on eight of its eleven serine residues (Sanz-García et al., 2011; Valbuena et al., 2008). VRK1 is thought to regulate Cajal body dynamics by phosphorylating coilin in a specific phosphorylation pattern in a cell-cycle dependent manner (Cantarero et al., 2015). Interestingly, VRK1 phosphorylation of coilin has also been shown to inhibit ubiquitination of coilin and its targeting for proteasome degradation. Knockdown of VRK1 or inhibition of its kinase activity by serum starvation causes a decrease in Cajal bodies, which can be

rescued by the proteasome inhibitor MG132 (Cantarero et al., 2015). This suggests the effect of coilin phosphorylation on Cajal body regulation is two-fold: (1) It influences the binding affinity for the self-interaction of coilin or between coilin and other Cajal body proteins and (2) coilin phosphorylation stabilises the coilin protein and prevents its ubiquitination and export to the cytoplasm for proteasome degradation.

Based on the results presented in this chapter, there is evidence C9orf72 may be involved in cell proliferation and/or post-translational modifications of coilin. C9orf72 influenced Cajal bodes in HEK293 and iAstrocyte cells, both of which are proliferating cell lines. No effect was observed however in non-dividing primary neurons. Interestingly, other DENN proteins which act as GEFs for Rab proteins have been reported to have functions in cell cycle regulation and proliferation (Lim et al., 2004; Thayanidhi et al., 2012). Overexpression and knockdown of C9orf72 could be performed on synchronised HEK293 cells, which would allow any cell-cycle dependent effect of C9orf72 on Cajal bodies to be visualised more clearly. Perhaps coilin phosphorylation is changing in response to C9orf72 protein levels, leading to increased/decreased self-interaction and changing numbers of Cajal bodies. One way to test this would to be to implement mass spectrometry to determine which serine residues are phosphorylated in normal HEK293 cells or cells overexpressing or depleted in C9orf72. Another approach could be to subject whole cell lysates of HEK293 cells after C9orf72 overexpression/knockdown to electrophoresis on a PhosTag acrylamide gel which allows for greater separation of phosphoproteins based on their level of phosphorylation. Immunoblot with anti-coilin could compare coilin phosphoisoforms between samples. Basal numbers of Cajal bodies in the transformed versus primary cell lines in this chapter were different, with HEK293 cells commonly containing 3 or more Cajal bodies whereas the primary neurons typically had 1-2 Cajal bodies per cell. One study compared the phosphorylation levels of coilin in transformed HeLa cells versus WI-38 primary cells and found multiple coilin phosphoisoforms existed in the HeLa cells whereas coilin in the WI-38 cells was stably hyperphosphorylated (Hearst et al., 2009). This suggests the difference in Cajal body numbers could be due to differences in basal levels of coilin phosphorylation in transformed versus primary cells.

In addition, Cajal bodies have been linked with telomere maintenance. Telomeres are protein-bound, repeated DNA sequences on the end of linear chromosomes which shorten with each round of cell division (Smogorzewska and de Lange,

2004). Once telomeres reach a certain length, they signal to the cell to undergo apoptosis or enter senescence, hence their common association with aging (McEachern et al., 2000). To counter shortening, telomeres are maintained and synthesised by the telomerase enzyme, assembled from the human telomerase RNA (hTR) and human telomerase reverse transcriptase (hTERT) (Greider and Blackburn, 1989; Mitchell et al., 1999). The telomerase component hTR is found in Cajal bodies, which themselves have been shown to directly associate with telomeres (Jady et al., 2005; Tomlinson et al., 2006). Interestingly, coilin, which we have shown interacts with C9orf72, is responsible for the recruitment of telomerase to telomeres alongside telomere Cajal body protein 1 (TCAB1) (Stern et al., 2012). Telomere maintenance is essential in transformed cells such as HEK293s which undergo indefinite cell division, and it has been suggested their high basal levels of Cajal bodies are for telomerase assembly (Lafarga et al., 2017). It's possible the C9orf72-coilin interaction facilitates telomerase trafficking to telomeres and hence the Cajal body phenotype was not observed in primary neurons which are post-mitotic and do not rely on this pathway.

In the C9orf72-ALS iAstrocytes, the Cajal body phenotype was not rescued with C9orf72 protein expression, suggesting other mechanisms may also be at play. Overexpression of DPR proteins GR and PR in HeLa cells has been shown to cause disassembly of Cajal bodies, therefore increased Cajal bodies are unlikely to be caused by DPR expression in the iAstrocytes (Lee et al., 2016). On the other hand, it has also been shown that DPR proteins GR and PR can bind the U2 snRNP and inhibit it's assembly which may cause cells to upregulate snRNP production and indirectly increase Cajal bodies in response to the increase in snRNP demand (Yin et al., 2017). The RNA foci in C9-ALS bind and sequester RNA binding proteins and may contribute to splicing defects seen in disease (Cooper-Knock et al., 2014; Prudencio et al., 2015). Cells may therefore increase transcription of snRNA genes to compensate for the defective splicing and increased Cajal bodies are a visible phenotype of this. Astrocytes have been shown to exacerbate degeneration of motor neurons in ALS, therefore it is worthwhile understanding the molecular pathology in these cells (Haidet-Phillips et al., 2011). However, the induced neuronal progenitor cells from which the iAstrocytes are derived, can also be differentiated into iNeurons, which are more biologically relevant to ALS. Work is now being completed to perform Cajal body immunostaining on iNeurons differentiated from these same C9-ALS patients to look at the behaviour of Cajal bodies in these cells.

It is difficult trying to discern the function of C9orf72 from the Cajal body phenotype, which varies between different cell models. In HEK293 cells, the evidence is convincing that Cajal bodies can be influenced by the C9orf72 protein. The mechanism for how C9orf72 exerts this effect is unknown, but it is well documented that the snRNP processing pathway and the cell cycle are the biggest regulators of Cajal bodies, suggesting C9orf72 may be involved in one of these. Many of the genes and molecular pathways linked to ALS highlight defective RNA processing as a major player in disease pathogenesis. In addition, the known cytoplasmic function of C9orf72 as a protein component of trafficking complexes fits best with a role of C9orf72 in the snRNP/splicing pathway. We therefore chose to investigate the effect of C9orf72 on splicing, which is the focus of the next chapter.

6 Investigating the effect of C9orf72 on splicing

6.1 Introduction

We have shown that C9orf72 interacts with coilin and can influence Cajal body numbers in HEK293 cells. Cajal bodies were also changed in C9orf72-ALS patient cells, although this was not proven to be a result of haploinsufficiency of the C9orf72 protein. Cajal bodies are enriched in proteins and RNAs that process and remodel snRNPs. These snRNPs form the splicing machinery, or 'spliceosome', which functions often co-transcriptionally to remove introns from pre-mRNA.

There are two splicing pathways, each with its own distinct spliceosome. The major spliceosome, containing the U1, U2, U4, U5 and U6 snRNPs, splices major or 'U2-type' introns from mRNA (Will and Luhrmann, 2011). Approximately 99% of all the introns in the human genome are U2-type introns. The major splicing pathway is efficient and the splice site and branch point sequences of U2-type introns, which act as splicing signals, can be highly degenerate (Burset et al., 2001; Mount, 2000). Conversely, the minor spliceosome removes minor, or 'U12-type' introns, and is comprised of the U11, U12, U4atac, U5 and U6atac snRNPs (Patel and Steitz, 2003). Only 1% of all introns are U12 introns, and unlike U2 introns, efficiency of splicing is low and the sequences which define U12 introns are highly conserved (Burge et al., 1998; Patel et al., 2002).

Defective splicing is a common theme which runs throughout ALS and other neurodegenerative diseases. ALS-linked genes FUS and TDP-43 are both RNA binding proteins in which mutations lead to defects in RNA processing (Arnold et al., 2013; Reber et al., 2016). Deficiency of the Cajal body protein SMN has been shown to cause failings in the minor splicing of U12 introns and in C9orf72-ALS patient brains, alternative splicing defects, including the retention of introns, has been detected (Boulisfane et al., 2011; Prudencio et al., 2015). Importantly, splicing defects in C9orf72-ALS correlate with disease severity, suggesting failures in the splicing pathway/machinery may play a significant role in disease pathogenesis (Cooper-Knock et al., 2015a).

In this chapter, we investigated the effect of loss of the C9orf72 protein on splicing, to determine whether haploinsufficiency of C9orf72 could be contributing to the splicing defects observed in C9orf72 ALS. Reporter splicing assays, which involve the transfection of an intron-containing construct into cells, have been described in the literature as an effective method of investigating the splicing ability of cells (Cooper, 2005). We therefore implemented a reporter splicing assay using pCI-neo and P120 constructs, which contain a major and minor intron respectively, and which have both been used successfully in other studies (Hall and Padgett, 1996; Whittom et al., 2008). The constructs were transfected into cells depleted in C9orf72, and the removal of their introns was used as a measure of splicing efficiency.

6.2 Results

6.2.1 C9orf72 siRNA transfection decreases splicing in HEK293 cells

The reporter splicing assay involves transfection of a reporter which contains an intron which can be retained or spliced. RNA is extracted from cells and RT-PCR performed using primers that flank either side of the intron which generates a spliced or unspliced product depending on whether the intron was retained or not. To investigate major splicing, a pCI-neo construct was used, which contains a U2-type intron. To look at minor splicing, a P120 minigene was used. The P120 minigene contained exons 6 and 7 from the P120 gene flanking a U12-type intron and was generously provided by Professor Richard Padgett at the Cleveland Clinic (Hall and Padgett, 1996). PCR primers were designed that flanked regions either side of the intron. Thus, if the reporter was unspliced, a PCR product containing the intron plus flanking sequences was amplified. However, if the reporter was spliced, a smaller PCR product lacking the intron was generated. Specifically, the pCI-neo intron was 133 base pairs (bps). Primers were designed to bind in the flanking pCI-neo construct such that the unspliced product was 388 bps and spliced product was 255 bps (Figure 6.1A). The P120 U12-intron was 99 base pairs, and the primers were designed to bind in flanking exons 6 and 7 such that the unspliced and spliced PCR products were 250 bps and 151 bps respectively (Figure 6.1B). The P120 plasmid also contained an additional two U2-type introns that were not analysed.

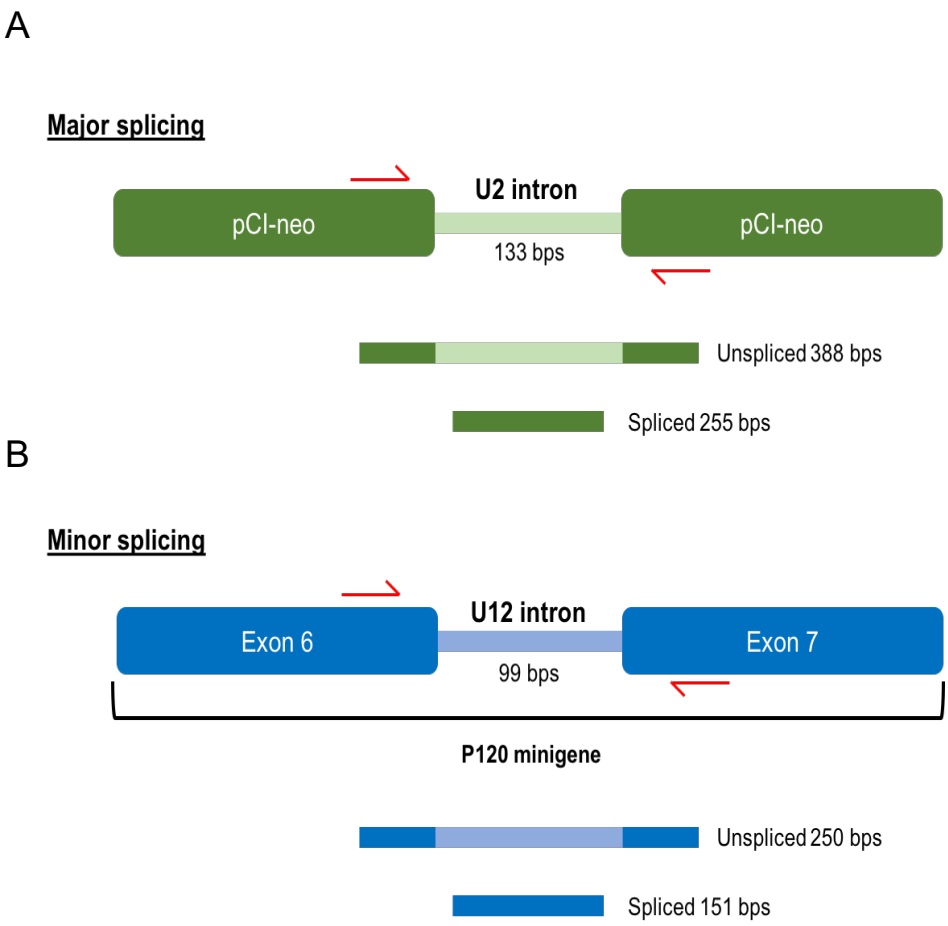


Figure 6.1. The reporter splicing assay. (A) pCI-neo was used as the reporter to measure major splicing. The pCI-neo plasmid contains a 133 base pair U2-intron. Primers were designed to bind in the flanking pCI-neo construct such that the unspliced product was 388 bps and spliced product was 255 bps. (B) A P120 minigene was used as the reporter to measure minor splicing. The P120 minigene contains a 99 base pair U12-intron. Primers were designed to bind to exon 6 and exon 7 of the P120 minigene such that the unspliced product was 250 bps and spliced product was 151 bps.

The primers were tested to confirm PCR products of the expected size were amplified from the constructs. A PCR was run using the pCI-neo or P120 construct as the DNA template with their respective primers for intron amplification. The pCI-neo primers successfully amplified a PCR product of expected 388 bp size from the pCI-neo construct, but did not generate product from a water control (Figure 6.2A). Similarly, P120 primers successfully amplified a PCR product of the expected 250 bp size from the P120 minigene but not from the water control (Figure 6.2B). The

PCRs confirmed the primers amplified the correct products and could be taken forward for use in the reporter splicing assay.

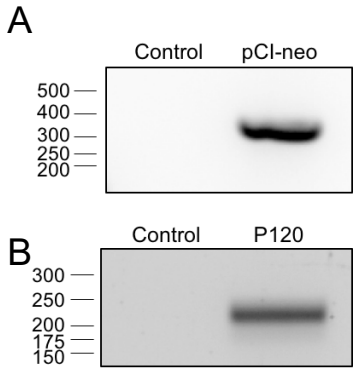


Figure 6.2. PCR amplification of introns from pCI-neo and P120 constructs. (A) Water (control) or pCI-neo plasmid was used as a template to test primers for PCR amplification of the pCI-neo intron (product size = 388 bp) (B) Water (control) or the P120 mini-gene plasmid was used as a template to test primers for PCR amplification of the P120 intron (product size = 250 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide.

Splicing of the pCI-neo intron in C9orf72-depleted HEK293 cells was used to investigate the effect of C9orf72 on major splicing. Coilin knockdown has been shown to decrease splicing of the pSI-CHECK intron in a similar reporter assay, and we therefore included it as a positive control (Whittom et al., 2008). HEK293 cells were transfected with NTC, pooled C9orf72 or pooled coilin siRNAs. After 72 hours of knockdown, cells were transfected with pCI-neo. RNA was extracted and poly(A) RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking pCI-neo primers. PCR products were resolved by agarose gel electrophoresis which showed a faint unspliced product at 388 bps and a spliced product at 255 bps (Figure 6.3A). PCR controls included water only and pCI-neo plasmid as DNA templates. There appeared to be more unspliced product in C9orf72 siRNA cells in comparison to NTC, suggesting there was less splicing of the pCI-neo intron after C9orf72 knockdown (Figure 6.3A). Densitometry was performed on the unspliced and spliced bands and the ratio of spliced/unspliced product was generated as a measure of splicing efficiency. The densitometry analysis confirmed there was less pCI-neo splicing in the C9orf72 knockdown cells (Figure 6.3B). Interestingly, in our experiments, knockdown of coilin had no effect

on the splicing of the pCI-neo intron (Figure 6.3B). C9orf72 and coilin knockdown was confirmed by measuring their mRNA levels by RT-qPCR. C9orf72 and coilin mRNA was reduced by approximately 80% and 70% across the four experiments respectively (Figure 6.3C-6.3D). Unexpectedly, C9orf72 knockdown appeared to cause a small reduction in coilin mRNA (Figure 6.3D). The results suggest that loss of the C9orf72 protein may decrease major splicing.

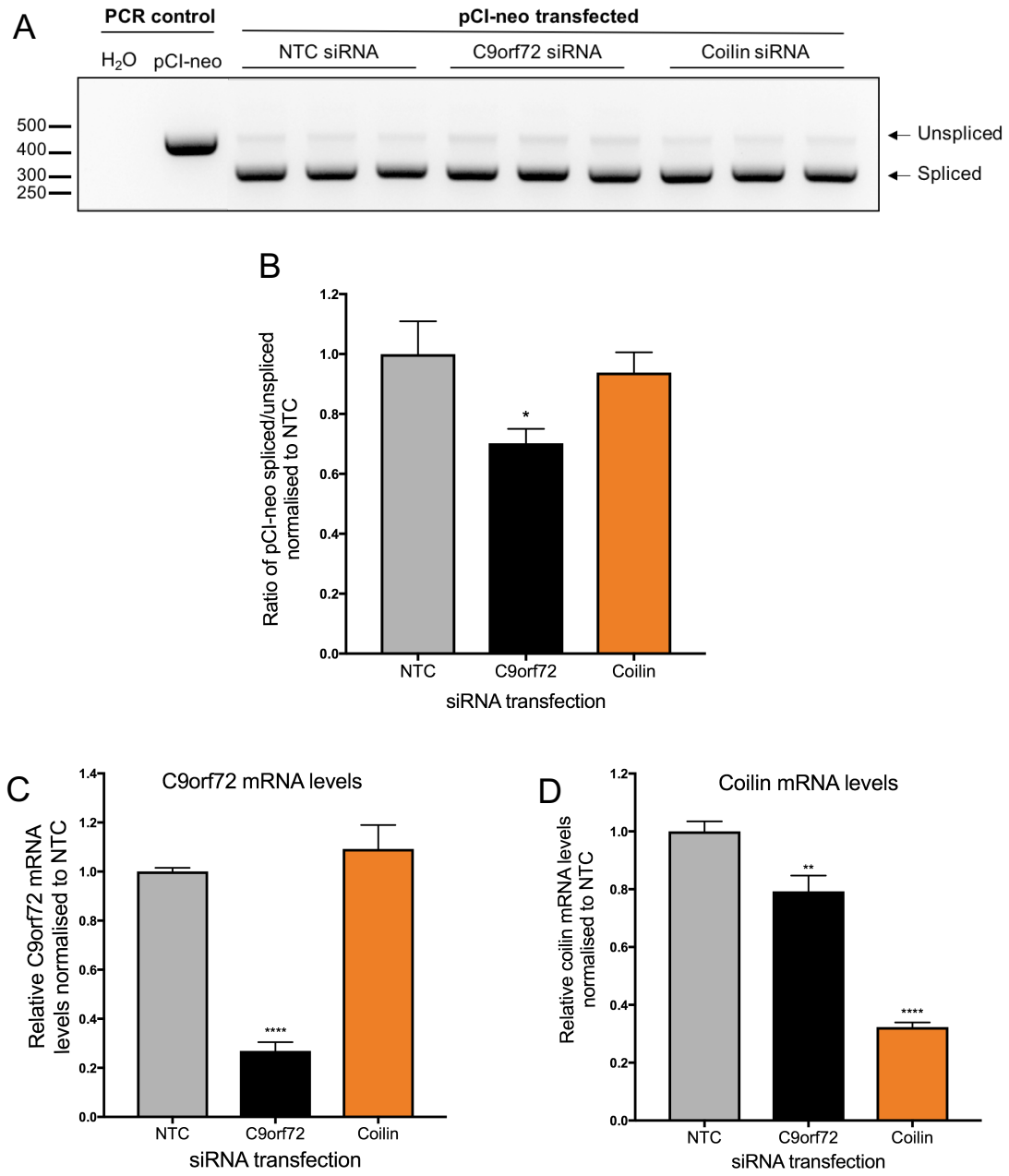


Figure 6.3. C9orf72 knockdown decreases splicing of the pCI-neo intron in HEK293 cells. (A) In vivo splicing of the pCI-neo intron in HEK293 cells. Cells were transfected with NTC, C9orf72 or coilin siRNA. Post-knockdown, cells were transfected with pCI-neo plasmid. RNA extractions were performed followed by RT-PCR from poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (388 bp) or spliced (255 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to NTC. C9orf72 (C) and coilin (D) knockdown was confirmed by RT-qPCR. All data represent mean ± SEM; One-way ANOVA with Dunnett's test, * $p \leq 0.05$, ** $p \leq 0.01$, **** $p \leq 0.0001$. n=4 experiments.

C9orf72 knockdown appeared to decrease major splicing. A very small subset of introns are U12-type introns, and are spliced by the minor spliceosome. The rarity of U12-type introns, combined with their high level of splice signal conservation, suggests they may have essential functions, which are not yet understood (Turunen et al., 2013). The P120 construct was used to investigate the effect of C9orf72 on minor splicing. HEK293 cells were transfected with NTC, pooled C9orf72 or pooled coilin siRNAs. After 72 hours' protein knockdown, cells were transfected with the P120 minigene. RNA was extracted and poly(A) RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking P120 primers. PCR products were resolved by agarose gel electrophoresis which showed a darker unspliced product at 250 bps and a spliced product at 155 bps (Figure 6.4A). The unspliced band from the C9orf72 siRNA cells appeared darker than the NTC unspliced band, suggesting there was less splicing of the P120 intron after C9orf72 knockdown (Figure 6.4A). Splicing efficiency was quantified as the ratio of the densitometry of spliced/unspliced bands (Figure 6.4B). The quantification confirmed there was less splicing of the P120 intron in C9orf72 knockdown cells, but coilin knockdown had no effect on P120 splicing (Figure 6.4B). RT-qPCR confirmed C9orf72 and coilin mRNA was reduced by approximately 70-75% across the experiments (Figure 6.4C-Figure 6.4D). Again, C9orf72 knockdown appeared to cause a slight knockdown of coilin mRNA (Figure 6.4D). The results suggest knockdown of C9orf72 in HEK293 cells can decrease splicing of minor introns, in addition to major.

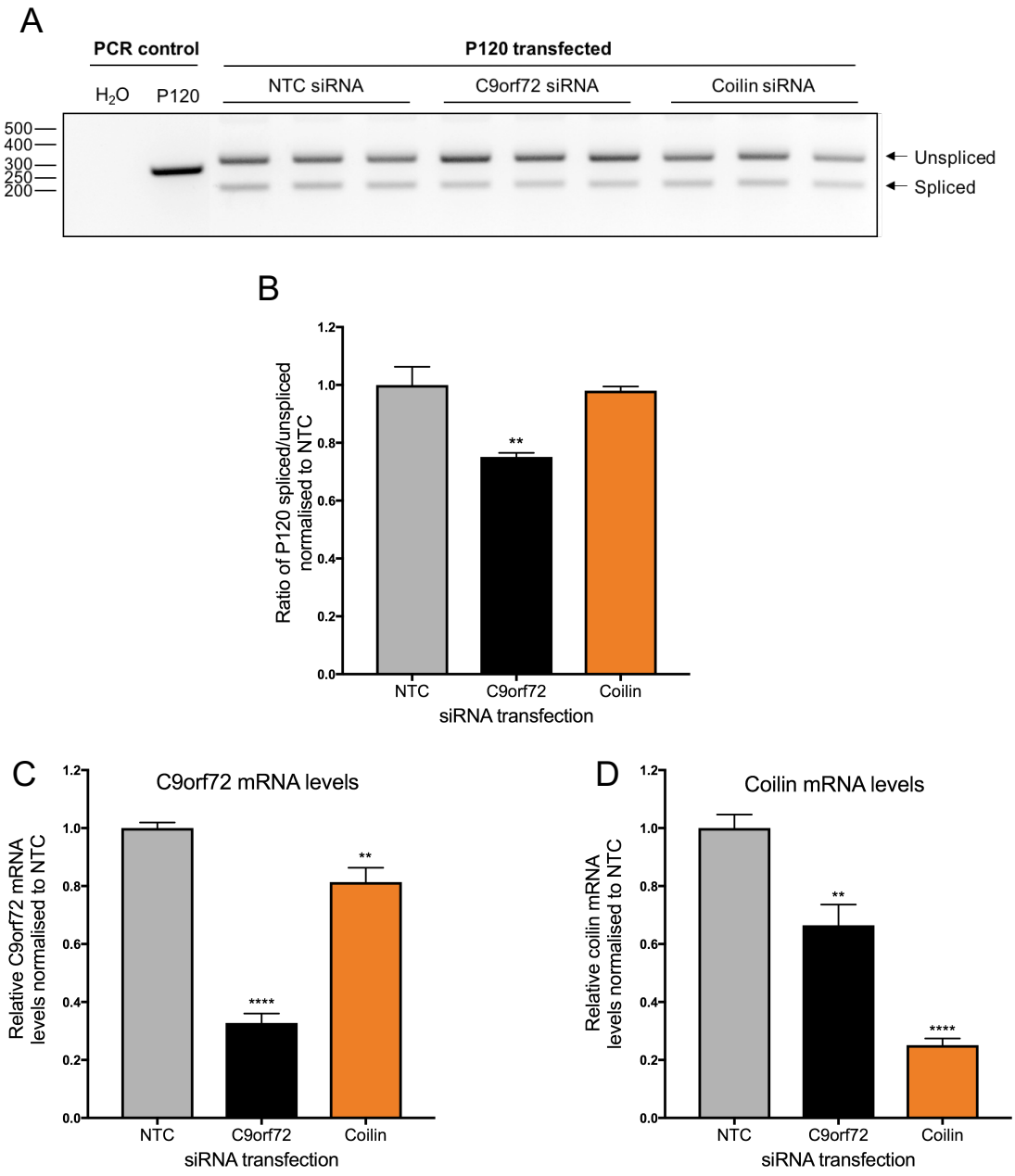


Figure 6.4. C9orf72 knockdown decreases splicing of the P120 intron in HEK293 cells. (A) In vivo splicing of the P120 minigene intron in HEK293 cells. Cells were transfected with NTC, C9orf72 or coilin siRNA. Post-knockdown, cells were transfected with P120 minigene plasmid. RNA extractions were performed followed by RT-PCR from poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (250 bp) or spliced (151 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to NTC. C9orf72 (C) and coilin (D) knockdown was confirmed by RT-qPCR. All data represent mean ± SEM; One-way ANOVA with Dunnett's test, ** $p \leq 0.01$, **** $p \leq 0.0001$. n=4 experiments.

We have shown that knockdown of C9orf72 using pooled siRNA decreased splicing of the U12 intron in the exogenous P120 minigene. The full length P120 protein is expressed from the human nucleolar protein P120 (NOL1) gene. Therefore, the P120 primers were utilised to determine if C9orf72 knockdown caused defective splicing of the endogenous P120 minor intron. HEK293 cells were transfected with NTC, C9orf72 pooled or coilin pooled siRNAs. After 72 hours' protein knockdown, RNA was extracted and RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking P120 primers. PCR products were resolved by agarose gel electrophoresis which showed a faint unspliced product at 250 bps and a darker spliced product at 155 bps (Figure 6.5A). The unspliced band in the C9orf72 knockdown samples appeared similar in intensity to the NTC unspliced bands, suggesting C9orf72 knockdown did not influence splicing of the endogenous P120 intron (Figure 6.5A). Densitometry was performed on the unspliced and spliced bands and the ratio of spliced/unspliced product was generated as a measure of splicing efficiency. The densitometry analysis confirmed the amount of endogenous P120 splicing after both C9orf72 and coilin knockdown was the same as in the NTC (Figure 6.5B). Both C9orf72 and coilin mRNA was reduced by approximately 70% across the experiments (Figure 6.5C-Figure 6.5D). C9orf72 knockdown in these experiments also appeared to reduce coilin mRNA levels by approximately 30% (Figure 6.5D). Taken together, the reporter splicing assays suggest knockdown of C9orf72 can decrease both major and minor splicing of an exogenous intron, but endogenous splicing is unchanged. It also appeared that loss of C9orf72 could cause a small decrease in coilin mRNA.

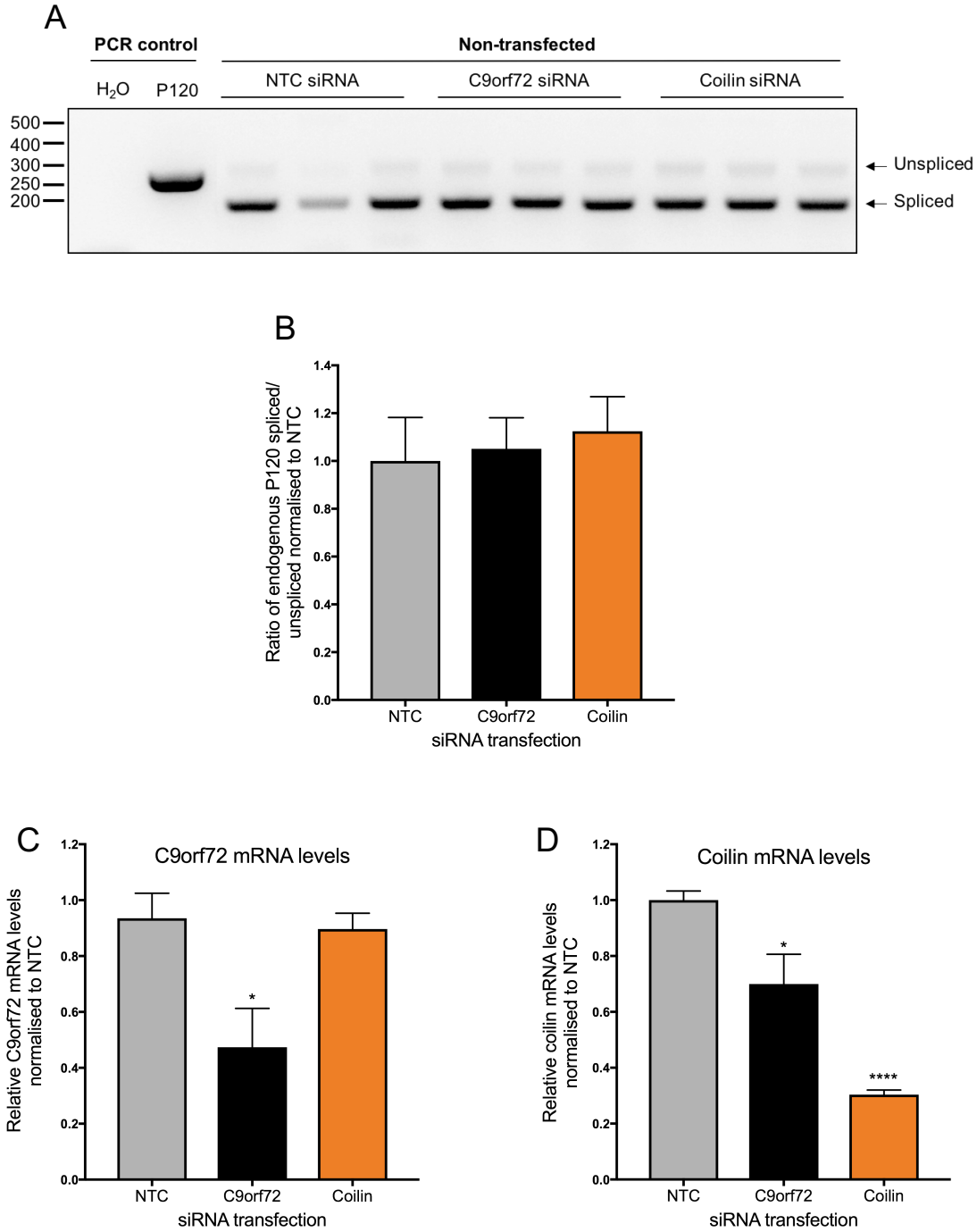


Figure 6.5. C9orf72 knockdown has no effect on endogenous P120 splicing in HEK293 cells. (A) In vivo splicing of the endogenous P120 intron in HEK293 cells. Cells were transfected with NTC, C9orf72 or coilin siRNA. RNA extractions were performed followed by RT-PCR from poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (250 bp) or spliced (151 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to NTC. C9orf72 (C) and coilin (D) knockdown was confirmed by RT-qPCR. All data represent mean ± SEM; One-way ANOVA with Dunnett's test, ** $p \leq 0.01$, **** $p \leq 0.0001$. n=4 experiments.

6.2.2 Rescue transfection of C9orf72 in the P120 splicing assays

C9orf72 knockdown led to a decrease in splicing of the U2 and U12 introns in the pCI-neo and P120 constructs respectively. To confirm the defective splicing was specific to a loss of the C9orf72 protein, rescue transfections with the siRNA resistant myc-C9orf72 constructs described in chapter 3, section 3.2.3 were performed. First, the myc-C9orf72 constructs were subjected to PCRs using the pCI-neo and P120 primers to ensure the primers would not amplify product from them. Unfortunately, the pCI-neo PCR generated a product from both pRK5-myc-C9orf72 siRNA resistant constructs (Figure 6.6A). Therefore, we could not perform rescue transfections in the pCI-neo splicing assay. However, the P120 PCR was clean, with no products generated from either myc-C9orf72 constructs (Figure 6.6B). The myc-C9orf72 siRNA constructs were therefore taken forward to perform rescue transfections in the P120 splicing assays.

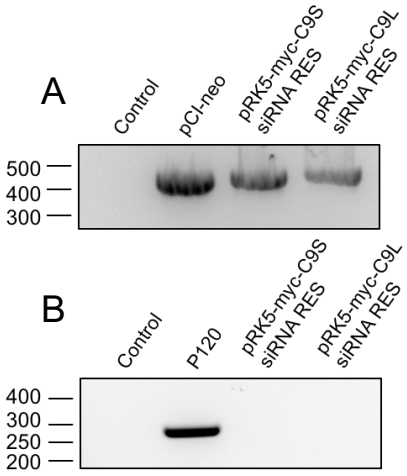


Figure 6.6. pCI-neo but not P120 specific primers generate a PCR product from pRK5-myc-C9orf72 siRNA resistant constructs. Water (control), pCI-neo or P120 minigene, pRK5-myc-C9orf72S siRNA resistant and pRK5-mycC9orf72L siRNA plasmids were used as templates in a PCR using pCI-neo specific primers (A) or P120 specific primers (B). PCR products were resolved on an agarose gel and stained with ethidium bromide.

To establish if the decrease in exogenous P120 splicing was caused by C9orf72 knockdown, changes were made to the P120 reporter splicing assay. First, knockdown of C9orf72 was achieved by individual C9orf72 siRNAs in addition to pooled. Secondly, siRNA resistant C9orf72 constructs were transfected after knockdown to determine if the splicing defect could be rescued. Coilin knockdown was not included as there was no observable effect on either major or minor splicing in previous experiments. HEK293 cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pooled siRNA. After 72 hours' protein knockdown, cells were co-transfected with the P120 minigene and either pCI-neo (empty vector control), siRNA resistant myc-C9orf72S or siRNA resistant myc-C9orf72L. RNA was extracted and poly(A) RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking P120 primers. PCR products were resolved by agarose gel electrophoresis which showed a darker unspliced product at 250 bps and a fainter spliced product at 155 bps (Figure 6.7A). In all co-transfection conditions, the unspliced product appeared darker after C9orf72 #D and C9orf72 pooled siRNA transfection, but not the C9orf72 #2 siRNA. Densitometry was performed on the unspliced and spliced bands and the ratio of spliced/unspliced product was generated as a measure of splicing efficiency. The quantification showed there was no statistically significant difference between any of the conditions. However, there was a small decrease in P120 intron splicing that occurred after C9orf72 #D or C9orf72 pooled siRNA transfections but not with the C9orf72 #2 siRNA (Figure 6.7B). C9orf72 mRNA was reduced by approximately 50-60% with all C9orf72 siRNAs however, as measured by RT-qPCR (Figure 6.7C). This suggests the observed defect in P120 splicing may have been caused by an off-target effect of the C9orf72 #D siRNA and not from a loss of C9orf72 protein. Interestingly however, transfection of myc-C9orf72S into cells after knockdown of C9orf72 by C9orf72 #D or pooled siRNA, appeared to restore splicing of the P120 intron to normal levels (Figure 6.7B). Transfection with myc-C9orf72L had no effect. The overexpression of the C9orf72 short and long isoforms was confirmed by RT-qPCR (Figure 6.7C).

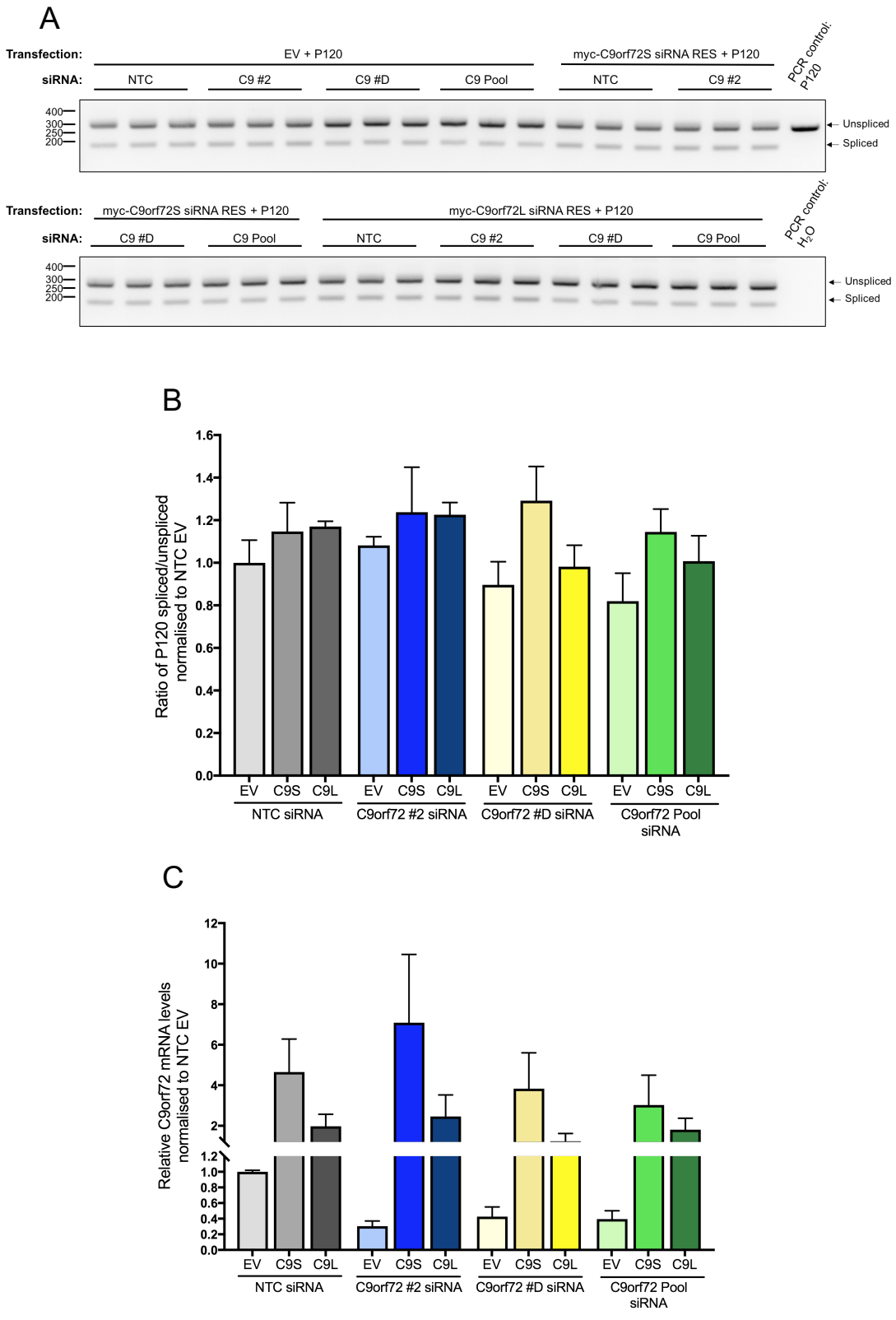
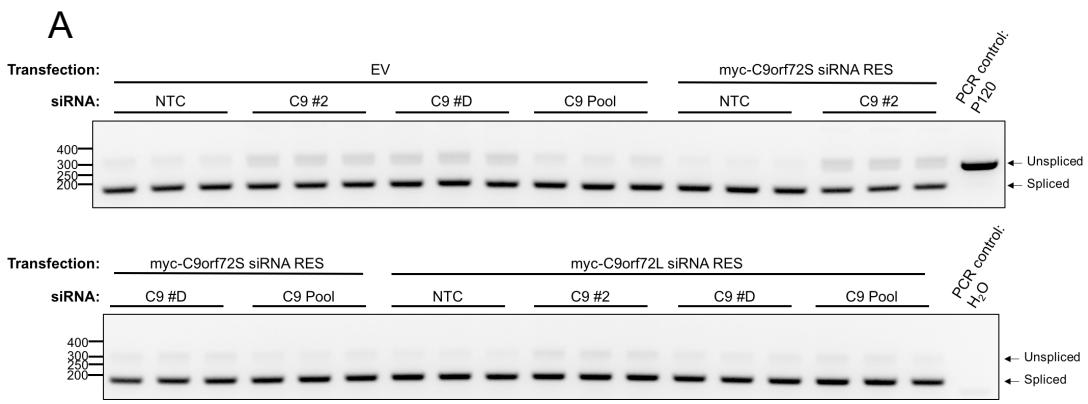
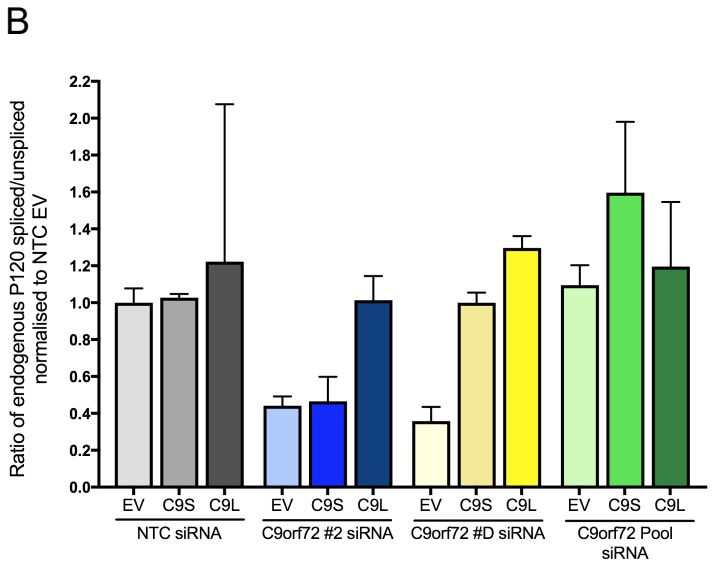

A
Transfection:
EV + P120
myc-C9orf72S siRNA RES + P120
siRNA:
NTC
C9 #2
C9 #D
C9 Pool
NTC
C9 #2
PCR control: P120
400
300
250
200
Unspliced
Spliced
Transfection:
myc-C9orf72S siRNA RES + P120
myc-C9orf72L siRNA RES + P120
siRNA:
C9 #D
C9 Pool
NTC
C9 #2
C9 #D
C9 Pool
PCR control: H2O
400
300
250
200
Unspliced
Spliced
B
Ratio of P120 spliced/unspliced normalised to NTC EV
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
EV C9S C9L
NTC siRNA
C9orf72 #2 siRNA
C9orf72 #D siRNA
C9orf72 Pool siRNA
C
Relative C9orf72 mRNA levels normalised to NTC EV
12
10
8
6
4
2
1.2
1.0
0.8
0.6
0.4
0.2
0.0
EV C9S C9L
NTC siRNA
C9orf72 #2 siRNA
C9orf72 #D siRNA
C9orf72 Pool siRNA

Figure 6.7. Decreased P120 minigene splicing in HEK293 cells is C9orf72 #D siRNA specific. (A) In vivo splicing of the P120 minigene intron in HEK293 cells following individual C9orf72 siRNA transfection and reintroduction of siRNA resistant myc-C9orf72. Cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pool siRNA. Post-knockdown, cells were co-transfected with P120 minigene + pCI-neo (EV), myc-C9S siRNA resistant or myc-C9L siRNA resistant. RNA extractions were performed followed by RT-PCR from poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (250 bp) or spliced (151 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to NTC EV. (C) C9orf72 knockdown and overexpression was confirmed by RT-qPCR. All data represent mean ± SEM; One-way ANOVA with Tukey's test. n=3 experiments.

The experiment was repeated to investigate splicing of the endogenous P120 minor intron. HEK293 cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pooled siRNAs. After 72 hours' protein expression, cells were transfected with pCI-neo (empty vector control), siRNA resistant myc-C9orf72S or siRNA resistant myc-C9orf72L. RNA was extracted and poly(A) RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking P120 primers. PCR products were resolved by agarose gel electrophoresis which showed a faint unspliced product at 250 bps and a darker spliced product at 155 bps (Figure 6.8A). The unspliced band in empty vector transfected cells appeared visibly darker after C9orf72 knockdown using individual siRNAs, but not pooled (Figure 6.8A). Densitometry was performed on the unspliced and spliced bands and the ratio of spliced/unspliced product was generated as a measure of splicing efficiency. The quantification showed there was no statistically significant difference between any of the conditions. However, the ratio of spliced/unspliced product after C9orf72 #2 and C9orf72 #D siRNA transfection was decreased, but not after C9orf72 pooled siRNA transfection (Figure 6.8B). Expression of myc-C9orf72S rescued splicing in the C9orf72 #D siRNA cells but not in the C9orf72 #2 siRNA cells. Expression of myc-C9orf72L rescued splicing in both C9orf72 siRNA conditions (Figure 6.8B). The knockdown and overexpression of C9orf72 was confirmed by RT-qPCR (Figure 6.8C).


A
Transfection:
EV
myc-C9orf72S siRNA RES
siRNA:
NTC
C9 #2
C9 #D
C9 Pool
NTC
C9 #2
PCR control: P120
400
300
250
200
Unspliced
Spliced
Transfection:
myc-C9orf72S siRNA RES
myc-C9orf72L siRNA RES
siRNA:
C9 #D
C9 Pool
NTC
C9 #2
C9 #D
C9 Pool
PCR control: H2O
400
300
250
200
Unspliced
Spliced


B
Ratio of endogenous P120 spliced/unspliced normalised to NTC EV
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
EV C9S C9L
NTC siRNA
C9orf72 #2 siRNA
C9orf72 #D siRNA
C9orf72 Pool siRNA

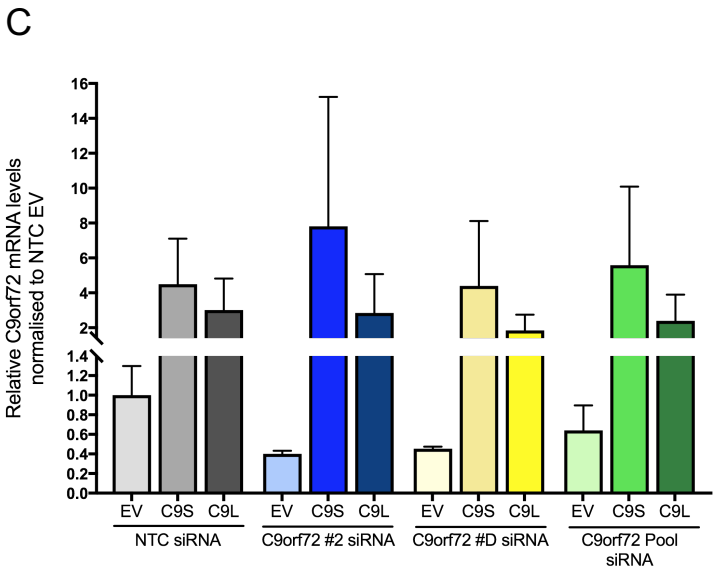

C
Relative C9orf72 mRNA levels normalised to NTC EV
16
14
12
10
8
6
4
2
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
EV C9S C9L
NTC siRNA
C9orf72 #2 siRNA
C9orf72 #D siRNA
C9orf72 Pool siRNA

Figure 6.8. Endogenous P120 splicing in HEK293 cells following individual C9orf72 siRNA transfection. (A) In vivo splicing of the endogenous P120 intron in HEK293 cells following individual C9orf72 siRNA transfection and reintroduction of siRNA resistant myc-C9orf72. Cells were transfected with NTC, C9orf72 #2, C9orf72 #D or C9orf72 pool siRNA. Post-knockdown, cells were transfected with pCI-neo (EV), myc-C9S siRNA resistant or myc-C9L siRNA resistant. RNA extractions were performed followed by RT-PCR from poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (250 bp) or spliced (151 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to NTC EV. (C) C9orf72 knockdown and overexpression was confirmed by RT-qPCR. All data represent mean ± SEM; One-way ANOVA with Tukey's test. n=3 experiments.

6.2.3 Investigating splicing in a stable C9orf72 knockout cell line

In HEK293 cells, there was a small decrease in exogenous P120 splicing after knockdown of C9orf72 using the C9orf72 #D siRNA, but not the C9orf72 #2 siRNA, suggesting the effect could be from an siRNA off-target effect. Expression of the C9orf72 protein in these cells appeared to show some rescue of the splicing defect however, which suggests loss of the C9orf72 protein could be responsible. The high variability observed in the HEK293 cells makes it difficult to draw solid conclusions about the involvement of C9orf72 in splicing. Therefore, we performed the reporter splicing assays in the B2 C9orf72 knockout cell line, to remove the need for siRNA transfection. HEK293 cells (control) or B2 cells were transfected with pCI-neo to investigate the C9orf72 proteins involvement in major splicing. After 24 hours, RNA was extracted and poly(A) RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking pCI-neo primers. PCR products were resolved by agarose gel electrophoresis which showed a faint unspliced product at 388 bps and a spliced product at 255 bps (Figure 6.9A). PCR controls included water only and pCI-neo plasmid as DNA templates. There was no visible difference in band intensity in either unspliced or spliced bands between HEK293 or B2 cells (Figure 6.9A). Densitometry was performed on the bands to quantify the ratio of spliced/unspliced product, as a measure of splicing efficiency. Quantification showed there was no difference in pCI-neo intron splicing in HEK293 and B2 cells

(Figure 6.9B), suggesting in the B2 cell line, loss of C9orf72 has no effect on major splicing.

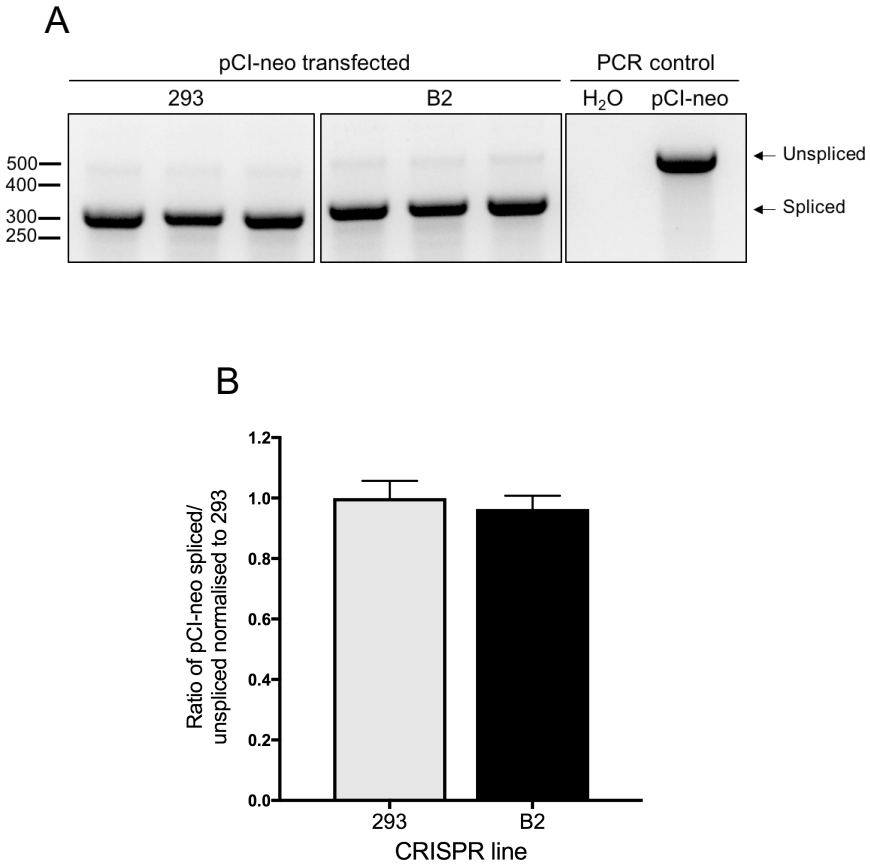


Figure 6.9. pCI-neo splicing is unaffected in the B2 C9orf72 knockout cell line. (A) In vivo splicing of the pCI-neo intron in HEK293 and B2 cells. Cells were transfected with pCI-neo plasmid and RNA extracted followed by RT-PCR on poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (388 bp) or spliced (255 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to 293. Data represent mean ± SEM; Unpaired t-test. n=3 experiments.

The P120 reporter splicing assay was performed on the B2 cell line to determine what effect a chronic loss of the C9orf72 protein had on minor splicing. HEK293 cells (control) or B2 cells were transfected with the P120 minigene. After 24 hours, RNA was extracted and poly(A) RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking P120 primers. PCR products were resolved by agarose gel electrophoresis which showed a darker unspliced product at 250 bps and a faint spliced product at 155 bps (Figure 6.10A). The unspliced band in the B2 cells appeared visibly darker versus the unspliced product from the HEK293 cells, suggesting there was less splicing of the P120 intron in the C9orf72 knockout cells. The decrease in splicing was confirmed by performing densitometry on the bands and quantifying the ratio of spliced/unspliced product (Figure 6.10B). There was a 60% decrease in splicing of the P120 intron in B2 cells in comparison to HEK293 cells, however this result was highly variable and not statistically significant. To determine if the decrease in splicing was specific to loss of C9orf72, rescue transfections of the C9orf72 isoforms were performed. Co-transfection of the myc-C9orf72 constructs with P120 did not visibly appear to reduce the strength of the unspliced product (Figure 6.10A). Quantification of the ratio of spliced/unspliced product showed expression of both myc-C9orf72S and myc-C9orf72L did not change splicing of exogenous P120 in the B2 cells (Figure 6.10B). Expression of myc-C9orf72S and myc-C9orf72L in the B2 cells was confirmed by anti-myc immunostaining on coverslip-plated cells which received the same transfection mix (Figure 6.10C). Unfortunately, the images are limited to showing myc-C9orf72 transfection efficiency only, and we cannot determine the co-transfection efficiency of P120 with myc-C9orf72. Interestingly, the results suggest minor splicing may be decreased in the B2 C9orf72 knockout cell line, but expression of C9orf72 in these cells does not restore splicing to normal levels.

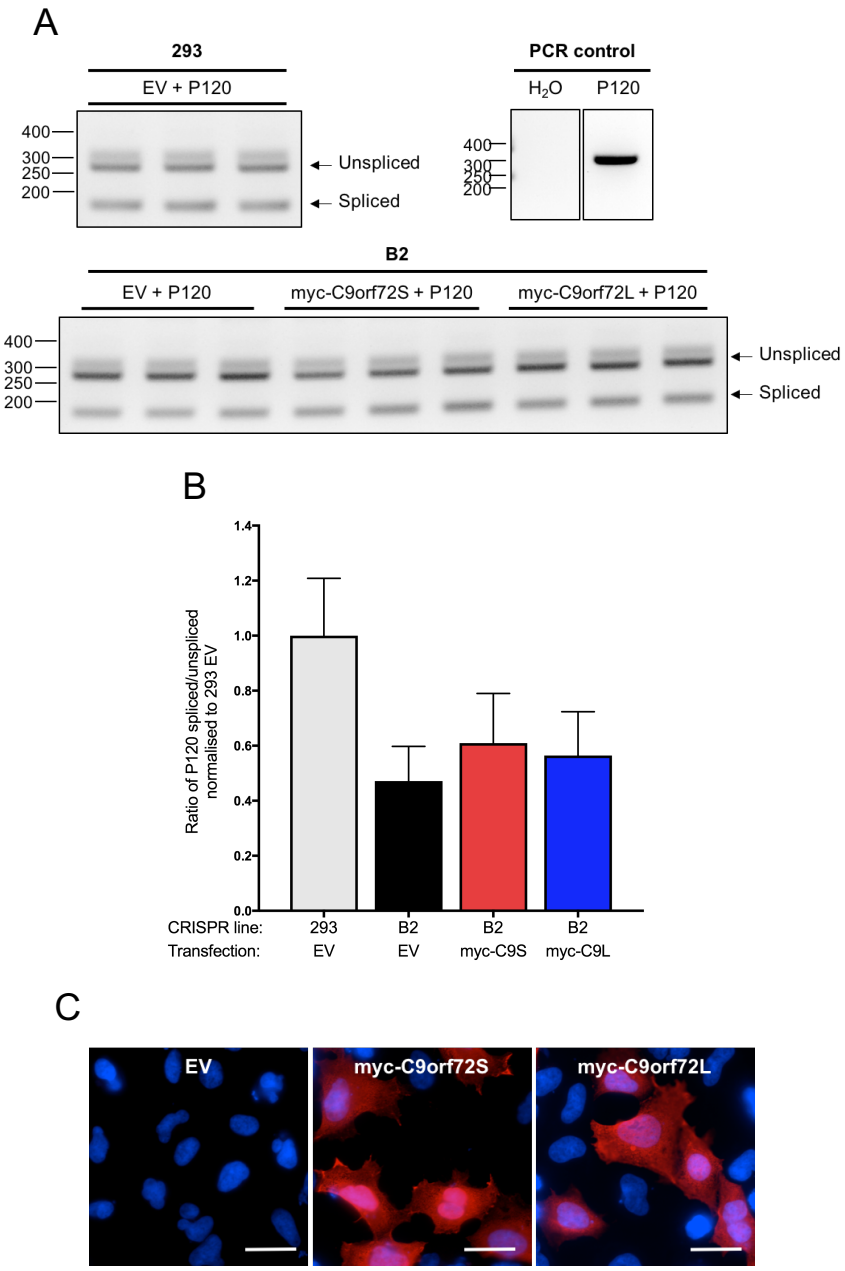


Figure 6.10. Splicing of the P120 minigene intron may be decreased in the B2 C9orf72 knockout line. (A) In vivo splicing of the P120 minigene intron in HEK293 and B2 CRISPR cells with myc-C9orf72 overexpression. Cells were co-transfected with P120 minigene + pCI-neo (EV), myc-C9orf72S or myc-C9orf72L. RNA extractions were performed followed by RT-PCR on poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (250 bp) or spliced (151 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to 293 EV. Data represent mean ± SEM; One-way ANOVA with Tukey's test. n=3 experiments. (C) myc-C9orf72 overexpression was confirmed by immunostaining on separately fixed cells using an anti-myc antibody (9B11) (red). Scale bar = 10 µm.

Splicing of the exogenous P120 intron appeared to be decreased in the B2 cell line. We therefore next investigated the splicing of the endogenous P120 minor intron in B2 cells. HEK293 or B2 cells were transfected with pCI-neo, myc-C9orf72S or myc-C9orf72L. After 24 hours, RNA was extracted and poly(A) RT-PCR performed. Triplicate cDNA samples were subjected to PCR using the intron-flanking P120 primers. PCR products were resolved by agarose gel electrophoresis which showed a faint unspliced product at 250 bps and a prominent spliced product at 155 bps (Figure 6.11A). The unspliced band in the B2 cells appeared visibly darker than the unspliced band in HEK293 cells. Expression of both C9orf72 isoforms appeared to decrease this unspliced band in the B2 cell line (Figure 6.11A). Densitometry was performed on the bands to quantify the ratio of spliced/unspliced product, as a measure of splicing efficiency. There was a decrease in the ratio of spliced/unspliced product in the B2 cells in comparison to HEK293 cells, but the experiment was not adequately powered for the high levels of variability observed, thus there was no statistical difference (Figure 6.11B). Splicing of the endogenous P120 intron after myc-C9orf72S and myc-C9orf72L expression in B2 cells was to a similar level as in the HEK293 cells (Figure 6.11B). Expression of myc-C9orf72S and myc-C9orf72L in the B2 cells was confirmed by anti-myc immunostaining on coverslip-plated cells which received the same transfection mix (Figure 6.11C). The immunostaining showed not all cells were successfully transfected and thus non-transfected cells will have been included in the analysis of endogenous P120 intron splicing in the rescue conditions. Again, high variability in the experiments make it difficult to conclude whether there is a minor splicing defect in the B2 C9orf72 knockout cell line, but there is a slight trend in the data which suggests loss of C9orf72 could be affecting the minor splicing of P120.

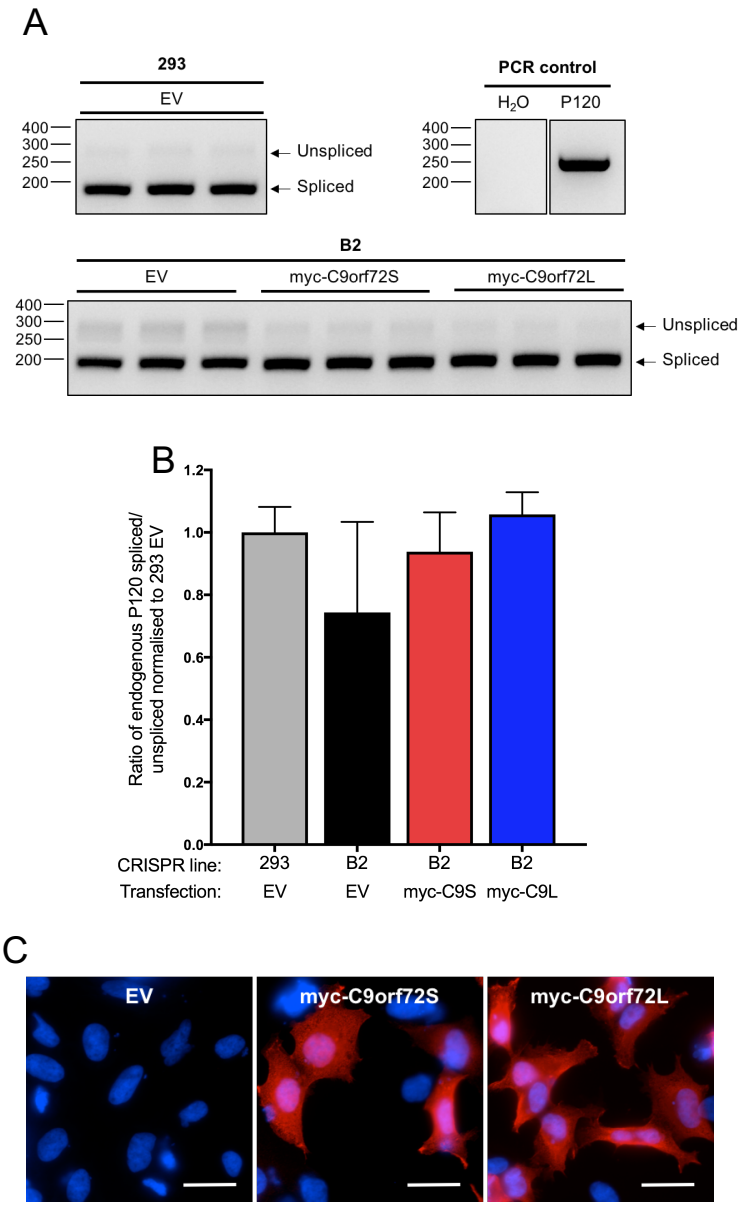


Figure 6.11. Endogenous P120 splicing in the B2 C9orf72 knockout CRISPR line. (A) In vivo splicing of the endogenous P120 intron in HEK293 and B2 CRISPR cells with myc-C9orf72 overexpression. Cells were transfected with pCI-neo (EV), myc-C9orf72S or myc-C9orf72L. RNA extractions were performed followed by RT-PCR on poly(A) mRNA. Subsequent PCR amplification of the intron produced two PCR products: Unspliced (250 bp) or spliced (151 bp). PCR products were resolved on an agarose gel and stained with ethidium bromide. (B) Splicing efficiency of the intron was quantified as the ratio of spliced product/unspliced product and normalised to 293 EV. Data represent mean ± SEM; One-way ANOVA with Tukey's test. n=3 experiments. (C) myc-C9orf72 overexpression was confirmed by immunostaining on separately fixed cells using an anti-myc antibody (9B11) (red). Scale bar = 10 μm.

6.3 Discussion

The aim of this chapter was to determine if the C9orf72 protein was involved in major and minor splicing. The hypothesis was that loss of C9orf72 in HEK293 cells influences Cajal bodies, and this would extend to an effect on splicing. Two reporter splicing assays were used, which have been described in the literature as a method of semi-quantitatively measuring the splicing efficiency of cells. A pCI-neo reporter, which contained a U2 intron, was used to investigate major splicing and a P120 minigene reporter, which contained a U12 intron, was used to investigate minor splicing. The P120 gene is expressed in human cells, therefore we also investigated the splicing of the endogenous P120 U12 intron. PCR primers which flanked either side of the intron amplified two products, a larger unspliced product if the intron was present, and a smaller spliced product when the intron had been removed. Changes in the densitometry of these bands was indicative of more or less splicing, and calculating the ratio of spliced/unspliced product allowed us to compare the amount of splicing in comparison to controls.

In the pCI-neo reporter assay, there was very little unspliced product in comparison to spliced, suggesting splicing of the U2 intron was highly efficient. The snRNPs that make up the major spliceosome are abundant in cells, which may facilitate the fast co-transcriptional removal of introns (Khodor et al., 2011). In HEK293 cells, transfection of C9orf72 pooled siRNAs caused a decrease in the major splicing of the pCI-neo intron, suggesting C9orf72 could play a role in splicing or in a splicing-associated pathway (Figure 6.3B). Unfortunately, we could not perform rescue transfections of C9orf72 in the pCI-neo reporter assay because the PCR primers amplified product from the myc-C9orf72 constructs. BLAST analysis of the primer sequences and alignment with the pRK5-myc-C9orf72 sequence showed there should be no primer binding to these constructs. pCI-neo is one of the most frequently prepared and used constructs within our lab group due to its use as an empty vector control. Therefore, it is plausible that there is pCI-neo contamination of other plasmid preps and equipment in the lab, leading to pCI-neo template availability in the PCR. However, the pCI-neo reporter assay was performed in the B2 C9orf72 stable knockout cell line which showed there was no effect on splicing of the pCI-neo intron in the B2 cell line versus HEK293 cells. This suggests the decrease in pCI-neo splicing observed in HEK293 cells was an artefact of siRNA transfection, and the C9orf72 protein is not involved in major splicing. Measuring the

splicing efficiency of a selection of endogenous major introns in the B2 cell line would support this result.

The P120 reporter assay was used to investigate the effect of C9orf72 on minor splicing. In these assays, there were considerably higher levels of unspliced product due to the slow removal of U12 introns by the minor spliceosome (Patel et al., 2002; Niemela et al., 2014). The minor snRNPs are found in much lower numbers in cells versus major spliceosome snRNPs, however one study suggests this is not causal of minor splicing inefficiency (Pessa et al., 2006). In addition to investigating splicing of the exogenous P120 intron, the splicing of the endogenous P120 minor intron was also examined. There was very little unspliced product generated from endogenous P120, suggesting the minor spliceosome efficiently removes endogenous introns, which are few in number, but is less able to cope with the large amount of exogenous P120.

The results from the P120 reporter splicing assays were highly variable and sometimes contradictory. In Figures 6.4 and 6.5, C9orf72 pooled siRNA transfection appeared to cause a small decrease in splicing of exogenous P120, but not endogenous. These results suggest the C9orf72 protein could have some involvement in minor splicing but loss of the C9orf72 protein is not overly detrimental; the splicing defect is only observed when introducing large amounts of an exogenous intron thereby increasing the splicing demand, but endogenous splicing is unaffected.

Rescue transfections of C9orf72 short and long isoforms after C9orf72 knockdown were performed to determine if expression of C9orf72 could restore splicing. In the NTC siRNA samples, expression of myc-C9orf72S and myc-C9orf72L did not appear to change splicing, suggesting although knockdown of C9orf72 may be able to decrease splicing, overexpression of the protein has no effect. It has been shown that overexpression of splicing factor proteins SF2 and hnRNPA1 leads to changes in the splicing of a reporter intron (Cáceres et al., 1994). Thus, if the C9orf72 protein was a splicing factor, its overexpression would be expected to influence splicing in the reporter splicing assays. This suggests C9orf72 is unlikely directly involved in splicing or RNA processing.

However, the rescue transfection experiments suggest the splicing defect may be an siRNA off-target effect rather than specific to C9orf72 knockdown (Figure 6.7-

6.8). It is worth noting that there was statistically no significant difference between the different conditions in the rescue experiments and therefore this discussion is based on small changes in the data only. In the P120 reporter splicing assay with C9orf72 rescue, transfection of the C9orf72 #2 siRNA did not decrease splicing, but C9orf72 mRNA levels were reduced by 70%. There was a non-significant decrease in splicing after transfection with C9orf72 #D and C9orf72 pooled siRNAs however. These results suggest the decrease in splicing could be caused by an off-target effect of the C9orf72 #D siRNA. BLAST analysis of the C9orf72 #D siRNA sequence suggests there should be no non-specific binding to other genes. However, nucleotides 2-7 within an siRNA can act as a 'seed region', facilitating the binding to other genes and exhibiting behaviour similar to microRNAs (Kamola et al., 2015). To determine whether the C9orf72 #D siRNA is targeting other genes through a seed region, a common method is to mutate siRNA base pairs downstream of the seed region, thus making it unable to knockdown C9orf72. The reporter splicing assay could then be performed on cells transfected with the mutant C9orf72 #D siRNA and if the splicing defect is observed, it can be concluded it is likely through binding of the siRNA seed region to another gene, thus generating an off-target effect.

Interestingly however, expression of myc-C9orf72S in cells after C9orf72 #D siRNA transfection increased exogenous P120 splicing, suggesting there is some involvement of the C9orf72 protein in minor splicing. Expression of myc-C9orf72L had a lesser effect, but the C9orf72 mRNA levels show myc-C9orf72L was not expressed as highly as myc-C9orf72S. In addition, transfection of both C9orf72 #2 and C9orf72 #D siRNAs appeared to decrease endogenous P120 splicing in the rescue experiment. Pooled C9orf72 siRNA did not decrease endogenous P120 splicing, but the experiment was performed only twice and the knockdown of C9orf72 mRNA in one of these experiments was only 20%, which may explain why there was no effect on splicing in this sample. There also appeared to be some rescue of splicing with C9orf72 expression in the endogenous P120 experiment. Taken together, these results do implicate a role for C9orf72 in minor splicing.

There was contrasting evidence from the HEK293 experiments as to how the decrease in splicing was occurring; knockdown of C9orf72 or an off-target siRNA effect. Transfection of siRNAs also appeared to introduce variability into the experiments as the knockdown of C9orf72 was inconsistent across repeated experiments. Therefore, the reporter splicing assays were performed in the B2

stable C9orf72 knockout cell line, thus removing need for siRNA transfections. As with the previous assays, the results in the B2 cell line were highly variable, and all changes were non-significant. However, there was an observable increase in unspliced product from both exogenous and endogenous P120 in B2 cells in comparison to HEK293 cells. The ratio of the densitometry of spliced/unspliced bands confirmed there was a decrease in P120 minor intron splicing in the B2 cell line (Figure 6.10B-6.11B). This suggests that loss of the C9orf72 protein can cause a defect in minor splicing. However, splicing of exogenous P120 was not rescued with co-transfection of myc-C9orf72 protein. In co-transfecting P120 and myc-C9orf72, we assume cells are transfected with both constructs. However, we are unable to confirm that the cells transfected with P120 were successfully co-transfected with myc-C9orf72. Therefore, there is a chance that we are analysing P120 splicing in cells that do not express C9orf72, which could explain the lack of splicing rescue. It is also possible that the minor splicing defect is caused by an off-target effect introduced with the CRISPR editing of the cell line. However, as discussed in Chapter 3, Section 3.2.1, two guide RNAs were used in the CRISPR-Cas 9 nickase editing of the C9orf72 gene, which significantly reduces the probability of off-target effects (Shen et al., 2014).

As discussed, C9orf72 overexpression has no effect on splicing, therefore is unlikely to be actively involved in the removal of introns. The interaction with minor spliceosome components could be investigated to consider whether C9orf72 is exerting its effect on minor splicing through a direct interaction with the splicing machinery. The C9orf72 protein, unlike the other ALS-linked genes that have been shown to interact with the minor snRNPs, does not appear to be an RNA binding protein (RBP) because of its lack of a RNA-recognition motif, supporting that C9orf72 is unlikely to be directly involved in RNA processing. In keeping with the Cajal body data, a decrease in splicing could be expected to occur if C9orf72 is involved in the trafficking of snRNAs and snRNPs. Loss of the C9orf72 protein could cause an accumulation of immature snRNPs in the nucleus, causing an increase in Cajal bodies and fewer functional snRNPs available for splicing. The more efficient major spliceosome may have more scope for compensating for a loss of C9orf72, whereas the less efficient minor spliceosome may be affected. In addition, overexpression of C9orf72 would not be expected to influence splicing if involved in snRNP trafficking. However, this is supposition based on non-significant data and more experiments are required to narrow-down the actual mechanism for how C9orf72 loss generates the Cajal body and minor splicing phenotype.

Other groups who have utilised the same assay often report 8 biological replicates, whereas we had only 3-4 repeats (Reber et al., 2016; Whittom et al., 2008). Therefore, it appears the assay is subject to high variation and we were likely lacking power to observe statistically significant changes. More repeats would allow us to say with more confidence if we are observing a true effect or not. More quantitative methods of measuring splicing, such as qPCR or RNA sequencing, would help support the data in this chapter. These methods are also more high-throughput, which would allow investigation of multiple minor introns after loss of the C9orf72 protein, as opposed to the single P120 intron. It would also be of value to repeat the reporter splicing assays with a more appropriate positive control. Coilin was initially included as a control, as it has been reported that knockdown decreases splicing (Whittom et al., 2008). However, it has also been shown that coilin depletion can increase splicing of a reporter gene in plants (Kanno et al., 2016). In our studies, coilin knockdown had no effect on the splicing of either pCI-neo or P120, hence a different positive control is required. Knockdown of a splicing factor, such as the SF2 protein, would be an ideal positive control as it should give a definite decrease in both major and minor splicing and would confirm the assays are functional.

7 Discussion

7.1 Summary

The GGGGCC hexanucleotide repeat expansion in the C9orf72 gene is the leading genetic cause of familial and sporadic ALS (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Three non-mutually exclusive disease mechanisms are proposed to occur from the mutation. First, toxic RNA foci sequester RNA-binding proteins and disrupt RNA metabolism. Second, RAN translation of the repeat expansion generates DPR proteins which form cytoplasmic and nuclear aggregates and have been shown to disrupt multiple cellular functions such as the UPS, nucleocytoplasmic transport and stress granule formation. Third, the expansion causes an approximate 50% reduction in C9orf72 mRNA and haploinsufficiency of the C9orf72 protein. Loss-of-function of the C9orf72 protein might cause disruption to the pathways it is normally involved in. Our group and others have discovered a role for the C9orf72 protein in autophagy (Sellier et al., 2016; Webster et al., 2016). There are also suggestions in the literature that C9orf72 may have a distinct nuclear function; the C9orf72 interaction network includes nuclear proteins and immunostaining of C9orf72 has found it to exhibit some nuclear localisation (Blokhuis et al., 2016; Sellier et al., 2016; Xiao et al., 2015). Many of the nuclear proteins identified from the different screens as C9orf72 interactors have functions in RNA metabolism and gene expression, for example splicing factors and heterogeneous nuclear RNPs are specifically enriched in the list (See appendix). However, a function for C9orf72 in the nucleus remains elusive.

The aim of this project was to better understand the nuclear function of C9orf72 and its relevance to ALS. Chapter 4 focused on characterising the interaction between C9orf72 and the nuclear protein coilin. Coilin was initially identified as an interacting partner of the C9orf72 long protein in an unbiased yeast two-hybrid screen (Table 4.1). The interaction was confirmed by applying three different methods for investigating protein-protein interactions. An *in vitro* GST-binding assay confirmed coilin directly binds to both the C9orf72 short and long protein isoforms (Figure 4.4). Co-immunoprecipitation experiments showed that the interaction could be detected in HEK293 cells and is therefore physiologically relevant (Figure 4.5). Lastly, the proximity ligation assay allowed visualisation of the close proximity between

C9orf72 and coilin *in situ*, which suggested C9orf72 and coilin interact in both the nucleus and the cytoplasm (Figure 4.7).

Coilin is the major protein component of Cajal bodies, nuclear suborganelles which are linked with snRNP processing, snRNA gene regulation and telomere maintenance. Chapter 5 therefore investigated the relationship between C9orf72 and Cajal bodies. The number of Cajal bodies was increased in HEK293 cells following transient C9orf72 knockdown and in the B2 stable C9orf72 knockout line (Figure 5.2, Figure 5.5). Overexpression of C9orf72 appeared to reduce the number of Cajal bodies (Figure 5.4-Figure 5.5). In contrast, overexpression or knockdown of C9orf72 in rat cortical or hippocampal neurons did not change the number of Cajal bodies (Figure 5.6-Figure 5.7). Lastly, Cajal bodies were increased in C9orf72-ALS iAstrocyte cells, but rescue of Cajal body numbers with C9orf72 expression was unsuccessful, suggesting a mechanism unrelated to C9orf72 haploinsufficiency was causal of the Cajal body increase in these cells (Figure 5.8-Figure 5.10). The results presented in this chapter suggest the C9orf72 protein can influence Cajal bodies and may be involved in a pathway linked with their regulation in HEK293 cells.

Finally, Chapter 6 investigated the effect of C9orf72 on the major and minor splicing pathways. The best-known function of Cajal bodies is the processing of the snRNPs that form the splicing machinery. In addition, RNA splicing defects have been reported in C9orf72-ALS patients (Cooper-Knock et al., 2015a; Prudencio et al., 2015). We therefore investigated whether changes in splicing in HEK293 cells occurred in response to loss of the C9orf72 protein. Transient knockdown of C9orf72 in HEK293 cells generated mixed results. C9orf72 knockdown appeared to cause a small decrease in major and minor splicing, but there was evidence to suggest this may have occurred from an siRNA off-target effect (Figure 6.3-Figure 6.8). In the B2 stable C9orf72 knockout line, major splicing was unaffected but there was a non-significant decrease in minor splicing (Figure 6.9-Figure 6.11). However, C9orf72 expression in B2 cells did not rescue the minor splicing defect (Figure 6.10-Figure 6.11). The results presented in Chapter 6 suggested there was a trend towards decreased minor splicing in response to loss of the C9orf72 protein. Unfortunately, high variability across the assay combined with a lack of power in the biological repeat number limits the confidence with which statements can be made from the data.

7.2 Discerning C9orf72 function from phenotype

7.2.1 Nucleocytoplasmic transport and RNA processing

We have shown that C9orf72 interacts with coilin and can influence the number of Cajal bodies in HEK293 cells. This hints at a potential role for C9orf72 in snRNP processing or a similarly related pathway. Published evidence demonstrates that the C9orf72 protein may be involved in nucleocytoplasmic transport. The endogenous C9orf72 short protein displays nuclear envelope localisation in motor neurons and both C9orf72 isoforms exhibit both nuclear and cytoplasmic localisation in transfected cell lines (Xiao et al., 2015). The results presented in Chapter 4 support this finding. Overexpression of either C9orf72 protein resulted in one of three cellular localisation patterns: equal diffusion throughout the whole cell, strong nuclear localisation or strong cytoplasmic localisation (Figure 4.1). These localisation patterns support a role for C9orf72 actively moving between the nucleus and cytoplasm. Indeed, the C9orf72S staining pattern in motor neurons co-localises with known nucleocytoplasmic transport proteins, and co-immunoprecipitation experiments in N2a cells showed C9orf72S interacted with Ran-GTPase and Importin-β (Xiao et al., 2015). The interaction between C9orf72 and Ran-GTPase was also identified in a mass spectrometry screen of C9orf72 protein interactors (See appendix). Interestingly, the C9orf72S nuclear speckles we observed in a subset of cortical neurons are similar in appearance to the staining pattern of the Ran-GTPase activating protein 1 (RanGAP1) using an anti-RanGAP1 antibody (ab204650) (unpublished). These results suggest the C9orf72 protein may be involved in Ran-GTPase mediated nucleocytoplasmic transport. The Ran protein exists in two forms, GTP-bound or GDP-bound. In the nucleus, Ran-GTP binds exportin proteins, for example CRM1, and facilitates the translocation of their cargo through the nuclear pore complex. In the cytoplasm, Ran-GTP interacts with RanGAP1 and hydrolyses the GTP to GDP (Görlich and Mattaj, 1996; Nigg, 1997). Ran-GDP is then imported back into the nucleus where it interacts with the RCC1 protein. The RCC1 protein acts as a GEF for Ran, and mediates the substitution of GTP for GDP, generating Ran-GTP and allowing the cycle to begin again. For translocation of cargo from the cytoplasm into the nucleus, importin proteins bind their cargo and mediate translocation across the NPC directly. In the nucleus, an interaction with Ran-GTP releases the cargo and facilitates translocation of the importin proteins back to the cytoplasm in the manner described above (Görlich and

Mattaj, 1996; Nigg, 1997; Weis, 2003). Ran is part of the same family of GTPases as Rab proteins, and they share evolutionary similarities in their structure and function (Wennerberg et al., 2005). C9orf72 has been reported to bind Rabs preferentially in their GTP-bound conformation, thus behaving as an effector (Webster et al., 2016). We therefore reason that C9orf72 may be involved in nucleocytoplasmic shuttling via an interaction with Ran-GTP.

Chapter 3 showed that both C9orf72 isoforms interact with the Cajal body protein coilin. In addition, we have demonstrated that both acute and chronic loss of C9orf72 in HEK293 cells led to an increase in the number of Cajal bodies, whereas its overexpression caused a decrease in the number of Cajal bodies. As reviewed in Chapter 1, snRNPs go through a long maturation pathway which involves entry into and processing in Cajal bodies. Many of the coilin-interacting proteins in Cajal bodies are involved in the snRNP processing pathway. We have therefore analysed the snRNP processing pathway to determine whether C9orf72 could fit in this pathway based on the results presented in this thesis.

One of the early steps in the snRNP maturation pathway is the assembly of the snRNA export complex. Nascent snRNAs are bound on their 5' end by the cap binding complex (CBC) and are recruited to Cajal bodies. Here, the export proteins PHAX and CRM1 are assembled sequentially onto the snRNA to form the export complex (Boulon et al., 2004). After the snRNA export complex is assembled, it exits Cajal bodies into the nucleoplasm where it is bound by Ran-GTP to be translocated to the cytoplasm for further modifications. Both the PHAX and CRM1 proteins have been detected in Cajal bodies, whereas Ran-GTP is only found in the nucleoplasm. PHAX binding of snRNA is essential for the snRNA to exit Cajal bodies, suggesting the Cajal body behaves as a quality checkpoint. We suggest that the evidence presented in Chapters 4 and 5 support a role for C9orf72 in trafficking the snRNA export complex from Cajal bodies to the cytoplasm via an interaction with coilin and Ran-GTP (Figure 7.1). A study in frog oocytes has demonstrated coilin can shuttle between the nucleus and cytoplasm, and the PLA experiments in Chapter 4 suggest C9orf72 and coilin interact in both compartments, suggesting they both may be translocating across the nuclear membrane as part of the same complex (Bellini and Gall, 1999). We did not detect C9orf72 in Cajal bodies as is the case for Ran-GTP, suggesting its position in the pathway is post-snRNA export complex formation in Cajal bodies. Depletion of C9orf72, if indeed it is involved in the trafficking of snRNA from Cajal bodies to the cytoplasm, may lead

to an accumulation of snRNA in the nucleus, which in turn could cause the Cajal body phenotype we report in this work (Figure 7.1). This is supported by a similar Cajal body phenotype observed after depletion of PHAX in *Xenopus* oocytes. Inhibition of PHAX activity inhibits snRNA export to the cytoplasm and causes an increase in their accumulation in Cajal bodies (Suzuki et al., 2010). In addition, knockdown of PHAX using siRNA in HeLa cells has been demonstrated to cause mislocalisation of coilin to the nucleolus (Takata et al., 2012). These studies are evidence that inhibition of snRNA export can influence the localisation of coilin and appearance of Cajal bodies, and the phenotype can be cell-type specific.

Interestingly, if snRNA trafficking is a function of C9orf72, it could also be involved in the intranuclear trafficking of snoRNA from Cajal bodies to the nucleolus. snoRNAs briefly visit Cajal bodies for RNA processing before trafficking to the nucleolus, where they catalyse RNA modifications of ribosomal RNA (Narayanan et al., 1999a, 1999b). In addition to interacting with coilin, C9orf72 has also been shown to interact with nucleolus proteins fibrallarin and nucleolin in both in house and published proteomic screens (Blokhuis et al., 2016). Cajal bodies often associate with the nucleolus and coilin also shows nucleolar localisation, reviewed in (Trinkle-Mulcahy and Sleeman, 2017). As is the case for snRNA export, the PHAX and CRM1 proteins are responsible for routing snoRNA from Cajal bodies to the nucleolus. Knockdown of the CRM1 protein leads to accumulation of snoRNA in Cajal bodies (Boulon et al., 2004).

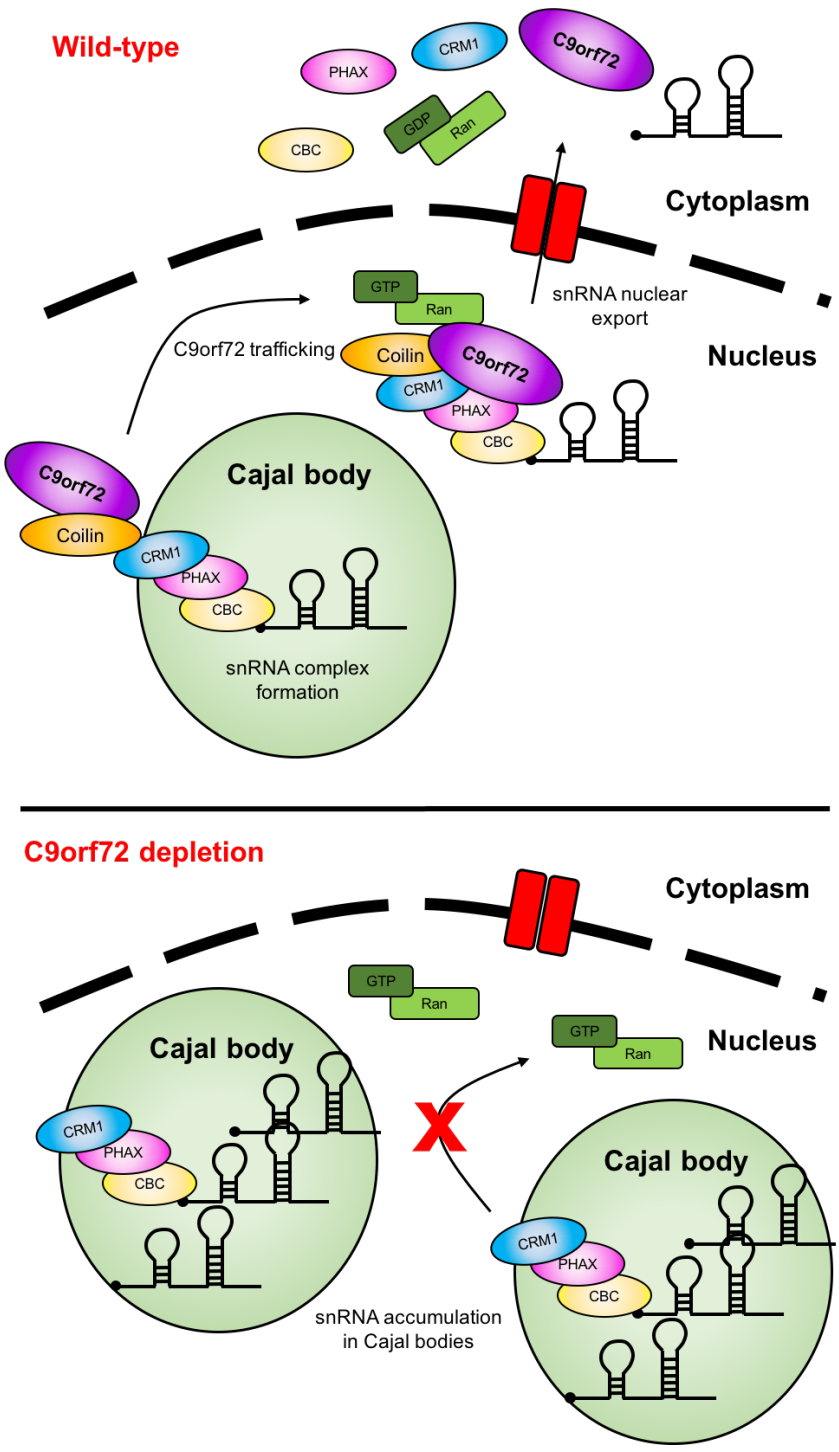


Figure 7.1. Hypothesised role for C9orf72 in snRNA trafficking. In wild-type cells, C9orf72 may be involved in the trafficking of the snRNA export complex. After PHAX and CRM1 bind in Cajal bodies, an interaction with C9orf72 which may be mediated by coilin, allows for trafficking to the nuclear pore complex and interaction with Ran-GTP for nuclear export. In C9orf72 depleted cells, snRNAs accumulate in the nucleus and Cajal bodies. The accumulation of snRNA nucleates new Cajal bodies thus leading to the increased Cajal body phenotype.

In Chapter 6, we investigated the effect of C9orf72 depletion on major and minor splicing. Transfection of C9orf72 siRNA appeared to have some off-target effects on splicing, therefore I have chosen to focus on discussing the results from the B2 C9orf72 stable knockout cell line and how they may link with C9orf72 in nucleocytoplasmic transport and the snRNP processing pathway. Statistically, the reporter splicing assays showed there was no effect of loss of C9orf72 on splicing. However, there was a non-statistically significant decrease in minor splicing in the B2 cells. As previously discussed, with an n number of 3 we were likely underpowered to observe a statistically significant result. Performing a power calculation suggests a minimum sample size of 5 would be required to give an 80% probability of observing an effect in the splicing assay. If haploinsufficiency of C9orf72 does disrupt the snRNP processing pathway, fewer functional snRNPs would be available and a decrease in splicing could be expected. Indeed, knockdown of the PHAX protein has been shown to cause a decrease in snRNPs which in turn led to mis-splicing events in a *Drosophila* model, confirming disruptions to snRNP biogenesis at the snRNA export step can lead to dysfunctional RNA processing (Garcia et al., 2016). Interestingly however, the splicing defect we observed in this project was minor splicing specific. It may be that in our cell model, the major spliceosome has greater capacity to cope with reductions in snRNP biogenesis. The minor spliceosome which has a low abundance of snRNPs to begin with, may be more vulnerable to changes in snRNP levels (Pessa et al., 2006). In addition, studies that have investigated inhibition of snRNP biogenesis on splicing have reported the effect on splicing to be specific to a subset of introns and exons, as opposed to global changes (Garcia et al., 2016; Pleiss et al., 2007; Saltzman et al., 2011). Therefore, it may be coincidental in our experiments that the pCI-neo (major) intron is unaffected by the decrease in snRNPs, whereas the P120 (minor) intron is targeted. We would need to analyse more major and minor introns to determine if the effect is transcript or spliceosome specific.

Conversely, there is also evidence against the hypothesis that C9orf72 may function in snRNP biogenesis. First, the snRNA export complex has been thoroughly researched and its molecular pathway is well understood. There are no suggestions in the literature for a role of another unidentified protein in this pathway. Second, this pathway is presumably ubiquitous across different cell types, therefore one could expect the Cajal body phenotype to be replicated in primary neurons as well as HEK293 cells. It is worth noting however, that the experiments investigating the molecular pathway of snRNA maturation have mostly been performed in

transformed cell lines and differences in the pathway in neurons is possible. Thirdly, a 2012 study demonstrated depletion of CRM1 and PHAX export proteins caused loss of Cajal bodies and mislocalisation of coilin to the nucleolus, which contradicts the increased Cajal body phenotype we observe (Takata et al., 2012). Lastly, if the C9orf72 protein is involved in the snRNA export complex, one would expect it to indirectly interact with the PHAX and CRM1 proteins yet neither of these proteins have been identified in the C9orf72 proteomic screens (See appendix).

7.2.2 Cell cycle

A second plausible role for nuclear C9orf72 could be an involvement in the cell cycle. The results presented in Chapter 5 indicated C9orf72 regulation of Cajal bodies was specific to dividing cells, suggesting its function may be linked with cell proliferation. Neurons are regarded as permanently post-mitotic, hence the absence of an effect of C9orf72 protein overexpression or knockdown in the primary rat neurons may be explained by differences in the cell-cycle biochemistry between differentiated and proliferating cells. The number of Cajal bodies is increased during mid-G_1, followed by their disassembly at entry into mitosis (Andrade et al., 1993; Carmo-Fonseca et al., 1993). It has been hypothesised that the changing number of Cajal bodies is in response to the transcription and splicing demand of the cell. During mitosis, transcription and splicing are halted, therefore there is less of a requirement for snRNP biogenesis. On the other hand, transcription and splicing are prevalent during G_1, thus Cajal bodies are increased. The regulation of Cajal body numbers during cell cycle progression is controlled by phosphorylation of coilin (Carmo-Fonseca et al., 1993; Chan et al., 1994; Hearst et al., 2009). Low levels of coilin phosphorylation, as detected in G_1, favour Cajal body assembly and hyperphosphorylation of coilin during mitosis causes Cajal body disassembly (Figure 7.2) .

In chapter 4, C9orf72 was shown to directly interact with coilin. We therefore reason C9orf72 may be involved in regulating Cajal body numbers via its interaction with coilin, and could do so by affecting coilin phosphorylation and/or protein levels. The VRK1 protein is a cell-cycle regulated nuclear kinase that is involved in coilin phosphorylation (Sanz-García et al., 2011). Phosphorylation of coilin by VRK1 promotes Cajal body assembly from two predicted mechanisms. First, specific coilin phosphorylation patterns enhance protein-protein interactions and increase Cajal

body nucleation events. Second, coilin phosphorylation prevents its ubiquitination by the Mdm2 E3 ubiquitin ligase and degradation by the UPS. Coilin can also be phosphorylated by the CDK2-cyclin E complex, which shows localisation to Cajal bodies and is responsible for transition from the G_1 to S phase (Liu et al., 2000). There is little evidence to currently support a role for C9orf72 in directly regulating coilin phosphorylation levels. Hyperphosphorylation or complete dephosphorylation of coilin completely abolishes its ability to form Cajal bodies, therefore it is probable that coilin in the C9orf72 knockout and knockdown HEK293 cells exists in the hypophosphorylated state that favours Cajal body formation.

From our current knowledge of C9orf72, the protein does not exhibit kinase or phosphatase enzymatic activity. This would therefore suggest the protein kinases or phosphatases responsible for managing coilin phosphorylation are influenced by loss of C9orf72. Because these proteins appear to be regulated by the cell-cycle, we cannot say whether they are affected directly through loss of C9orf72 protein, e.g. loss of a C9orf72-kinase interaction, or by a change in the cell cycle, e.g. stalling in a particular phase of the cell cycle. The table of proposed nuclear interactors of C9orf72 in the appendix shows a lack of detected interactions between C9orf72 and nuclear serine/threonine kinases. However, analysis of the C9orf72 protein sequence shows it contains four potential sites for CDK2-cyclin binding, which follow the XK/RXL motif: $_{248}$VRTL, $_{407}$RKAL, $_{427}$FKSL, $_{430}$LRNL (Takeda et al., 2001; Wohlschlegel et al., 2001). As mentioned, the CDK2-cyclin E complex can phosphorylate coilin and joins Cajal bodies at the G_1/S boundary, presumably where it performs phosphorylation of coilin and other Cajal body-localised substrates (Liu et al., 2000). Hence, C9orf72 may interact with both coilin and the CDK2-cyclin E complex and mediate the G_1 phosphorylation of coilin. The CDK2-cyclin binding motifs found in C9orf72 are located within the central DENN and downstream DENN domains, and are therefore not present in the C9orf72 short isoform. We have shown that both the C9orf72 short and long protein isoforms interact with coilin, suggesting the longin domain of C9orf72 is responsible for mediating this interaction. It is therefore possible for C9orf72 to interact with coilin via its longin domain and bridge an interaction with the CDK2-cyclin E complex via its C-terminal domain. In this scenario, the C9orf72 short protein would lack this function. However, in the experiments presented in this thesis, transfection of both C9orf72 isoforms rescued Cajal body numbers, questioning the relevance of the CKD2-cyclin binding motifs to Cajal body regulation. It may be that the motifs are

not involved in cyclin binding but are hydrophobic stretches of residues that contribute to proper dDENN domain folding.

It is not illogical to assume the regulation of Cajal bodies by C9orf72 in HEK293 cells requires the interaction with coilin. However, there is evidence to suggest there is a link between the cell cycle regulation of Cajal bodies and C9orf72's previously discussed predicted functioning in Ran-GTP nucleocytoplasmic transport. Disruptions to the Ran-GTPase network has been found to disrupt multiple steps in the cell cycle (Moore, 2000). Of most relevance to this project, depletion of the RCC1 protein, which is a GEF for Ran-GTP, can cause cells to arrest at the G_1/S boundary (Nishimoto et al., 1978). In addition, expression of a mutant Ran-GTPase protein that cannot catalyse the conversion of GTP to GDP, also leads to cell cycle arrest at the G_1/S boundary or G_2/M boundary (Ren et al., 1994). In line with this, one could expect to observe an increase in the number of Cajal bodies in cells arrested at the G_1/S transition. One interpretation of these results is that loss of C9orf72 inhibits Ran-GTP driven nucleocytoplasmic transport, as proposed in the previous section. It has been hypothesised that loss of nucleocytoplasmic transport may then indirectly influence the cell cycle (Dasso, 1995; Moore, 2000). Faults in the nuclear export/import pathways can lead to mislocalisation of both RNA and protein, including the CDK-cyclin complexes required for cell cycle regulation. Loss of these proteins then signal for cells to arrest at the boundary between G_1 and S (Figure 7.2).

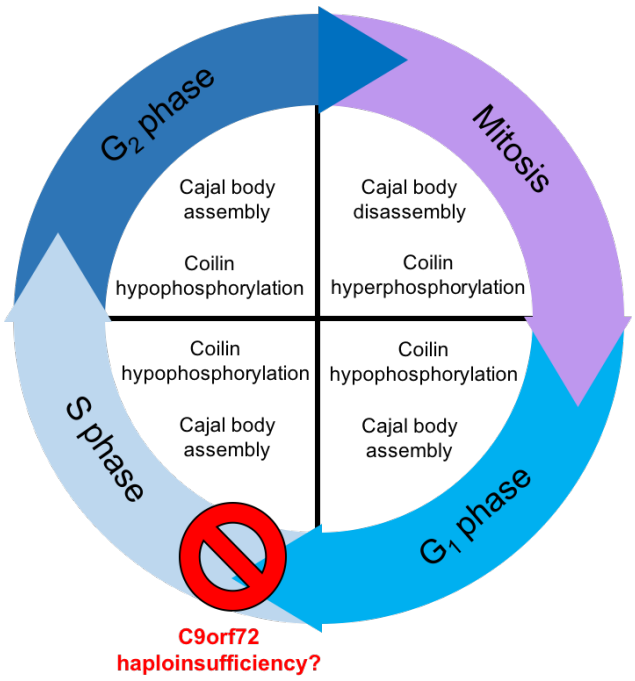


Figure 7.2. Haploinsufficiency of C9orf72 may lead to cell cycle arrest at the G_1/S transition. Coilin is hypophosphorylated during the G_1, S, and G_2 phases of interphase which promotes Cajal body assembly. Mitotic coilin is hyperphosphorylated which causes Cajal body disassembly during mitosis. Loss of C9orf72 causes an increase in the number of Cajal bodies and a decrease in minor splicing, suggestive of cell-cycle arrest at the G_1/S transition.

There is also some evidence to suggest the minor spliceosome has some cell-cycle regulated activity, perhaps explaining why we detect a decrease in minor but not major splicing following C9orf72 depletion, should the above hypothesis prove true. A study in 2007 reported that the minor and major spliceosomes may be spatially segregated in vertebrates, with minor splicing in the cytoplasm and major splicing in the nucleus (König et al., 2007). As reviewed in Chapter 1, Section 1.7.2.1, the inefficiency of the minor spliceosome in comparison to the major spliceosome, often causes generation of mRNA transcripts that contain a single minor intron. The König paper reports that these minor-intron mRNAs can undergo nuclear export and minor snRNPs can catalyse the removal of the intron in the cytoplasmic environment. Interestingly, splicing activity of the P120 minor intron was detected in cells arrested in mitosis when transcription and major splicing are typically halted, and depletion of the minor snRNPs caused cell cycle arrest at the G_1/S transition, suggestive of a role for minor splicing in the cell cycle (König et al., 2007). Taken together, these results suggest minor splicing may be regulated by and/or have

regulatory control of the cell cycle. We have suggested that loss of C9orf72 nucleocytoplasmic transport may lead to cell cycle arrest in the G_1/S boundary, which in turn may decrease minor splicing activity which may have a different cell cycle activity and localisation pattern to major splicing. It may also be that loss of C9orf72 nucleocytoplasmic transport leads to a decrease in snRNP biogenesis as discussed in the previous section. Loss of the snRNPs may then signal for cells to arrest before DNA synthesis can occur.

In summary, we reason the Cajal body phenotype and minor splicing defect we have reported in this thesis could occur through multiple converging mechanisms which comply with C9orf72 having a function in nucleocytoplasmic transport (Figure 7.3).

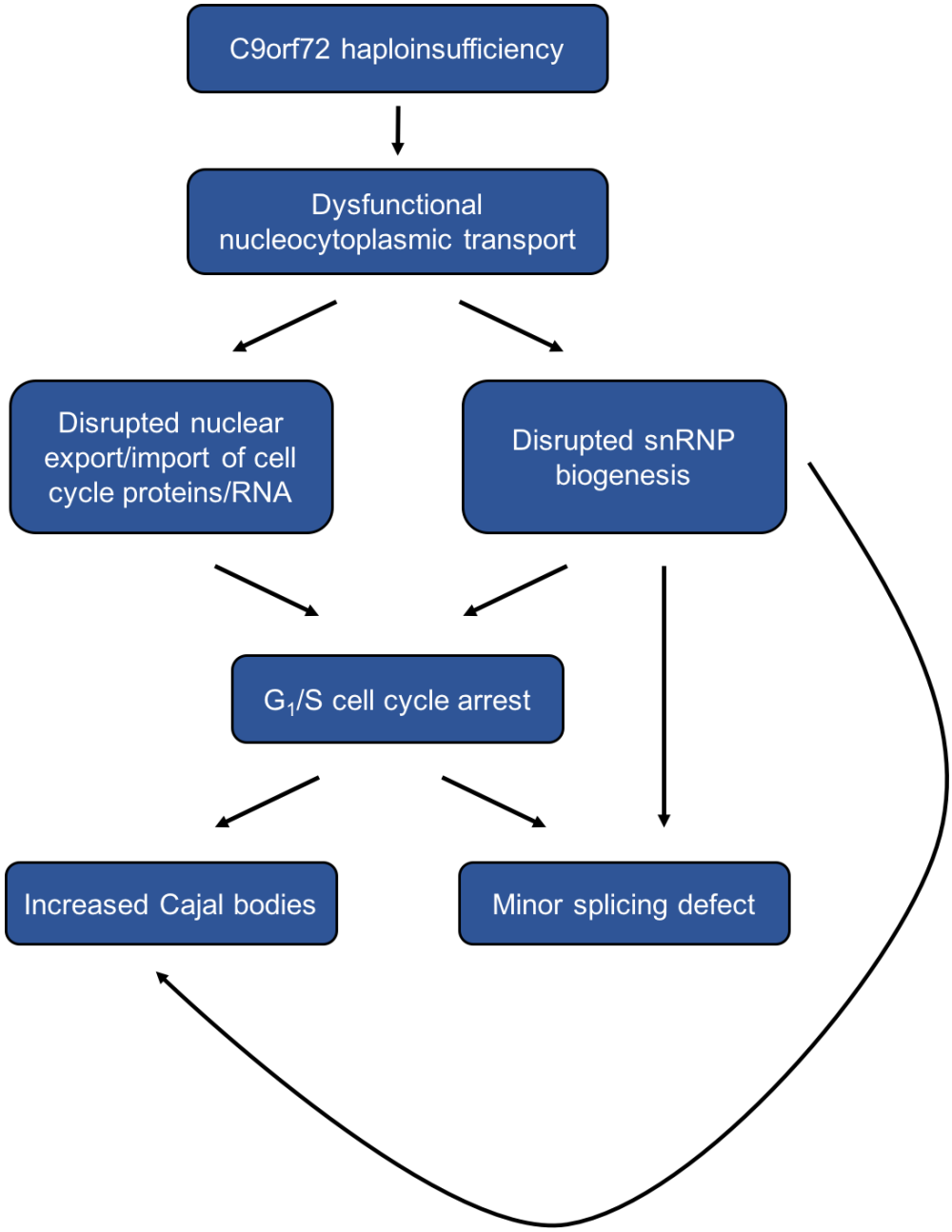


Figure 7.3. Converging disease mechanisms caused by C9orf72 haploinsufficiency. There is evidence to suggest the C9orf72 protein has an involvement in nucleocytoplasmic transport based on its interaction with Ran-GTP and importin-B, and it's subcellular localisation pattern. C9orf72 haploinsufficiency therefore may cause dysfunctional nucleocytoplasmic transport. This may lead to mislocalisation of RNA and proteins involved in cell cycle regulation which in turn signals for cell cycle arrest. Cajal bodies and minor splicing can both be regulated by cell cycle, and will therefore be affected by stalling at the G_1/S phase. Dysfunctional nucleocytoplasmic transport may also affect snRNP biogenesis, which relies on nuclear export and import for the maturation pathway. This may lead to an accumulation of immature snRNA in the nucleus of cells, which could nucleate Cajal body assembly. We also propose that a decrease in available snRNPs could directly lead to a minor splicing defect and can also cause cell cycle arrest.

7.3 Cajal bodies in ALS

One of the main findings of this project was that loss of the C9orf72 protein can lead to an increase in Cajal bodies. A proposed mechanism for how this occurs has been discussed (Figure 7.3). In Chapter 5 we observed that the number of Cajal bodies was increased in C9orf72-ALS patient iAstrocytes in comparison to their age- and sex-matched healthy controls. This suggests an increase in the number of Cajal bodies may be a pathological hallmark of C9orf72-ALS.

Cajal bodies have links with neurodegenerative disorders, however it is normally their loss which is thought to be pathogenic, and the loss of snRNP biogenesis/splicing that goes with this. It is not known whether an increase in Cajal bodies can have related toxic side effects, therefore more work is needed to investigate the downstream consequences of an increase in the number of Cajal bodies in cells. From the current knowledge of the function of Cajal bodies, we hypothesise that the increase in Cajal bodies themselves are not pathogenic, but are a visible consequence of a separate pathomechanism.

Interestingly, the number of Cajal bodies in the patient iAstrocyte cell lines was not rescued by exogenous C9orf72 expression, suggesting in these cells there may be another pathway involved that leads to the same Cajal body phenotype as C9orf72 protein haploinsufficiency. We have hypothesised that the C9orf72 protein may have a role in nucleocytoplasmic transport, and Figure 7.3 demonstrates how a disruption to this function may lead to misregulation of Cajal bodies. As reviewed in Chapter 1, both the RNA foci and DPRs generated from the C9orf72 mutation have been reported to interfere with nucleocytoplasmic transport. RNA foci are hypothesised to exert toxicity via sequestration of RNA-binding proteins. Interestingly, the RanGAP1 protein which functions within the Ran-GTP network, interacts with the C9orf72 repeat expansion RNA and colocalises with RNA foci in C9orf72-ALS derived neurons and patient tissue (Zhang et al., 2015). The same study described a consequence of RanGAP1 sequestration was Ran-GTP mislocalisation to the cytoplasm and significant disruption to the translocation of proteins across the nuclear membrane (Zhang et al., 2015). In another study, expression of 58 GGGGCC hexanucleotide repeats in both flies and cells caused the nuclear accumulation of RNA, suggestive of disrupted RNA export (Freibaum et al., 2015). The authors reported only a fraction of the retained RNA was mRNA, suggesting other small non-coding RNAs may also have been retained. There are a

number of publications suggesting arginine-containing DPRs can also directly interact with the NPC and disrupt RNA and protein nucleocytoplasmic transport (Boeynaems et al., 2016; Jovičić et al., 2015; Shi et al., 2017). Knockdown of both the RCC1 and RanGAP1 proteins in a *Drosophila* model expressing the Poly-GP protein enhanced neurodegeneration of the fly eye and toxicity (Boeynaems et al., 2016). In C9orf72-ALS derived induced neurons, the RCC1 protein was also shown to be mislocalised to the cytoplasm (Jovičić et al., 2015). Taken together, disruptions to the Ran-GTP network and nucleocytoplasmic transport may be the end target from three individual but converging disease mechanisms that occur from the C9orf72 mutation in ALS. If this hypothesis proves true, it may be expected that in the presence of C9orf72 haploinsufficiency, blocking DPR or foci generation might also fail to rescue Cajal bodies and Ran-mediated nucleocytoplasmic transport.

Disruption to the nuclear export and import of RNA and protein could have widespread repercussions. The changes to alternative splicing in C9orf72-ALS patients have been attributed to RNA foci sequestration of splicing factor proteins, but the mislocalisation of snRNA and proteins involved in splicing from disrupted nucleocytoplasmic transport may also be a contributing factor (Cooper-Knock et al., 2015a; Prudencio et al., 2015). In addition, researchers have questioned whether defective nucleocytoplasmic transport in ALS could also be causal of the mislocalisation of proteins such as TDP-43. Interestingly, inhibition of nuclear import leads to mislocalisation of FUS into cytoplasmic aggregates and overexpression of the transportin 1 protein, responsible for FUS-nuclear import, suppresses DPR-related toxicity in both yeast and flies (Boeynaems et al., 2016; Reber et al., 2016; Sun et al., 2015). Pathogenic FUS-related splicing defects have been reported to occur through loss of the nuclear function of FUS and sequestration of snRNPs by FUS aggregates (Reber et al., 2016; Sun et al., 2015). Hence, defective nucleocytoplasmic transport may be an underlying causal factor for FUS pathology in ALS and FTD. In support, DPR pathology, and presumably also RNA foci formation, has been shown to precede TDP-43 neuronal cytoplasmic inclusion pathology in C9orf72-FTLD patients (Baborie et al., 2015; Proudfoot et al., 2014). ALS is an aging disease, and there is evidence to suggest nucleocytoplasmic transport efficiency decreases with age (Mertens et al., 2015). Motor neurons, which do not undergo mitosis, do not have regular break down of the nuclear envelope and nuclear pore complex disassembly. A consequence of this is the proteins involved in nucleocytoplasmic transport in neurons are recycled at a much slower

rate and therefore could be more prone to damage and deterioration (D'Angelo et al., 2009; Toyama et al., 2013). Motor neurons, which have a higher dependency on splicing and metabolism than other post-mitotic cells, may then be more vulnerable to defects in RNA and protein import/export.

The iAstrocyte Cajal body results are interesting when thinking about the implications for ALS and future treatments. Anti-coilin antibodies that exhibit both high sensitivity and specificity are commercially available. We reason that after further validation studies to confirm increased Cajal bodies are a consequence of dysfunctional nucleocytoplasmic transport, their immunostaining could be used as a useful biomarker for investigating the usefulness of drugs targeting this pathway. For example, if the increased Cajal bodies in patient iAstrocytes are caused by RNA foci sequestration of the RanGAP1 protein and disrupted nucleocytoplasmic transport, then a decline in Cajal bodies could be a relatively high-throughput readout to measure the efficiency of drugs targeting RNA foci. In line with this, it would be particularly interesting to perform Cajal body immunostaining from different cells, animals and patient tissue that model the three different C9orf72-ALS disease mechanisms to investigate this further. As ALS causes the specific degeneration of motor neurons, the regulation of Cajal bodies in these cells would be most important. Unfortunately, we did not observe a change in the number of Cajal bodies in primary neurons following C9orf72 protein knockdown, although it is worth noting these were not motor neurons. In addition, the level of C9orf72 knockdown in neurons was determined by measuring C9orf72 mRNA levels, which had limitations. The miRNA transduced into neurons to knockdown C9orf72 was tagged with GFP, and immunofluorescence images suggest there was 100% transduction efficiency, whilst C9orf72 mRNA was reduced approximately by 50%, suggesting miRNA may not be the most effective method for C9orf72 knockdown. Immunoblot of C9orf72 from neuronal lysate following C9orf72 miRNA transduction would allow for better characterisation of C9orf72 knockdown in the primary cells. Should the C9orf72 protein be reduced in neurons without the Cajal body phenotype, this suggests either (1) C9orf72 haploinsufficiency alone does not cause disruption enough to generate the Cajal body phenotype in neurons or (2) loss of C9orf72 is pathogenic in neurons but differences in their biochemistry means Cajal body numbers do not respond as they would in different cell models.

7.4 C9orf72 haploinsufficiency in ALS

If the Cajal body phenotype is a consequence of a disrupted molecular pathway in the iAstrocyte patient cells, then the experiments included in this thesis suggest overexpression of the C9orf72 protein alone may not be enough to restore normal function. This would suggest, at least in the iAstrocyte cells, that haploinsufficiency of the C9orf72 protein is not the only cause of ALS pathogenesis. This would need to be confirmed however by rescuing iAstrocyte C9orf72 protein levels whilst simultaneously blocking DPRs and RNA foci.

In chapter 6, we found that loss of the C9orf72 protein may directly lead to a small defect in minor splicing, suggesting loss of function may have some pathogenic consequences, although these results are variable and need to be further validated. We cannot conclude from the reporter splicing assays the size of the effect of C9orf72 haploinsufficiency on splicing. However, the decrease in minor splicing was small, and there was no effect to major splicing. It is therefore unlikely to be solely responsible for the widespread changes to alternative splicing events found in C9orf72-ALS patient brains, but may still contribute to the intron retention (Prudencio et al., 2015).

In summary, we propose that loss of this nuclear function of C9orf72 does not have a large enough effect to cause severe pathogenesis by itself. This is supported by the relatively minor neurodegeneration in C9orf72 loss of function animal models in comparison to C9orf72 gain-of-function models (Jiang et al., 2016; Koppers et al., 2015; Liu et al., 2016; O'Rourke et al., 2016; Peters et al., 2015). However, in combination with gain of toxic functions from RNA foci and DPRs, loss of C9orf72 could have a detrimental effect by enhancing the neurodegeneration caused by these disease mechanisms, such as RNA processing and nucleocytoplasmic transport. That is not to say reintroduction of the C9orf72 protein would not provide some advantage to patients, as there is compelling evidence to suggest it could be beneficial to restoring C9orf72-haploinsufficiency related autophagy problems (Shi et al., 2018; Webster et al., 2016), but its impact on nuclear-associated disease pathways, such as RNA processing and nucleocytoplasmic transport, is still to be determined. Overexpression of nucleocytoplasmic transport proteins, such as Ran-GAP1 or KPNA3, have been reported to partially alleviate the neurodegeneration caused by the toxic gain-of-function mechanisms suggesting restoration of this

pathway has therapeutic benefit (Jovičić et al., 2015; Zhang et al., 2015). However, the literature suggests several proteins in nuclear export/import can be affected in C9orf72-ALS. As such, expression of a singular protein may provide some therapeutic benefit, but would be limited to a partial rescue. In addition to this, expression of wild-type proteins in motor neurons to overcome their loss-of-function would require implementing gene therapy techniques which has had limited success so far in modern treatments. Antisense oligonucleotides that target the C9orf72 repeat expansion RNA have shown promise and could inhibit the generation of both RNA foci and DPRs (Lagier-Tourenne et al., 2013; Zhang et al., 2015).

7.5 Future directions

The work completed towards this thesis has set the groundwork for some interesting future experiments to build upon. The two most important things to lead on from this project are to (1) validate the splicing results presented in Chapter 6 and (2) design experiments to determine a mechanistic understanding of the nuclear function of C9orf72.

Chapter 4 showed an interaction between C9orf72 and coilin. We have shown that C9orf72 interacts with coilin via its longin domain, as the interaction was detected with both the short and long protein isoforms. It would be interesting to confirm this result by including a C9orf72 mutant protein lacking the longin domain in the co-immunoprecipitation experiment. Coilin deletion mutants could also be used to investigate where on coilin C9orf72 is binding. Deletion mutants of particular interest would be coilin lacking the N-terminal self-association domain, the C-terminal tudor domain, the C-terminal RG-box and a double tudor domain/RG box deletion mutant. Both C9orf72 protein isoforms were shown to rescue the number of Cajal bodies in Chapter 5, suggesting the C9orf72 longin domain is sufficient for Cajal body regulation. Overexpression of a ∆longin-C9orf72 protein in the B2 C9orf72 knockout cells would not be expected to rescue the number of Cajal bodies. Currently, the significance of the C9orf72-coilin interaction for the ability of C9orf72 to influence Cajal body numbers is not clear. If we could further narrow down the interaction to the exact binding residues on C9orf72, we could generate a C9orf72 mutant that cannot bind coilin. It would then be interesting to determine if expression of the C9orf72 protein that lacks coilin binding can regulate the number of Cajal bodies.

In chapter 6, the effect of loss of C9orf72 on major and minor splicing was presented. Issues with variability in the reporter splicing assays has already been discussed in detail. Others who have implemented these assays have often included 8 biological repeats, therefore we suggest performing sufficient repeats of our assays to determine if C9orf72 has a significant effect on minor splicing; the power calculation referred to earlier in this discussion would suggest another two repeats are needed to observe a significant effect (Reber et al., 2016; Whittom et al., 2008). In addition, we were limited in these assays by analysing the splicing of just one major and minor intron, therefore making it difficult to conclude if the decrease in splicing was transcript specific or loss of spliceosome-specific splicing. Therefore the assay should be performed using additional major and minor reporters such as SMS, MAPK3, and MAPK12, which contain both minor and major introns and can therefore be used as reporters for both splicing pathways (Markmiller et al., 2014). RNA sequencing after C9orf72 depletion would generate the most informative data on the effect of C9orf72 haploinsufficiency on splicing. Mis-splicing of targets shared between the transient C9orf72 knockdown and C9orf72 knockout cells are likely to be true mis-splicing events caused by loss of C9orf72 protein rather than an off-target effect from siRNA transfection or CRISPR/Cas9 editing. Confirmation of these effects in cerebellum and cortex from the brains of C9orf72 haploinsufficient mice, as has been performed on C9orf72-ALS patient brains (Prudencio et al., 2015), would allow separation of the effect of C9orf72 protein haploinsufficiency on splicing from RNA foci and DPR contributions.

The effect of C9orf72 on Cajal bodies and minor splicing, combined with the localisation pattern of C9orf72 in Chapter 3, led to predictions of C9orf72 having a nuclear function in nucleocytoplasmic transport and snRNA export. There is evidence to suggest C9orf72 interacts with Ran-GTPase. To investigate C9orf72-Ran-GTPase binding, one could perform an in vitro GST-binding assay to measure the direct interaction between the proteins (Xiao et al., 2015). Further from this, to determine if the interaction is physiologically relevant, co-immunoprecipitation experiments between C9orf72 and export adaptor proteins PHAX and CRM1 could determine if an interaction is present in cells. If C9orf72 haploinsufficiency does cause a defect in nucleocytoplasmic transport, one could expect an accumulation of RNA in the nucleus, as has been detected in HEK293 cells following DPR transfection (Freibaum et al., 2015). To test whether C9orf72 depletion causes an RNA export defect, one could perform RNA *in situ* hybridisation of C9orf72-knockdown HEK293 cells to measure RNA accumulation in the nucleus. To

investigate whether said RNA accumulation contributes to the increase in the number of Cajal bodies, coilin immunostaining may be performed to determine if there is a correlation between RNA accumulation and Cajal body numbers. Many of the studies into nucleocytoplasmic transport in ALS suggest disruption to the Ran-GTP gradient across the nuclear membrane is disrupted (Zhang et al., 2015). To test if there is a disruption to the nuclear to cytoplasmic ratio of Ran in C9orf72 depleted cells, one could perform immunofluorescence or nuclear fractionation of cells followed by immunoblot for Ran to measure the relative Ran proteins levels in the different subcellular localisations.

Earlier we discussed a mechanism for how C9orf72 haploinsufficiency could cause cell cycle arrest in HEK293 cells. This was based on the Cajal body and minor splicing results, both of which have links with cell-cycle regulation (Andrade et al., 1991; Carmo-Fonseca et al., 1993; König et al., 2007). However, this hypothesis is based on circumstantial evidence only and needs to be assessed further. One way of investigating this would be determining the frequency of wild-type versus C9orf72 knockout cells at different stages in the cell cycle. As discussed, the cyclin E protein localises to Cajal bodies in mid G_1 and exits during the S phase (Liu et al., 2000). Co-staining for coilin and the cyclin E protein after C9orf72 siRNA transfection or in the B2 C9orf72 knockout cells could therefore inform if cells are stalled at the G_1/S by an observable increase in the percentage of cells with coilin-cyclin E colocalisation. Similarly, DNA incorporation of BrdU is frequently implemented as a marker of DNA synthesis during the S phase of the cell cycle (Nowakowski et al., 1989). We also predict coilin in C9orf72 depleted cells to show a different phosphorylation pattern to coilin from wild-type cells. As coilin has 11 phosphorylation sites, it would prove difficult using phospho-site specific antibodies to track the post-translational modifications to all individual sites. Mass spectrometry analysis on immunoprecipitated coilin from C9orf72 knockout cells would be the most suitable option for investigating this further.

Further investigation is also required into the relevance of C9orf72 haploinsufficiency in ALS. Chapter 5 included interesting results in C9orf72-ALS iAstrocytes which displayed elevated numbers of Cajal bodies. Work is now ongoing to determine if the Cajal body phenotype is also present in iNeurons derived from the same patient fibroblasts. This result will hopefully answer some of the questions arising from the lack of effect of C9orf72 knockdown on Cajal bodies in primary neurons. If Cajal bodies are increased in the iNeurons, this would more strongly

suggest motor neurons in C9orf72-ALS patients may also exhibit the same phenotype.

8 Appendix

Gene	Protein	Uniprot protein function	Reference
ADCY10	Adenylate cyclase type 10	cAMP biosynthesis	In house
ANXA7	Annexin A7	Ca/GTP regulated exocytosis	Yeast two-hybrid
BANF1	Barrier-to-autointegration factor	Nuclear assembly; Chromatin organisation	In house
BEX1	Protein BEX1	Signalling adaptor molecule	Yeast two-hybrid
C9orf72	Chromosome 9 open reading frame 72	Autophagy; Trafficking	In house, Sullivan et al. 2016
CACTIN	Cactin	Transcriptional activator	Yeast two-hybrid
CAD	CAD protein	Pyrimidine pathway	In house; Blokhuis et al. 2016
CAND1	Cullin-associated NEDD8-dissociated protein 1	F-box protein exchange factor	In house
CBX3	Chromobox protein homolog 3	Chromatin organisation	In house
CENPF	Centromere protein F	Chromosome segregation; DNA synthesis regulation	In house
CXXC5	CXXC-type zinc finger protein 5	Transcription regulation; DNA damage response	Yeast two-hybrid
DDX21	Nucleolar RNA helicase 2	RNA helicase	In house
DDX50	ATP-dependent RNA helicase DDX50	RNA helicase	In house
DEK	Protein DEK	Chromatin organisation	In house
EEF1A1	Elongation factor 1- alpha 1	Transcription regulation; Chaperone mediated autophagy	In house
ERH	Enhancer of rudimentary homolog	Cell cycle	In house
ERI1	3’-5’ exoribonuclease 1	RNA exonuclease	Yeast two-hybrid
EXOSC10	Exosome component 10	3’-5’ exoribonuclease activity; Maturation of rRNA, snRNA and snoRNA	Yeast two-hybrid
FBL	Fibrallarin	rRNA processing; Nucleolar protein	In house
FUS	RNA-binding protein FUS	RNA processing	In house
HDGFRP2	Hepatoma-derived growth factor-related protein 2	mRNA processing co-factor; Cyclin D1 expression regulation	In house
HMGA1	High mobility group protein HMG-I/HMG-Y	Nucleosome phasing; Transcription regulation	In house
HNRNPA1	Heterogenous nuclear ribonucleoprotein A1	RNA processing	In house
HNRNPC	Heterogenous nuclear ribonucleoprotein C	RNA processing	In house
HNRNPF	Heterogenous nuclear ribonucleoprotein F	RNA processing	Blokhuis et al. 2016
HNRNPH	Heterogenous nuclear ribonucleoprotein H	RNA processing	Blokhuis et al. 2016

HNRNPK	Heterogenous nuclear ribonucleoprotein K	RNA processing	In house
HNRNPM	Heterogenous nuclear ribonucleoprotein M	RNA processing	In house
HNRNPU	Heterogenous nuclear ribonucleoprotein U	RNA processing	In house
HUWE1	E3 ubiquitin-protein ligase HUWE1	Protein ubiquitination	Sellier et al. 2016
ILF2/3	Interleukin enhancer-binding factor 2/3	IL2 transcription regulation	In house
KDM4A	Lysine-specific demethylase 4A	Transcription regulation; Histone demethylase	In house
KDM4B	Lysine-specific demethylase 4B	Transcription regulation; Histone demethylase	In house
MATR3	Matrin-3	Transcription regulation; RNA processing	In house
NCL	Nucleolin	Major nucleolar proteins; Pre-rRNA transcription; Ribosome assembly	In house, Blokhuis et al. 2016
NOTCH3	Neurogenic locus notch homolog protein 3	Transcriptional activator	In house
NPM1	Nucleophosmin	Ribsome biogenesis; Centromsome duplication	In house
NR1D2	Nuclear receptor subfamily 1 group D member 2	Circadian rhythm	Yeast two-hybrid
PA2G4	Proliferation-associated protein 2G4	Transcription regulation; rRNA processing	In house
PKM	Pyruvate kinase PKM	Glycolysis; Cell death induction	In house
PPP1CA	Serine/Threonine protein phosphatase PP1-alpha catalytic subunit	Protein phosphatase	Blokhuis et al. 2016
PRKAA1	5'AMP-activated protein kinase catalytic subunit alpha-1	Transcription regulation	Sellier et al. 2016
PSIP1	PC4 and SFRS1-interacting protein	RNA processing	In house
PSMA4	Proteasome subunit alpha type-4	Protein degradation	Sellier et al. 2016
PSMD8	26S proteasome non-ATPase regulatory subunit 8	Protein degradation	Sellier et al. 2016
PSMD10	26S proteasome non-ATPase regulatory subunit 10	Protein degradation	Sellier et al. 2016
PTBP1	Polypyrimidine tract-binding protein 1	RNA processing	In house
RAD54B	DNA repair and recombination protein RAD54B	DNA repair; Mitotic recombination	Yeast two-hybrid
RAN	GTP-binding nuclear protein Ran	Nucleocytoplasmic transport	In house
RB1CC1	RB1-inducible coilin-coil protein 1	Autophagy; Transcription regulation	In house; Sellier et al. 2016
RBMX	RNA-binding motif protein, X chromosome	Regulation of pre- and post- transcriptional processes	In house
RLIM	E3 ubiquitin-protein	Protein ubiquitination	Sellier et al. 2016

	ligase RLIM		
RNF126	E3 ubiquitin-protein ligase RNF126	Protein ubiquitination	Sellier et al. 2016
RPS3	40S ribosomal protein S3	DNA damage response; Transcription regulation; Neuronal apoptosis	In house
RSL1D1	Ribosomal L1 domain-containing protein 1	Cellular senescence; Apoptosis	In house
SENP6	Sentrin-specific protease 6	Protein desumoylation	Yeast two-hybrid
SETX	Probable helicase senataxin	RNA/DNA helicase; RNA metabolism	Sellier et al. 2016
SFPQ	Splicing factor, proline- and glutamine-rich	RNA processing; DNA repair	In house
SNX6	Sorting nexin-6	Intracellular trafficking	Sellier et al. 2016
SQSTM1	Sequestosome-1	Autophagy	Sellier et al. 2016
SRSF3	Serine/arginine rich splicing factor 3	RNA processing	In house
SRSF7	Serine/arginine rich splicing factor 7	RNA processing	In house
STK38	Serine/Threonine protein kinase 38	MAP3K1/2 signalling	In house
STUB1	E3 ubiquitin-protein ligase CHIP	Protein ubiquitination	Sellier et al. 2016
TARDBP	TAR DNA-binding protein 43	Transcription and splicing regulation	Sellier et al. 2016
TCP1	T-complex protein 1 subunit alpha	Molecular chaperone; Telomerase regulation	Yeast two-hybrid, Blokhuis et al. 2016
TJP2	Tight junction protein ZO-2	Tight junctions and adhere junctions	Yeast two-hybrid
TOX4	TOX high mobility group box family member 4	Chromatin structure; Cell cycle	Blokhuis et al. 2016
UACA	Uveal autoantigen with coiled-coil domains and ankyrin repeats	Apoptosis signalling	In house
UBR4	E3 ubiquitin-protein ligase UBR4	Protein ubiquitination	In house; Sellier et al. 2016
USP48/8	Ubiquitin carboxyl-terminal hydrolase 48/8	Protein deubiquitination	Yeast two-hybrid
WAC	WW domain-containing adaptor protein with coiled-coil	Transcription regulation; DNA damage response	Yeast two-hybrid
XRCC6	X-ray repair cross-complementing protein 6	DNA helicase; DNA damage response	In house
YBX1	Nuclease-sensitive element binding protein 1	RNA processing	In house
ZC3H14	Zinc finger CCCH domain-containing protein 14	Poly(A) RNA binding and regulation	In house
ZNF551	Zinc finger protein 551	Transcription regulation	In house
ZNF585A	Zinc finger protein 585A	Transcription regulation	Yeast two-hybrid

Table 8.1. C9orf72 nuclear interactors. Where 'In house' refers to mass spectrometry identification of C9orf72 protein interactors after immunoprecipitation of exogenous C9orf72 protein from HEK293 cell lysates. Performed by Dr Guillaume Hautbergue and Dr Matthew Walsh. Where 'Yeast two-hybrid' refers to the screen detailed in Section 4.2.3.

9 References

Abràmoff, M. D., Magalhães, P. J. and Ram, S. J. (2004). Image Processing with ImageJ. *Biophotonics International*.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *The EMBO journal*, 18 (20), pp.5789–5802.

Aguirre, N., Beal, M. F., Matson, W. R. and Bogdanov, M. B. (2005). Increased oxidative damage to DNA in an animal model of amyotrophic lateral sclerosis. *Free Radical Research*, 39 (4), pp.383–388.

Aizawa, H., Kimura, T., Hashimoto, K., Yahara, O., Okamoto, K. and Kikuchi, K. (2000). Basophilic cytoplasmic inclusions in a case of sporadic juvenile amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, 176 (2), pp.109–113.

Al-Sarraj, S., King, A., Troakes, C., Smith, B., Maekawa, S., Bodi, I., Rogelj, B., Al-Chalabi, A., Hortobagyi, T. and Shaw, C. E. (2011). p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. *Acta Neuropathologica*, 122 (6), pp.691–702.

Alami, N. H., Smith, R. B., Carrasco, M. A., Williams, L. A., Winborn, C. S., Han, S. S. W., Kiskinis, E., Winborn, B., Freibaum, B. D., Kanagaraj, A., et al. (2014). Axonal Transport of TDP-43 mRNA Granules Is Impaired by ALS-Causing Mutations. *Neuron*, 81 (3), pp.536–543.

Albo, F., Pieri, M. and Zona, C. (2004). Modulation of AMPA receptors in spinal motor neurons by the neuroprotective agent riluzole. *Journal of Neuroscience Research*, 78 (2), pp.200–207.

Alexander, G. M., Deitch, J. S., Seeburger, J. L., Del Valle, L. and Heiman-Patterson, T. D. (2000). Elevated cortical extracellular fluid glutamate in transgenic mice expressing human mutant (G93A) Cu/Zn superoxide dismutase. *Journal of Neurochemistry*, 74 (4), pp.1666–1673.

Alshikho, M. J., Zürcher, N. R., Loggia, M. L., Cernasov, P., Chonde, D. B., Izquierdo Garcia, D., Yasek, J. E., Akeju, O., Catana, C., Rosen, B. R., et al. (2016). Glial activation colocalizes with structural abnormalities in amyotrophic lateral sclerosis. *Neurology*, 87 (24), pp.2554–2561.

Andersen, P. M. (2006). Amyotrophic Lateral Sclerosis Associated with Mutations in the CuZn Superoxide Dismutase Gene. *Current Neurology and Neuroscience Reports*, 6, pp.37–46.

Andrade, L. E., Chan, E. K., Raska, I., Peebles, C. L., Roos, G. and Tan, E. M. (1991). Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. *The Journal of Experimental Medicine*, 173 (6), pp.1407–1419.

Andrade, L. E., Tan, E. M. and Chan, E. K. (1993). Immunocytochemical analysis of the coiled body in the cell cycle and during cell proliferation. *Proceedings of the National Academy of Sciences*, 90 (5), pp.1947–1951.

Aoki, Y., Manzano, R., Lee, Y., Dafinca, R., Aoki, M., Douglas, A. G. L., Varela, M. A., Sathyaprakash, C., Scaber, J., Barbagallo, P., et al. (2017). C9orf72 and RAB7L1 regulate vesicle trafficking in amyotrophic lateral sclerosis and frontotemporal dementia. *Brain*, 140 (4), pp.887–897.

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y., et al. (2006). TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochemical and Biophysical Research Communications*, 351 (3), pp.602–611.

Arnold, E. S., Ling, S.-C., Huelga, S. C., Lagier-Tourenne, C., Polymenidou, M., Ditsworth, D., Kordasiewicz, H. B., McAlonis-Downes, M., Platoshyn, O., Parone, P. A., et al. (2013). ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *Proceedings of the National Academy of Sciences*, 110 (8), pp.E736–E745.

Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul 3rd, J. W., Rademakers, R., et al. (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, 77 (4), pp.639–646.

Atanasio, A., Decman, V., White, D., Ramos, M., Ikiz, B., Lee, H.-C., Siao, C.-J., Brydges, S., LaRosa, E., Bai, Y., et al. (2016). C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production and glomerulonephropathy in mice. *Scientific Reports*, 6 (1), p.23204.

Baborie, A., Griffiths, T. D., Jaros, E., Perry, R., McKeith, I. G., Burn, D. J., Masuda-Suzukake, M., Hasegawa, M., Rollinson, S., Pickering-Brown, S., et al. (2015). Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in *C9ORF72* gene. *Neuropathology and Applied Neurobiology*, 41 (5), pp.601–612.

Baillat, D., Hakimi, M.-A., Näär, A. M., Shilatifard, A., Cooch, N. and Shiekhattar, R. (2005). Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. *Cell*, 123 (2), pp.265–276.

Barmada, S. J., Serio, A., Arjun, A., Bilican, B., Daub, A., Ando, D. M., Tsvetkov, A., Pleiss, M., Li, X., Peisach, D., et al. (2014). Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. *Nature Chemical Biology*, 10 (8), pp.677–685.

Barmada, S. J., Skibinski, G., Korb, E., Rao, E. J., Wu, J. Y. and Finkbeiner, S. (2010). Cytoplasmic Mislocalization of TDP-43 Is Toxic to Neurons and Enhanced by a Mutation Associated with Familial Amyotrophic Lateral Sclerosis. *Journal of Neuroscience*, 30 (2), pp.639–649.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129 (4), pp.823–837.

Bártová, E., Foltánková, V., Legartová, S., Sehnalová, P., Sorokin, D. V, Suchánková, J. and Kozubek, S. (2014). Coilin is rapidly recruited to UVA-induced DNA lesions and γ-radiation affects localized movement of Cajal bodies. *Nucleus*, 5 (3), pp.460–468.

Bäumer, D., Hilton, D., Paine, S. M. L., Turner, M. R., Lowe, J., Talbot, K. and Ansorge, O. (2010). Juvenile ALS with basophilic inclusions is a FUS proteinopathy with FUS mutations. *Neurology*, 75 (7), pp.611–618.

Bayer, T. S., Booth, L. N., Knudsen, S. M. and Ellington, A. D. (2005). Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA*, 11 (12), pp.1848–1857.

Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J. and Gygi, S. P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nature Biotechnology*, 24 (10), pp.1285–1292.

Bell, M., Schreiner, S., Damianov, A., Reddy, R. and Bindereif, A. (2002). p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *The EMBO Journal*, 21 (11), pp.2724–2735.

Bellini, M. and Gall, J. G. (1998). Coilin can form a complex with the U7 small nuclear ribonucleoprotein. *Molecular biology of the cell*, 9 (10), pp.2987–3001.

Bellini, M. and Gall, J. G. (1999). Coilin shuttles between the nucleus and cytoplasm in Xenopus oocytes. *Molecular biology of the cell*, 10 (10), pp.3425–3434.

Belzil, V. V, Bauer, P. O., Prudencio, M., Gendron, T. F., Stetler, C. T., Yan, I. K., Pregent, L., Daughrity, L., Baker, M. C., Rademakers, R., et al. (2013). Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathologica*, 126 (6), pp.895–905.

Berget, S. M., Moore, C. and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74 (8), pp.3171–3175.

Bhardwaj, A., Myers, M. P., Buratti, E. and Baralle, F. E. (2013). Characterizing TDP-43 interaction with its RNA targets. *Nucleic Acids Research*, 41 (9), pp.5062–5074.

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, 72 (1), pp.291–336.

Blokhuis, A. M., Koppers, M., Groen, E. J. N., van den Heuvel, D. M. A., Dini Modigliani, S., Anink, J. J., Fumoto, K., van Diggelen, F., Snelting, A., Sodaar, P., et al. (2016). Comparative interactomics analysis of different ALS-associated proteins identifies converging molecular pathways. *Acta neuropathologica*, 132 (2), pp.175–196.

Boeynaems, S., Bogaert, E., Michiels, E., Gijselinck, I., Sieben, A., Jovičić, A., De Baets, G., Scheveneels, W., Steyaert, J., Cuijt, I., et al. (2016). Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Scientific Reports*, 6 (1), p.20877.

Bogdanov, M., Brown, R. H., Matson, W., Smart, R., Hayden, D., O'Donnell, H., Flint Beal, M. and Cudkowicz, M. (2000). Increased oxidative damage to DNA in ALS patients. *Free Radical Biology & Medicine*, 29 (7), pp.652–658.

Bohmann, K., Ferreira, J. A. and Lamond, A. I. (1995). Mutational analysis of p80 coilin indicates a functional interaction between coiled bodies and the nucleolus. *Journal of Cellular Biochemistry*, 131 (4), pp.817–831.

Boisvert, F. M., Cote, J., Boulanger, M. C., Cleroux, P., Bachand, F., Autexier, C. and Richard, S. (2002). Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. *The Journal of Cell Biology*, 159 (6), pp.957–969.

Bosco, D. A., Lemay, N., Ko, H. K., Zhou, H., Burke, C., Kwiatkowski, T. J., Sapp, P., McKenna-Yasek, D., Brown, R. H. and Hayward, L. J. (2010). Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. *Human Molecular Genetics*, 19 (21), pp.4160–4175.

Boulisfane, N., Choleza, M., Rage, F., Neel, H., Soret, J. and Bordonné, R. (2011). Impaired minor tri-snRNP assembly generates differential splicing defects of U12-type introns in lymphoblasts derived from a type I SMA patient. *Human Molecular Genetics*, 20 (4), pp.641–648.

Boulon, S., Verheggen, C., Jady, B. E., Girard, C., Pescia, C., Paul, C., Ospina, J. K., Kiss, T., Matera, A. G., Bordonné, R., et al. (2004). PHAX and CRM1 Are Required Sequentially to Transport U3 snoRNA to Nucleoli. *Molecular Cell*, 16 (5), pp.777–787.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72 (1–2), pp.248–254.

Brahms, H., Meheus, L., de Brabandere, V., Fischer, U. and Lührmann, R. (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA*, 7 (11), pp.1531–1542.

Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L. and Lührmann, R. (2000). The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *The Journal of biological chemistry*, 275 (22), pp.17122–17129.

Brenner, D., Müller, K., Wieland, T., Weydt, P., Böhm, S., Lulé, D., Hübers, A., Neuwirth, C., Weber, M., Borck, G., et al. (2016). NEK1 mutations in familial amyotrophic lateral sclerosis. *Brain*, 139 (5), pp.e28–e28.

Brenner, D., Yilmaz, R., Müller, K., Grehl, T., Petri, S., Meyer, T., Grosskreutz, J., Weydt, P., Ruf, W., Neuwirth, C., et al. (2018). Hot-spot KIF5A mutations cause familial ALS. *Brain*, p.awx370.

Brettschneider, J., Del Tredici, K., Toledo, J. B., Robinson, J. L., Irwin, D. J., Grossman, M., Suh, E., Van Deerlin, V. M., Wood, E. M., Baek, Y., et al. (2013). Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. *Annals of Neurology*, 74 (1), pp.20–38.

Broome, H. J. and Hebert, M. D. (2012). In vitro RNase and nucleic acid binding activities implicate coilin in U snRNA processing. *PLoS One*, 7 (4), p.e36300.

Bruijn, L. I., Houseweart, M. K., Kato, S., Anderson, K. L., Anderson, S. D., Ohama, E., Reaume, A. G., Scott, R. W. and Cleveland, D. W. (1998). Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1. *Science*, 281 (5384), pp.1851–1854.

Buratti, E. and Baralle, F. E. (2001). Characterization and Functional Implications of the RNA Binding Properties of Nuclear Factor TDP-43, a Novel Splicing Regulator of *CFTR* Exon 9. *Journal of Biological Chemistry*, 276 (39), pp.36337–36343.

Buratti, E., Brindisi, A., Giombi, M., Tisminetzky, S., Ayala, Y. M. and Baralle, F. E. (2005). TDP-43 Binds Heterogeneous Nuclear Ribonucleoprotein A/B through Its C-terminal Tail. *Journal of Biological Chemistry*, 280 (45), pp.37572–37584.

Buratti, E., De Conti, L., Stuani, C., Romano, M., Baralle, M. and Baralle, F. (2010). Nuclear factor TDP-43 can affect selected microRNA levels. *FEBS Journal*, 277 (10), pp.2268–2281.

Burge, C. B., Burge, C. B., Tuschl, T., Sharp, P. A., Burge, C. B., Tuschl, T. and Sharp, P. A. (1999). Splicing of Precursors to mRNAs by the Spliceosomes. *Nucleic Acids Research*.

Burge, C. B., Padgett, R. A. and Sharp, P. A. (1998). Evolutionary fates and origins of U12-type introns. *Molecular Cell*, 2 (6), pp.773–785.

Burghes, A. H. M. and Beattie, C. E. (2009). Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nature Reviews Neuroscience*, 10 (8), pp.597–609.

Burguete, A. S., Almeida, S., Gao, F.-B., Kalb, R., Akins, M. R. and Bonini, N. M. (2015). GGGGCC microsatellite RNA is neuritically localized, induces branching defects, and perturbs transport granule function. *eLife*, 4, p.e08881.

Burset, M., Seledtsov, I. A. and Solovyev, V. V. (2001). SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Research*, 29 (1), pp.255–259.

Bylund, L., Kytölä, S., Lui, W.-O., Larsson, C. and Weber, G. (2004). Analysis of the cytogenetic stability of the human embryonal kidney cell line 293 by cytogenetic and STR profiling approaches. *Cytogenetic and genome research*, 106 (1), pp.28–32.

Cáceres, J. F., Stamm, S., Helfman, D. M. and Krainer, A. R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science*, 265 (5179), pp.1706–1709.

Cajal, S. R. (1903). Un sencillo metodo de coloracion selectiva del reticulo protoplasmico y sus efectos en los diversos organos nerviosos de vertebrados e invertebrados. *Trab Lab Investig Biol Univ Madr*, 2, pp.129–221.

Cantarero, L., Sanz-García, M., Vinograd-Byk, H., Renbaum, P., Levy-Lahad, E. and Lazo, P. A. (2015). VRK1 regulates Cajal body dynamics and protects coilin from proteasomal degradation in cell cycle. *Scientific Reports*, 5 (1), p.10543.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298 (5595), pp.1039–1043.

Carlsson, L., Nyström, L. E., Sundkvist, I., Markey, F. and Lindberg, U. (1977). Actin polymerizability is influenced by profilin, a low molecular weight protein in non-muscle cells. *Journal of Molecular Biology*, 115 (3), pp.465–483.

Carmo-Fonseca, M., Ferreira, J. and Lamond, A. I. (1993). Assembly of snRNP-containing coiled bodies is regulated in interphase and mitosis--evidence that the coiled body is a kinetic nuclear structure. *The Journal of Cell Biology*, 120 (4), pp.841–852.

Carmo-Fonseca, M., Pepperkok, R., Carvalho, M. T. and Lamond, A. I. (1992). Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *The Journal of cell biology*, 117 (1), pp.1–14.

Carrero, Z. I., Velma, V., Douglas, H. E. and Hebert, M. D. (2011). Coilin Phosphomutants Disrupt Cajal Body Formation, Reduce Cell Proliferation and Produce a Distinct Coilin Degradation Product. Bridger, J. M. (Ed). *PLoS One*, 6 (10), p.e25743.

Carvalho, T., Almeida, F., Calapez, A., Lafarga, M., Berciano, M. T. and Carmo-Fonseca, M. (1999). The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. *The Journal of cell biology*, 147 (4), pp.715–728.

Castillo, E. F., Dekonenko, A., Arko-Mensah, J., Mandell, M. A., Dupont, N., Jiang, S., Delgado-Vargas, M., Timmins, G. S., Bhattacharya, D., Yang, H., et al. (2012). Autophagy protects against active tuberculosis by suppressing bacterial burden and inflammation. *Proceedings of the National Academy of Sciences*, 109 (46), pp.E3168–E3176.

Chan, E. K., Takano, S., Andrade, L. E., Hamel, J. C. and Matera, A. G. (1994). Structure, expression and chromosomal localization of human p80-coilin gene. *Nucleic Acids Research*, 22 (21), pp.4462–4469.

Chang, Y.-J., Jeng, U.-S., Chiang, Y.-L., Hwang, I.-S. and Chen, Y.-R. (2016). The Glycine-Alanine Dipeptide Repeat from C9orf72 Hexanucleotide Expansions Forms Toxic Amyloids Possessing Cell-to-Cell Transmission Properties. *The Journal of biological chemistry*, 291 (10), pp.4903–4911.

Chen, L.-C. and Casadevall, A. (1999). Labeling of Proteins with [35S]Methionine and/or [35S]Cysteine in the Absence of Cells. *Analytical Biochemistry*, 269 (1), pp.179–188.

Chen, Y., Craigen, W. J. and Riley, D. J. (2009). Nek1 regulates cell death and mitochondrial membrane permeability through phosphorylation of VDAC1. *Cell Cycle*, 8 (2), pp.257–267.

Chen, Y. Z., Bennett, C. L., Huynh, H. M., Blair, I. P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M. L., Rabin, B. A., et al. (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *American journal of human genetics*, 74 (6), pp.1128–1135.

Cheng, W., Wang, S., Mestre, A. A., Fu, C., Makarem, A., Xian, F., Hayes, L. R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., et al. (2018). C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nature Communications*, 9 (1), p.51.

Cherfils, J. and Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiological Reviews*, 93 (1), pp.269–309.

Chew, J., Gendron, T. F., Prudencio, M., Sasaguri, H., Zhang, Y.-J., Castanedes-Casey, M., Lee, C. W., Jansen-West, K., Kurti, A., Murray, M. E., et al. (2015). Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science*, 348 (6239), pp.1151–1154.

Chiò, A., Restagno, G., Brunetti, M., Ossola, I., Calvo, A., Mora, G., Sabatelli, M., Monsurrò, M. R., Battistini, S., Mandrioli, J., et al. (2009). Two Italian kindreds with familial amyotrophic lateral sclerosis due to FUS mutation. *Neurobiology of Aging*, 30 (8), pp.1272–1275.

Chow, C. Y., Landers, J. E., Bergren, S. K., Sapp, P. C., Grant, A. E., Jones, J. M., Everett, L., Lenk, G. M., McKenna-Yasek, D. M., Weisman, L. S., et al. (2009). Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. *American journal of human genetics*, 84 (1), pp.85–88.

Chow, L. T., Gelinas, R. E., Broker, T. R. and Roberts, R. J. (1977). An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. *Cell*, 12 (1), pp.1–8.

Cioce, M., Boulon, S., Matera, A. G. and Lamond, A. I. (2006). UV-induced fragmentation of Cajal bodies. *The Journal of Cell Biology*, 175 (3), pp.401–413.

Cioce, M. and Lamond, A. I. (2005). CAJAL BODIES: A Long History of Discovery. *Annual Review of Cell and Developmental Biology*, 21 (1), pp.105–131.

Cirulli, E. T., Lasseigne, B. N., Petrovski, S., Sapp, P. C., Dion, P. A., Leblond, C. S., Couthouis, J., Lu, Y.-F., Wang, Q., Krueger, B. J., et al. (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science*, 347 (6229), pp.1436–1441.

Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A. and Kabashi, E. (2013). Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Annals of Neurology*, 74 (2), pp.180–187.

Cleary, J. D. and Ranum, L. P. (2013). Repeat-associated non-ATG (RAN) translation in neurological disease. *Human Molecular Genetics*, 22 (R1), pp.R45-51.

Clement, A. M., Nguyen, M. D., Roberts, E. A., Garcia, M. L., Boillée, S., Rule, M., McMahon, A. P., Doucette, W., Siwek, D., Ferrante, R. J., et al. (2003). Wild-Type Nonneuronal Cells Extend Survival of SOD1 Mutant Motor Neurons in ALS Mice. *Science*, 302 (5642), pp.113–117.

Cleveland, D. W. and Williamson, T. L. (1999). Slowing of axonal transport is a very early event in the toxicity of ALS-linked SOD1 mutants to motor neurons. *Nature Neuroscience*, 2 (1), pp.50–56.

Collard, J.-F., Côté, F. and Julien, J.-P. (1995). Defective axonal transport in a transgenic mouse model of amyotrophic lateral sclerosis. *Nature*, 375 (6526), pp.61–64.

Colombrita, C., Onesto, E., Megiorni, F., Pizzuti, A., Baralle, F. E., Buratti, E., Silani, V. and Ratti, A. (2012). TDP-43 and FUS RNA-binding Proteins Bind Distinct Sets of Cytoplasmic Messenger RNAs and Differently Regulate Their Post-transcriptional Fate in Motoneuron-like Cells. *Journal of Biological Chemistry*, 287 (19), pp.15635–15647.

Conlon, E. G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N. A. and Manley, J. L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *eLife*, 5.

Cooper-Knock, J., Bury, J. J., Heath, P. R., Wyles, M., Higginbottom, A., Gelsthorpe, C., Highley, J. R., Hautbergue, G., Rattray, M., Kirby, J., et al. (2015a). C9ORF72 GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis. Cai, H. (Ed). *PLoS One*, 10 (5), p.e0127376.

Cooper-Knock, J., Higginbottom, A., Stopford, M. J., Highley, J. R., Ince, P. G., Wharton, S. B., Pickering-Brown, S., Kirby, J., Hautbergue, G. M. and Shaw, P. J. (2015b). Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathologica*, 130 (1), pp.63–75.

Cooper-Knock, J., Walsh, M. J., Higginbottom, A., Robin Highley, J., Dickman, M. J., Edbauer, D., Ince, P. G., Wharton, S. B., Wilson, S. A., Kirby, J., et al. (2014). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*, 137 (Pt 7), pp.2040–2051.

Cooper, T. A. (2005). Use of minigene systems to dissect alternative splicing elements. *Methods*, 37 (4), pp.331–340.

Cuervo, A. M., Bergamini, E., Brunk, U. T., Dröge, W., Ffrench, M. and Terman, A. (2005). Autophagy and aging: the importance of maintaining 'clean' cells. *Autophagy*, 1 (3), pp.131–140.

D'Angelo, M. A., Raices, M., Panowski, S. H. and Hetzer, M. W. (2009). Age-Dependent Deterioration of Nuclear Pore Complexes Causes a Loss of Nuclear Integrity in Postmitotic Cells. *Cell*, 136 (2), pp.284–295.

Darzacq, X., Jády, B. E., Verheggen, C., Kiss, A. M., Bertrand, E. and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *The EMBO Journal*, 21 (11), pp.2746–2756.

Dasso, M. (1995). The role of the Ran GTPase pathway in cell cycle control and interphase nuclear functions. In: *Progress in Cell Cycle Research*. pp.163–172.

Daughters, R. S., Tuttle, D. L., Gao, W., Ikeda, Y., Moseley, M. L., Ebner, T. J., Swanson, M. S. and Ranum, L. P. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genetics*, 5 (8), p.e1000600.

Davidson, Y. S., Barker, H., Robinson, A. C., Thompson, J. C., Harris, J., Troakes, C., Smith, B., Al-Saraj, S., Shaw, C., Rollinson, S., et al. (2014). Brain distribution of dipeptide repeat proteins in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathologica Communications*, 2 (1), p.70.

Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H. and Housman, D. E. (1997). Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proceedings of the National Academy of Sciences*, 94 (14), pp.7388–7393.

De Vos, K. J., Chapman, A. L., Tennant, M. E., Manser, C., Tudor, E. L., Lau, K.-F., Brownlees, J., Ackerley, S., Shaw, P. J., McLoughlin, D. M., et al. (2007). Familial amyotrophic lateral sclerosis-linked SOD1 mutants perturb fast axonal transport to reduce axonal mitochondria content. *Human Molecular Genetics*, 16 (22), pp.2720–2728.

Decatur, W. A. and Fournier, M. J. (2003). RNA-guided nucleotide modification of ribosomal and other RNAs. *The Journal of biological chemistry*, 278 (2), pp.695–698.

DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, 72 (2), pp.245–256.

Deng, H. X., Chen, W., Hong, S. T., Boycott, K. M., Gorrie, G. H., Siddique, N., Yang, Y., Fecto, F., Shi, Y., Zhai, H., et al. (2011). Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature*, 477 (7363), pp.211–215.

Deng, H. X., Hentati, A., Tainer, J. A., Iqbal, Z., Cayabyab, A., Hung, W. Y., Getzoff, E. D., Hu, P., Herzfeldt, B., Roos, R. P., et al. (1993). Amyotrophic lateral sclerosis and structural defects in Cu,Zn superoxide dismutase. *Science*, 261 (5124), pp.1047–1051.

Deryusheva, S. and Gall, J. G. (2004). Dynamics of coilin in Cajal bodies of the Xenopus germinal vesicle. *Proceedings of the National Academy of Sciences*, 101 (14), pp.4810–4814.

Dietrich, R. C., Fuller, J. D. and Padgett, R. A. (2005). A mutational analysis of U12-dependent splice site dinucleotides. *RNA*, 11 (9), pp.1430–1440.

Dietrich, R. C., Incorvaia, R. and Padgett, R. A. (1997). Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Molecular Cell*, 1 (1), pp.151–160.

DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P. and Aronin, N. (1997). Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science*, 277 (5334), pp.1990–1993.

Doble, A. (1996). The pharmacology and mechanism of action of riluzole. *Neurology*, 47 (6 Suppl 4), pp.S233-41.

Doble, A. (1999). The role of excitotoxicity in neurodegenerative disease: implications for therapy. *Pharmacology & therapeutics*, 81 (3), pp.163–221.

Dols-Icardo, O., Garcia-Redondo, A., Rojas-Garcia, R., Sanchez-Valle, R., Noguera, A., Gomez-Tortosa, E., Pastor, P., Hernandez, I., Esteban-Perez, J., Suarez-Calvet, M., et al. (2014). Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. *Human Molecular Genetics*, 23 (3), pp.749–754.

Donnelly, C. J., Zhang, P. W., Pham, J. T., Haeusler, A. R., Mistry, N. A., Vidensky, S., Daley, E. L., Poth, E. M., Hoover, B., Fines, D. M., et al. (2013). RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron*, 80 (2), pp.415–428.

Dormann, D., Rodde, R., Edbauer, D., Bentmann, E., Fischer, I., Hruscha, A., Than, M. E., Mackenzie, I. R. A., Capell, A., Schmid, B., et al. (2010). ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *The EMBO Journal*, 29 (16), pp.2841–2857.

Drepper, C., Herrmann, T., Wessig, C., Beck, M. and Sendtner, M. (2011). C-terminal FUS/TLS mutations in familial and sporadic ALS in Germany. *Neurobiology of Aging*, 32 (3), p.548.e1-548.e4.

Dundr, M., Hebert, M. D., Karpova, T. S., Stanek, D., Xu, H., Shpargel, K. B., Meier, U. T., Neugebauer, K. M., Matera, A. G. and Misteli, T. (2004). In vivo kinetics of Cajal body components. *The Journal of cell biology*, 164 (6), pp.831–842.

Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D. and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. *Journal of Biological Methods*, 1 (2), p.10.

Enwerem, I. I., Wu, G., Yu, Y. T. and Hebert, M. D. (2015). Cajal Body Proteins Differentially Affect the Processing of Box C/D scaRNPs. Bardoni, B. (Ed). *PLoS One*, 10 (4), p.e0122348.

Epstein, F. H., Lipton, S. A. and Rosenberg, P. A. (1994). Excitatory Amino Acids as a Final Common Pathway for Neurologic Disorders. *New England Journal of Medicine*, 330 (9), pp.613–622.

Fang, X., Lin, H., Wang, X., Zuo, Q., Qin, J. and Zhang, P. (2015). The NEK1 interactor, C21ORF2, is required for efficient DNA damage repair. *Acta Biochimica et Biophysica Sinica*, 47 (10), pp.834–841.

Farg, M. A., Sundaramoorthy, V., Sultana, J. M., Yang, S., Atkinson, R. A., Levina, V., Halloran, M. A., Gleeson, P. A., Blair, I. P., Soo, K. Y., et al. (2014). C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Human Molecular Genetics*, 23 (13), pp.3579–3595.

Fay, M. M., Anderson, P. J. and Ivanov, P. (2017). ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions In Vitro and in Cells. *Cell reports*, 21 (12), pp.3573–3584.

Fecto, F., Yan, J., Vemula, S. P., Liu, E., Yang, Y., Chen, W., Zheng, J. G., Shi, Y., Siddique, N., Arrat, H., et al. (2011). SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Archives of Neurology*, 68 (11), pp.1440–1446.

Ferrante, R. J., Browne, S. E., Shinobu, L. A., Bowling, A. C., Baik, M. J., MacGarvey, U., Kowall, N. W., Brown, R. H. and Beal, M. F. (1997). Evidence of increased oxidative damage in both sporadic and familial amyotrophic lateral sclerosis. *Journal of Neurochemistry*, 69 (5), pp.2064–2074.

Ferri, A., Cozzolino, M., Crosio, C., Nencini, M., Casciati, A., Gralla, E. B., Rotilio, G., Valentine, J. S. and Carri, M. T. (2006). Familial ALS-superoxide dismutases associate with mitochondria and shift their redox potentials. *Proceedings of the National Academy of Sciences*, 103 (37), pp.13860–13865.

Figlewicz, D. A., Krizus, A., Martinoli, M. G., Meininger, V., Dib, M., Rouleau, G. A. and Julien, J. P. (1994). Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. *Human Molecular Geneticsolecular genetics*, 3 (10), pp.1757–1761.

Fischer, U. and Lührmann, R. (1990). An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science*, 249 (4970), pp.786–790.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T. and Lührmann, R. (1993). Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *The EMBO journal*, 12 (2), pp.573–583.

Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I. W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, 90 (6), pp.1051–1060.

Fratta, P., Mizielinska, S., Nicoll, A. J., Zloh, M., Fisher, E. M. C., Parkinson, G. and Isaacs, A. M. (2012). C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Scientific Reports*, 2, p.1016.

Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., Kim, N. C., Almeida, S., Lee, K.-H., Badders, N., Valentine, M., Miller, B. L., Wong, P. C., et al. (2015). GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature*, 525 (7567), pp.129–133.

Frey, M. R., Bailey, A. D., Weiner, A. M. and Matera, A. G. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. *Current Biology*, 9 (3), pp.126–135.

Frey, M. R. and Matera, A. G. (1995). Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 92 (13), pp.5915–5919.

Fujii, R., Okabe, S., Urushido, T., Inoue, K., Yoshimura, A., Tachibana, T., Nishikawa, T., Hicks, G. G. and Takumi, T. (2005). The RNA Binding Protein TLS Is Translocated to Dendritic Spines by mGluR5 Activation and Regulates Spine Morphology. *Current Biology*, 15 (6), pp.587–593.

Fujita, K., Ito, H., Nakano, S., Kinoshita, Y., Wate, R. and Kusaka, H. (2008). Immunohistochemical identification of messenger RNA-related proteins in basophilic inclusions of adult-onset atypical motor neuron disease. *Acta Neuropathologica*, 116 (4), pp.439–445.

Gal, J., Zhang, J., Kwinter, D. M., Zhai, J., Jia, H., Jia, J. and Zhu, H. (2011). Nuclear localization sequence of FUS and induction of stress granules by ALS mutants. *Neurobiology of Aging*, 32 (12), p.2323.e27-2323.e40.

Ganot, P., Jády, B. E., Bortolin, M. L., Darzacq, X. and Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Molecular and Cellular Biology*, 19 (10), pp.6906–6917.

Gao, K., Masuda, A., Matsuura, T. and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. *Nucleic Acids Research*, 36 (7), pp.2257–2267.

Garcia, E. L., Wen, Y., Praveen, K. and Matera, A. G. (2016). Transcriptomic comparison of Drosophila snRNP biogenesis mutants reveals mutant-specific changes in pre-mRNA processing: implications for spinal muscular atrophy. *RNA (New York, N.Y.)*, 22 (8), pp.1215–1227.

Gendron, T. F., Bieniek, K. F., Zhang, Y.-J., Jansen-West, K., Ash, P. E. A., Caulfield, T., Daughrity, L., Dunmore, J. H., Castanedes-Casey, M., Chew, J., et al. (2013). Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathologica*, 126 (6), pp.829–844.

Gerety, S. S. and Wilkinson, D. G. (2011). Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development. *Developmental Biology*, 350 (2), pp.279–289.

Gertz, B., Wong, M. and Martin, L. J. (2012). Nuclear Localization of Human SOD1 and Mutant SOD1-Specific Disruption of Survival Motor Neuron Protein Complex in Transgenic Amyotrophic Lateral Sclerosis Mice. *Journal of Neuropathology & Experimental Neurology*, 71 (2), pp.162–177.

Ghiasi, P., Hosseinkhani, S., Noori, A., Nafissi, S. and Khajeh, K. (2012). Mitochondrial complex I deficiency and ATP/ADP ratio in lymphocytes of amyotrophic lateral sclerosis patients. *Neurological Research*, 34 (3), pp.297–303.

Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S., et al. (2012). A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurology*, 11 (1), pp.54–65.

Gilder, A. S., Do, P. M., Carrero, Z. I., Cosman, A. M., Broome, H. J., Velma, V., Martinez, L. A. and Hebert, M. D. (2011). Coilin participates in the suppression of RNA polymerase I in response to cisplatin-induced DNA damage. *Molecular biology of the cell*, 22 (7), pp.1070–1079.

Gomez-Deza, J., Lee, Y., Troakes, C., Nolan, M., Al-Sarraj, S., Gallo, J.-M. and Shaw, C. E. (2015). Dipeptide repeat protein inclusions are rare in the spinal cord and almost absent from motor neurons in C9ORF72 mutant amyotrophic lateral sclerosis and are unlikely to cause their degeneration. *Acta Neuropathologica Communications*, 3 (1), p.38.

Görlich, D. and Mattaj, I. W. (1996). Nucleocytoplasmic transport. *Science (New York, N.Y.)*, 271 (5255), pp.1513–1518.

Greenway, M. J., Andersen, P. M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., et al. (2006). ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nature Genetics*, 38 (4), pp.411–413.

Greider, C. W. and Blackburn, E. H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature*, 337 (6205), pp.331–337.

Grigg, J. C., Shumayrikh, N. and Sen, D. (2014). G-quadruplex structures formed by expanded hexanucleotide repeat RNA and DNA from the neurodegenerative disease-linked C9orf72 gene efficiently sequester and activate heme. *PLoS One*, 9 (9), p.e106449.

Gu, J., Chen, F., Iqbal, K., Gong, C.-X., Wang, X. and Liu, F. (2017). Transactive response DNA-binding protein 43 (TDP-43) regulates alternative splicing of tau exon 10: Implications for the pathogenesis of tauopathies. *The Journal of biological chemistry*, 292 (25), pp.10600–10612.

Guo, W., Chen, Y., Zhou, X., Kar, A., Ray, P., Chen, X., Rao, E. J., Yang, M., Ye, H., Zhu, L., et al. (2011). An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nature Structural & Molecular Biology*, 18 (7), pp.822–830.

Gutierrez, M. G., Master, S. S., Singh, S. B., Taylor, G. A., Colombo, M. I. and Deretic, V. (2004). Autophagy Is a Defense Mechanism Inhibiting BCG and Mycobacterium tuberculosis Survival in Infected Macrophages. *Cell*, 119 (6), pp.753–766.

Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, 507 (7491), pp.195–200.

Haidet-Phillips, A. M., Hester, M. E., Miranda, C. J., Meyer, K., Braun, L., Frakes, A., Song, S., Likhite, S., Murtha, M. J., Foust, K. D., et al. (2011). Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. *Nature Biotechnology*, 29 (9), pp.824–828.

Hall, E. D., Oostveen, J. A. and Gurney, M. E. (1998). Relationship of microglial and astrocytic activation to disease onset and progression in a transgenic model of familial ALS. *Glia*, 23 (3), pp.249–256.

Hall, S. L. and Padgett, R. A. (1996). Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science*, 271 (5256), pp.1716–1718.

Hallais, M., Pontvianne, F., Andersen, P. R., Clerici, M., Lener, D., Benbahouche, N. E. H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al. (2013). CBC–ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. *Nature Structural & Molecular Biology*, 20 (12), pp.1358–1366.

Hao, L. thi, Fuller, H. R., Lam, L. T., Le, T. T., Burghes, A. H. and Morris, G. E. (2007). Absence of gemin5 from SMN complexes in nuclear Cajal bodies. *BMC Cell Biology*, 8 (1), p.28.

Hardin, J. H., Spicer, S. S. and Greene, W. B. (1969). The paranucleolar structure, accessory body of cajal, sex chromatin, and related structures in nuclei of rat trigeminal neurons: A cytochemical and ultrastructural study. *The Anatomical Record*, 164 (4), pp.403–431.

Hautbergue, G. M., Castelli, L. M., Ferraiuolo, L., Sanchez-Martinez, A., Cooper-Knock, J., Higginbottom, A., Lin, Y.-H., Bauer, C. S., Dodd, J. E., Myszczynska, M. A., et al. (2017). SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nature Communications*, 8, p.16063.

Hearst, S. M., Gilder, A. S., Negi, S. S., Davis, M. D., George, E. M., Whittom, A. A., Toyota, C. G., Husedzinovic, A., Gruss, O. J. and Hebert, M. D. (2009). Cajal-body formation correlates with differential coilin phosphorylation in primary and transformed cell lines. *Journal of Cell Science*, 122 (11), pp.1872–1881.

Hebert, M. D. and Matera, A. G. (2000). Self-association of coilin reveals a common theme in nuclear body localization. *Molecular biology of the cell*, 11 (12), pp.4159–4171.

Hebert, M. D., Shpargel, K. B., Ospina, J. K., Tucker, K. E. and Matera, A. G. (2002). Coilin methylation regulates nuclear body formation. *Developmental cell*, 3 (3), pp.329–337.

Hebert, M. D., Szymczyk, P. W., Shpargel, K. B. and Matera, A. G. (2001). Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. *Genes & Development*, 15 (20), pp.2720–2729.

Henkel, J. S., Beers, D. R., Siklós, L. and Appel, S. H. (2006). The chemokine MCP-1 and the dendritic and myeloid cells it attracts are increased in the mSOD1 mouse model of ALS. *Molecular and Cellular Neuroscience*, 31 (3), pp.427–437.

Henkel, J. S., Engelhardt, J. I., Siklós, L., Simpson, E. P., Kim, S. H., Pan, T., Goodman, J. C., Siddique, T., Beers, D. R. and Appel, S. H. (2004). Presence of dendritic cells, MCP-1, and activated microglia/macrophages in amyotrophic lateral sclerosis spinal cord tissue. *Annals of Neurology*, 55 (2), pp.221–235.

Hernandez, N. (2001). Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *The Journal of biological chemistry*, 276 (29), pp.26733–26736.

Highley, J. R., Kirby, J., Jansweijer, J. A., Webb, P. S., Hewamadduma, C. A., Heath, P. R., Higginbottom, A., Raman, R., Ferraiuolo, L., Cooper-Knock, J., et al. (2014). Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones. *Neuropathology & Applied Neurobiology*, 40 (6), pp.670–685.

Hirano, A., Donnenfeld, H., Sasaki, S. and Nakano, I. (1984). Fine structural observations of neurofilamentous changes in amyotrophic lateral sclerosis. *Journal of Neuropathology & Experimental Neurology*, 43 (5), pp.461–470.

Hodges, J. R. (2001). Frontotemporal dementia (Pick's disease): clinical features and assessment. *Neurology*, 56 (11 Suppl 4), pp.S6-10.

Hong, K., Li, Y., Duan, W., Guo, Y., Jiang, H., Li, W. and Li, C. (2012). Full-length TDP-43 and its C-terminal fragments activate mitophagy in NSC34 cell line. *Neuroscience Letters*, 530 (2), pp.144–149.

Hosler, B. A., Siddique, T., Sapp, P. C., Sailor, W., Huang, M. C., Hossain, A., Daube, J. R., Nance, M., Fan, C., Kaplan, J., et al. (2000). Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22. *JAMA*, 284 (13), pp.1664–1669.

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. and Lührmann, R. (1998). Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *The EMBO journal*, 17 (14), pp.4114–4126.

Hug, N., Longman, D. and Cáceres, J. F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research*, 44 (4), pp.1483–1495.

Hutton, M., Lendon, C. L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., et al. (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*, 393 (6686), pp.702–705.

Iguchi, Y., Katsuno, M., Niwa, J., Takagi, S., Ishigaki, S., Ikenaka, K., Kawai, K., Watanabe, H., Yamanaka, K., Takahashi, R., et al. (2013). Loss of TDP-43 causes age-dependent progressive motor neuron degeneration. *Brain*, 136 (5), pp.1371–1382.

Ito, H. (2014). Basophilic inclusions and neuronal intermediate filament inclusions in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neuropathology*, 34 (6), pp.589–595.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, 78 (4), pp.657–668.

Jackson, I. J. (1991). A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Research*, 19 (14), pp.3795–3798.

Jady, B. E., Bertrand, E. and Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *The Journal of cell biology*, 164 (5), pp.647–652.

Jady, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E. and Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *The EMBO journal*, 22 (8), pp.1878–1888.

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.

Jady, B. E., Richard, P., Bertrand, E. and Kiss, T. (2005). Cell Cycle-dependent Recruitment of Telomerase RNA and Cajal Bodies to Human Telomeres. *Molecular Biology of the Cell*, 17 (2), pp.944–954.

Ji, Y. J., Ugolino, J., Brady, N. R., Hamacher-Brady, A. and Wang, J. (2017). Systemic deregulation of autophagy upon loss of ALS- and FTD-linked C9orf72. *Autophagy*, 13 (7), pp.1254–1255.

Jiang, J., Zhu, Q., Gendron, T. F., Saberi, S., McAlonis-Downes, M., Seelman, A., Stauffer, J. E., Jafar-nejad, P., Drenner, K., Schulte, D., et al. (2016). Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. *Neuron*, 90 (3), pp.535–550.

Johnson, B. S., Snead, D., Lee, J. J., McCaffery, J. M., Shorter, J. and Gitler, A. D. (2009). TDP-43 Is Intrinsically Aggregation-prone, and Amyotrophic Lateral Sclerosis-linked Mutations Accelerate Aggregation and Increase Toxicity. *Journal of Biological Chemistry*, 284 (30), pp.20329–20339.

Johnson, J. O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V. M., Trojanowski, J. Q., Gibbs, J. R., Brunetti, M., Gronka, S., Wuu, J., et al. (2010). Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron*, 68 (5), pp.857–864.

Johnson, J. O., Pioro, E. P., Boehringer, A., Chia, R., Feit, H., Renton, A. E., Pliner, H. A., Abramzon, Y., Marangi, G., Winborn, B. J., et al. (2014). Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nature Neuroscience*, 17 (5), pp.664–666.

Jounai, N., Takeshita, F., Kobiyama, K., Sawano, A., Miyawaki, A., Xin, K.-Q., Ishii, K. J., Kawai, T., Akira, S., Suzuki, K., et al. (2007). The Atg5 Atg12 conjugate associates with innate antiviral immune responses. *Proceedings of the National Academy of Sciences*, 104 (35), pp.14050–14055.

Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., Paul, J. W., Sun, S., Herdy, J. R., Bieri, G., et al. (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nature Neuroscience*, 18 (9), pp.1226–1229.

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

Kabashi, E., Lin, L., Tradewell, M. L., Dion, P. A., Bercier, V., Bourgouin, P., Rochefort, D., Bel Hadj, S., Durham, H. D., Velde, C. Vande, et al. (2010). Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Human Molecular Genetics*, 19 (4), pp.671–683.

Kabashi, E., Valdmanis, P. N., Dion, P., Spiegelman, D., McConkey, B. J., Vande Velde, C., Bouchard, J. P., Lacomblez, L., Pochigaeva, K., Salachas, F., et al. (2008). TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature Genetics*, 40 (5), pp.572–574.

Kaiser, T. E., Intine, R. V and Dundr, M. (2008). De novo formation of a subnuclear body. *Science*, 322 (5908), pp.1713–1717.

Kamola, P. J., Nakano, Y., Takahashi, T., Wilson, P. A. and Ui-Tei, K. (2015). The siRNA Non-seed Region and Its Target Sequences Are Auxiliary Determinants of Off-Target Effects. Ioshikhes, I. (Ed). *PLOS Computational Biology*, 11 (12), p.e1004656.

Kanno, T., Lin, W.-D., Fu, J. L., Wu, M.-T., Yang, H.-W., Lin, S.-S., Matzke, A. J. M. and Matzke, M. (2016). Identification of Coilin Mutants in a Screen for Enhanced Expression of an Alternatively Spliced GFP Reporter Gene in Arabidopsis thaliana. *Genetics*, p.genetics.116.190751.

Kariya, S., Re, D. B., Jacquier, A., Nelson, K., Przedborski, S. and Monani, U. R. (2012). Mutant superoxide dismutase 1 (SOD1), a cause of amyotrophic lateral sclerosis, disrupts the recruitment of SMN, the spinal muscular atrophy protein to nuclear Cajal bodies. *Human Molecular Genetics*, 21 (15), pp.3421–3434.

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., et al. (2012). Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell*, 149 (4), pp.753–767.

Kawahara, Y. and Mieda-Sato, A. (2012). TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proceedings of the National Academy of Sciences*, 109 (9), pp.3347–3352.

Kawamata, H. and Manfredi, G. (2008). Different regulation of wild-type and mutant Cu,Zn superoxide dismutase localization in mammalian mitochondria. *Human Molecular Genetics*, 17 (21), pp.3303–3317.

Kawamata, T., Akiyama, H., Yamada, T. and McGeer, P. L. (1992). Immunologic reactions in amyotrophic lateral sclerosis brain and spinal cord tissue. *The American journal of pathology*, 140 (3), pp.691–707.

Kenna, K. P., van Doormaal, P. T. C., Dekker, A. M., Ticozzi, N., Kenna, B. J., Diekstra, F. P., van Rheenen, W., van Eijk, K. R., Jones, A. R., Keagle, P., et al. (2016). NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nature Genetics*, 48 (9), pp.1037–1042.

Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C.-H. A., Marr, M. T. and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes & Development*, 25 (23), pp.2502–2512.

Khusial, P., Plaag, R. and Zieve, G. W. (2005). LSm proteins form heptameric rings that bind to RNA via repeating motifs. *Trends in Biochemical Sciences*, 30 (9), pp.522–528.

Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R. and Zoing, M. C. (2011). Amyotrophic lateral sclerosis. *Lancet*, 377 (9769), pp.942–955.

Kim, H. J., Kim, N. C., Wang, Y.-D., Scarborough, E. A., Moore, J., Diaz, Z., MacLea, K. S., Freibaum, B., Li, S., Molliex, A., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature*, 495 (7442), pp.467–473.

Kiss, A. M., Jady, B. E., Bertrand, E. and Kiss, T. (2004). Human Box H/ACA Pseudouridylation Guide RNA Machinery. *Molecular and Cellular Biology*, 24 (13), pp.5797–5807.

Kiss, A. M., Jády, B. E., Darzacq, X., Verheggen, C., Bertrand, E. and Kiss, T. (2002). A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Research*, 30 (21), pp.4643–4649.

Kitao, S., Segref, A., Kast, J., Wilm, M., Mattaj, I. W. and Ohno, M. (2008). A compartmentalized phosphorylation/dephosphorylation system that regulates U snRNA export from the nucleus. *Molecular and Cellular Biology*, 28 (1), pp.487–497.

Klingauf, M., Stanek, D. and Neugebauer, K. M. (2006). Enhancement of U4/U6 Small Nuclear Ribonucleoprotein Particle Association in Cajal Bodies Predicted by Mathematical Modeling. *Molecular Biology of the Cell*, 17 (12), pp.4972–4981.

Komatsu, M., Waguri, S., Chiba, T., Murata, S., Iwata, J., Tanida, I., Ueno, T., Koike, M., Uchiyama, Y., Kominami, E., et al. (2006). Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature*, 441 (7095), pp.880–884.

Kong, J. and Xu, Z. (1998). Massive mitochondrial degeneration in motor neurons triggers the onset of amyotrophic lateral sclerosis in mice expressing a mutant SOD1. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18 (9), pp.3241–3250.

König, H., Matter, N., Bader, R., Thiele, W. and Müller, F. (2007). Splicing Segregation: The Minor Spliceosome Acts outside the Nucleus and Controls Cell Proliferation. *Cell*, 131 (4), pp.718–729.

Koppers, M., Blokhuis, A. M., Westeneng, H.-J., Terpstra, M. L., Zundel, C. A. C., Vieira de Sá, R., Schellevis, R. D., Waite, A. J., Blake, D. J., Veldink, J. H., et al. (2015). C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Annals of Neurology*, 78 (3), pp.426–438.

Kriz, J., Nguyen, M. D. and Julien, J.-P. (2002). Minocycline slows disease progression in a mouse model of amyotrophic lateral sclerosis. *Neurobiology of Disease*, 10 (3), pp.268–278.

Kruman, I. I., Pedersen, W. A., Springer, J. E. and Mattson, M. P. (1999). ALS-Linked Cu/Zn–SOD Mutation Increases Vulnerability of Motor Neurons to Excitotoxicity by a Mechanism Involving Increased Oxidative Stress and Perturbed Calcium Homeostasis. *Experimental Neurology*, 160 (1), pp.28–39.

Kunkel, G. R., Maser, R. L., Calvet, J. P. and Pederson, T. (1986). U6 small nuclear RNA is transcribed by RNA polymerase III. *Proceedings of the National Academy of Sciences of the United States of America*, 83 (22), pp.8575–8579.

Kurtzke, J. F. (1982). Epidemiology of amyotrophic lateral sclerosis. *Advances in Neurology*, 36, pp.281–302.

Kusaka, H., Matsumoto, S. and Imai, T. (1990). An adult-onset case of sporadic motor neuron disease with basophilic inclusions. *Acta Neuropathologica*, 80 (6), pp.660–665.

Kusaka, H., Matsumoto, S. and Imai, T. (1993). Adult-onset motor neuron disease with basophilic intraneuronal inclusion bodies. *Clinical neuropathology*, 12 (4), pp.215–218.

Kwiatkowski Jr., T. J., Bosco, D. A., Leclerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E. J., Munsat, T., et al. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science*, 323 (5918), pp.1205–1208.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y. and McKnight, S. L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science*, 345 (6201), pp.1139–1145.

Laaksovirta, H., Peuralinna, T., Schymick, J. C., Scholz, S. W., Lai, S. L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D. G., Gibbs, J. R., et al. (2010). Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurology*, 9 (10), pp.978–985.

Lacomblez, L., Bensimon, G., Leigh, P. N., Guillet, P. and Meininger, V. (1996). Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. *Lancet*, 347 (9013), pp.1425–1431.

Lafarga, M., Berciano, M. T., Garcia-Segura, L. M., Andres, M. A. and Carmo-Fonseca, M. (1998). Acute osmotic/stress stimuli induce a transient decrease of transcriptional activity in the neurosecretory neurons of supraoptic nuclei. *Journal of Neurocytology*, 27 (4), pp.205–217.

Lafarga, M., Tapia, O., Romero, A. M. and Berciano, M. T. (2017). Cajal bodies in neurons. *RNA Biology*, 14 (6), pp.712–725.

Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H. R., Jiang, J., Watt, A. T., Chun, S., Katz, M., et al. (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proceedings of the National Academy of Sciences*, 110 (47), pp.E4530-9.

Lagier-Tourenne, C., Polymenidou, M., Hutt, K. R., Vu, A. Q., Baughn, M., Huelga, S. C., Clutario, K. M., Ling, S.-C., Liang, T. Y., Mazur, C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nature Neuroscience*, 15 (11), pp.1488–1497.

Laird, F. M., Farah, M. H., Ackerley, S., Hoke, A., Maragakis, N., Rothstein, J. D., Griffin, J., Price, D. L., Martin, L. J. and Wong, P. C. (2008). Motor neuron disease occurring in a mutant dynactin mouse model is characterized by defects in vesicular trafficking. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28 (9), pp.1997–2005.

Lam, Y. W., Lyon, C. E. and Lamond, A. I. (2002). Large-scale isolation of Cajal bodies from HeLa cells. *Molecular Biology of the Cell*, 13 (7), pp.2461–2473.

Lattante, S., de Calbiac, H., Le Ber, I., Brice, A., Ciura, S. and Kabashi, E. (2015). Sqstm1 knock-down causes a locomotor phenotype ameliorated by rapamycin in a zebrafish model of ALS/FTLD. *Human Molecular Genetics*, 24 (6), pp.1682–1690.

Leblond, C. S., Kaneb, H. M., Dion, P. A. and Rouleau, G. A. (2014). Dissection of genetic factors associated with amyotrophic lateral sclerosis. *Experimental Neurology*, 262, pp.91–101.

Lee, K.-H., Zhang, P., Kim, H. J., Mitrea, D. M., Sarkar, M., Freibaum, B. D., Cika, J., Coughlin, M., Messing, J., Molliex, A., et al. (2016). C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell*, 167 (3), p.774–788.e17.

Lee, Y.-B., Baskaran, P., Gomez-Deza, J., Chen, H.-J., Nishimura, A. L., Smith, B. N., Troakes, C., Adachi, Y., Stepto, A., Petrucelli, L., et al. (2017). C9orf72 poly GA RAN-translated protein plays a key role in amyotrophic lateral sclerosis via aggregation and toxicity. *Human Molecular Genetics*, 26 (24), pp.4765–4777.

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

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, 80 (1), pp.155–165.

Lemm, I., Girard, C., Kuhn, A. N., Watkins, N. J., Schneider, M., Bordonné, R. and Lührmann, R. (2006). Ongoing U snRNP biogenesis is required for the integrity of Cajal bodies. *Molecular biology of the cell*, 17 (7), pp.3221–3231.

Levine, T. P., Daniels, R. D., Gatta, A. T., Wong, L. H. and Hayes, M. J. (2013). The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. *Bioinformatics*, 29 (4), pp.499–503.

Licht, K., Medenbach, J., Lührmann, R., Kambach, C. and Bindereif, A. (2008). 3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins. *RNA*, 14 (8), pp.1532–1538.

Lim, K. M., Yeo, W. S. and Chow, V. T. K. (2004). Antisense abrogation of DENN expression induces apoptosis of leukemia cells in vitro, causes tumor regression in vivo and alters the transcription of genes involved in apoptosis and the cell cycle. *International Journal of Cancer*, 109 (1), pp.24–37.

Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., et al. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*, 5, p.4767.

Liu, E. Y., Russ, J., Wu, K., Neal, D., Suh, E., McNally, A. G., Irwin, D. J., Van Deerlin, V. M. and Lee, E. B. (2014). C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathologica*, 128 (4), pp.525–541.

Liu, J.-L., Hebert, M. D., Ye, Y., Templeton, D. J., Kung, H.-J. and Materta, G. A. (2000). Cell cycle-dependent localization of the CDK2-cyclin E complex in Cajal (coiled) bodies. *Journal of Cell Science*, 113, pp.1543–1552.

Liu, J.-L., Wu, Z., Nizami, Z., Deryusheva, S., Rajendra, T. K., Beumer, K. J., Gao, H., Matera, A. G., Carroll, D. and Gall, J. G. (2009). Coilin is essential for Cajal body organization in Drosophila melanogaster. *Molecular biology of the cell*, 20 (6), pp.1661–1670.

Liu, S., Rauhut, R., Vornlocher, H.-P. and Lührmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA*, 12 (7), pp.1418–1430.

Liu, Y., Pattamatta, A., Zu, T., Reid, T., Bardhi, O., Borchelt, D. R., Yachnis, A. T. and Ranum, L. P. W. (2016). C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. *Neuron*, 90 (3), pp.521–534.

Logroscino, G., Traynor, B. J., Hardiman, O., Chio, A., Mitchell, D., Swingler, R. J., Millul, A., Benn, E., Beghi, E. and EURALS. (2010). Incidence of amyotrophic lateral sclerosis in Europe. *Journal of Neurology, Neurosurgery & Psychiatry*, 81 (4), pp.385–390.

Lund, E. and Dahlberg, J. E. (1992). Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science*, 255 (5042), pp.327–330.

Ma, X., Peterson, R. and Turnbull, J. (2011). Adenylyl cyclase type 3, a marker of primary cilia, is reduced in primary cell culture and in lumbar spinal cord in situ in G93A SOD1 mice. *BMC Neuroscience*, 12, p.71.

Machyna, M., Kehr, S., Straube, K., Kappei, D., Buchholz, F., Butter, F., Ule, J., Hertel, J., Stadler, P. F. and Neugebauer, K. M. (2014). The Coilin Interactome Identifies Hundreds of Small Noncoding RNAs that Traffic through Cajal Bodies. *Molecular Cell*, 56 (3), pp.389–399.

Mackenzie, I. R. A., Bigio, E. H., Ince, P. G., Geser, F., Neumann, M., Cairns, N. J., Kwong, L. K., Forman, M. S., Ravits, J., Stewart, H., et al. (2007). Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Annals of Neurology*, 61 (5), pp.427–434.

Mackenzie, I. R. A. and Feldman, H. H. (2005). Ubiquitin immunohistochemistry suggests classic motor neuron disease, motor neuron disease with dementia, and frontotemporal dementia of the motor neuron disease type represent a clinicopathologic spectrum. *Journal of Neuropathology & Experimental Neurology*, 64 (8), pp.730–739.

Mackenzie, I. R. A., Frick, P., Grässer, F. A., Gendron, T. F., Petrucelli, L., Cashman, N. R., Edbauer, D., Kremmer, E., Prudlo, J., Troost, D., et al. (2015). Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathologica*, 130 (6), pp.845–861.

Mackenzie, I. R., Arzberger, T., Kremmer, E., Troost, D., Lorenzl, S., Mori, K., Weng, S.-M., Haass, C., Kretzschmar, H. A., Edbauer, D., et al. (2013). Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathologica*, 126 (6), pp.859–879.

Mahmoudi, S., Henriksson, S., Weibrecht, I., Smith, S., Söderberg, O., Strömblad, S., Wiman, K. G. and Farnebo, M. (2010). WRAP53 Is Essential for Cajal Body Formation and for Targeting the Survival of Motor Neuron Complex to Cajal Bodies. *PLoS Biology*, 8 (11), p.e1000521.

Makarov, V., Rakitina, D., Protopopova, A., Yaminsky, I., Arutiunian, A., Love, A. J., Taliansky, M. and Kalinina, N. (2013). Plant coilin: structural characteristics and RNA-binding properties. *PLoS One*, 8 (1), p.e53571.

Makarova, O. V, Makarov, E. M., Liu, S., Vornlocher, H.-P. and Lührmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *The EMBO journal*, 21 (5), pp.1148–1157.

Malessa, S., Leigh, P. N., Bertel, O., Sluga, E. and Hornykiewicz, O. (1991). Amyotrophic lateral sclerosis: glutamate dehydrogenase and transmitter amino acids in the spinal cord. *Journal of neurology, neurosurgery, and psychiatry*, 54 (11), pp.984–988.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L. and Church, G. M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, 31 (9), pp.833–838.

Mann, D. M. A., Rollinson, S., Robinson, A., Bennion Callister, J., Thompson, J. C., Snowden, J. S., Gendron, T., Petrucelli, L., Masuda-Suzukake, M., Hasegawa, M., et al. (2013). Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathologica Communications*, 1, p.68.

Mao, Y. S., Sunwoo, H., Zhang, B. and Spector, D. L. (2011a). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nature Cell Biology*, 13 (1), pp.95–101.

Mao, Y. S., Zhang, B. and Spector, D. L. (2011b). Biogenesis and function of nuclear bodies. *Trends in genetics*, 27 (8), pp.295–306.

Marat, A. L. and McPherson, P. S. (2010). The connecdenn family, Rab35 guanine nucleotide exchange factors interfacing with the clathrin machinery. *The Journal of Biological Chemistry*, 285 (14), pp.10627–10637.

Marchetto, M. C. N., Muotri, A. R., Mu, Y., Smith, A. M., Cezar, G. G. and Gage, F. H. (2008). Non-Cell-Autonomous Effect of Human SOD1G37R Astrocytes on Motor Neurons Derived from Human Embryonic Stem Cells. *Cell Stem Cell*, 3 (6), pp.649–657.

Markmiller, S., Cloonan, N., Lardelli, R. M., Doggett, K., Keightley, M.-C., Boglev, Y., Trotter, A. J., Ng, A. Y., Wilkins, S. J., Verkade, H., et al. (2014). Minor class splicing shapes the zebrafish transcriptome during development. *Proceedings of the National Academy of Sciences of the United States of America*, 111 (8), pp.3062–3067.

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H., et al. (2010). Mutations of optineurin in amyotrophic lateral sclerosis. *Nature*, 465 (7295), pp.223–226.

Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I. W. and Dreyfuss, G. (2002). The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Molecular and Cellular Biology*, 22 (18), pp.6533–6541.

Mastrocola, A. S., Kim, S. H., Trinh, A. T., Rodenkirch, L. A. and Tibbetts, R. S. (2013). The RNA-binding protein fused in sarcoma (FUS) functions downstream of poly(ADP-ribose) polymerase (PARP) in response to DNA damage. *The Journal of biological chemistry*, 288 (34), pp.24731–24741.

Matera, A. G. and Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, 15 (2), pp.108–121.

Matera, A. G. and Ward, D. C. (1993). Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. *The Journal of cell biology*, 121 (4), pp.715–727.

Matsumoto, S., Kusaka, H., Murakami, N., Hashizume, Y., Okazaki, H. and Hirano, A. (1992). Basophilic inclusions in sporadic juvenile amyotrophic lateral sclerosis: an immunocytochemical and ultrastructural study. *Acta neuropathologica*, 83 (6), pp.579–583.

Mattiazzi, M., D'Aurelio, M., Gajewski, C. D., Martushova, K., Kiaei, M., Beal, M. F. and Manfredi, G. (2002). Mutated Human SOD1 Causes Dysfunction of Oxidative Phosphorylation in Mitochondria of Transgenic Mice. *Journal of Biological Chemistry*, 277 (33), pp.29626–29633.

May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grässer, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., et al. (2014). C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathologica*, 128 (4), pp.485–503.

McEachern, M. J., Krauskopf, A. and Blackburn, E. H. (2000). Telomeres and Their Control. *Annual Review of Genetics*, 34 (1), pp.331–358.

McMillan, C. T., Russ, J., Wood, E. M., Irwin, D. J., Grossman, M., McCluskey, L., Elman, L., Van Deerlin, V. and Lee, E. B. (2015). C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. *Neurology*, 84 (16), pp.1622–1630.

Meister, G., Bühler, D., Pillai, R., Lottspeich, F. and Fischer, U. (2001). A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nature Cell Biology*, 3 (11), pp.945–949.

Mertens, J., Paquola, A. C. M., Ku, M., Hatch, E., Böhnke, L., Ladjevardi, S., McGrath, S., Campbell, B., Lee, H., Herdy, J. R., et al. (2015). Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell*, 17 (6), pp.705–718.

Meyer, K., Ferraiuolo, L., Miranda, C. J., Likhite, S., McElroy, S., Renusch, S., Ditsworth, D., Lagier-Tourenne, C., Smith, R. A., Ravits, J., et al. (2014). Direct conversion of patient fibroblasts demonstrates non-cell autonomous toxicity of astrocytes to motor neurons in familial and sporadic ALS. *Proceedings of the National Academy of Sciences of the United States of America*, 111 (2), pp.829–832.

Miller, T. M., Pestronk, A., David, W., Rothstein, J., Simpson, E., Appel, S. H., Andres, P. L., Mahoney, K., Allred, P., Alexander, K., et al. (2013). An antisense oligonucleotide against SOD1 delivered intrathecally for patients with SOD1 familial amyotrophic lateral sclerosis: a phase 1, randomised, first-in-man study. *The Lancet Neurology*, 12 (5), pp.435–442.

Mitchell, J. R., Wood, E. and Collins, K. (1999). A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*, 402 (6761), pp.551–555.

Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., Moens, T., Norona, F. E., Woollacott, I. O., Pietrzyk, J., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science*, 345 (6201), pp.1192–1194.

Mizuno, Y., Amari, M., Takatama, M., Aizawa, H., Mihara, B. and Okamoto, K. (2006). Transferrin localizes in Bunina bodies in amyotrophic lateral sclerosis. *Acta Neuropathologica*, 112 (5), pp.597–603.

Mizusawa, H. (1993). Hyaline and Skein-like Inclusions in Amyotrophic Lateral Sclerosis. *Neuropathology*, 13 (3), pp.201–208.

Moller, A., Bauer, C. S., Cohen, R. N., Webster, C. P. and De Vos, K. J. (2017). Amyotrophic lateral sclerosis-associated mutant SOD1 inhibits anterograde axonal transport of mitochondria by reducing Miro1 levels. *Human Molecular Genetics*, 26 (23), pp.4668–4679.

Montzka, K. A. and Steitz, J. A. (1988). Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proceedings of the National Academy of Sciences of the United States of America*, 85 (23), pp.8885–8889.

Moore, J. D. (2000). The Ran-GTPase and cell-cycle control. *BioEssays*, 23 (1), pp.77–85.

Mori, K., Arzberger, T., Grässer, F. A., Gijselinck, I., May, S., Rentzsch, K., Weng, S.-M., Schludi, M. H., van der Zee, J., Cruts, M., et al. (2013a). Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathologica*, 126 (6), pp.881–893.

Mori, K., Weng, S.-M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., et al. (2013b). The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science*, 339 (6125), pp.1335–1338.

Morse, R., Shaw, D. J., Todd, A. G. and Young, P. J. (2007). Targeting of SMN to Cajal bodies is mediated by self-association. *Human Molecular Genetics*, 16 (19), pp.2349–2358.

Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A. G., Bertrand, E., Tazi, J. and Bordonné, R. (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.

Mount, S. M. (2000). Genomic sequence, splicing, and gene annotation. *American journal of human genetics*, 67 (4), pp.788–792.

Mroczek, S. and Dziembowski, A. (2013). U6 RNA biogenesis and disease association. *Wiley Interdisciplinary Reviews: RNA*, 4 (5), pp.581–592.

Nagai, M., Re, D. B., Nagata, T., Chalazonitis, A., Jessell, T. M., Wichterle, H. and Przedborski, S. (2007). Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. *Nature Neuroscience*, 10 (5), pp.615–622.

Nakagawa, I., Amano, A., Mizushima, N., Yamamoto, A., Yamaguchi, H., Kamimoto, T., Nara, A., Funao, J., Nakata, M., Tsuda, K., et al. (2004). Autophagy Defends Cells Against Invading Group A Streptococcus. *Science*, 306 (5698), pp.1037–1040.

Nakahira, K., Haspel, J. A., Rathinam, V. A. K., Lee, S.-J., Dolinay, T., Lam, H. C., Englert, J. A., Rabinovitch, M., Cernadas, M., Kim, H. P., et al. (2011). Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial DNA mediated by the NALP3 inflammasome. *Nature Immunology*, 12 (3), pp.222–230.

Narayanan, A., Lukowiak, A., Jády, B. E., Dragon, F., Kiss, T., Terns, R. M. and Terns, M. P. (1999a). Nucleolar localization signals of Box H/ACA small nucleolar RNAs. *The EMBO Journal*, 18 (18), pp.5120–5130.

Narayanan, A., Speckmann, W., Terns, R. and Terns, M. P. (1999b). Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Molecular biology of the cell*, 10 (7), pp.2131–2147.

Nejedla, M., Sadi, S., Sulimenko, V., de Almeida, F. N., Blom, H., Draber, P., Aspenström, P. and Karlsson, R. (2016). Profilin connects actin assembly with microtubule dynamics. *Molecular biology of the cell*, 27 (15), pp.2381–2393.

Nesic, D., Tanackovic, G. and Krämer, A. (2004). A role for Cajal bodies in the final steps of U2 snRNP biogenesis. *Journal of Cell Science*, 117 (Pt 19), pp.4423–4433.

Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., et al. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, 314 (5796), pp.130–133.

Niblock, M., Smith, B. N., Lee, Y.-B., Sardone, V., Topp, S., Troakes, C., Al-Sarraj, S., Leblond, C. S., Dion, P. A., Rouleau, G. A., et al. (2016). Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. *Acta Neuropathologica Communications*, 4 (1), p.18.

Niemela, E. H., Oghabian, A., Staals, R. H. J., Greco, D., Pruijn, G. J. M. and Frilander, M. J. (2014). Global analysis of the nuclear processing of transcripts with unspliced U12-type introns by the exosome. *Nucleic Acids Research*, 42 (11), pp.7358–7369.

Nigg, E. A. (1997). Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature*, 386 (6627), pp.779–787.

Nishimoto, T., Eilen, E. and Basilico, C. (1978). Premature of chromosome condensation in a ts DNA- mutant of BHK cells. *Cell*, 15 (2), pp.475–483.

Nishimura, A. L., Mitne-Neto, M., Silva, H. C., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J. R., Gillingwater, T., Webb, J., et al. (2004). A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *American journal of human genetics*, 75 (5), pp.822–831.

Nixon, R. A. (2013). The role of autophagy in neurodegenerative disease. *Nature Medicine*, 19 (8), pp.983–997.

Novotny, I., Blazikova, M., Stanek, D., Herman, P. and Malinsky, J. (2011). In vivo kinetics of U4/U6{middle dot}U5 tri-snRNP formation in Cajal bodies. *Molecular biology of the cell*, 22 (4), pp.513–523.

Novotný, I., Malinová, A., Stejskalová, E., Matějů, D., Klimešová, K., Roithová, A., Švéda, M., Knejzlík, Z. and Staněk, D. (2015). SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. *Cell reports*, 10 (3), pp.429–440.

Nowakowski, R. S., Lewin, S. B. and Miller, M. W. (1989). Bromodeoxyuridine immunohistochemical determination of the lengths of the cell cycle and the DNA-synthetic phase for an anatomically defined population. *Journal of Neurocytology*, 18 (3), pp.311–318.

O'Rourke, J. G., Bogdanik, L., Muhammad, A. K. M. G., Gendron, T. F., Kim, K. J., Austin, A., Cady, J., Liu, E. Y., Zarrow, J., Grant, S., et al. (2015). C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. *Neuron*, 88 (5), pp.892–901.

O'Rourke, J. G., Bogdanik, L., Yáñez, A., Lall, D., Wolf, A. J., Muhammad, A. K. M. G., Ho, R., Carmona, S., Vit, J. P., Zarrow, J., et al. (2016). C9orf72 is required for proper macrophage and microglial function in mice. *Science*, 351 (6279), pp.1324–1329.

Ogg, S. C. and Lamond, A. I. (2002). Cajal bodies and coilin--moving towards function. *The Journal of Cell Biology*, 159 (1), pp.17–21.

Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I. W. (2000). PHAX, a Mediator of U snRNA Nuclear Export Whose Activity Is Regulated by Phosphorylation. *Cell*, 101 (2), pp.187–198.

Okamoto, K., Hirai, S., Amari, M., Watanabe, M. and Sakurai, A. (1993). Bunina bodies in amyotrophic lateral sclerosis immunostained with rabbit anti-cystatin C serum. *Neuroscience Letters*, 162 (1–2), pp.125–128.

Olsen, J. V, Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127 (3), pp.635–648.

Orlacchio, A., Babalini, C., Borreca, A., Patrono, C., Massa, R., Basaran, S., Munhoz, R. P., Rogaeva, E. A., St George-Hyslop, P. H., Bernardi, G., et al. (2010). SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. *Brain*, 133 (Pt 2), pp.591–598.

Ospina, J. K., Gonsalvez, G. B., Bednenko, J., Darzynkiewicz, E., Gerace, L. and Matera, A. G. (2005). Cross-talk between snurportin1 subdomains. *Molecular biology of the cell*, 16 (10), pp.4660–4671.

Ou, S. H., Wu, F., Harrich, D., Garcia-Martinez, L. F. and Gaynor, R. B. (1995). Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *Journal of Virology*, 69 (6), pp.3584–3596.

Palacios, I., Hetzer, M., Adam, S. A. and Mattaj, I. W. (1997). Nuclear import of U snRNPs requires importin beta. *The EMBO journal*, 16 (22), pp.6783–6792.

Parkinson, N., Ince, P. G., Smith, M. O., Highley, R., Skibinski, G., Andersen, P. M., Morrison, K. E., Pall, H. S., Hardiman, O., Collinge, J., et al. (2006). ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). *Neurology*, 67 (6), pp.1074–1077.

Pasinelli, P., Belford, M. E., Lennon, N., Bacskai, B. J., Hyman, B. T., Trotti, D. and Brown, R. H. (2004). Amyotrophic Lateral Sclerosis-Associated SOD1 Mutant Proteins Bind and Aggregate with Bcl-2 in Spinal Cord Mitochondria. *Neuron*, 43 (1), pp.19–30.

Patel, A. A., McCarthy, M. and Steitz, J. A. (2002). The splicing of U12-type introns can be a rate-limiting step in gene expression. *The EMBO journal*, 21 (14), pp.3804–3815.

Patel, A. A. and Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. *Nature Reviews Molecular Cell Biology*, 4 (12), pp.960–970.

Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, 162 (5), pp.1066–1077.

Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. L., Fischbeck, K. H. and Pittman, R. N. (1997). Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron*, 19 (2), pp.333–344.

Pelegrini, A. L., Moura, D. J., Brenner, B. L., Ledur, P. F., Maques, G. P., Henriques, J. A. P., Saffi, J. and Lenz, G. (2010). Nek1 silencing slows down DNA repair and blocks DNA damage-induced cell cycle arrest. *Mutagenesis*, 25 (5), pp.447–454.

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.

Pena, E., Berciano, M. T., Fernandez, R., Ojeda, J. L. and Lafarga, M. (2001). Neuronal body size correlates with the number of nucleoli and Cajal bodies, and with the organization of the splicing machinery in rat trigeminal ganglion neurons. *The Journal of comparative neurology*, 430 (2), pp.250–263.

Pessa, H. K. J., Ruokolainen, A. and Frilander, M. J. (2006). The abundance of the spliceosomal snRNPs is not limiting the splicing of U12-type introns. *RNA*, 12 (10), pp.1883–1892.

Peters, O. M., Cabrera, G. T., Tran, H., Gendron, T. F., McKeon, J. E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J., et al. (2015). Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. *Neuron*, 88 (5), pp.902–909.

Platani, M., Goldberg, I., Lamond, A. I. and Swedlow, J. R. (2002). Cajal Body dynamics and association with chromatin are ATP-dependent. *Nature Cell Biology*, 4 (7), pp.502–508.

Platani, M., Goldberg, I., Swedlow, J. R. and Lamond, A. I. (2000). In vivo analysis of Cajal body movement, separation, and joining in live human cells. *The Journal of Cell Biology*, 151 (7), pp.1561–1574.

Pleiss, J. A., Whitworth, G. B., Bergkessel, M. and Guthrie, C. (2007). Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS biology*, 5 (4), p.e90.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K. R., Huelga, S. C., Moran, J., Liang, T. Y., Ling, S. C., Sun, E., Wancewicz, E., Mazur, C., et al. (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nature Neuroscience*, 14 (4), pp.459–468.

Poole, A. R., Hebert, M. D., Andrade, L. E. C., Tan, E. M., Chan, E. K. L., Bachand, F., Boisvert, F.-M., Cote, J., Richard, S., Autexier, C., et al. (2016). SMN and coilin negatively regulate dyskerin association with telomerase RNA. *Biology open*, 5 (6), pp.726–735.

Pradet-Balade, B., Girard, C., Boulon, S., Paul, C., Azzag, K., Bordonné, R., Bertrand, E. and Verheggen, C. (2011). CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *The EMBO journal*, 30 (11), pp.2205–2218.

Prinz, M. and Priller, J. (2014). Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nature Reviews Neuroscience*, 15 (5), pp.300–312.

Project MinE Press Release. (2013). Retrieved from https://www.projectminE.com/wp-content/uploads/2013/06/Press-release-project-minE_21June2013.pdf

Proudfoot, M., Gutowski, N. J., Edbauer, D., Hilton, D. A., Stephens, M., Rankin, J. and Mackenzie, I. R. A. (2014). Early dipeptide repeat pathology in a frontotemporal dementia kindred with C9ORF72 mutation and intellectual disability. *Acta Neuropathologica*, 127 (3), pp.451–458.

Prudencio, M., Belzil, V. V, Batra, R., Ross, C. A., Gendron, T. F., Pregent, L. J., Murray, M. E., Overstreet, K. K., Piazza-Johnston, A. E., Desaro, P., et al. (2015). Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nature Neuroscience*, 18 (8), pp.1175–1182.

Puls, I., Jonnakuty, C., LaMonte, B. H., Holzbaur, E. L. F., Tokito, M., Mann, E., Floeter, M. K., Bidus, K., Drayna, D., Oh, S. J., et al. (2003). Mutant dynactin in motor neuron disease. *Nature Genetics*, 33 (4), pp.455–456.

Raghunathan, P. L. and Guthrie, C. (1998). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science*, 279 (5352), pp.857–860.

Raker, V. A., Plessel, G. and Lührmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *The EMBO journal*, 15 (9), pp.2256–2269.

Ramirez, C., Piemonte, M. E., Callegaro, D. and Da Silva, H. C. (2008). Fatigue in amyotrophic lateral sclerosis: frequency and associated factors. *Amyotrophic Lateral Sclerosis*, 9 (2), pp.75–80.

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*, 154 (6), pp.1380–1389.

Ravits, J. M. and La Spada, A. R. (2009). ALS motor phenotype heterogeneity, focality, and spread: deconstructing motor neuron degeneration. *Neurology*, 73 (10), pp.805–811.

Re, D. B., Le Verche, V., Yu, C., Amoroso, M. W., Politi, K. A., Phani, S., Ikiz, B., Hoffmann, L., Koolen, M., Nagata, T., et al. (2014). Necroptosis Drives Motor Neuron Death in Models of Both Sporadic and Familial ALS. *Neuron*, 81 (5), pp.1001–1008.

Reaume, A. G., Elliott, J. L., Hoffman, E. K., Kowall, N. W., Ferrante, R. J., Siwek, D. R., Wilcox, H. M., Flood, D. G., Beal, M. F., Brown, R. H., et al. (1996). Motor neurons in Cu/Zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury. *Nature Genetics*, 13 (1), pp.43–47.

Rebelo, L., Almeida, F., Ramos, C., Bohmann, K., Lamond, A. I. and Carmo-Fonseca, M. (1996). The dynamics of coiled bodies in the nucleus of adenovirus-infected cells. *Molecular biology of the cell*, 7 (7), pp.1137–1151.

Reber, S., Stettler, J., Filosa, G., Colombo, M., Jutzi, D., Lenzken, S. C., Schweingruber, C., Bruggmann, R., Bachi, A., Barabino, S. M., et al. (2016). Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants. *The EMBO Journal*, 35 (14), pp.1504–1521.

Reddy, K., Zamiri, B., Stanley, S. Y., Macgregor Jr., R. B. and Pearson, C. E. (2013). The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *The Journal of Biological Chemistry*, 288 (14), pp.9860–9866.

Ren, M., Coutavas, E., D'Eustachio, P. and Rush, M. G. (1994). Effects of mutant Ran/TC4 proteins on cell cycle progression. *Molecular and cellular biology*, 14 (6), pp.4216–4224.

Renton, A. E., Chiò, A. and Traynor, B. J. (2013). State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience*, 17 (1), pp.17–23.

Renton, A. E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, 72 (2), pp.257–268.

Richard, P., Darzacq, X., Bertrand, E., Jády, B. E., Verheggen, C. and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *The EMBO Journal*, 22 (16), pp.4283–4293.

Rippon, G. A., Scarmeas, N., Gordon, P. H., Murphy, P. L., Albert, S. M., Mitsumoto, H., Marder, K., Rowland, L. P. and Stern, Y. (2006). An Observational Study of Cognitive Impairment in Amyotrophic Lateral Sclerosis. *Archives of Neurology*, 63 (3), p.345.

Rogelj, B., Easton, L. E., Bogu, G. K., Stanton, L. W., Rot, G., Curk, T., Zupan, B., Sugimoto, Y., Modic, M., Haberman, N., et al. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Scientific Reports*, 2 (1), p.603.

Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J. P., Deng, H. X., et al. (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature*, 362 (6415), pp.59–62.

Rossi, S., Serrano, A., Gerbino, V., Giorgi, A., Di Francesco, L., Nencini, M., Bozzo, F., Schininà, M. E., Bagni, C., Cestra, G., et al. (2015). Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS. *Journal of Cell Science*, 128 (9), pp.1787–1799.

Rothstein, J. D., Tsai, G., Kuncl, R. W., Clawson, L., Cornblath, D. R., Drachman, D. B., Pestronk, A., Stauch, B. L. and Coyle, J. T. (1990). Abnormal excitatory amino acid metabolism in amyotrophic lateral sclerosis. *Annals of Neurology*, 28 (1), pp.18–25.

Röttgers, K., Krohn, N. M., Lichota, J., Stemmer, C., Merkle, T. and Grasser, K. D. (2000). DNA-interactions and nuclear localisation of the chromosomal HMG domain protein SSRP1 from maize. *The Plant Journal*, 23 (3), pp.395–405.

Rowland, L. P. and Shneider, N. A. (2001). Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, 344 (22), pp.1688–1700.

Saltzman, A. L., Pan, Q. and Blencowe, B. J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. *Genes & Development*, 25 (4), pp.373–384.

Sanz-García, M., Vázquez-Cedeira, M., Kellerman, E., Renbaum, P., Levy-Lahad, E. and Lazo, P. A. (2011). Substrate profiling of human vaccinia-related kinases identifies coilin, a Cajal body nuclear protein, as a phosphorylation target with neurological implications. *Journal of Proteomics*, 75 (2), pp.548–560.

Scekic-Zahirovic, J., Sendscheid, O., El Oussini, H., Jambeau, M., Sun, Y., Mersmann, S., Wagner, M., Dieterlé, S., Sinniger, J., Dirrig-Grosch, S., et al. (2016). Toxic gain of function from mutant FUS protein is crucial to trigger cell autonomous motor neuron loss. *The EMBO Journal*, 35 (10), pp.1077–1097.

Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T. and Lührmann, R. (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *The EMBO Journal*, 23 (15), pp.3000–3009.

Schramm, L. and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes & development*, 16 (20), pp.2593–2620.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, 473 (7347), pp.337–342.

Sellier, C., Campanari, M., Julie Corbier, C., Gaucherot, A., Kolb-Cheynel, I., Oulad-Abdelghani, M., Ruffenach, F., Page, A., Ciura, S., Kabashi, E., et al. (2016). Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. *The EMBO Journal*, 35 (12), pp.1276–1297.

Seok Ko, H., Uehara, T., Tsuruma, K. and Nomura, Y. (2004). Ubiquilin interacts with ubiquitylated proteins and proteasome through its ubiquitin-associated and ubiquitin-like domains. *FEBS Letters*, 566 (1–3), pp.110–114.

Shalom, O., Shalva, N., Altschuler, Y. and Motro, B. (2008). The mammalian Nek1 kinase is involved in primary cilium formation. *FEBS Letters*, 582 (10), pp.1465–1470.

Shanbhag, R., Kurabi, A., Kwan, J. J. and Donaldson, L. W. (2010). Solution structure of the carboxy-terminal Tudor domain from human Coilin. *FEBS Letters*, 584 (20), pp.4351–4356.

Sharma, A., Lyashchenko, A. K., Lu, L., Nasrabady, S. E., Elmaleh, M., Mendelsohn, M., Nemes, A., Tapia, J. C., Mentis, G. Z. and Shneider, N. A. (2016). ALS-associated mutant FUS induces selective motor neuron degeneration through toxic gain of function. *Nature Communications*, 7, p.10465.

Shatunov, A., Mok, K., Newhouse, S., Weale, M. E., Smith, B., Vance, C., Johnson, L., Veldink, J. H., van Es, M. A., van den Berg, L. H., et al. (2010). Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurology*, 9 (10), pp.986–994.

Shav-Tal, Y., Blechman, J., Darzacq, X., Montagna, C., Dye, B. T., Patton, J. G., Singer, R. H. and Zipori, D. (2005). Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. *Molecular biology of the cell*, 16 (5), pp.2395–2413.

Shaw, P. J., Forrest, V., Ince, P. G., Richardson, J. P. and Wastell, H. J. (1995). CSF and plasma amino acid levels in motor neuron disease: elevation of CSF glutamate in a subset of patients. *Neurodegeneration : a journal for neurodegenerative disorders, neuroprotection, and neuroregeneration*, 4 (2), pp.209–216.

Shaw, P. J. and Ince, P. G. (1997). Glutamate, excitotoxicity and amyotrophic lateral sclerosis. *Journal of Neurology*, 244 (S2), pp.S3–S14.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature Methods*, 11 (4), pp.399–402.

Shevtsov, S. P. and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. *Nature Cell Biology*, 13 (2), pp.167–173.

Shi, C.-S., Shenderov, K., Huang, N.-N., Kabat, J., Abu-Asab, M., Fitzgerald, K. A., Sher, A. and Kehrl, J. H. (2012). Activation of autophagy by inflammatory signals limits IL-1β production by targeting ubiquitinated inflammasomes for destruction. *Nature Immunology*, 13 (3), pp.255–263.

Shi, K. Y., Mori, E., Nizami, Z. F., Lin, Y., Kato, M., Xiang, S., Wu, L. C., Ding, M., Yu, Y., Gall, J. G., et al. (2017). Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proceedings of the National Academy of Sciences of the United States of America*, 114 (7), pp.E1111–E1117.

Shi, Y., Lin, S., Staats, K. A., Li, Y., Chang, W.-H., Hung, S.-T., Hendricks, E., Linares, G. R., Wang, Y., Son, E. Y., et al. (2018). Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nature Medicine*.

Shuman, S. (1994). Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *The Journal of biological chemistry*, 269 (51), pp.32678–32684.

Simone, R., Balendra, R., Moens, T. G., Preza, E., Wilson, K. M., Heslegrave, A., Woodling, N. S., Niccoli, T., Gilbert-Jaramillo, J., Abdelkarim, S., et al. (2017). G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO molecular medicine*, p.e201707850.

Singh, R. and Reddy, R. (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 86 (21), pp.8280–8283.

Sleeman, J. E., Ajuh, P. and Lamond, A. I. (2001). snRNP protein expression enhances the formation of Cajal bodies containing p80-coilin and SMN. *Journal of Cell Science*, 114 (Pt 24), pp.4407–4419.

Sleeman, J. E. and Lamond, A. I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Current Biology*, 9 (19), pp.1065–1074.

Sleeman, J. E., Trinkle-Mulcahy, L., Prescott, A. R., Ogg, S. C. and Lamond, A. I. (2003). Cajal body proteins SMN and Coilin show differential dynamic behaviour in vivo. *Journal of Cell Science*, 116 (Pt 10), pp.2039–2050.

Smith, B. N., Ticozzi, N., Fallini, C., Gkazi, A. S., Topp, S., Kenna, K. P., Scotter, E. L., Kost, J., Keagle, P., Miller, J. W., et al. (2014). Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron*, 84 (2), pp.324–331.

Smith, B. N., Topp, S. D., Fallini, C., Shibata, H., Chen, H.-J., Troakes, C., King, A., Ticozzi, N., Kenna, K. P., Soragia-Gkazi, A., et al. (2017). Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral sclerosis. *Science translational medicine*, 9 (388), p.eaad9157.

Smith, K. P., Carter, K. C., Johnson, C. V. and Lawrence, J. B. (1995). U2 and U1 snRNA gene loci associate with coiled bodies. *Journal of Cellular Biochemistry*, 59 (4), pp.473–485.

Smith, K. P. and Lawrence, J. B. (2000). Interactions of U2 gene loci and their nuclear transcripts with Cajal (coiled) bodies: evidence for PreU2 within Cajal bodies. *Molecular biology of the cell*, 11 (9), pp.2987–2998.

Smith, R. A., Miller, T. M., Yamanaka, K., Monia, B. P., Condon, T. P., Hung, G., Lobsiger, C. S., Ward, C. M., McAlonis-Downes, M., Wei, H., et al. (2006). Antisense oligonucleotide therapy for neurodegenerative disease. *Journal of Clinical Investigation*, 116 (8), pp.2290–2296.

Smogorzewska, A. and de Lange, T. (2004). Regulation of Telomerase by Telomeric Proteins. *Annual Review of Biochemistry*, 73 (1), pp.177–208.

Smoliński, D. J., Wróbel, B., Noble, A., Zienkiewicz, A. and Górska-Brylass, A. (2011). Periodic expression of Sm proteins parallels formation of nuclear Cajal bodies and cytoplasmic snRNP-rich bodies. *Histochemistry and cell biology*, 136 (5), pp.527–541.

Sofroniew, M. V and Vinters, H. V. (2010). Astrocytes: biology and pathology. *Acta Neuropathologica*, 119 (1), pp.7–35.

Song, S., Miranda, C. J., Braun, L., Meyer, K., Frakes, A. E., Ferraiuolo, L., Likhite, S., Bevan, A. K., Foust, K. D., McConnell, M. J., et al. (2016). Major histocompatibility complex class I molecules protect motor neurons from astrocyte-induced toxicity in amyotrophic lateral sclerosis. *Nature Medicine*, 22 (4), pp.397–403.

Spector, D. L. and Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harbor perspectives in Biology*, 3 (2).

Spector, D. L., Lark, G. and Huang, S. (1992). Differences in snRNP localization between transformed and nontransformed cells. *Molecular biology of the cell*, 3 (5), pp.555–569.

Sprangers, R., Groves, M. R., Sinning, I. and Sattler, M. (2003). High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *Journal of Molecular Biology*, 327 (2), pp.507–520.

Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J. C., Williams, K. L., Buratti, E., et al. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science*, 319 (5870), pp.1668–1672.

Staněk, D. and Neugebauer, K. M. (2004). Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. *The Journal of cell biology*, 166 (7), pp.1015–1025.

Staněk, D., Rader, S. D., Klingauf, M. and Neugebauer, K. M. (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *The Journal of Cell Biology*, 160 (4), pp.505–516.

Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. *Nature Reviews Molecular Cell Biology*, 10 (8), pp.513–525.

Stern, J. L., Zyner, K. G., Pickett, H. A., Cohen, S. B. and Bryan, T. M. (2012). Telomerase recruitment requires both TCAB1 and Cajal bodies independently. *Molecular and Cellular Biology*, 32 (13), pp.2384–2395.

Stewart, M. D., Li, J. and Wong, J. (2005). Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. *Molecular and Cellular Biology*, 25 (7), pp.2525–2538.

Stoneley, M. and Willis, A. E. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene*, 23 (18), pp.3200–3207.

Stopford, M. J., Higginbottom, A., Hautbergue, G. M., Cooper-Knock, J., Mulcahy, P. J., De Vos, K. J., Renton, A. E., Pliner, H., Calvo, A., Chio, A., et al. (2017). C9ORF72 hexanucleotide repeat exerts toxicity in a stable, inducible motor neuronal cell model, which is rescued by partial depletion of Pten. *Human Molecular Genetics*, 26 (6), pp.1133–1145.

Stricker, J., Falzone, T. and Gardel, M. L. (2010). Mechanics of the F-actin cytoskeleton. *Journal of Biomechanics*, 43 (1), pp.9–14.

Strzelecka, M., Oates, A. C. and Neugebauer, K. (2010). Dynamic control of Cajal body number during zebrafish embryogenesis. *Nucleus*, 1 (1), pp.96–108.

Sullivan, P. M., Zhou, X., Robins, A. M., Paushter, D. H., Kim, D., Smolka, M. B., Hu, F., Achi, E., Rudnicki, S., Atanasio, A., et al. (2016). The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. *Acta Neuropathologica Communications*, 4 (1), p.51.

Sun, S., Ling, S.-C., Qiu, J., Albuquerque, C. P., Zhou, Y., Tokunaga, S., Li, H., Qiu, H., Bui, A., Yeo, G. W., 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, p.6171.

Suzuki, N., Maroof, A. M., Merkle, F. T., Koszka, K., Intoh, A., Armstrong, I., Moccia, R., Davis-Dusenbery, B. N. and Eggan, K. (2013). The mouse C9ORF72 ortholog is enriched in neurons known to degenerate in ALS and FTD. *Nature Neuroscience*, 16 (12), pp.1725–1727.

Suzuki, T., Izumi, H. and Ohno, M. (2010). Cajal body surveillance of U snRNA export complex assembly. *The Journal of Cell Biology*, 190 (4), pp.603–612.

Swinnen, B., Bento-Abreu, A., Gendron, T. F., Boeynaems, S., Bogaert, E., Nuyts, R., Timmers, M., Scheveneels, W., Hersmus, N., Wang, J., et al. (2018). A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta Neuropathologica*, pp.1–17.

Takata, H., Nishijima, H., Maeshima, K. and Shibahara, K. -i. (2012). The integrator complex is required for integrity of Cajal bodies. *Journal of Cell Science*, 125 (1), pp.166–175.

Takeda, D. Y., Wohlschlegel, J. A. and Dutta, A. (2001). A bipartite substrate recognition motif for cyclin-dependent kinases. *The Journal of biological chemistry*, 276 (3), pp.1993–1997.

Tan, C.-F., Eguchi, H., Tagawa, A., Onodera, O., Iwasaki, T., Tsujino, A., Nishizawa, M., Kakita, A. and Takahashi, H. (2007). TDP-43 immunoreactivity in neuronal inclusions in familial amyotrophic lateral sclerosis with or without SOD1 gene mutation. *Acta Neuropathologica*, 113 (5), pp.535–542.

Tanaka, Y., Nonaka, T., Suzuki, G., Kametani, F. and Hasegawa, M. (2016). Gain-of-function profilin 1 mutations linked to familial amyotrophic lateral sclerosis cause seed-dependent intracellular TDP-43 aggregation. *Human Molecular Genetics*, 25 (7), pp.1420–1433.

Taneja, K. L., McCurrach, M., Schalling, M., Housman, D. and Singer, R. H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *The Journal of Cell Biology*, 128 (6), pp.995–1002.

Tao, Z., Wang, H., Xia, Q., Li, K., Li, K., Jiang, X., Xu, G., Wang, G. and Ying, Z. (2015). Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Human Molecular Genetics*, 24 (9), pp.2426–2441.

Tapia, O., Bengoechea, R., Berciano, M. T. and Lafarga, M. (2010). Nucleolar targeting of coilin is regulated by its hypomethylation state. *Chromosoma*, 119 (5), pp.527–540.

Tapia, O., Narcís, J. O., Riancho, J., Tarabal, O., Piedrafita, L., Calderó, J., Berciano, M. T. and Lafarga, M. (2017). Cellular bases of the RNA metabolism dysfunction in motor neurons of a murine model of spinal muscular atrophy: Role of Cajal bodies and the nucleolus. *Neurobiology of Disease*, 108, pp.83–99.

Tarn, W. Y. and Steitz, J. A. (1996). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science*, 273 (5283), pp.1824–1832.

Teyssou, E., Takeda, T., Lebon, V., Boillée, S., Doukouré, B., Bataillon, G., Sazdovitch, V., Cazeneuve, C., Meininger, V., LeGuern, E., et al. (2013). Mutations in SQSTM1 encoding p62 in amyotrophic lateral sclerosis: genetics and neuropathology. *Acta Neuropathologica*, 125 (4), pp.511–522.

Thandapani, P., O'Connor, T. R., Bailey, T. L. and Richard, S. (2013). Defining the RGG/RG Motif. *Molecular Cell*, 50 (5), pp.613–623.

Thayanidhi, N., Liang, Y., Hasegawa, H., Nycz, D. C., Oorschot, V., Klumperman, J. and Hay, J. C. (2012). R-SNARE ykt6 resides in membrane-associated protease-resistant protein particles and modulates cell cycle progression when over-expressed. *Biology of the Cell*, 104 (7), pp.397–417.

Theimer, C. A., Jády, B. E., Chim, N., Richard, P., Breece, K. E., Kiss, T. and Feigon, J. (2007). Structural and Functional Characterization of Human Telomerase RNA Processing and Cajal Body Localization Signals. *Molecular Cell*, 27 (6), pp.869–881.

Therrien, M., Rouleau, G. A., Dion, P. A. and Parker, J. A. (2013). Deletion of C9ORF72 results in motor neuron degeneration and stress sensitivity in C. elegans. *PLoS One*, 8 (12), p.e83450.

Timney, B. L., Raveh, B., Mironska, R., Trivedi, J. M., Kim, S. J., Russel, D., Wente, S. R., Sali, A. and Rout, M. P. (2016). Simple rules for passive diffusion through the nuclear pore complex. *The Journal of cell biology*, 215 (1), pp.57–76.

Tomlinson, R. L., Ziegler, T. D., Supakorndej, T., Terns, R. M. and Terns, M. P. (2006). Cell cycle-regulated trafficking of human telomerase to telomeres. *Molecular biology of the cell*, 17 (2), pp.955–965.

Tomonaga, M., Saito, M., Yoshimura, M., Shimada, H. and Tohgi, H. (1978). Ultrastructure of the Bunina bodies in anterior horn cells of amyotrophic lateral sclerosis. *Acta Neuropathologica*, 42 (2), pp.81–86.

Toyama, B. H., Savas, J. N., Park, S. K., Harris, M. S., Ingolia, N. T., Yates, J. R. and Hetzer, M. W. (2013). Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell*, 154 (5), pp.971–982.

Toyota, C. G., Davis, M. D., Cosman, A. M. and Hebert, M. D. (2010). Coilin phosphorylation mediates interaction with SMN and SmB′. *Chromosoma*, 119 (2), pp.205–215.

Tradewell, M. L., Yu, Z., Tibshirani, M., Boulanger, M.-C., Durham, H. D. and Richard, S. (2012). Arginine methylation by PRMT1 regulates nuclear-cytoplasmic localization and toxicity of FUS/TLS harbouring ALS-linked mutations. *Human Molecular Genetics*, 21 (1), pp.136–149.

Tran, H., Almeida, S., Moore, J., Gendron, T. F., Chalasani, U., Lu, Y., Du, X., Nickerson, J. A., Petrucelli, L., Weng, Z., et al. (2015). Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron*, 87 (6), pp.1207–1214.

Trinkle-Mulcahy, L. and Sleeman, J. E. (2017). The Cajal body and the nucleolus: 'In a relationship' or 'It's complicated'? *RNA Biology*, 14 (6), pp.739–751.

Tsai, G. C., Stauch-Slusher, B., Sim, L., Hedreen, J. C., Rothstein, J. D., Kuncl, R. and Coyle, J. T. (1991). Reductions in acidic amino acids and N-acetylaspartylglutamate in amyotrophic lateral sclerosis CNS. *Brain research*, 556 (1), pp.151–156.

Tsuiji, H., Iguchi, Y., Furuya, A., Kataoka, A., Hatsuta, H., Atsuta, N., Tanaka, F., Hashizume, Y., Akatsu, H., Murayama, S., et al. (2013). Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. *EMBO molecular medicine*, 5 (2), pp.221–234.

Tucker, K. E., Berciano, M. T., Jacobs, E. Y., LePage, D. F., Shpargel, K. B., Rossire, J. J., Chan, E. K., Lafarga, M., Conlon, R. A. and Matera, A. G. (2001). Residual Cajal bodies in coilin knockout mice fail to recruit Sm snRNPs and SMN, the spinal muscular atrophy gene product. *The Journal of Cell Biology*, 154 (2), pp.293–307.

Turner, M. ., Cagnin, A., Turkheimer, F. ., Miller, C. C. ., Shaw, C. ., Brooks, D. ., Leigh, P. . and Banati, R. . (2004). Evidence of widespread cerebral microglial activation in amyotrophic lateral sclerosis: an [11C](R)-PK11195 positron emission tomography study. *Neurobiology of Disease*, 15 (3), pp.601–609.

Turunen, J. J., Niemelä, E. H., Verma, B. and Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley interdisciplinary reviews. RNA*, 4 (1), pp.61–76.

Tycowski, K. T., Shu, M.-D., Kukoyi, A. and Steitz, J. A. (2009). A Conserved WD40 Protein Binds the Cajal Body Localization Signal of scaRNP Particles. *Molecular Cell*, 34 (1), pp.47–57.

Uenal, H., Rosenbohm, A., Kufeldt, J., Weydt, P., Goder, K., Ludolph, A., Rothenbacher, D., Nagel, G. and Group, and the A. registry S. (2014). Incidence and Geographical Variation of Amyotrophic Lateral Sclerosis (ALS) in Southern Germany – Completeness of the ALS Registry Swabia. Raoul, C. (Ed). *PLoS ONE*, 9 (4), p.e93932.

van Blitterswijk, M., DeJesus-Hernandez, M. and Rademakers, R. (2012). How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? *Current opinion in neurology*, 25 (6), pp.689–700.

Van Den Bosch, L., Van Damme, P., Bogaert, E. and Robberecht, W. (2006). The role of excitotoxicity in the pathogenesis of amyotrophic lateral sclerosis. *Biochimica et Biophysica Acta (BBA)*, 1762 (11–12), pp.1068–1082.

van Rheenen, W., Shatunov, A., Dekker, A. M., McLaughlin, R. L., Diekstra, F. P., Pulit, S. L., van der Spek, R. A. A., Võsa, U., de Jong, S., Robinson, M. R., et al. (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nature Genetics*, 48 (9), pp.1043–1048.

Valbuena, A., López-Sánchez, I. and Lazo, P. A. (2008). Human VRK1 Is an Early Response Gene and Its Loss Causes a Block in Cell Cycle Progression. Williams, S. (Ed). *PLoS ONE*, 3 (2), p.e1642.

Vance, C., Al-Chalabi, A., Ruddy, D., Smith, B. N., Hu, X., Sreedharan, J., Siddique, T., Schelhaas, H. J., Kusters, B., Troost, D., et al. (2006). Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain*, 129 (Pt 4), pp.868–876.

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K. J., Nishimura, A. L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science*, 323 (5918), pp.1208–1211.

Vande Velde, C., Miller, T. M., Cashman, N. R. and Cleveland, D. W. (2008). Selective association of misfolded ALS-linked mutant SOD1 with the cytoplasmic face of mitochondria. *Proceedings of the National Academy of Sciences*, 105 (10), pp.4022–4027.

Venteicher, A. S., Abreu, E. B., Meng, Z., McCann, K. E., Terns, R. M., Veenstra, T. D., Terns, M. P. and Artandi, S. E. (2009). A Human Telomerase Holoenzyme Protein Required for Cajal Body Localization and Telomere Synthesis. *Science*, 323 (5914), pp.644–648.

Verheggen, C., Lafontaine, D. L. J., Samarsky, D., Mouaikel, J., Blanchard, J.-M., Bordonné, R. and Bertrand, E. (2002). Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *The EMBO Journal*, 21 (11), pp.2736–2745.

Vogel, C., Abreu, R. de S., Ko, D., Le, S.-Y., Shapiro, B. A., Burns, S. C., Sandhu, D., Boutz, D. R., Marcotte, E. M. and Penalva, L. O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular systems biology*, 6 (1), p.400.

Wagner, E. J. and Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology*, 13 (2), pp.115–126.

Waite, A. J., Baumer, D., East, S., Neal, J., Morris, H. R., Ansorge, O. and Blake, D. J. (2014). Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiology of Aging*, 35 (7), p.1779.e5-1779.e13.

Walker, C., Herranz-Martin, S., Karyka, E., Liao, C., Lewis, K., Elsayed, W., Lukashchuk, V., Chiang, S.-C., Ray, S., Mulcahy, P. J., et al. (2017). C9orf72 expansion disrupts ATM-mediated chromosomal break repair. *Nature Neuroscience*, 20 (9), pp.1225–1235.

Wang, I.-F., Wu, L.-S., Chang, H.-Y. and Shen, C.-K. J. (2008). TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor. *Journal of Neurochemistry*, 105 (3), pp.797–806.

Wang, I. F., Reddy, N. M. and Shen, C. K. (2002). Higher order arrangement of the eukaryotic nuclear bodies. *Proceedings of the National Academy of Sciences*, 99 (21), pp.13583–13588.

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

Wang, W., Li, L., Lin, W.-L., Dickson, D. W., Petrucelli, L., Zhang, T. and Wang, X. (2013a). The ALS disease-associated mutant TDP-43 impairs mitochondrial dynamics and function in motor neurons. *Human Molecular Genetics*, 22 (23), pp.4706–4719.

Wang, W. Y., Pan, L., Su, S. C., Quinn, E. J., Sasaki, M., Jimenez, J. C., Mackenzie, I. R. A., Huang, E. J. and Tsai, L.-H. (2013b). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. *Nature Neuroscience*, 16 (10), pp.1383–1391.

Wang, X., Schwartz, J. C. and Cech, T. R. (2015). Nucleic acid-binding specificity of human FUS protein. *Nucleic Acids Research*, 43 (15), pp.7535–7543.

Warita, H., Hayashi, T., Murakami, T., Manabe, Y. and Abe, K. (2001). Oxidative damage to mitochondrial DNA in spinal motoneurons of transgenic ALS mice. *Brain research. Molecular brain research*, 89 (1–2), pp.147–152.

Webster, C. P., Smith, E. F., Bauer, C. S., Moller, A., Hautbergue, G. M., Ferraiuolo, L., Myszczynska, M. A., Higginbottom, A., Walsh, M. J., Whitworth, A. J., et al. (2016). The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *The EMBO Journal*, 11 (12), p.e201694401.

Weis, K. (2003). Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell*, 112 (4), pp.441–451.

Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., Lin, S., Shneider, N. A., Monaghan, J., Pandey, U. B., et al. (2014). Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron*, 84 (6), pp.1213–1225.

Wennerberg, K., Rossman, K. L. and Der, C. J. (2005). The Ras superfamily at a glance. *Journal of Cell Science*, 118 (Pt 5), pp.843–846.

Whittom, A. A., Xu, H. and Hebert, M. D. (2008). Coilin levels and modifications influence artificial reporter splicing. *Cellular and Molecular Life Sciences*, 65 (7–8), pp.1256–1271.

Wiedemann, F. R., Manfredi, G., Mawrin, C., Beal, M. F. and Schon, E. A. (2002). Mitochondrial DNA and respiratory chain function in spinal cords of ALS patients. *Journal of Neurochemistry*, 80 (4), pp.616–625.

Will, C. L. and Luhrmann, R. (2011). Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, 3 (7), pp.a003707–a003707.

Wils, H., Kleinberger, G., Janssens, J., Pereson, S., Joris, G., Cuijt, I., Smits, V., Ceuterick-de Groote, C., Van Broeckhoven, C. and Kumar-Singh, S. (2010). TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proceedings of the National Academy of Sciences*, 107 (8), pp.3858–3863.

Winer, L., Srinivasan, D., Chun, S., Lacomis, D., Jaffa, M., Fagan, A., Holtzman, D. M., Wancewicz, E., Bennett, C. F., Bowser, R., et al. (2013). SOD1 in Cerebral Spinal Fluid as a Pharmacodynamic Marker for Antisense Oligonucleotide Therapy. *JAMA Neurology*, 70 (2), p.201.

Wohlschlegel, J. A., Dwyer, B. T., Takeda, D. Y. and Dutta, A. (2001). Mutational analysis of the Cy motif from p21 reveals sequence degeneracy and specificity for different cyclin-dependent kinases. *Molecular and cellular biology*, 21 (15), pp.4868–4874.

Wong, P. C., Pardo, C. A., Borchelt, D. R., Lee, M. K., Copeland, N. G., Jenkins, N. A., Sisodia, S. S., Cleveland, D. W. and Price, D. L. (1995). An adverse property of a familial ALS-linked SOD1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria. *Neuron*, 14 (6), pp.1105–1116.

Wu, C. H., Fallini, C., Ticozzi, N., Keagle, P. J., Sapp, P. C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D. M., et al. (2012a). Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature*, 488 (7412), pp.499–503.

Wu, L., Cheng, W.-C. and Shen, C.-K. J. (2012b). Targeted depletion of TDP-43 expression in the spinal cord motor neurons leads to the development of amyotrophic lateral sclerosis-like phenotypes in mice. *The Journal of biological chemistry*, 287 (33), pp.27335–27344.

Wu, Q. and Krainer, A. R. (1997). Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs. *RNA*, 3 (6), pp.586–601.

Xiao, S., MacNair, L., McGoldrick, P., McKeever, P. M., McLean, J. R., Zhang, M., Keith, J., Zinman, L., Rogaeva, E. and Robertson, J. (2015). Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. *Annals of Neurology*, 78 (4), pp.568–583.

Xiao, S., Sanelli, T., Dib, S., Sheps, D., Findlater, J., Bilbao, J., Keith, J., Zinman, L., Rogaeva, E. and Robertson, J. (2011). RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Molecular and Cellular Neuroscience*, 47 (3), pp.167–180.

Xu, H., Pillai, R. S., Azzouz, T. N., Shpargel, K. B., Kambach, C., Hebert, M. D., Schumperli, D. and Matera, A. G. (2005). The C-terminal domain of coilin interacts with Sm proteins and U snRNPs. *Chromosoma*, 114 (3), pp.155–166.

Xu, Z., Henderson, R. D., David, M. and McCombe, P. A. (2016). Neurofilaments as Biomarkers for Amyotrophic Lateral Sclerosis: A Systematic Review and Meta-Analysis. *PLoS One*, 11 (10), p.e0164625.

Yamada, M., Wood, J. D., Shimohata, T., Hayashi, S., Tsuji, S., Ross, C. A. and Takahashi, H. (2001). Widespread occurrence of intranuclear atrophin-1 accumulation in the central nervous system neurons of patients with dentatorubral-pallidoluysian atrophy. *Annals of Neurology*, 49 (1), pp.14–23.

Yamakawa, M., Ito, D., Honda, T., Kubo, K. I., Noda, M., Nakajima, K. and Suzuki, N. (2014). Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Human Molecular Genetics*, 24 (6), pp.1630–1645.

Yamazaki, T., Chen, S., Yu, Y., Yan, B., Haertlein, T. C., Carrasco, M. A., Tapia, J. C., Zhai, B., Das, R., Lalancette-Hebert, M., et al. (2012). FUS-SMN protein interactions link the motor neuron diseases ALS and SMA. *Cell reports*, 2 (4), pp.799–806.

Yang, L., Gal, J., Chen, J. and Zhu, H. (2014). Self-assembled FUS binds active chromatin and regulates gene transcription. *Proceedings of the National Academy of Sciences*, 111 (50), pp.17809–17814.

Yang, Y., Hentati, A., Deng, H. X., Dabbagh, O., Sasaki, T., Hirano, M., Hung, W. Y., Ouahchi, K., Yan, J., Azim, A. C., et al. (2001). The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nature Genetics*, 29 (2), pp.160–165.

Yasuda, K., Clatterbuck-Soper, S. F., Jackrel, M. E., Shorter, J. and Mili, S. (2017). FUS inclusions disrupt RNA localization by sequestering kinesin-1 and inhibiting microtubule detyrosination. *The Journal of cell biology*, 216 (4), pp.1015–1034.

Yin, S., Lopez-Gonzalez, R., Kunz, R. C., Gangopadhyay, J., Borufka, C., Gygi, S. P., Gao, F.-B. and Reed, R. (2017). Evidence that C9ORF72 Dipeptide Repeat Proteins Associate with U2 snRNP to Cause Mis-splicing in ALS/FTD Patients. *Cell reports*, 19 (11), pp.2244–2256.

Young, P. J., Le, T. T., Dunckley, M., thi Man, N., Burghes, A. H. M. and Morris, G. E. (2001). Nuclear Gems and Cajal (Coiled) Bodies in Fetal Tissues: Nucleolar Distribution of the Spinal Muscular Atrophy Protein, SMN. *Experimental Cell Research*, 265 (2), pp.252–261.

Young, P. J., Le, T. T., thi Man, N., Burghes, A. H. and Morris, G. E. (2000). The relationship between SMN, the spinal muscular atrophy protein, and nuclear coiled bodies in differentiated tissues and cultured cells. *Experimental Cell Research*, 256 (2), pp.365–374.

Yu, Y., Chi, B., Xia, W., Gangopadhyay, J., Yamazaki, T., Winkelbauer-Hurt, M. E., Yin, S., Eliasse, Y., Adams, E., Shaw, C. E., et al. (2015). U1 snRNP is mislocalized in ALS patient fibroblasts bearing NLS mutations in FUS and is required for motor neuron outgrowth in zebrafish. *Nucleic Acids Research*, 43 (6), pp.3208–3218.

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.

Zaric, B., Chami, M., Rémigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F. K. and Kambach, C. (2005). Reconstitution of Two Recombinant LSm Protein Complexes Reveals Aspects of Their Architecture, Assembly, and Function. *Journal of Biological Chemistry*, 280 (16), pp.16066–16075.

Zhang, D., Iyer, L. M., He, F. and Aravind, L. (2012). Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. *Frontiers in Genetics*, 3, p.283.

Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., Daley, E. L., Miller, S. J., Cunningham, K. M., Vidensky, S., et al. (2015). The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*, 525 (7567), pp.56–61.

Zhang, Y. J., Gendron, T. F., Grima, J. C., Sasaguri, H., Jansen-West, K., Xu, Y.-F., Katzman, R. B., Gass, J., Murray, M. E., Shinohara, M., et al. (2016). C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nature Neuroscience*, 19 (5), pp.668–677.

Zhang, Y. J., Jansen-West, K., Xu, Y. F., Gendron, T. F., Bieniek, K. F., Lin, W. L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., et al. (2014). Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathologica*, 128 (4), pp.505–524.

Zhou, Y., Liu, S., Liu, G., Öztürk, A. and Hicks, G. G. (2013). ALS-Associated FUS Mutations Result in Compromised FUS Alternative Splicing and Autoregulation. Scott, H. S. (Ed). *PLoS Genetics*, 9 (10), p.e1003895.

Zhu, Y., Tomlinson, R. L., Lukowiak, A. A., Terns, R. M. and Terns, M. P. (2004). Telomerase RNA accumulates in Cajal bodies in human cancer cells. *Molecular biology of the cell*, 15 (1), pp.81–90.

Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. *Proceedings of the National Academy of Sciences*, 108 (1), pp.260–265.

Zu, T., Liu, Y., Banez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., et al. (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proceedings of the National Academy of Sciences*, 110 (51), pp.E4968-77.