In vitro and pathological investigation of dipeptide repeat proteins in *C9orf72* linked frontotemporal dementia and amyotrophic lateral sclerosis

Lauren Marie Gittings

2018

Department of Neurodegenerative Diseases

UK Dementia Research Institute at UCL

UCL Queen Square Institute of Neurology

University College London

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy from University College London

Declaration

I, Lauren Marie Gittings, confirm that the work presented in this thesis is my own. Where information has been derived from other sources or work has been produced collaboratively, I confirm that this has been indicated in the text. This thesis is not substantially the same as any other that may have been submitted for a qualification at any other university or similar institution. It does not exceed the limit of 100,000 words.

_____

Acknowledgements

Much of the work in this thesis has been performed using brain tissue or cells that have been generously donated from patients who have experienced the effects of neurodegenerative diseases. I would like to thank these people for their generous donations and hope that the research performed using their gifts will help in finding treatments for these diseases. I would like to dedicate this thesis to all those whose lives have been affected by dementia, especially my granddad, John Houghton.

My PhD experience has been as much about building my resilience as it has been about expanding my scientific knowledge and I couldn't have done it without the support, guidance and love of many people, to whom I am extremely grateful.

Firstly, I would like to thank my primary supervisor, Professor Adrian Isaacs, for his supervisory guidance and enthusiasm for science. Thank you for giving me the opportunity to work with your group on several exciting projects. I am grateful for the time you have dedicated to supervising me, the tasks you have challenged me with, and for your valuable scientific input to my projects.

Next, I would like to say thank you to my secondary supervisor, Dr Tammaryn Lashley. I am thankful that I was able to be supervised by you and have loved working with you and your group. Thank you for always making the time for me, no matter how busy you are or how small my problem may be. I am always grateful to receive your advice and I am continuously inspired by the passion, dedication and love you have for your work and family. Thank you for all the laughter and encouragement you have provided me with throughout my PhD and for always believing in my abilities, even when I don't believe in them myself. I will be forever grateful to you for helping me re-discover my passion for science…and an obsession with step counting!

I would also like to say thank you to my initial supervisors, Professor Mike Cheetham and Professor Alison Hardcastle. It was a great experience working with you both and I am grateful for the kindness and care you gave me during the first year of my PhD. Thank you for your supervision and for supporting me in my decision to change projects.

friendship and support over the years. My thanks also go to Dr Jim Bellingham and Dr Katarina Jovanovic for helping with the molecular cloning in Chapter 3.

I would like to say a special thank you to Dr Agnieszka Ule for proof reading sections of my thesis and for giving me feedback comments designed to make me laugh during the doom and gloom of thesis writing.

Thank you, also, to everyone else who has helped and supported me in many other ways over the years. I owe a great deal of thanks to the many post-docs, technicians and students I have worked with in various research groups during my time at UCL. I feel incredibly lucky to have worked with such an amazing group of scientists and even luckier that I have formed close friendships with many of these people. In particular, I would like to thank Martha Foiani, Chris Lovejoy, Dr Giulia Tzyzack, and Dr Selina Wray.

Thank you to the Wolfson Foundation for providing the funding to make my PhD possible, and to Professor Nick Wood and Professor Nick Fox for leading the Wolfson Neurodegeneration PhD program. I'm very grateful to you for giving me the opportunity to pursue research I love, and I have valued your support and words of encouragement throughout my PhD. Thank you also to all the other Wolfson funded PhD students for being a great network of support.

I owe a great deal of thanks to Elizabeth Halton for her administrative and pastoral support over the last four years. I don't think I could have made it to the end of my PhD without your kindness, help and encouragement. Thank you for your patience and the many hours you have spent listening to me and helping me solve my problems. I am extremely grateful for all of your insightful guidance regarding my PhD, life & yoga. Thank you for your friendship and the many fun times we spent together. Thank you also for the brilliant job you do of co-ordinating the Wolfson PhD programme. This has been of invaluable help to me and I admire your dedication to helping your students succeed.

Thank you to Professor Elizabeth Barley for helping equip me with strategies to deal with the stresses of a PhD, and life in general.

Thank you to all of my wonderful friends for being incredibly supportive throughout my PhD, for reminding me there is a life full of laughter and enjoyment outside of the lab and for being so understanding whenever I had to change plans at the last minute to 'feed my cells'. I would like to especially thank the friends that have put up with living with me at some point over the last four years; Lee Carter, Demelza Smeeth, Jenny Doyle and Lucy Bielby. I feel incredibly lucky to have been supported throughout my PhD by loving friends that I can come home to and have fun with after a long day in the lab.

Finally, thank you to my wonderfully loving family; Michèle Harkins, Phil Gittings and Tom Gittings (with a special mention to Sophie Lee). I couldn't have gone this far in my education without your love and support. You have been my constant source of encouragement, inspiration and motivation. I feel very lucky to have been able to escape my PhD focused world to enjoy weekends with you back in the New Forest full of laughter, cycling and cheesecake. Thank you for everything that you have done for me to make achieving my PhD possible and for always being there to remind me that doing my best is all that I can do.

And last, but not least, thank you to my favourite fluffy dog, Lacey, for all the therapeutic walks, distractions and cuddles whilst writing this thesis. You are the best (although, also the laziest) thesis writing companion I could wish for!

Abstract

An intronic hexanucleotide repeat expansion in the *C9orf72* gene is the most common genetic cause of two clinically, genetically and pathologically overlapping neurodegenerative disorders, frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS). The pathogenic mechanism of disease is currently unknown, but it is hypothesised that toxicity may, at least in part, be attributed to dipeptide repeat (DPR) proteins that are produced as a result of unconventional repeat associated non-ATG initiated (RAN) translation of the hexanucleotide repeat. In this thesis, cellular models and post-mortem tissue from *C9orf72* mutation patients have been used to further characterise various properties of the DPR proteins that have been previously shown to be the most toxic; poly-glycine-alanine (poly-GA), poly-proline-arginine (poly-PR) and poly-glycine-arginine (poly-GR). Overexpression of DPR proteins in SK-N-SH cells was used to assess the ability of a molecular chaperone to reduce the aggregation of these proteins, and CRISPR-Cas9 technology was used to fluorescently tag endogenous DPR proteins in *C9orf72* patient induced pluripotent stem cells. In addition to cellular models, post-mortem brain tissue from *C9orf72* patients was used to identify the presence and characterise the pathological significance of methylated forms of the arginine-containing DPR proteins, poly-PR and poly-GR. Additionally, this thesis contains research on another pathological subtype of FTD; FTLD-FUS. Immunohistochemistry was used to identify and characterise two abundant RNA binding proteins, hnRNP R and Q, which were found to accumulate in pathological inclusions specifically in FTLD-FUS patient post-mortem tissue.

Impact Statement

Frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) are two clinically, genetically and pathologically overlapping neurodegenerative diseases for which there is currently no effective treatment or cure. Not only do these diseases have debilitating consequences for the patients and carers of those suffering from the disease, they also place a large social and economic burden on society. Understanding the biological mechanism responsible for causing a disease is crucial in the development of effective therapies. In 2011, the most common genetic cause of FTD and ALS was identified as a hexanucleotide repeat expansion mutation in the *C9orf72* gene, however the biological mechanism by which the *C9orf72* mutation causes FTD and ALS is currently unknown, although three non-mutually exclusive hypotheses have been proposed. The work in this thesis primarily focuses on one of these hypotheses and aims to aid in understanding the role of dipeptide repeat (DPR) proteins in *C9orf72*-linked FTD and ALS.

A significant part of this research project was the development of a unique induced pluripotent stem cell model with endogenous *C9orf72* DPR proteins tagged with a fluorescent protein. The methodology developed in this project and the cell model generated will provide the Isaacs lab, and many international collaborating groups, with the first patient-derived tool to study endogenously produced DPR proteins in real-time. This could directly lead to screens for new therapies for *C9orf72*-linked FTD and ALS. Other work presented in this thesis that will also benefit the research community investigating the mechanism of *C9orf72* FTD/ALS disease include; the production of fluorescently tagged DPR protein overexpression constructs, investigation into the ability of a molecular chaperone to reduce aggregation of these tagged constructs, and the identification of DPR protein post-translational modification in *C9orf72* patient post-mortem tissue.

Finally, this thesis also explores the pathology of an additional form of FTD; FTLD-FUS. In this work, two novel RNA binding proteins were discovered in pathological inclusions specifically in FTLD-FUS cases, indicating that these proteins may play a central role in the development of disease. The identification of these proteins will likely impact the research of several groups who focus on the biological cause of FTLD-FUS because the presence of these two proteins cannot be explained by the current hypothesis used to explain the disease mechanism.

In summary, the work in this thesis adds to the understanding of two debilitating neurodegenerative diseases which are associated with large social and economic costs. It is hoped that the findings, methodology and models generated from the research presented in this thesis will make a significant contribution to the field of FTD and ALS research and will, in the longer-term, help enable the development of effective therapies, which would have a major impact on public health.

Table of Contents

List of Figures

List of Tables

Abbreviations

ABC	avidin-biotin complex
AAD	age at death
AAO	age at onset
AAV	adeno-associated virus
AD	Alzheimer's disease
ADMA	asymmetric dimethylarginine
ADMA-GR	asymmetrically dimethylated poly-glycine-arginine
ADMA-PR	asymmetrically dimethylated poly-proline-arginine
AdOx	adenosine dialdehyde
aFTLD-U	atypical frontotemporal lobar degeneration
AGD	argyrophilic grain disease
ANOVA	analysis of variance
ALS	amyotrophic lateral sclerosis
ALS-FUS	amyotrophic lateral sclerosis with FUS inclusions
ALS-SOD1	amyotrophic lateral sclerosis with SOD1 inclusions
ALS-TDP	amyotrophic lateral sclerosis with TDP-43 inclusions
ASO	antisense oligonucleotides
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BFP	blue fluorescent protein
BIBD	basophilic inclusion body disease
BSA	bovine serum albumin
bvFTD	behavioural variant frontotemporal dementia
bp	base pair
C9orf72	chromosome 9 open reading frame 72
CBD	corticobasal degeneration
cDNA	complementary deoxyribonucleic acid
CI	confidence interval
CRISPR	clustered regularly interspaced short palindromic repeat
CSF	cerebrospinal fluid
Ctrl	control
DAB	3,3'-diaminobenzidine
DAPI	4',6-diamidino-2-phenylindole
ddH$_2$O	double-distilled water

DENN	Differentially Expressed in Normal and Neoplasia
DM1	myotonic dystrophy type 1
DMD	Duchenne muscular dystrophy
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DPR	dipeptide repeat protein
dsDNA	double stranded deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
EWS	Ewing's sarcoma
fALS	familial amyotrophic lateral sclerosis
FACS	fluorescence activated cell sorting
FCS	fetal calf serum
FISH	fluorescent *in situ* hybridisation
FLuc	firefly luciferase
FTD	frontotemporal dementia
FTLD	frontotemporal lobar degeneration
FTLD-FUS	frontotemporal lobar degeneration with FUS inclusions
FTLD-tau	frontotemporal lobar degeneration with tau pathology
FTLD-TDP	frontotemporal lobar degeneration with TDP-43 inclusions
FTLD-UPS	frontotemporal lobar degeneration with ubiquitinated inclusions
FUS	fused in sarcoma
FXTAS	fragile X-associated tremor/ataxia syndrome
GDP	guanosine diphosphate
GEF	guanine exchange factor
GFP	green fluorescent protein
GGT	globular glial tauopathy
GRN	progranulin
gRNA	guide ribonucleic acid
GTP	guanine triphosphate
GWAS	genome wide association study
HD	Huntington's disease
HDR	homology directed repair
hnRNP	heterogenous nuclear ribonucleoprotein

HSR	heat shock response
HPRT	hypoxanthine-guanine phosphoribosyltransferase
HRP	horseradish peroxidase
Hsp	heat shock protein
HSPB8	small heat shock protein B8
ICC	immunocytochemistry
iCLIP	individual nucleotide-resolution cross-linking and immunoprecipitation
iPSC	induced pluripotent stem cell
IF	immunofluorescence
IHC	immunohistochemistry
IOP	King's College Institute of Psychiatry Brain Bank
iPSC	induced pluripotent stem cell
ivTRT	in vitro transcription reverse transcription
kB	kilobase
kDa	kilodalton
LB	Luria-Broth
LCD	low complexity domain
LLPS	liquid-liquid phase separation
lvPPA	logopenic variant primary progressive aphasia
m^7G	5'7-methylguanosine
Man	Manchester Brain Bank
MMA	monomethylarginine
MND	motor neuron disease
MRI	magnetic resonance imaging
nfvPPA	non-fluent variant primary progressive aphasia
NHEJ	non-homologous end joining
NIFID	neuronal intermediate filament inclusion disease
nt	nucleotides
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline with tween
PCR	polymerase chain reaction
PD	Parkinson's disease
PET	positron emission tomography
PFA	paraformarldehyde
PI	propidium iodide

PM	post mortem
PNFA	progressive non-fluent aphasia
Poly-AP	poly-alanine-proline
Poly-GA	poly-glycine-alanine
Poly-GR	poly-glycine-arginine
Poly-GP	poly-glycine-proline
Poly-PR	poly-proline-arginine
PPA	primary progressive aphasia
PSP	progressive supranuclear palsy
PRMT	protein arginine methyltransferase
PY-NLS	proline-tyrosine nuclear localisation signal
QSBB	Queen Square Brain Bank
RAN	repeat associated non-ATG
RBP	RNA binding protein
RIPA	radioimmunoprecipitation assay
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein
ROCK	Rho-associated, coiled-coil containing protein kinase
ROI	region of interest
RRM	rna recognition motif
RT	reverse transcription
sALS	sporadic amyotrophic lateral sclerosis
SAM	S-adenosylmethionine
SD	semantic dementia
SDS	sodium dodecyl sulfate
SDMA	symmetric dimethylarginine
SDMA-GR	symmetrically dimethylated poly-glycine-arginine
SDMA-PR	symmetrically dimethylated poly-proline-arginine
SEM	standard error of mean
SFM	serum free media
sHSP	small heat shock protein
SMA	spinal muscular atrophy
SMN	survival motor neuron
SOC	super optimal broth with catabolite repression
SOD1	superoxide dismutase 1

ssDNA	single stranded deoxyribonucleic acid
ssODN	single stranded oliogonucleotide
svPPA	semantic variant primary progressive aphasia
TAF15	TATA-binding protein-associated factor 15
TARDBP	gene encoding TAR DNA-binding protein 43
TBS	tris-buffered saline
TBS-T	tris-buffered saline with tween
TBE	tris-boric acid-ethylenediaminetetraacetic acid
TDP-43	TAR DNA-binding protein 43
TE	tris-ethylenediaminetetraacetic acid
TLS	translocated in liposarcoma
TRN1	transportin
TSA	tyramide signal amplification
UIM	ubiquitin interacting motif
UPR	unfolded protein response
UV	ultraviolet
VCP	valosin containing protein

Publications produced during PhD

Moens, T.G., Niccoli, T., Wilson, K.M., Atilano, M.L., Birsa, N., **Gittings, L.M**., Holbling, B.V., Dyson, M., Thoeng, A.T., Neeves, J, Glaria, I., Yu, L, Bussmann, Julia., Storkebaum, E, Pardo, M., Choudhary, J.C., Fratta, P., Partridge, L., and Isaacs, A.M. 2018. C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is rescued by eIF1A. *Acta Neuropathol.*

Hock, E.M., Z. Maniecka, M. Hruska-Plochan, S. Reber, F. Laferriere, M.K.S. Sahadevan, H. Ederle, **L. Gittings**, L. Pelkmans, L. Dupuis, T. Lashley, M.D. Ruepp, D. Dormann, and M. Polymenidou. 2018. Hypertonic Stress Causes Cytoplasmic Translocation of Neuronal, but Not Astrocytic, FUS due to Impaired Transportin Function. *Cell reports*. 24:987-1000 e1007.

Zarouchlioti, C., D.A. Parfitt, W. Li, **L.M. Gittings**, and M.E. Cheetham. 2018. DNAJ Proteins in neurodegeneration: essential and protective factors. *Philos Trans R Soc Lond B Biol Sci*. 373.

Chapter 1. Introduction

1.1 Overview

Frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) are two clinically, genetically and pathological heterogenous neurodegenerative disorders that severely reduce the quality of life for both the patient living with the disease and those caring for them. The two diseases present high economic costs to society in the form of health and social care, as current treatment options typically only focus on alleviating symptoms and do little to delay disease progression. The major risk factor for both diseases, as well as many other neurodegenerative disorders, is advancing age. Thus, as the global population grows and ages, the proportion of individuals living with these neurodegenerative disorders is predicated to dramatically rise, increasing their psychological and socio-economic impact. Within the scientific research community, efforts are being made to understand the molecular pathogenesis of these diseases in the hope of finding potential treatments to alleviate this burden.

A major breakthrough in FTD and ALS research occurred in 2011 when two independent groups identified a mutation in the *C9orf72* gene, that is now known to be the most common genetic cause of both these diseases (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Although several hypotheses have been proposed, the molecular mechanism of disease remains unknown and this has led to a flurry of exciting research within the FTD/ALS field in recent years. To add to this, the work in this thesis describes the generation and use of cellular models and post-mortem brain tissue to investigate molecular mechanisms underlying *C9orf72*-linked FTD and ALS, specifically focusing on the unique dipeptide repeat protein (DPR) pathology associated with the disease.

1.2 Clinical features of FTD and ALS

1.2.1 Frontotemporal dementia

FTD describes a clinically and pathologically heterogeneous group of progressive neurodegenerative disorders that arise from frontotemporal lobar degeneration (FTLD); selective atrophy of the frontal and/or temporal cortices, with a non-Alzheimer's disease pathology. FTD disease accounts for 5 – 15 % of all cases of dementia making it the second most common form of young onset dementia, after Alzheimer's disease (Rademakers et al., 2012).

It is a common cause of young-onset dementia with disease onset most often occurring between the ages of 45 – 65 years, with a median duration of 6 – 8 years (Thompson et al., 2005). Estimates of population prevalence range from 4 to 12 per 100,000 before the age of 65 in European and US epidemiological studies (Hogan et al., 2016; Rabinovici and Miller, 2010). The incidence of disease is equally distributed between men and women (Hogan et al., 2016; Thompson et al., 2005). Approximately 50 % of all FTD cases have a heritable form of the disease caused by autosomal dominant mutations in one of several FTD associated genes (familial FTD), however the remaining cases have no currently known genetic cause (sporadic FTD) (Rademakers et al., 2012).

Clinically, FTD can be divided into two main subcategories; behavioural variant frontotemporal dementia (bvFTD) and primary progressive asphasia (PPA), which can be further subdivided into semantic variant PPA (svPPA), non-fluent variant PPA (nfvPPA) and logopenic variant PPA (lvPPA) (Figure 1.1) (Neary et al., 1998; Woollacott and Rohrer, 2016).
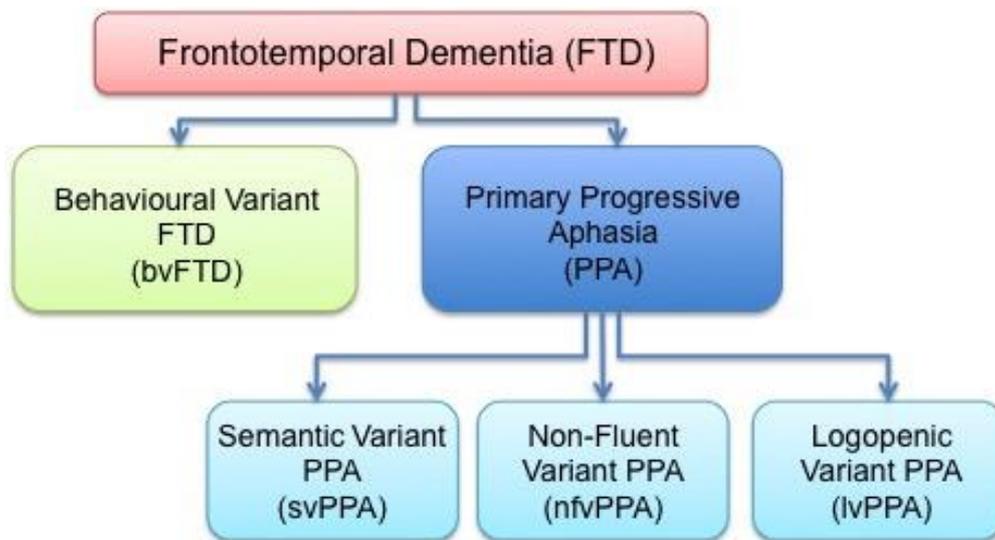
Figure 1.1 Clinical subtypes of frontotemporal dementia

Frontotemporal dementia can be separated into two clinical categories; behavioural variant FTD (bvFTD) and primary progressive aphasia (PPA). PPA can be further subdivided into three variants; semantic (svPPA), non-fluent (nfvPPA) and logopenic (lvPPA).

bvFTD is the most common clinical syndrome within the FTD clinical range, accounting for nearly 60% of cases (Onyike and Huey, 2013; Rohrer et al., 2009). It is attributed to pronounced degeneration of the frontal and anterior temporal lobes, which tends to be asymmetric between hemispheres (Fratta et al., 2013). Patients typically present with progressive decline in social skills, changes in behaviour, and difficulties with planning and higher level thinking due to executive dysfunction (Fratta et al., 2013). Other symptoms can include a reduced ability to see another person's point of view, altered perceptions of sound or music, altered perception of pain, and neuropsychiatric symptoms, such as hallucinations or delusions (Rascovsky et al., 2011; Woollacott and Rohrer, 2016; Woolley et al., 2011).

PPA is a form of FTD that initially and predominantly affects language. It is defined as progressive and insidious language decline affecting at least one of speech

production, object naming, syntax, or word comprehension (Gorno-Tempini et al., 2011). Other cognitive or behavioural symptoms may also develop over the course of the disease, but language is the most severely affected function. The absence or preservation of certain language features can be used to differentiate between the three PPA variants.

svPPA is associated with predominant left anterior temporal lobe atrophy and is characterised by loss of semantic or conceptual knowledge of words, objects and concepts, which affects both spoken and written language. Patients have word-finding difficulties and will often have to ask the meaning of familiar words (Woollacott and Rohrer, 2016). In contrast, patients presenting with nfvPPA have intact word comprehension and semantic memory but struggle with speech fluency, articulation and grammar (Fratta et al., 2013; Gorno-Tempini et al., 2011). nfvPPA typically results from atrophy of the left frontal lobe and insular cortex (Rollinson et al., 2009). lvPPA patients typically present with frequent word-finding pauses, anomia and an impairment in sentence and phrase repetition. These patients have problems with word retrieval but semantic knowledge and memory remains intact (Henry et al., 2008). Pathologically, lvPPA is typically associated with left posterior temporoparietal atrophy (Gorno-Tempini et al., 2004; Rohrer et al., 2013). Patients can also present with a mixture of the clinical subtypes, and aspects of both behavioural and language deficits often co-occur in the later stages of disease (Neary et al., 1998; Sieben et al., 2012).

1.2.2 Amyotrophic lateral sclerosis

ALS is the most prevalent form of motor neuron disease (MND), accounting for more than 75 % of all MND cases. It is a neurodegenerative disease that affects both the

upper motor neurons in the motor cortex and lower motor neurons in the brainstem and anterior horn of the spinal cord.

Clinically, this manifests as progressive weakness and atrophy of voluntary skeletal muscles, fasciculations and spasticity (Robberecht and Philips, 2013). As motor neurons are affected segmentally, initial clinical presentation varies depending on the initial neurons affected and therefore the site of onset often classifies disease phenotype. Patients can present with limb-onset (65 %) or bulbar-onset (30 %), and in rare cases, with respiratory-onset (5 %) (Hardiman et al., 2011) The site of onset usually defines the symptoms a patient presents with (Mitchell and Borasio, 2007). Patients typically die of respiratory failure due to progressive weakening of respiratory muscles (Kiernan et al., 2011; Kurian et al., 2009).

The age of disease onset is variable but is almost always after the fourth decade of life; juvenile onset ALS is rare (Kiernan et al., 2011; Robberecht and Philips, 2013). Disease progression is also variable with 50 % of patients dying within 3 years of onset but 20 % of patients surviving between 5 – 10 years after symptom onset (Kiernan et al., 2011; Mitchell and Borasio, 2007). The incidence of ALS in the European population is approximately 2.1 cases per 100,000, with males more likely to be diagnosed than women (Logroscino et al., 2010).

ALS is traditionally classified into two clinically indistinguishable categories; familial ALS (fALS) and sporadic ALS (sALS). The majority of cases are sporadic with unknown aetiology but approximately 10 % of ALS cases are classified as familial, usually with an autosomal dominant pattern of inheritance (Renton et al., 2014). Advancements in sequencing technologies and genetic studies have meant that in the past two decades, significant progress has been made in identifying genetic mutations underlying many cases of ALS, with at least 25 genes linked to ALS being

identified, and the genetic aetiology now known for two-thirds of fALS cases and approximately 11 % of sALS cases (Marangi and Traynor, 2015; Renton et al., 2014).

1.2.3 Clinical overlap between FTD and ALS

Historically, FTD and ALS were considered separate neurodegenerative disorders, however increasing evidence has shown that these diseases have significant clinical, pathological and genetic overlap indicating that they are, in fact, likely to be part of single disease spectrum (Lillo and Hodges, 2009).

Cognitive and behavioural deficits, such as personality changes, irritability and deficits in executive functioning tasks, have frequently been reported in ALS patients (Lomen-Hoerth et al., 2002; Neary et al., 2000; Portet et al., 2001). Historically these impairments were often overlooked but it is now estimated that between 20 – 50 % of ALS patients would fulfil the criteria for diagnosis of probable or definite FTD (Ringholz et al., 2005). Conversely, up to 40 % of FTD patients exhibit mild motor system dysfunction, while it is estimated 15 % of FTD patients exhibit motor dysfunction that meets ALS diagnosis criteria (Van Langenhove et al., 2012). The motor deficits experience by FTD patients typically resemble classical ALS, while the cognitive changes observed in ALS patients are most commonly like bvFTD, with rare cases of nfvPPA or svPPA with ALS also reported (Van Langenhove et al., 2012). Patients exhibiting features of combined FTD-ALS typically have a poor prognosis with an average survival of 2 – 3 years after symptom onset (Broe et al., 2003; Whitwell et al., 2011).

In addition to the clinical overlap between FTD and ALS, several neuropathological and genetic discoveries in recent years have reinforced the link between the two diseases. A particularly notable discovery was the identification of transactive

response DNA-binding protein 43 (TDP-43) as the major pathological protein in the majority of neuronal inclusions in both FTD and ALS cases (See Section 1.3.1) (Arai et al., 2006; Neumann et al., 2007; Neumann et al., 2006). Additionally, mutations in the gene encoding TDP-43 (*TARDBP*) have been identified as causing approximately 3 % of familial ALS and rare cases of FTD (Benajiba et al., 2009; Borroni et al., 2009; Lattante et al., 2013). In addition to this, mutations in a number of different genes have also been identified as the underlying cause of FTD, ALS, or a combination of the two, further unifying the diseases (See Section 1.4) (Guerreiro et al., 2015). Perhaps the strongest evidence to date that FTD and ALS form part of the same disease spectrum is the identification that a repeat expansion mutation in the *C9orf72* gene which is now known to be the most common cause of FTD and ALS in patients with familial disease (Van Langenhove et al., 2012).

1.3 Pathological features of FTLD and ALS

FTD and ALS have similar and unique features in their underlying pathology. On a macroscopic level, ALS pathology refers to the selective atrophy of motor neurons in the cortex, corticospinal and corticobulbar tracts (Saberi et al., 2015), whilst the pathology of FTD, commonly referred to as FTLD, is characterised by the relatively selective degeneration of the frontal and temporal lobes (Neary et al., 1998). Microscopically, both FTLD and ALS are characterised by the presence of abnormal intracellular protein aggerates within neurons and glia. Like the clinical presentation of these diseases, the protein constituents of these aggregates are also heterogenous. The primary protein component of these aggregates is used to sub-classify both ALS and FTLD. FTLD is subdivided into four groups; FTLD-tau, FTLD-TDP, FTLD-FUS and FTLD-UPS, whilst ALS is grouped into three categories; ALS-TDP, ALS-SOD1 and ALS-FUS (Figure 1.2).
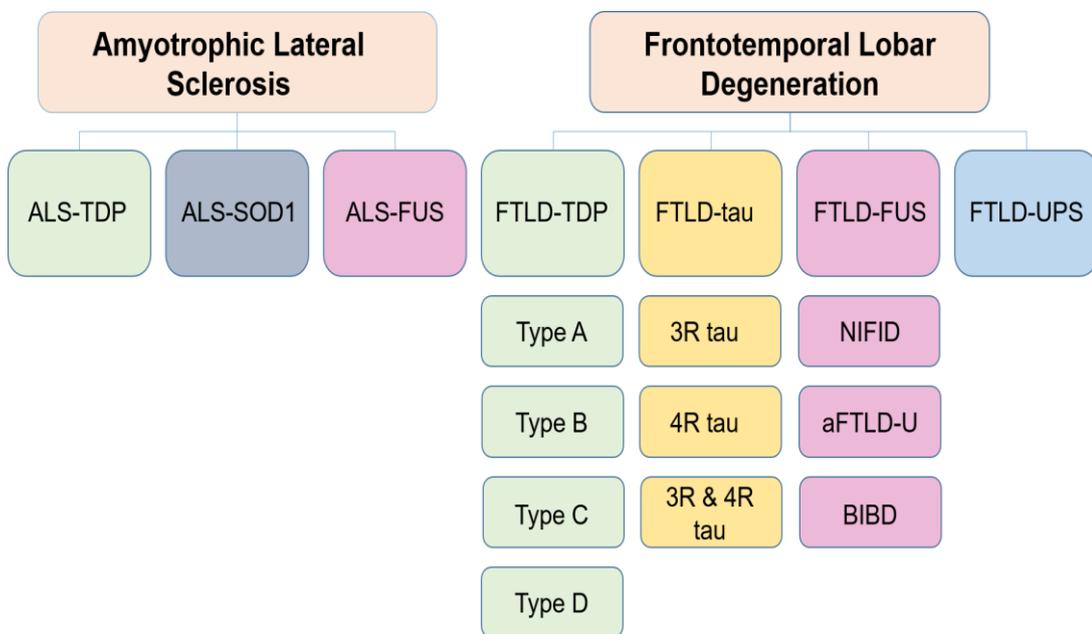
Figure 1.2 Pathological sub-groups of amyotrophic lateral sclerosis and frontotemporal lobar degeneration.

1.3.1 FTLD-TDP and ALS-TDP

In 2006, TDP-43 was identified as the major protein component in tau-negative, ubiquitin-positive inclusions in the majority of FTLD cases, and in most cases of ALS (Neumann et al., 2006). Pathological accumulation of TDP-43 is thought to account for approximately 50 % of FTLD cases and over 95 % of both familial and sporadic ALS cases (Neumann et al., 2007; Neumann et al., 2006). Additionally, mutations in the gene encoding TDP-43, TARDBP, are known to cause approximately 3 % of fALS cases (Lattante et al., 2013).

TDP-43 is a highly conserved and ubiquitously expressed multi-functional DNA/RNA binding protein. It is predominantly localized to the nucleus but shuttles continuously between cytoplasm and nucleus to carry out a variety of functions associated with RNA metabolism including transcription, splicing, transport and stabilization (Buratti and Baralle, 2008; Ratti and Buratti, 2016).

In FTLD-TDP and ALS-TDP, TDP-43 often becomes depleted from the nucleus and mislocalises into protein deposits in the cytoplasm or nuclei of affected neurons and glia, forming pathological aggregates. These aggregates of TDP-43 are associated with several post-translational modifications including hyperphosphorylation, ubiquitination, oxidation, lysine acetylation and C-terminal cleavage (Lashley et al., 2015; Mackenzie and Neumann, 2016).

In ALS-TDP, TDP-43 neuronal cytoplasmic inclusions are primarily observed within neurons and glia in the motor cortex, spinal cord, and brainstem motor nuclei, as well as the white matter tracts associated with these areas (Neumann et al., 2007). The distribution and morphology of TDP-43 inclusions in FTLD-TDP is more heterogenous and can be sub-classified into four pathological subtypes: A, B, C and D (Mackenzie et al., 2011). It is currently unknown whether the differences in underlying TDP

pathology are a result of a single pathogenic mechanism or whether multiple, distinct mechanisms are involved in each subtype. FTLD-TDP type A is characterised by numerous compact, oval or crescent shaped neuronal cytoplasmic inclusions, short dystrophic neurites and variable numbers of neuronal intranuclear inclusions, that are most abundant in layer II of the cortex. FTLD cases with TDP type A pathology typically present clinically with bvFTD or nfvPPA (Mackenzie and Neumann, 2016). Type B TDP pathology is associated with numerous neuronal cytoplasmic inclusions in both the superficial and deeper cortical layers, with relatively few dystrophic neurites and neuronal intranuclear inclusions. Almost exclusive to Type B TDP pathology is the presence of neuronal cytoplasmic inclusion in lower motor neurons, even in the absence of clinical ALS features. The majority of cases with clinical features of both FTD and ALS have type B TDP pathology (Mackenzie and Neumann, 2016). Type C TDP pathology is characterised by abundant long dystrophic neurites, that often have a corkscrew appearance, throughout the cortical layers. This subtype typically has few or no neuronal cytoplasmic inclusions and neuronal intranuclear inclusions. Type C TDP is the most common pathology associated with cases presenting clinically with svPPA. Finally, type D TDP pathology is present only in cases with mutations in the valosin containing protein (*VCP*) gene. This subtype is characterised by frequent lentiform neuronal intranuclear inclusions and short dystrophic neurites in the cortex and rare neuronal cytoplasmic inclusions (Forman et al., 2006; Neumann et al., 2007).

1.3.2 FTLD-FUS and ALS-FUS

Mutations in the fused in sarcoma (*FUS*) gene were discovered to cause a small proportion of ALS cases in 2009 (Kwiatkowski et al., 2009; Vance et al., 2009). Subsequent pathological studies in ALS and FTLD patients later identified the FUS protein as a major component of pathological inclusions in a sub-set of cases

(Neumann et al., 2009a). The protein is found in inclusions in approximately 5 – 10 % of tau and TDP-43 negative, ubiquitin positive FTLD cases (Lashley et al., 2015), and mutations in the *FUS* gene are responsible for approximately 4 % of fALS cases (Van Langenhove et al., 2010).

Like TDP-43, FUS is a highly conserved DNA/RNA binding protein that is associated with a wide range of functions, including DNA damage repair, transcriptional regulation and the splicing of thousands of mRNAs (Efimova et al., 2017; Ratti and Buratti, 2016). The functions of FUS are discussed further in Chapter 6. In a similar manner to TDP-43, FUS primarily localises to the nucleus but shuttles continuously back and forth into the cytoplasm via interaction with nuclear importers and exporters (Aman et al., 1996; Andersson et al., 2008).

In FTLD-FUS and ALS-FUS, the FUS protein is observed in pathological aggregates in the cytoplasm and nucleus of neurons and glia. However, in contrast to TDP-43 inclusions, cytoplasmic FUS aggregates and non-pathogenic nuclear FUS are not mutually exclusive and can be observed within the same cell (Saberi et al., 2015).

FTLD-FUS is sub-categorised into three groups based on the morphological features and anatomical location of FUS-positive inclusions (Lashley et al., 2011). The first category of FTLD-FUS is neuronal intermediate filament inclusion disease (NIFID) (Neumann et al., 2009b). These cases typically show FUS-positive neuronal cytoplasmic inclusions that have a variety of morphologies including, round, crescent-shaped, annular and granular. The FUS inclusions are present throughout many brain regions, including the cerebral cortex, hippocampus, thalamus and numerous brainstem nuclei. Occasionally neuronal intranuclear FUS positive inclusions are observed in NIFID cases, but these are generally restricted to the hippocampus. The defining feature of NIFID pathology is the presence of α-internexin positive inclusions,

however these are far less abundant than FUS inclusions. The second category of FTLD-FUS pathology is atypical FTLD (aFTLD-U) (Neumann et al., 2009a). This subtype has far less FUS-positive inclusions compared to the NIFID subtype, but the pathology of the inclusions is more uniform. aFTLD-U is characterised by compact, round or bean-shaped neuronal cytoplasmic inclusions and vermiform neuronal intranuclear inclusions that are abundant throughout the frontal and temporal cortices and the granule cells of the dentate gyrus of the hippocampus (Lashley et al., 2011). The final subtype of FTLD-FUS is basophilic inclusion body disease (BIBD) (Munoz et al., 2009). The defining feature of this subtype is basophilic neuronal cytoplasmic inclusions on haematoxylin and eosin stained sections of the cerebral cortices. FUS immunohistochemistry in these cases demonstrates widespread neuronal cytoplasmic inclusions with a spectrum of morphologies similar to NIFID cases in the cortex, basal ganglia and brainstem. Neuronal intranuclear inclusions are consistently absent from BIBD cases.

In all three FTLD-FUS subtypes, FUS aggregates have been found to consistently contain the nuclear importer of FUS, transportin, and two additional RNA-binding proteins, TAF15 and EWS (Brelstaff et al., 2011; Neumann et al., 2012) . Along with FUS, these RNA binding proteins make up the FET protein family and will be discussed further in relation to FTLD-FUS pathology in Chapter 6.

Unlike FTLD-FUS, inclusions of FUS in ALS-FUS cases are negative for transportin and other FET proteins (Neumann et al., 2012). Characterisation of a small group of ALS-FUS cases suggested that different FUS mutations may be associated with distinct pathological phenotypes (Mackenzie et al., 2011), however, in general patients with ALS-FUS show severe motor neuron loss in the spinal cord and mild to moderate motor neuron loss in the motor cortex (Johnson et al., 2012). In many

cases, large globular and elongated cytoplasmic aggregates of mutant FUS are found in spinal cord motor neurons, as well as dystrophic neurites and occasional neuronal intranuclear inclusions (Johnson et al., 2012; Saberi et al., 2015).

1.3.3 ALS-SOD1

Mutations in the superoxide-dismutase-1 (*SOD1*) gene were the first identified to be associated with ALS and are thought to account for approximately 20 % of familial ALS cases (Rosen et al., 1993; Saberi et al., 2015). Aggregates of the misfolded SOD1 protein were later identified in inclusions in these familial ALS patients, and in around 1 % of sporadic ALS cases (Forsberg et al., 2010; Kato et al., 2000; Pasinelli and Brown, 2006). The SOD1 protein is anti-oxidant enzyme that catalyses the conversion of reactive oxygen species to hydrogen peroxide and oxygen, thereby protecting cells against oxidative stress. Despite being the most heavily researched protein in the ALS field, it remains unknown how almost any mutation in the protein can cause the disease. In cases with ALS-SOD1, misfolded skeins or rounded inclusions of SOD1, that are TDP-43 and FUS negative, can be observed in motor neurons in the spinal cord but are typically absent in the Betz cells of the motor cortex (Saberi et al., 2015).

1.3.4 FTLD-tau

Hyperphosphorylation and aggregation of the microtubule associated protein, tau, is responsible for a major group of FTLD, designated as FTLD-tau. The accumulation of hyperphosphorylated tau is believed to be the cause of approximately 40 % of FTLD cases, and various mutations in the *MAPT* gene, which encodes tau, are a known cause of some familial forms of FTLD (Lashley et al., 2011; Sleegers et al., 2004).

FTLD-tau is sub-categorised based on the biochemical composition of the tau inclusions. Alternative splicing of the *MAPT* gene can generate six different isoforms of the tau protein containing either three (3R) or four (4R) repeating microtubule domains, thus FTLD-tau may be subdivided into disorders with inclusions containing predominantly 3R, 4R tau or a mixture of both (Lee and Trojanowski, 2001; Spillantini and Goedert, 2013). FTLD-tau subtypes include Pick's diseases, corticobasal degeneration (CBD), progressive supranuclear palsy (PSP), globular glial tauopathy (GGT) and argyrophilic grain disease (AGD) (Lashley et al., 2015; Mackenzie and Neumann, 2016). Each of these FTLD-tau subtypes is pathologically diagnosed based on 3R-tau and 4R-tau immunohistochemistry, the morphological features and anatomical distribution of the tau inclusions (Lashley et al., 2015; Mackenzie and Neumann, 2016).

1.3.5 FTLD-UPS

Before identification of TDP-43 and FUS as the major protein component of pathological aggregates in FTLD, all cases that were tau-negative and showed the presence of ubiquitinated inclusions were classified as 'FTLD-U'. The majority of these cases have now been re-classified into either FTLD-TDP or FTLD-FUS subtypes, however there are number of cases that do not fit this classification for which the protein components of the ubiquitinated inclusions has not been identified. These cases are now re-classified as FTLD-UPS (Lashley et al., 2015; Mackenzie and Neumann, 2016). A key example are rare cases of familial FTD linked to mutations in the *CHMP2B* gene. The pathology of these cases is typically associated with ubiquitin and p62-positive neuronal cytoplasmic inclusions, which are tau, TDP-43 and FUS negative, within neurons of the frontal cortex and the granular cell layer of the hippocampus (Isaacs et al., 2011; Piggott et al., 2007).

1.4 Genetics of FTD and ALS

Disease causing mutations have been identified in all the genes encoding the major aggregating proteins in FTLD and ALS pathology These include *TARDBP*, which encodes the TDP-43 protein, *MAPT*, which encodes that tau protein, *FUS*, which encodes the FUS protein, and *SOD1*, which encodes the SOD1 protein (Gitcho et al., 2008; Hutton et al., 1998; Kabashi et al., 2008; Kwiatkowski et al., 2009; Rosen et al., 1993; Vance et al., 2009). Mutations in these genes each account for a small proportion of some familial forms of FTD and ALS. Several mutations have also been identified in a number of other genes, but these do not typically encode proteins found in the underlying disease pathology and are generally only responsible for rare cases. The most common genetic cause of both familial FTD and ALS is a repeat expansion mutation in the *C9orf72* gene (discussed in detail in Section 1.5),

Some genes have been found to be specific to causing either FTD or ALS. For example, mutations in *MAPT, CHMP2B* and *GRN* have only been found to cause FTD, whilst mutations in *SOD1* and *FUS* have only been found to cause ALS (Cruts et al., 2006; Hutton et al., 1998; Rosen, 1993; Skibinski et al., 2005; Vance et al., 2009). Other genes are known to have mutations that can cause either FTD, ALS or a phenotype spanning a combination of these two diseases. This includes genes such as *C9orf72*, *UBQLN2*, *SQSTM1, HNRNPA2B1* and *HNRNPA1* (DeJesus-Hernandez et al., 2011; Deng et al., 2011; Kim et al., 2013b; Renton et al., 2011; Rubino et al., 2012). Interestingly, many of the genes identified are involved in similar cellular pathways, which helps to identify potentially important mechanisms underlying these diseases (Guerreiro et al., 2015). For example, several genes are involved in protein clearance pathways, such as *VCP*, *CHMP2B*, *UBQLN2* and *SQSTM1*, and a number of genes are associated with RNA metabolism, such as *FUS*, *TARDBP*, *HNRNPA2B1* and *HNRNPA1*.

Despite the identification of disease-causing mutations in several genes, many cases of FTD and ALS have no known genetic cause. A large proportion of these cases have sporadic FTD or ALS, for which finding a genetic cause is more difficult given the lack of family history. However, there are still a large proportion of familial FTD and ALS that also lack a known genetic cause. Sequencing and genetic studies of these cases are ongoing and continuously identifying novel genes associated with the disease.

1.5 *C9orf72* repeat expansion mutation

1.5.1 Identification of the *C9orf72* repeat expansion mutation

In an attempt to identify new genetic causes of FTD and ALS, multiple groups conducted extensive sequencing and genome-wide linkage analysis studies in families with hereditary ALS, FTD or combined FTD/ALS with unknown causal mutations (Boxer et al., 2011; Gijselinck et al., 2010; Le Ber et al., 2009; Luty et al., 2008; Morita et al., 2006; Pearson et al., 2011; Pillai et al., 2006; Valdmanis et al., 2007). These studies linked the diseases to a 3.7 Mb region on the short arm of chromosome 9. Simultaneously, several genome-wide association studies (GWAS) performed on both FTD (Van Deerlin et al., 2010) and ALS patients (Laaksovirta et al., 2010; Shatunovskii and Ruban, 2010; van Es et al., 2009) identified a risk haplotype in the same linkage region. With several independent lines of evidence linking this region to the FTD/ALS disease spectrum, sequencing studies were conducted to identify potential pathogenic mutations however initial sequencing of the protein coding regions of the genes within this region found no exonic mutations linked to disease (Pearson et al., 2011). Subsequent sequencing and careful examination of the non-coding regions led two groups to independently identify a GGGGCC hexanucleotide repeat expansion in a non-protein-coding region of the *C9orf72* gene as a genetic cause of familial ALS and FTD (DeJesus-Hernandez et

al., 2011; Renton et al., 2011). This was the first genetic abnormality to be identified as a common cause of both FTD and ALS, and its identification confirmed the genetic and clinical overlap between these diseases.

1.5.2 Genetic characteristics of the *C9orf72* repeat expansion

The *C9orf72* gene is located on the short arm of chromosome 9 at position p21.2. The gene is alternatively spliced to produce three different transcripts that encode two protein isoforms (Figure 1.3). Transcript variants 2 and 3 encode the 481 amino acid long protein isoform C9orf72-L, while transcript variant 1 encodes the shorter 222 amino acid short isoform C9orf72-S (DeJesus-Hernandez et al., 2011; Xiao et al., 2016). The hexanucleotide repeat sequence is located within the promotor region of transcript variant 2 but lies within the first intron, between non-coding exons 1a and 1b, of transcript variants 1 and 3 (DeJesus-Hernandez et al., 2011; Renton et al., 2011). The location of the hexanucleotide repeat between two exons in transcript variants 2 and 3 means that it is possible for the repeat to undergo transcription.

The length of the hexanucleotide repeat is variable within the general population, with more than 90 % of the European population having between two and ten GGGGCC repeat units (Renton et al., 2011). However, larger repeats of between 20 - 30 repeat units have also been identified in healthy individuals (Beck et al., 2013; DeJesus-Hernandez et al., 2011; Meisler et al., 2013; van der Zee et al., 2013). In contrast, in *C9orf72* FTD/ALS affected individuals, the number of repeat units is typically expanded to several hundred or thousands of repeats (Beck et al., 2013; DeJesus-Hernandez et al., 2011; van Blitterswijk et al., 2013). The smallest hexanucleotide expansion that confers a risk of *C9orf72* FTD/ALS is not currently known. It should be noted, however, that the expanded repeats are somatically unstable which can result in a range of expanded repeat sizes across various tissues within one individual. This

means that an individual with an intermediate expansion size detected in blood sample-derived DNA might have a much larger expansion in the affected tissue of the CNS (Fratta et al., 2015; van Blitterswijk et al., 2013). Generally, repeat lengths in patients appear shorter in blood and cerebellum compared to frontal cortex (van Blitterswijk et al., 2013). Additionally, due to the repetitive nature and GC-rich content of the repeat expansion, sequencing of expansions comprising of more than around 30-50 repeat units cannot be reliably distinguished using repeat primed PCR. It is therefore recommended that Southern blotting is used to confirm large expansions, however this method can only provide an approximation of the repeat size, or range of repeat sizes, present in an individual. The variation in repeat size across tissues and difficulties in confirming repeat length, have made it difficult to determine a minimum repeat expansion length that confers pathogenicity. The length of the *C9orf72* repeat expansion has not been shown to consistently correlate with clinical phenotypes or survival, however two studies have identified a positive correlation between age of disease onset and expansion length (Beck et al., 2013; van Blitterswijk et al., 2013).
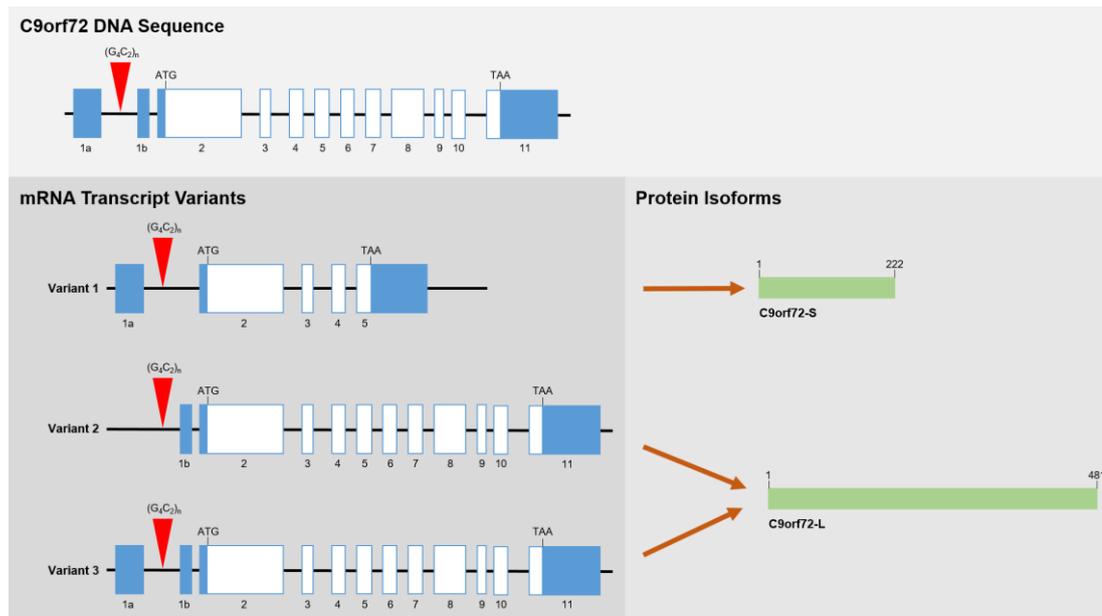
Figure 1.3 Genomic structure, mRNA transcript variants and protein isoforms of C9orf72

Schematic representation of the structure of the *C9orf72* gene, its three major mRNA transcript variants and two protein isoforms. Blue boxes represent untranslated exons and white boxes represent translated exons. The location of the hexanucleotide repeat expansion, $(G4C2)_n$, is represented by a red triangle. The repeat expansion is located within the promoter region of transcript variant 2 and between exons 1a and 1b of transcript variants 1 and 3. C9orf72 transcripts can be translated into two protein isoforms. *C9orf72* variant 1 encodes the 222 amino acid short protein isoform (C9orf72-S). *C9orf72* variants 2 and 3 encode the 481 amino acid protein isoform (C9orf72-L).

1.5.3 Epidemiology of *C9orf72* FTD/ALS

The hexanucleotide repeat expansion in *C9orf72* is currently the most common known genetic cause of FTD, ALS or combined FTD/ALS, however the prevalence depends largely on ethnic background and therefore varies geographically (Sabatelli et al., 2012). The expansion mutation has the highest prevalence in European, North American and Australian Caucasian populations. One study found the *C9orf72* expansion mutation to account for 7 % of sALS, 39 % fALS, 6 % of sporadic FTD and 25 % of familial FTD cases (Sabatelli et al., 2012). There is a particularly high frequency of the expansion found in certain Scandinavian populations such as Finland, Denmark and Sweden, with 28 % of FTD and 29 % of ALS patients in Finland

found to be positive for the expansion (Renton et al., 2011; van der Zee et al., 2013).

In contrast, *C9orf72* repeat expansion mutations are rare in India and East Asia, found

in only 0 - 3 % of familial ALS patients in these regions (He et al., 2016; Konno et al.,

2013; Mori et al., 2013a; Mukherjee et al., 2015; Zou et al., 2013). However, Taiwan

and the Kii peninsula in Japan are an exception to this as these are regions where

the frequency of the repeat expansion is relatively high in Asia (Ishiura et al., 2012;

Tsai et al., 2012). Data on the frequency of *C9orf72* expansions in South American,

South Asian and Middle Eastern populations is currently sparse (Mizielinska et al.,

2014).

1.5.4 Clinical characteristics of *C9orf72* FTD/ALS

A wide variety of clinical phenotypes have been associated with the *C9orf72* repeat

expansion mutation. The majority of these are forms of FTD, ALS or an overlapping

combination of the two diseases. However, the mutation has also been identified in

several other neurodegenerative diseases, including Parkinson's disease,

Huntington's disease and Alzheimer's disease, albeit at a low frequency (Beck et al.,

2013; Cacace et al., 2013; Hensman Moss et al., 2014; Kohli et al., 2013; Lesage et

al., 2013; Theuns et al., 2014).

The most common clinical FTD phenotype among *C9orf72* repeat expansion carriers

is bvFTD, accounting for approximately 87 % of cases, however patients presenting

with nfvPPA and svPPA have also been reported (Boeve et al., 2012; Devenney et

al., 2014; Hsiung et al., 2012; Mahoney et al., 2012b; Van Langenhove et al., 2013).

In addition to typical bvFTD symptoms, many *C9orf72* FTD patients are reported to

present with psychiatric symptoms, such as anxiety, agitation and psychosis, much

more frequently than FTD patients without the repeat expansion (Arighi et al., 2012;

Floris et al., 2012; Kertesz et al., 2013; Mahoney et al., 2012a; Rollinson et al., 2012).

Psychotic symptoms often occur simultaneously or before the manifestation of behavioural disturbances and can often lead to an initial psychiatric diagnosis (Galimberti et al., 2013; Rollinson et al., 2012).

Clinically, *C9orf72* ALS cases are typically characterised by an early age of onset, rapid disease progression and cognitive or behavioural impairments (Chi et al., 2016). The motor features displayed by *C9orf72* ALS patients are indistinguishable from typical ALS, usually with a limb onset (Cooper-Knock et al., 2012; Rutherford et al., 2012). However, a higher proportion of bulbar onset has been observed in *C9orf72* repeat expansion carriers compared to ALS patients without the expansion (Chi et al., 2016). Similarly, the co-occurrence of cognitive or behavioural impairment, or both, is more common among ALS patients with *C9orf72* expansion mutations than other ALS cohorts, occurring in nearly half of cases (Millecamps et al., 2012; Montuschi et al., 2015; Rutherford et al., 2012). Such cognitive and behavioural impairments are often associated with the behavioural symptoms of FTD and can also occur with psychiatric symptoms (Majounie et al., 2012; Rutherford et al., 2012). Rare cases linking the *C9orf72* repeat expansion mutation to some other forms of motor neuron disease, including progressive muscular atrophy and primary lateral sclerosis have also been described however the frequencies are low (van Rheenen et al., 2012), and other forms of MNDs, such as hereditary spastic paraplegia, have not been detected to harbour the repeat expansion (Hubers et al., 2014; Luo et al., 2014; Nielsen et al., 2014).

Disease penetrance in patients with the *C9orf72* repeat expansion is almost complete, with 50 % penetrance in ALS and FTD patients at 58 years of age, and almost full penetrance by 80 years of age (Sabatelli et al., 2012). The age of disease onset typically occurs within the fifth decade of life (Wen et al., 2017), however

disease onset can be variable with cases of disease occurring as late as the ninth decade of life, and as early as the twentieth decade (Sabatelli et al., 2012). Similarly, disease duration varies greatly between patients and depends heavily on the clinical manifestation of disease with *C9orf72* ALS patients typically experiencing much shorter disease duration than *C9orf72* FTD patients (Rohrer et al., 2015). Some studies suggest there is evidence of anticipation in *C9orf72* FTD/ALS, with offspring developing disease on average around 5 - 7 years before their affected parent, however these studies need to be validated in further multi-generational cohorts, (Benussi et al., 2014; Chio et al., 2012; Gijselinck et al., 2016; Van Mossevelde et al., 2017).

1.5.5 Pathological features of *C9orf72* FTD/ALS

1.5.5.1 TDP-43

Almost all reported *C9orf72* FTD/ALS cases demonstrate TDP-43 pathology, irrespective of their clinical phenotype (Cooper-Knock et al., 2012; Hsiung et al., 2012; Mahoney et al., 2012a; Murray et al., 2011; Simon-Sanchez et al., 2012; Snowden et al., 2012; Stewart et al., 2012). Of the four sub-categories of TDP-43 pathology, *C9orf72* repeat expansion carriers typically exhibit pathological features in line with TDP type A and type B, showing the presence of frequent cytoplasmic neuronal TDP-43 containing inclusions, dystrophic neurites and occasional neuronal intranuclear inclusions (Mackenzie et al., 2013; Mahoney et al., 2012a; Murray et al., 2011; Snowden et al., 2012). Rare exceptions to this subclassification have been reported in a few cases that exhibit TDP type C pathology, and in individuals who died prior to the onset of symptoms associated with the *C9orf72* repeat expansion, which show very sparse or no TDP-43 pathology (Baborie et al., 2015; Josephs et al., 2013; Mackenzie et al., 2013; Murray et al., 2011; Proudfoot et al., 2014; Vatsavayai et al., 2016).

TDP-43 pathology in *C9orf72* cases is observed in several different brain regions, including the frontal, temporal and primary motor cortices, hippocampus, amygdala, thalamus and midbrain (Hsiung et al., 2012; Mackenzie et al., 2013; Mahoney et al., 2012a; Murray et al., 2011). The brain regions with the most abundant TDP-43 pathology, however, correlate with regions of neurodegeneration associated with the clinical phenotype. For example, cases presenting with pure FTD have more abundant TDP-43 in the frontal cortex than those with pure ALS, whilst more TDP-43 pathology is seen in the lower motor neurons of pure ALS cases than those with FTD (Mackenzie et al., 2013; Mackenzie et al., 2015).

1.5.5.2 Dipeptide repeat proteins

In addition to TDP-43 pathology, *C9orf72* repeat expansion carriers exhibit a unique pathological feature that is a defining characteristic of the disease. This is the presence of abundant p62- and ubiquitin-positive inclusions that are TDP-43 negative, within neurons of the cortex, hippocampal granule cell layer and cerebellum (Al-Sarraj et al., 2011; Brettschneider et al., 2012; Pikkarainen et al., 2010; Troakes et al., 2012). In addition to their unique immunoreactivity profile, these inclusions also have characteristic morphological features with neuronal cytoplasmic inclusions forming 'star-like' structures and intranuclear inclusions forming 'dot-like' aggregates. In 2013, the major protein constituent of these inclusions was identified as translation products of the *C9orf72* repeat expansion, now known as dipeptide repeat (DPR) proteins (Ash et al., 2013; Gendron et al., 2013; Mori et al., 2013a; Mori et al., 2013c; Zu et al., 2013). The production and pathology of these proteins will be discussed further in Section 1.6.4.

1.5.5.3 RNA foci

C9orf72 FTD/ALS patient brain tissue also exhibits aggregates of RNA, known as RNA foci, which arise from both the sense (GGGGCC) and antisense (CCCCGG) mRNA transcripts of the *C9orf72* repeat expansion (Figure 1.5). Several studies employing RNA fluorescence in situ hybridisation (FISH) techniques have demonstrated that both sense and antisense RNA foci are widely distributed throughout the CNS with high abundance in neuronal nuclei in frontal and motor cortices, hippocampus, cerebellum and in the spinal cord (DeJesus-Hernandez et al., 2017a; Gendron et al., 2013; Lagier-Tourenne et al., 2013; Mizielinska et al., 2013; Zu et al., 2013). They are also less frequently observed in the cytoplasm of neurons and within glial cells, including astrocytes, microglia and oligodendrocytes (Lagier-Tourenne et al., 2013; Mizielinska et al., 2013). In general, sense foci are reported to be more abundant than antisense foci, but neither species have been found to consistently correlate well with region-specific neurodegeneration observed in patient post-mortem tissue (DeJesus-Hernandez et al., 2017b; Mizielinska et al., 2013). The role of RNA foci in *C9orf72* FTD/ALS pathogenesis will be discussed further in Section 1.6.3.

1.6 Potential disease mechanisms in *C9orf72* FTD/ALS

The mechanism by which the *C9orf72* hexanucleotide repeat expansion causes FTD/ALS is currently unknown, however three non-mutually exclusive mechanisms of toxicity have been proposed (Figure 1.4).

1) Loss of function of the C9orf72 protein

2) RNA toxicity via the sequestration of RNA binding proteins

3) Toxicity caused by the production of dipeptide repeat proteins

Each of these mechanisms will now be discussed in detail with reference to pathological and clinical data, as well as evidence from *in vitro* and *in vivo* models of the disease.
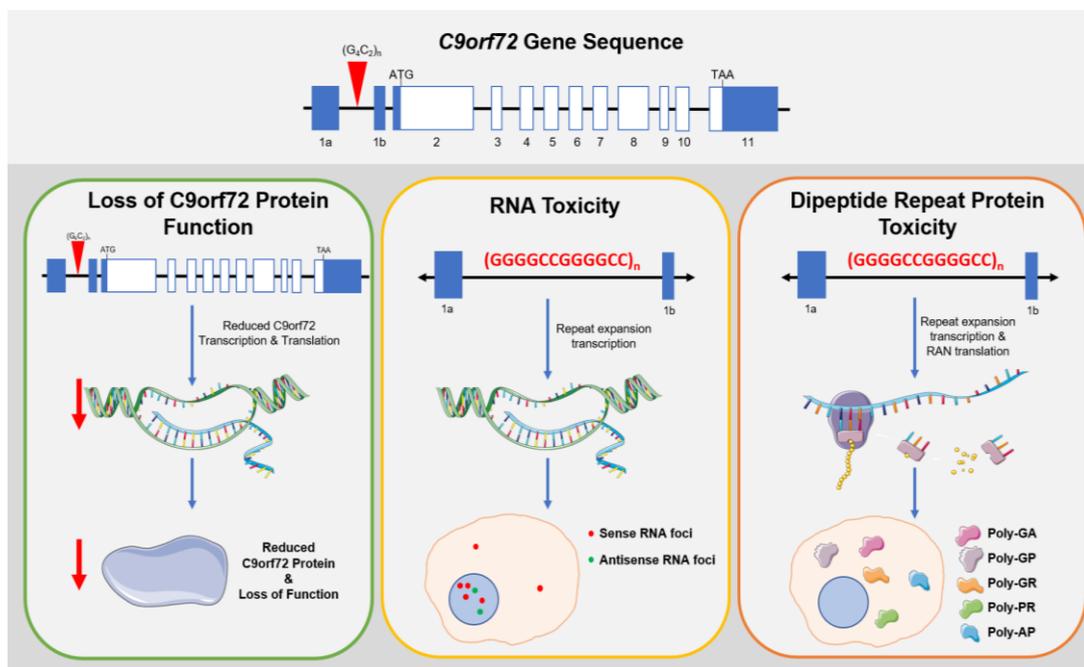
Figure 1.4 Three proposed mechanisms of C9orf72 repeat expansion toxicity.
Three non-mutually exclusive mechanisms of toxicity have been proposed to cause *C9orf72* repeat expansion associated neurodegeneration. These include: loss of function of C9orf72 protein, RNA toxicity via sequestration of RNA binding proteins into sense and antisense RNA foci, and toxicity caused from RAN translation of the RNA into 5 dipeptide repeat proteins. Some parts of this figure were obtained from SMART (Servier medical art).

1.6.1 Loss of C9orf72 protein function

The loss of protein function due to genetic mutation is a common mechanism of disease and is observed in several repeat expansion disorders, such as fragile X-syndrome and Friedreich's ataxia (Delatycki et al., 2000; Penagarikano et al., 2007). Given that *C9orf72* FTD/ALS also arises from a repeat expansion mutation, several studies have investigated whether loss of function of C9orf72 protein could contribute to the pathogenic mechanism of disease.

1.6.2 Normal function of C9orf72 protein

Since the discovery of the *C9orf72* repeat expansion, extensive research has been conducted to deduce the normal function of the C9orf72 protein. Bioinformatic studies have indicated that C9orf72 is homologous to members of the differentially expressed in normal and neoplasia (DENN) domain protein family. These are a family of GDP/GTP exchange factors (GEFs), which activate Rab GTPases and regulate vesicular trafficking (Levine et al., 2013; Zhang et al., 2012). Consistent with this, C9orf72 has been shown to co-localise and interact with various Rab proteins in neuronal cell lines, primary rat cortical neurons, iPSC-derived motor neurons and in human post-mortem spinal cord sections (Aoki et al., 2017; Farg et al., 2014; Frick et al., 2018; Webster et al., 2016). Several groups have demonstrated that the C9orf72 protein forms a complex with WDR41 and SMCR8, which is thought to participate in the initiation of autophagy, via interaction with Unc-51-like kinase 1 (ULK1) autophagy initiation complex (Amick et al., 2016; Blokhuis et al., 2016; Sellier et al., 2016; Sullivan et al., 2016; Ugolino et al., 2016; Webster et al., 2016; Yang et al., 2016). This is supported by the observation that autophagic function is impaired by knockdown of C9orf72 in human cell lines and primary neurons (Sellier et al., 2016; Webster et al., 2016), and some *C9orf72* knockout mice models are reported to show defects in autophagy, although there is conflicting evidence as to whether C9orf72

knockout decreases or enhances autophagy (Sullivan et al., 2016; Ugolino et al., 2016). Additionally, *C9orf72* patient derived neurons have been shown to have impaired basal autophagy (Aoki et al., 2017; Webster et al., 2016). It should be noted that the interaction with SMCR8 and role of C9orf72 in autophagy has largely been linked to the long isoform of the protein, C9orf72-L, and it has been proposed that the C9orf72-S isoform may have a distinct function within the cell linked to its predominantly nuclear membrane localisation (Xiao et al., 2016). Other studies have implicated a role for C9orf72 in other cellular processes, including stress granule formation, mitochondrial function and axonal growth (Blokhuis et al., 2016; Maharjan et al., 2016; Sivadasan et al., 2016a).

1.6.2.1 Evidence for C9orf72 loss of function

The main supporting evidence for loss-of-function of C9orf72 protein causing disease is the numerous reports of reduced levels of all three C9orf72 mRNA transcripts in post-mortem *C9orf72* FTD/ALS patient brains and patient-derived neurons (Belzil et al., 2013; Ciura et al., 2013; DeJesus-Hernandez et al., 2011; Donnelly et al., 2013; Fratta et al., 2013; Gijselinck et al., 2012; van Blitterswijk et al., 2015; Waite et al., 2014; Xi et al., 2013). In line with reduced mRNA transcripts, several studies have demonstrated that the C9orf72 protein is also reduced in *C9orf72* FTD/ALS patient brains (Frick et al., 2018; Sivadasan et al., 2016a; Waite et al., 2014; Xiao et al., 2016).

The observed reduction in C9orf72 mRNA and protein levels in *C9orf72* repeat expansion carriers has led to the hypothesis that haploinsufficiency of the C9orf72 protein results in disease pathogenies. However, no disease causing mutations have been found in the coding regions of the *C9orf72* gene, with one exception, suggesting that it is more likely to be the repeat expansion that is important to disease

pathogenesis and not the C9orf72 protein itself (Harms et al., 2013). The rare exception to this is one sporadic ALS patient who lacked the *C9orf72* repeat expansion but carried a splice-site mutation in exon 5 of the *C9orf72* gene, potentially capable of causing heterozygous loss of function (Liu et al., 2016a). However, no familial history of neurodegeneration was reported in this case, suggesting this is probably a rare cause of disease and unlikely to be the mechanism responsible for most familial cases of FTD and ALS.

Further evidence to support the notion that loss of C9orf72 protein function is not a major mechanism of disease comes from a case study reporting a patient that was homozygous for the *C9orf72* repeat expansion. This rare case was not observed to have clinical or pathological features more severe than heterozygote cases, which would be expected if the pathological mechanism of disease were purely loss of protein function (Fratta et al., 2013).

A number of *in vitro* and *in vivo* models have been produced to investigate the impact of haploinsufficiency of the C9orf72 protein (Table 1.1). This summary highlights the variability in phenotypes resulting from loss of C9orf72 protein function, but a common theme emerging among many models is defects in autophagy and immune system function.

As previously described in Section 1.6.2, knockdown of C9orf72 in cell lines or primary neurons consistently results in an impairment in autophagy and the formation of p62-positive puncta, speculated to recapitulate *C9orf72* FTD/ALS pathology (Sellier et al., 2016; Webster et al., 2016; Yang et al., 2016).

Several groups have utilised pharmacological or genetic techniques to knockdown expression of the mouse ortholog of C9orf72 (3110043O21Rik), either in a tissue-

specific or ubiquitous manner. Regardless of the knockdown technique, degree of knockdown, timing of knockdown or tissue specificity, none of these models demonstrate neurodegeneration, mislocalisation of TDP-43 or any motor or behavioural phenotypes associated with *C9orf72* linked FTD and ALS. Instead, nearly all *C9orf72* whole-genome knockout mouse models exhibit immune-system related pathology, typically showing features of lymphadenopathy (enlarged lymph nodes), splenomegaly (enlarged spleen) and elevated levels of autoantibodies (Atanasio et al., 2016; Burberry et al., 2016; Jiang et al., 2016; O'Rourke et al., 2016; Sudria-Lopez et al., 2016; Sullivan et al., 2016; Ugolino et al., 2016). This effect was not observed in neuronal-specific knockdown or knockout models (Koppers et al., 2015; Lagier-Tourenne et al., 2013). These models clearly suggest that loss of C9orf72 protein function is insufficient to cause the neurodegeneration seen in *C9orf72* repeat expansion carriers, however the effect on the immune system is interesting given the increasing evidence to support immunological dysfunction in a range of neurodegenerative diseases (Doty et al., 2015).

Overall, the majority of evidence suggests that, although *C9orf72* FTD/ALS patients typically have reduced levels of *C9orf72* mRNA and protein, loss of function of the C9orf72 protein alone is not sufficient to cause *C9orf72*-linked FTD or ALS phenotypes. However, it is clear that a reduction in C9orf72 function may have important implications for essential autophagic processes and immune system regulation, both of which may contribute to or exacerbate disease progression in combination with gain of function mechanisms.

Table 1.1 Description of studies that have induced knockout or knockdown of C9orf72 orthologs in model organisms.

Reference	Model Organism	C9orf72 Ortholog	Method of knockdown	Level of Knockdown	Reported phenotype
Therrien *et al.* (2013)	*Caenorhabditis elegans*	alfa-1	Deletion of exons 3 and 4 / RNAi silencing	Complete knockdown. No detectable alfa-1 expresssion.	Age dependent motility defects. Paralysis. Degeneration of GABAergic motor neurons.
Ciura *et al.* (2013)	*Danio rerio*	zC9orf72	Antisense morpholino injection into fertilised eggs.	Not specified.	Disrupted neuronal arborisation. Shortening of motor neurons. Motor deficits. Reduced spontaneous swimming of larvae.
Lagier-Tourenne *et al.* (2013)			Intracerebroventricular injection of C9orf72 antisense oligonucleotide.	60 - 70 % knockdown in spinal cord and brain.	No evidence of TDP-43 pathology. No behavioural or motor deficits.
Koppers *et al.* (2015)			Brain-specific Cre-loxP mediated excision of exons 4 and 5.	Not specified.	No evidence of motor neuron degeneration, TDP-43 or ubquitin pathology. No motor deficits. No reduction in survival.
Atanasio *et al.* (2016)			Replacement of exons 2 - 11 and introns with a *lacZ* reporter.	Complete mRNA knockdown in homozygotes. 50 % reduction in heterozygotes.	Homozygote knockouts showed reduced motor function, reduced survival and evidence of immune system dysregulation.
Burberry *et al.* (2016)			Replacement of exons 2 - 6 with a *lacZ* reporter.	Significant, dose-dependent reduction in heterozygotes and homozygfotes.	Reduced survival in homozygotes and heterozygotes. No degeneration of spinal motor neurons. Evidence of immune system dysregulation and cortical neural inflammation.
			CRISPR/Cas9 mediated induction of deletion mutations in exon 4.		Reduced survival. Evidence of immune system dysregulation.
Jiang *et al.* (2016)	*Mus musculus*	3110043O21Rik	Replacement of exons 2 - 6 with a *lacZ* reporter.	Complete mRNA knockdown in homozygotes. 50 % reduction in heterozygotes.	Homozygotes showed reduced survival, evidence of immune system dysregulation, mild social and motor deficits.
O'Rourke *et al.* (2016)			Replacement of exons 2 - 6 with a *lacZ* reporter.	Complete mRNA knockdown in homozygotes. 50 % reduction in heterozygotes.	No motor or behavioural deficits. No evidence of neurodegeneration. Evidence of immune system dysregulation in homozygotes.
			Zinc finger nuclease mediated removal of exon 2 start codon.	Complete protein knockdown in homozygotes.	
Sudria-Lopez *et al.* (2016)			Whole genome Cre-loxP mediated excision of exons 4 and 5.	Not specified.	Reduced survival. No evidence of motor neuron degeneration or TDP-43 pathology. No motor deficits. Evidence of immune system dysregulation.
Sullivan *et al.* (2016)			CRISPR/Cas9 mediated frameshift mutation in exon 2 causing premature stop codon affecting isoforms 1 and 3.	Complete protein knockdown of isoform 1 in homozygotes. Isoforms 2 and 3 undetectable.	Evidence of immune system dysregulation.
Ugolino *et al.* (2016)			Replacement of exons 2 - 6 with a *lacZ* reporter.	Complete protein knockdown of isoform 1 in homozygotes. Isoforms 2 and 3 undetectable.	Reduced survival in homozygotes and heterozygotes. Evidence of immune system dysregulation in homozygotes. No evidence of neurodegeneration.

1.6.3 RNA toxic gain of function

The *C9orf72* hexanucleotide repeat expansion can be bi-directionally transcribed in both the sense and antisense direction (DeJesus-Hernandez et al., 2011; Gendron et al., 2013; Mizielinska et al., 2013; Zu et al., 2013). Due to the guanine-rich sequence of the expansion, the DNA and RNA form a highly stable secondary structure called a G-quadruplex (Fratta et al., 2012; Haeusler et al., 2014; Reddy et al., 2013). The RNA forms a parallel G-quadruplex structure (Figure 1.5 A) (Fratta et al., 2012; Reddy et al., 2013), while the DNA mostly forms an anti-parallel structure (Haeusler et al., 2014).

As described in Section 1.5.5.3, both sense and antisense RNA transcripts from repeat expanded *C9orf72* form foci in patient brain tissue and patient-derived cell lines (Figure 1.5 B) (Cooper-Knock et al., 2015; DeJesus-Hernandez et al., 2011; Donnelly et al., 2013; Gendron et al., 2013; Lagier-Tourenne et al., 2013; Lee et al., 2013; Mizielinska et al., 2013; Zu et al., 2013). Initially only foci corresponding to the sense strand of hexanucleotide repeat RNA were described, however it was later discovered that the repeat could be bi-directionally transcribed and foci consisting of RNA from transcription of the antisense strand were identified throughout the CNS, although these are generally less abundant. The RNA foci are typically localised to the nucleus of cells but have also been shown to form in the cytoplasm (Lagier-Tourenne et al., 2013; Mizielinska et al., 2013) and, although foci have predominantly been identified in neurons, they are also present to a lesser extent in glial cells in patient tissue (Gendron et al., 2013; Lagier-Tourenne et al., 2013; Mizielinska et al., 2013).
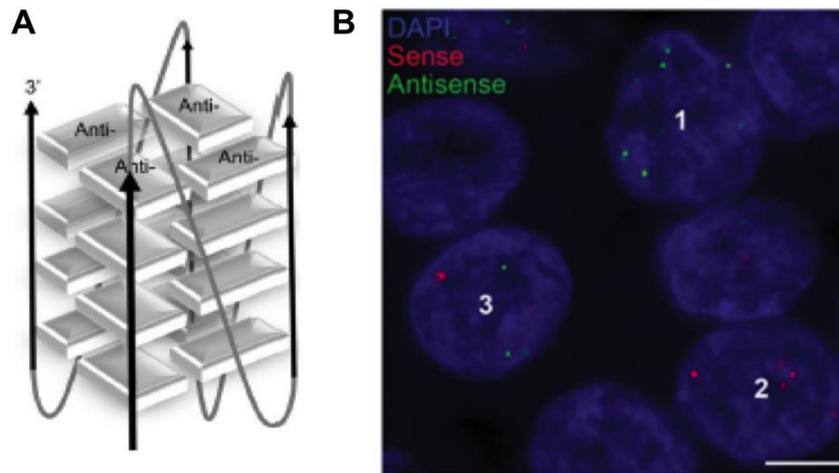
Figure 1.5 RNA G-Quadruplex and RNA foci in patient tissue.

A) Schematic representation of the parallel stranded RNA G-quadruplex secondary structure formed by *C9orf72* repeat RNA. Figure taken from Fratta *et al.* (2012). **B**) Confocal microscopy image of RNA *in-situ* hybridisation in a *C9orf72* FTD patient frontal cortex, showing DAPI stained cell nuclei (blue), sense RNA foci (red) and antisense foci (green). Antisense foci only are present in cell 1, sense foci only are present in cell 2, and both sense and antisense foci are present in cell 3. Figure taken form Mizielinska *et al.* (2013).

1.6.3.1 RNA binding protein sequestration

RNA foci are a feature of some other non-coding repeat expansion disorders, such as myotonic dystrophy type 1 (DM1), fragile X-associated tremor/ataxia syndrome (FXTAS) and various spinocerebellar ataxias (Shi et al., 2017). The proposed mechanism of toxicity in these diseases is the sequestration of essential RNA-binding proteins by RNA foci, leading to multiple defects in various cellular systems. For example, RNA foci in DM1 are known to sequester the essential RNA binding protein, muscleblind-like protein 1, which is responsible for the splicing of the muscle-specific chloride channel, CLCN1. This results in the mis-splicing and reduced expression of CLCN1 in skeletal muscle, giving rise to symptoms of myotonia experienced by DM1 patients (Mankodi et al., 2002; Miller et al., 2000). Toxicity induced by sequestration of RNA-binding proteins into RNA foci has been proposed as a similar mechanism of disease pathogenesis in *C9orf72* FTD and ALS.

Several groups have used various biochemical techniques, including RNA pulldowns, proteome arrays and co-localisation studies, to identify binding partners of the expanded repeat RNA that may be sequestered into RNA foci in several *in vitro* and *in vivo* models (Cooper-Knock et al., 2014a; Donnelly et al., 2013; Fay et al., 2017; Haeusler et al., 2014; Venditti et al., 2017; Xu et al., 2013). These studies have identified many potential foci-interacting proteins, the majority of which contain RNA recognition motifs and are involved in splicing, mRNA nuclear export and/or translation. Many of the proteins identified belong to the family of heterogenous nuclear ribonucleoproteins (hnRNPs), a group of proteins that regulate pre-mRNA splicing. Thus far hnRNP A1, A3, H/F, K, U have been shown to bind *C9orf72* repeat RNA (Cooper-Knock et al., 2015; Cooper-Knock et al., 2014b; Haeusler et al., 2014; Lee et al., 2013; Mori et al., 2013b; Sareen et al., 2013). Another common mechanism among proteins sequestered by *C9orf72* RNA foci is proteins involved in nuclear mRNA trafficking and nucleocytoplasmic transport. Ran-GAP1, ALYREF and Pur-α, proteins that play a role in the transport of mRNA nuclear export or import, have been shown to bind *C9orf72* repeat (Cooper-Knock et al., 2014b; Rossi et al., 2015; Sareen et al., 2013; Zhang et al., 2015).

Despite many proteins being identified, there is little reported overlap between studies, which likely reflects the diversity of models and methodologies used. Additionally, few of the identified proteins have been shown to co-localise consistently with RNA foci in patient post-mortem tissue. Proteins that have been shown to co-localise with a proportion of RNA foci in *C9orf72* patient brain tissue include hnRNP A1, hnRNP H, ALYREF, SRSF2, nucleolin and ADARB2 (Cooper-Knock et al., 2014a; Donnelly et al., 2013; Gendron et al., 2013; Haeusler et al., 2014; Hautbergue et al., 2017). However, the amount of co-localisation reported varies between studies,

possibly due to patient variability or co-localisation studies being performed in different brain regions.

The range of proteins identified to bind to *C9orf72* repeat RNA and the variation in the validation of these protein in human tissue doesn't point to an obvious protein candidate responsible for causing disease via sequestration into RNA foci. Additionally, many of the proteins identified have multiple cellular target and functions thus their sequestration could have knock-on effects for the regulation and expression of many other proteins, making it difficult to decipher specific pathogenic defects. Therefore, although *C9orf72* repeat RNA is clearly capable of binding many protein targets, and several of these have been shown to co-localise in diseased tissue, further research into the specific functional effects of protein sequestration into RNA foci is required to determine the role this may play in *C9orf72* FTD/ALS pathogenesis. Additionally, the potential toxicity of diffuse *C9orf72* repeat RNA that is not contained with foci should be investigated.

1.6.3.2 RNA toxicity models

A number of *in vivo* models have been generated to assess the contribution of *C9orf72* repeat RNA to disease pathogenesis independently of DPR protein toxicity (Discussed further in Section 1.6.4). Two independent *Drosophila melanogaster* models have demonstrated that the presence of nuclear sense RNA foci is insufficient to cause neurodegeneration alone (Mizielinska et al., 2014; Tran et al., 2015). One study found that, although RNA foci and a neurodegenerative phenotype were observed when the expanded GGGGCC repeat was overexpressed in *Drosophila* adult neurons or eyes, this phenotype was abolished when the repeats were interrupted with stop codons to prevent the translation of DPR proteins, suggesting that it is DPR proteins that mediate this neurodegeneration (Mizielinska et al., 2014).

Similarly, in a model where sense *C9orf72* RNA was ubiquitously overexpressed from the intron of a transgene in *Drosophila,* frequent RNA foci were observed in neurons and glia, but the *Drosophila* showed no evidence of neurodegeneration or reduced survival. This was attributed to efficient splicing of intronic repeat RNA preventing its nuclear export and translation (Tran et al., 2015). To date, only one study has investigated the contribution of antisense repeat RNA to neurodegeneration independently of DPR protein toxicity in *Drosophila.* In agreement with the findings for the sense RNA, adult neuronal expression of antisense repeat RNA that was interrupted with stop codons resulted in the formation of frequent RNA foci but did not significantly alter lifespan or neurodegeneration (Simone et al., 2018).

However, there are a small number of studies which imply that RNA toxicity, at least in part, contributes to disease. Recent work demonstrated that *in vitro* transcribed sense and antisense *C9orf72* repeat RNA can induce motor axonopathy independently of DPR proteins, which were not detected, when injected into developing zebrafish embryos (Swinnen et al., 2018). This toxicity was still observed when stop-codon-interrupted RNA was injected, suggesting that the observed phenotype was primarily induced by *C9orf72* repeat RNA. A degenerative effect of RNA was also observed in rat primary cortical and motor neurons transfected with 42 GGGGCC repeats located inside an artificial intron of eGFP (Wen et al., 2014). These neurons contained abundant RNA foci and demonstrated reduced survival in the absence of detectable DPR protein expression, indicating this phenotype was RNA mediated. Furthermore, co-expression of arginine-containing DPR proteins and the intronic expanded GGGGCC repeats led to a synergistic effect of decreased neuronal survival. Similarly, RNA toxicity was the attributable cause of age-dependent disruption in eye morphology in a *Drosophila* model with neuronal expression of 30 *C9orf72* hexanucleotide repeats interrupted by a 6-base pair restriction endonuclease

cut site (Xu et al., 2013; Zhang et al., 2015). However, it should be noted that the contribution of DPR proteins to toxicity in this *Drosophila* model, and in the transfected primary neurons mentioned previously, cannot be excluded as lack of DPR detection does not rule out the possibility that they are being produced. Indeed, detection of DPR proteins was subsequently demonstrated in this *Drosophila* model upon strong overexpression, thereby demonstrating that the phenotypes observed may not be independent of DPR protein induced toxicity (Zhang et al., 2015).

Overall the findings of various models investigating *C9orf72* repeat RNA toxicity indicate that the expanded repeat RNA alone is unlikely to be the main pathogenic cause of disease, but the expression and accumulation of the RNA may be a potential contributor to toxicity, depending on the model system employed. It is important to note, however, that the length of the RNA repeats analysed in these models is typically much shorter than what is observed in patients. Additionally, very few studies have assessed the effect of antisense RNA (Simone et al., 2018; Swinnen et al., 2018). Thus, further characterisation of existing models and generation of newer models expressing more disease-relevant repeat length is required to fully evaluate the contribution of RNA toxicity to *C9orf72* FTD/ALS disease pathogenesis.

1.6.4 Dipeptide repeat protein toxicity

The third mechanism of toxicity relates to the production and accumulation of unique proteins that are generated from the *C9orf72* expanded repeat RNA transcripts.

1.6.4.1 RAN translation

In 2011, an unconventional method of protein translation was discovered in a study investigating the CAG repeat expansion that occurs in DM1 and spinocerebellar ataxia type 8. The authors described a process whereby the CAG repeat was translated in the absence of an ATG start codon in all reading frames to produce three homopolymeric proteins; poly-glutamine, poly-alanine and poly-serine (Nichols et al., 2011). This non-canonical mechanism of translation was termed repeat associated non-ATG initiated (RAN) translation and has subsequently been shown to occur in other repeat expansion disorders including FXTAS, Huntington's disease, and myotonic dystrophy type 2 (Banez-Coronel et al., 2015; Cleary and Ranum, 2014; Todd et al., 2013; Wojciechowska et al., 2014a).

Following the discovery of RAN translation and the identification of the *C9orf72* hexanucleotide repeat expansion, in 2013 several groups demonstrated that the repeat is bidirectionally transcribed and translated by RAN translation in all reading frames from both the sense and antisense transcripts (Figure 1.6) (Ash et al., 2013; Gendron et al., 2013; Himuro et al., 2013; Mori et al., 2013a; Zu et al., 2013). The product of this translation is five different DPR proteins. Glycine-alanine (poly-GA) and glycine-arginine (poly-GR) are produced from translation of the sense transcript, and alanine-proline (poly-AP) and proline-arginine (poly-PR) are produced from translation of the antisense transcript. Glycine-proline (poly-GP) is produced from both transcripts.
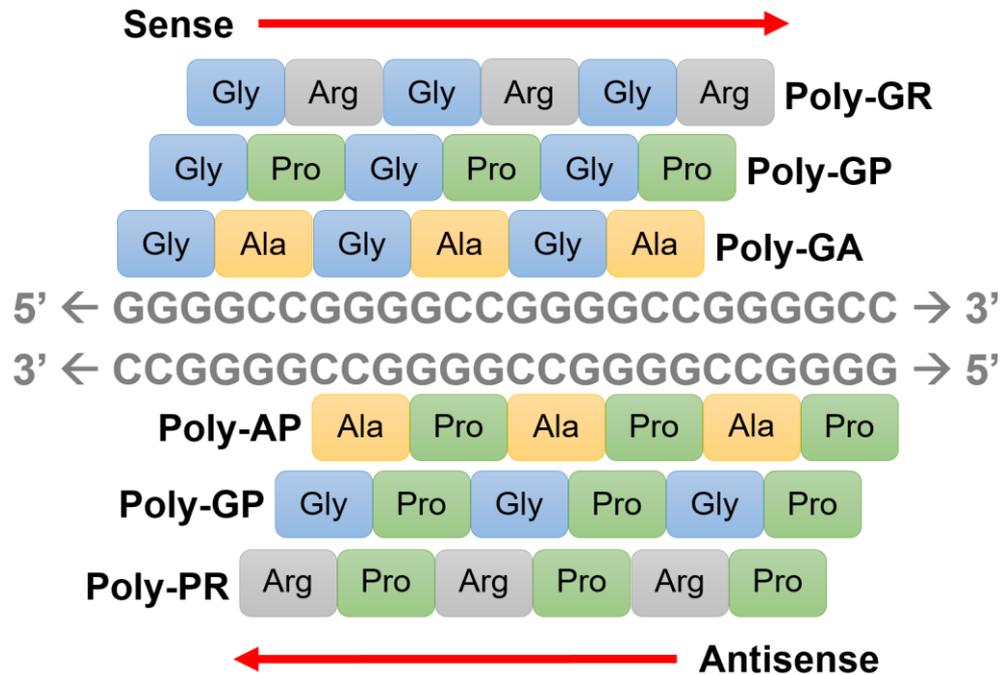
Figure 1.6 Bidirectional RAN translation of the C9orf72 repeat expansion.
Schematic representation of the possible DPR proteins produced by RAN translation of expanded GGGGCC (sense) and CCCCGG (antisense) repeats in all reading frames. Poly-GA, poly-GP and poly-GR are produced in the sense direction, and poly-AP, poly-GP and poly-PR are produced in the antisense direction.

The mechanism(s) of RAN translation initiation has recently been investigated in several repeat expansion disorders. For both the *C9orf72* repeat in the sense direction and the CCG repeat in FXTAS, RAN translation is hypothesised to initiate at a near-cognate start codon upstream of the repeat sequence. It is also believed to depend on canonical ribosomal translational machinery and is a mRNA 5'7-methylguanosine (m7G) cap dependent process (Green et al., 2017; Kearse et al., 2016; Sellier et al., 2017; Tabet et al., 2018). One study has also demonstrated that RAN translation of the *C9orf72* repeat can occur independently of m7G transcript capping, however this is at a much lower efficiency than cap-dependent translation (Cheng et al., 2018). For the *C9orf72* repeats, RAN translation of all three sense DPR proteins has been proposed to initiate at a CUG codon located upstream of the repeat

that is in frame with poly-GA. Translation of the other two sense DPR proteins is thought to arise from ribosomal frameshifting, a phenomenon that can be induced by structural abnormalities in mRNA, such as RNA G-quadruplexes, and has been reported to occur in other repeat expansion disorders (Wojciechowska et al., 2014b; Yu et al., 2014). However, it is important to note that use of the near-cognate start codon has only been demonstrated in experimental systems and the presence of the peptide sequence that would be predicted to lie upstream of the DPR proteins has not yet been detected in patient post-mortem tissue.

The initiation of the RAN translation process in the antisense direction of the *C9orf72* repeat has not yet been investigated. However, it has been noted that there are several ATG start codons in the 5' sequence upstream of the antisense repeat in frame with poly-GP and poly-PR, raising the possibility that ATG-initiated canonical translation is responsible for the production of these DPR proteins (Zu et al., 2013). This does not, however, explain the production of poly-AP. Further studies are required to investigate the mechanism of RAN translation initiation of the *C9orf72* repeat in the antisense direction.

1.6.4.2 Detection of DPR proteins in patient tissue

Inclusions formed of DPR proteins are observed in *C9orf72* patient post-mortem brain tissue, but have not been detected in other tissues, except for the Sertoli cells of the testes (Ash et al., 2013). All five of the DPR proteins form widespread neuronal cytoplasmic, and occasionally intranuclear, inclusions in *C9orf72* FTD/ALS patient brain, with high abundance in the cortex, hippocampus and cerebellum (Ash et al., 2013; Gendron et al., 2013; Himuro et al., 2013; Mori et al., 2013a; Zu et al., 2013). The DPR inclusions are predominantly localised to neurons but have also been detected, albeit rarely, in glial cells in the spinal cord (Schludi et al., 2015). These

inclusions frequently colocalise with p62 but not with TDP-43, and often have a unique star-like morphology (Ash et al., 2013; Himuro et al., 2013; Mackenzie et al., 2013).

The abundance of inclusions containing each DPR protein has been well characterised with inclusions formed from the sense strand DPR proteins more frequently observed than those derived from the antisense strand. Inclusions containing poly-GA and poly-GP are the most commonly observed, poly-GR containing inclusions are less common, and poly-AP and poly-PR are rarely observed (Davidson et al., 2016; Gendron et al., 2013; Gomez-Deza et al., 2015; Mackenzie et al., 2013; Mackenzie et al., 2015; Saberi et al., 2018; Schludi et al., 2015). Although the abundance of total DPR inclusions varies in different brain regions, the different DPR species do not segregate by brain region, with poly-GA and poly-GP consistently being observed most frequently (Davidson et al., 2016; Gomez-Deza et al., 2015; Mackenzie et al., 2013; Mackenzie et al., 2015; Mori et al., 2013c)

Several pathological studies have examined the distribution of DPR protein inclusions and their correlation with pathology, however there has been no consistent correlations found between the frequency of DPR protein inclusions, clinical disease severity, TDP-43 pathology or neurodegeneration in any brain region (Davidson et al., 2016; Davidson et al., 2014; Mackenzie et al., 2013; Mackenzie et al., 2015; Schludi et al., 2015). One exception has been reported in a recent study where an increased abundance of poly-GR, but not other DPR proteins, was observed in brain regions associated with neurodegeneration compared to unaffected areas (Saberi et al., 2018). Additionally, it should be noted that, due to their low abundance, many pathological studies have not attempted to correlate poly-PR and poly-AP with regional neurodegeneration. Most studies indicate no difference in DPR protein inclusion abundance or distribution between *C9orf72* FTD, ALS or FTD-ALS cases

(Davidson et al., 2014; Mackenzie et al., 2015; Schludi et al., 2015). However, one study has reported higher levels of soluble poly-GP in the cerebellum of *C9orf72* FTD and FTD-ALS cases compared to *C9orf72* ALS cases (Gendron et al., 2015). Although the majority of studies conclude that there is not a clear and consistent association between DPR protein abundance and neurodegeneration, each study reports slight variation in findings. These discrepancies likely reflect differences in antibody specificity, brain processing, post-mortem delay, brain regions analysed, and methodology used to determine DPR protein abundance. Caution should therefore be taken when comparing these studies.

Pathological assessment of a few rare *C9orf72* expansion carriers, where death has occurred prior to the end-stage of disease, have indicated that the accumulation of DPR proteins is likely to occur before the onset TDP-43 pathology. A handful of studies report symptomatic *C9orf72* cases that have extensive DPR protein pathology but have minimal accumulation or lack of TDP-43 pathology (Baborie et al., 2015; Gijselinck et al., 2012; Proudfoot et al., 2014; Rollinson et al., 2012; Vatsavayai et al., 2016). Additionally, there has been a report of a *C9orf72* mutation carrier exhibiting DPR protein pathology in surgically resected tissue prior to the onset of FTD or ALS (Vatsavayai et al., 2016). This indicates that accumulation of the DPR proteins occurs early in disease pathogenesis and that the aggregation of TDP-43 is not essential for clinical manifestation of the disease.

Little work has been done to characterise the post-translational modifications of the DPR protein in post-mortem tissue. One exception is the identification of co-localisation of poly-GR containing inclusions and immunoreactivity of an antibody raised against asymmetrically-dimethylated arginine (Boeynaems et al., 2016; Sakae et al., 2018), suggesting that this DPR protein may undergo post-translational arginine

methylation. It should be noted, however, that neither of these studies used antibodies specific to methylation of poly-GR, thus it is unknown whether the DPR protein itself is methylated or whether it co-localises with another protein that is methylated. The methylation of poly-GR in *C9orf72* FTD/ALS will be discussed further in Chapter 5.

1.6.4.3 DPR protein toxicity models

The relative toxicity of each DPR protein in isolation has been investigated in variety of cellular and animal model systems. The most common methods for studying the proteins independently include either synthesising and purifying individual DPR proteins *in vitro* and applying these directly onto cells or, exploiting codon degeneracy and altering the repetitive GGGGCC sequence to non-repetitive codons to allow expression of individual DPR proteins in an ATG-dependent manner.

The main criticism of these studies is that they rely on the overexpression of the DPR proteins which may not reflect levels typically found in patients. Similarly, the validity of the length of the DPR proteins used in these studies is another limitation. The length of DPR proteins produced in *C9orf72* patients is currently unknown but given that patients can have several hundred or thousands of copies of the GGGGCC repeat, if RAN translation extends all the way across the repeat then the DPR proteins produced in these patients will likely be much longer than those assessed in model systems.

Comparison of the five DPR proteins in these model systems has indicated that the arginine containing proteins (poly-PR and poly-GR) are the most toxic, with poly-GA also causing toxicity in a select few models (Flores et al., 2016; Jovicic et al., 2015; Lee et al., 2016; Mizielinska et al., 2014; Schludi et al., 2015; Swaminathan et al.,

2018; Swinnen et al., 2018; Wen et al., 2014; Yamakawa et al., 2015; Zhang et al., 2014).

In contrast, poly-AP and poly-GP are rarely demonstrated to elicit toxicity and these will not be discussed further. The toxicity of each of the DPR proteins in various model systems is summarised in Table 1.2., which clearly demonstrates that the arginine-containing DPR proteins as the most toxic DPR protein species across a wide variety of models. Poly-GA clearly induces moderate toxicity but only in select models, whereas poly-AP and poly-GP do not induce toxicity in any model tested.

Table 1.2 Comparison of studies that investigate the toxicity of individual DPR proteins in different model systems.

Reference	Model System	GR	PR	GA	AP	GP
Kwon *et al.* (2014)	U2OS cells / human astrocytes	20 *	20 *			
May *et al.* (2014)	HEK293 cells	149	175	175	175	80
Mizielinska *et al.* (2014)	*Drosophila melanogaster*	36; 100	36; 100	36; 100	36; 100	
Wen *et al.* (2014)	*Drosophila melanogaster*		50	50	50	
Wen *et al.* (2014)	Rat primary neurons	25 - 400	25 - 200	25 - 400	25 - 200	25; 50
Wen *et al.* (2014)	Human iPSC derived neurons		50	50		
Zhang *et al.* (2014)	HEK293 cells / Rat primary neurons	50	50	50	50	47
Freibaum *et al.* (2015)	*Drosophila melanogaster*	50		50		47
Jovicic *et al.* (2015)	*Saccharomyces cerevisiae*	50; 100	50	50	50	
Schuldi *et al.* (2015)	Rat primary neurons	149	175	175		80
Tao *et al.* (2015)	HEK293 / NSC34 cells	30	30	30	30	30
Yamakawa *et al.* (2015)	Neuro-2a cells	100	100	100	100	100
Yang *et al.* (2015)	*Drosophila melanogaster*	80	80	80		
Boeynaems *et al.* (2016)	*Drosophila melanogaster*	50	25;50	25;50	25;50	
Chang *et al.* (2016)	Human neuroblastoma BE(2) cells			15 *		
Flores *et al.* (2016)	Rat primary neurons	3; 6 *		3; 6 *		3; 6 *
Kanekura *et al.* (2016)	NSC34 cells	20 *	20 *	21 *		
Khosravi *et al.* (2016)	Rat primary neurons	149	175	149		
Lopez-Gonzalez *et al.* (2016)	Human iPSC derived neurons	80		80		
Lee *et al.* (2016)	Neuro-2a cells	50	50	50	50	47
Lee *et al.* (2016)	*Drosophila melanogaster*	50	50	50	50	47
Lee *et al.* (2017)	HEK293 cells	125	125	125	125	125
Lee *et al.* (2017)	*Gallus gallus*			125	125	
Ohki *et al.* (2017)	*Danio rerio*			80		
Rudich *et al.* (2017)	*Caenorhabditis elegans*	50	50	50	50	
Gupta *et al.* (2017)	Mixed spinal cord neuron cultures	20 *	20 *			
Swinnen *et al.* (2018)	*Danio rerio*	50	50	50	50	50
Xu *et al.* (2018)	*Drosophila melanogaster*	36	36			

Key	
(red)	Toxic
(orange)	Mild or length dependent toxicity
(green)	Not toxic
(pale yellow)	Details on toxicity not provided
(grey)	Not studied

Numbers indicate the length of the repeat(s) tested. * denotes that DPR peptides were exogenously applied rather than overexpressed. Note: In some studies, the DPR proteins are epitope tagged

The abundance of poly-GA inclusions in *C9orf72* patient post-mortem brain has led many groups to investigate its toxicity in model systems (Chang et al., 2016; Flores et al., 2016; May et al., 2014; Mizielinska et al., 2014; Schludi et al., 2017; Schludi et al., 2015; Yamakawa et al., 2015; Zhang et al., 2016; Zhang et al., 2014). Overexpression of poly-GA in transfected cells results in the formation of insoluble cytoplasmic, and occasionally intranuclear, inclusions that co-localise with p62, similar to those seen in patient brain (May et al., 2014; Schludi et al., 2015; Yamakawa et al., 2015; Zhang et al., 2014). Several groups have reported that overexpression of various lengths of poly-GA (50 – 175 repeats) is toxic in human cell lines and cultured primary neurons (Barker et al., 2017; May et al., 2014; Yamakawa et al., 2015; Zhang et al., 2014). Similarly, application of synthetic poly-GA to cultured cells was able to reduce cell viability in a dose-dependent manner (Chang et al., 2016; Flores et al., 2016). The mechanism of this poly-GA induced toxicity remains debated but has been attributed to various factors including dysfunction of the ubiquitin-proteasome system, sequestration of Unc119 and the disruption of the nucleocytoplasmic transport of TDP-43 (Khosravi et al., 2016; May et al., 2014; Yamakawa et al., 2015; Zhang et al., 2014). Other cellular models, however, have not found poly-GA capable of inducing toxicity (Kanekura et al., 2016; Tao et al., 2015; Wen et al., 2014; Yang et al., 2015).

Similar discrepancy is seen in reports of poly-GA toxicity in *in vivo* models. One study found that poly-GA moderately reduces *Drosophila* lifespan when expressed in adult neurons (Mizielinska et al., 2014), but the majority of *Drosophila* models demonstrate that poly-GA has no neurodegenerative effect (Boeynaems et al., 2016; Freibaum et al., 2015; Lee et al., 2016; Wen et al., 2014; Yang et al., 2015). Ubiquitous expression of poly-GA in zebrafish embryos was highly toxic but this was due to pericardial edema and not a neuronal or motor phenotype (Ohki et al., 2017). An additional

zebrafish model has found poly-GA to induce locomotor defects (Swaminathan et al., 2018). Two mice models have been generated to investigate poly-GA toxicity (Schludi et al., 2017; Zhang et al., 2016). Both models displayed accumulation of poly-GA in neuronal cytoplasmic inclusions, albeit in differing regions of the CNS, and exhibited some form of motor dysfunction. However, only one displayed evidence of neurodegeneration, and little or no TDP-43 pathology was observed in either model. In summary, the studies on poly-GA toxicity suggest that this DPR protein is likely to have mild toxic effects but the expression of poly-GA alone is insufficient to recapitulate the neurodegeneration seen in *C9orf72* patients.

In contrast to poly-GA, nearly all studies investigating the arginine containing DPR proteins suggest that poly-PR and poly-GR are the most toxic species in all model systems tested. Overexpression or synthetic application of poly-PR and poly-GR is highly toxic to a range of neuronal and non-neuronal cell lines, primary neuronal cultures and cultures human astrocytes (Flores et al., 2016; Kanekura et al., 2016; Kwon et al., 2014; Lee et al., 2016; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2015). Additionally, expression of the arginine-rich DPR proteins is toxic to yeast, induces reduced mobility and eventual paralysis in *Caenorhabditis elegans*, leads to motor axonopathy in Zebrafish, and causes reduced lifespan, locomotor phenotypes and severe degeneration of the eye when expressed in various *Drosophila* models (Boeynaems et al., 2016; Freibaum et al., 2015; Jovicic et al., 2015; Lee et al., 2016; Mizielinska et al., 2014; Rudich et al., 2017; Swaminathan et al., 2018; Swinnen et al., 2018; Wen et al., 2014; Xu and Xu, 2018; Yang et al., 2015)

The mechanism by which the arginine-rich DPR proteins induce toxicity has been widely investigated. Several groups have implicated dysfunction in nucleocytoplasmic transport as a potential mechanism with four independent groups identifying various

nuclear pore proteins and regulators of nucleocytoplasmic transport as modifiers of toxicity in yeast and *Drosophila* (Boeynaems et al., 2016; Freibaum et al., 2015; Jovicic et al., 2015; Zhang et al., 2015). Two of the groups specifically linked the defect in nucleocytoplasmic transport to arginine-rich DPR protein toxicity (Boeynaems et al., 2016; Jovicic et al., 2015), whilst the other two linked it to general expression of the *C9orf72* hexanucleotide repeat (Freibaum et al., 2015; Zhang et al., 2015), making it difficult to dissect DPR protein toxicity from RNA toxicity. In support of nucleocytoplasmic transport defects being a cause of toxicity, abnormal nuclear pore protein staining has been identified in *C9orf72* ALS patient post-mortem tissue, and nucleocytoplasmic transport defects have been confirmed in *C9orf72* patient-derived fibroblasts and induced pluripotent stem cells (iPSCs) (Li et al., 2018; Zhang et al., 2015). Whether these defects are directly caused by the arginine-rich DPR proteins remains to be determined however, all these studies clearly highlight dysfunctional nucleocytoplasmic transport as a potential pathogenic mechanism in *C9orf72*-linked disease.

Another common mechanism hypothesised to be the cause of poly-GR and poly-PR toxicity is impaired ribosomal function and consequently impaired protein synthesis, which may be linked to nucleolar stress. In most cellular studies, both poly-PR and poly-GR localise to the nucleus, possibly due to their arginine-rich sequences resembling a nuclear localisation signal. Within the nucleus, the arginine-containing DPR proteins often localise to the nucleolus, a sub-nuclear compartment that is the primary site of ribosome biogenesis and assembly, and several studies have found this to induce nucleolar enlargement, cause translocation of nucleolar markers or cause disrupted processing of ribosomal subunits (Kwon et al., 2014; Mizielinska et al., 2017; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2015). Given that the nucleolus plays a key role in the synthesis of ribosomes, it follows that nucleolar

stress seen in poly-GR and poly-PR models could be a cause or consequence of ribosomal impairment. This hypothesis fits with several studies that have published the interactomes of DPR proteins which have observed ribosomal proteins and RNA-binding proteins as major interactors of the arginine-containing DPR proteins (Boeynaems et al., 2017; Kanekura et al., 2016; Lee et al., 2016; Lin et al., 2016; Lopez-Gonzalez et al., 2016; Shi et al., 2018; Yin et al., 2017). Additionally, recent work on an overexpressing adeno-associated virus (AAV) mouse model of poly-GR, which exhibited behavioural deficits and neurodegeneration, found that poly-GR co-localised with ribosomal subunits and altered expression of genes encoding ribosomal subunits. The authors of the study also demonstrated that poly-GR impaired protein translation in HEK293 cells and in neurons containing poly-GR in the cortex of the poly-GR AAV mice (Shi et al., 2018).

In line with their localisation to ribosomes and the nucleolus, poly-GR and poly-PR have been demonstrated to undergo reversible spontaneous liquid-liquid phase separation (LLPS) *in vitro* (Boeynaems et al., 2017; Lee et al., 2016; Lin et al., 2016). This is a naturally occurring phenomenon that underlies the formation of various membrane-less organelles, such as the nucleolus and stress granules, and occurs when proteins demix from an aqueous solution and form dynamic liquid-like droplets. This process is believed to be mediated by low complexity domains (LCD) within the phase separating protein (Brangwynne et al., 2015; Kato et al., 2012; Lin et al., 2015; Molliex et al., 2015). Several studies have demonstrated that expression of poly-PR and poly-GR in immortalised human cell lines can lead to the formation of stress granules, further supporting the notion that these proteins undergo phase separation (Boeynaems et al., 2017; Kanekura et al., 2016; Lee et al., 2016; Wen et al., 2014). Interestingly, as well as being able to phase separate themselves, several proteomic studies have found poly-PR and poly-GR to associate with many LCD-containing

proteins that are likely capable of LLPS (Boeynaems et al., 2017; Lee et al., 2016; Lin et al., 2016). Furthermore, these studies have demonstrated that the binding of poly-PR and poly-GR to these proteins disrupts their ability to reversibly phase separate and impairs the dynamics of membraneless organelles. Both poly-PR and poly-GR have been shown to strengthen the multivalent interactions that comprise the liquid phase of stress granules in living cells, and the addition of arginine-containing DPR proteins can reduce the required concentrations of other proteins to phase separate (Lee et al., 2016). This suggests that the arginine-containing DPR proteins are able to both promote and disturb phase transition. This may alter the assembly, dynamics and function of membrane-less organelles and has been proposed to play a role in disease by upsetting the internal functional balance of membrane-less organelles. This hypothesis links together the localisation of poly-PR and poly-GR to various membrane-less organelles, such as the nucleolus and stress granules, and draws together many of the functions impaired by these proteins, such as nucleolar dysfunction, disrupted ribosomal biogenesis and impaired protein translation.

Taken together, these studies imply that the arginine-containing DPR proteins are the most toxic of the *C9orf72* DPR protein species and that their toxicity is likely a result of their biophysical properties causing broad disruption to the dynamics and functions of several important cellular pathways.

1.6.5 Combined models of *C9orf72* toxicity

Although the models that separate RNA and DPR protein toxicity are useful in providing information about the respective toxicities of these two species, it is important to remember that *C9orf72* repeat expansion carriers exhibit both RNA foci and DPR inclusions, raising the possibility that their mechanisms of toxicity may not be mutually exclusive. Therefore, to better recapitulate disease pathogenesis, several

mouse and patient-derived cellular models have been generated to assess RNA and DPR protein toxicity in combination. Depending on how they are designed, some of these models also factor in the contribution of C9orf72 protein haploinsufficiency, but the majority focus on combined toxicity derived from the repeat RNA and DPR proteins.

1.6.5.1 Patient-derived cellular models

Patient iPSC-derived neurons or induced neurons are a powerful tool for modelling neurodegeneration diseases *in vitro* as they provide a unique system where human genes and proteins are expressed at physiologically relevant levels in disease-relevant cell types. They also have the added benefit that they have the capacity to model early pathogenic events and so can provide a useful insight to disease that may not be gained from disease end-point post-mortem tissue.

Many groups have generated iPSCs from fibroblasts derived from *C9orf72* ALS, FTD/ALS and FTD patients, and have subsequently differentiated these into cortical or motor neurons to model disease. Other groups have directly reprogrammed patient fibroblasts into induced neurons. Neurons derived from these cells typically recapitulate *C9orf72* disease pathology as they exhibit RNA foci and endogenous expression of some DPR proteins (Almeida et al., 2013; Dafinca et al., 2016; Donnelly et al., 2013; Sareen et al., 2013; Simone et al., 2018; Su et al., 2014b; Westergard et al., 2016). There is conflicting evidence, however, to suggest that these patient-derived cellular models recapitulate haploinsufficiency of the C9orf72 protein. Some groups observe no reduction in *C9orf72* transcript and protein levels compared to controls (Sareen et al., 2013; Selvaraj et al., 2017), whilst other have reported reduced transcript abundance of some of the *C9orf72* variants (Almeida et al., 2013; Donnelly et al., 2013; Shi et al., 2018).

Compared to neurons derived from control patients, *C9orf72* patient-derived neurons do not show reduced viability (Devlin et al., 2015; Sareen et al., 2013; Shi et al., 2018), however, a variety of cellular defects have been described and attributed to the repeat expansion mutation. Changes to expression of genes and proteins involved in neuronal excitability have been reported in some studies (Devlin et al., 2015; Sareen et al., 2013), while others have demonstrated that *C9orf72* neurons are more susceptible to glutamate induced excitotoxicity (Donnelly et al., 2013; Shi et al., 2018). Other alterations observed include increased sensitivity to autophagy inhibitors (Almeida et al., 2013), altered calcium signaling (Dafinca et al., 2016) and changes to gene expression (Donnelly et al., 2013; Sareen et al., 2013; Selvaraj et al., 2017). Although various defects have been reported, there has been no common phenotype reported across all the *C9orf72* patient-derived neuron studies. This may be explained by differences in differentiation protocols or patient-specific variability between lines. Similarly, there has been no consensus on mechanism of toxicity responsible for causing observed phenotypes. Several groups have shown that cellular defects observed in *C9orf72* patient-derived neurons can be corrected using antisense oligonucleotides or small molecules targeted against the repeat RNA, suggesting that the observed phenotypes are a result of gain of function mechanisms, although it is unclear whether the phenotype amelioration is linked to RNA or DPR protein toxicity as both of these were reduced (Donnelly et al., 2013; Sareen et al., 2013; Simone et al., 2018; Su et al., 2014b). In contrast, another group has rescued *C9orf72* patient-derived neuron phenotypes by restoring levels of the C9orf72 protein or augmenting its function, implicating haploinsufficiency as the primary mechanism of toxicity (Shi et al., 2018). It is possible that all three proposed mechanisms of toxicity are acting in *C9orf72* patient-derived cellular models, however further studies are required to deduce this.

1.6.5.2 *C9orf72* repeat expansion transgenic mouse models

Several groups have attempted to model the *C9orf72* repeat expansion in mice, either by using AAV-driven overexpression of the hexanucleotide repeats (Chew et al., 2015; Herranz-Martin et al., 2017), or by the integration of human bacterial artificial chromosomes (BACs) containing *C9orf72* repeat expanded DNA (Jiang et al., 2016; Liu et al., 2016b; O'Rourke et al., 2015; Peters et al., 2015) (Table 1.3). In all of these studies, repeat-expanded RNA foci and at least some of the DPR proteins were detected in the mice, however the degree of neurodegeneration and phenotypes observed varied significantly.

AAV overexpression of 66 GGGGCC repeats in the CNS led to accumulation of RNA foci, inclusions of several DPR protein species and TDP-43 pathology, as well as cortical neuron loss and the development of abnormal behaviours similar to clinical symptoms of *C9orf72* FTD/ALS, such as hyperactivity, anxiety and motor deficits (Chew et al., 2015).

A mixture of phenotypes has been observed in the *C9orf72* BAC mice, all of which exhibit widespread RNA foci and DPR protein pathology. Two models show no evidence of neurodegeneration, motor or cognitive abnormalities (O'Rourke et al., 2015; Peters et al., 2015), one model expressing 450 repeats showed age-dependent impairment in cognitive function (Jiang et al., 2016), and the final BAC mouse model displayed a mixture of acute of progressive phenotypes characterised by widespread neurodegeneration, reduced survival, and behavioural and motor deficits (Liu et al., 2016b). A possible explanation for these phenotypic differences could be the genetic background of each mouse model. All four *C9orf72* BAC mice were produced in different mouse strains and this has previously been shown to cause different phenotypes in other mouse models of other neurodegenerative diseases (Mohajeri et al., 2004; Sedelis et al., 2000; Strong et al., 2012). Another explanation could be

differences in amount of the *C9orf72* gene included and the length of 3' and 5' *C9orf72* flanking sequences present in each BAC, as this may have affected the expression or regulation of the gene. This, however, is unlikely as the BACs from non-phenotypic and phenotypic mice overlapped. It is more likely that differences in the level of transgene expression could explain the phenotypic differences. Models that had strong overexpression of even relatively short repeat lengths demonstrated a neurodegenerative phenotype (Chew et al., 2015; Liu et al., 2016b), whereas models that included longer repeat lengths but were expressed at the level of the endogenous *C9orf72* ortholog were non-phenotypic (O'Rourke et al., 2015; Peters et al., 2015). Furthermore, one study showed that RNA foci and DPR protein pathology positively correlated with the expression levels of RNA containing the repeat expansion (Jiang et al., 2016). This suggests that the non-phenotypic models may not have developed neurodegeneration or associated phenotypes because they did not produce RNA foci or DPR proteins at a level sufficient to induce toxicity. The mice models also highlight the importance of TDP-43 in causing disease-relevant phenotypes as it is only the models that exhibit TDP-43 inclusions or increased TDP-43 phosphorylation which show the most prominent motor or behavioural phenotypes (Chew et al., 2015; Jiang et al., 2016; Liu et al., 2016b).

Despite the neurodegenerative and phenotypic variability between these models, the fact that *C9orf72* repeat is capable of inducing motor and behavioural abnormalities similar to clinical symptoms of *C9orf72* FTD/ALS in at least some models strongly suggests that gain of function mechanisms is sufficient alone to cause pathogenesis. Whether this is due to the repeat expanded RNA, DPR proteins or a combination of both, however, remains to be determined.

Table 1.3 Comparison of transgenic C9orf72 repeat expansion mouse models.

Reference	Background	Transgenic Type	Repeat length	Phenotype	TDP-43 Pathology	Sense Foci	Antisense Foci	GP	GA	GR	AP	PR
Chew et al. (2015)	C57BL/6J	AAV based overexpression with chicken beta actin promoter. Intracerebroventricular injection at P0.	66 repeats	Behavioural, social and motor impairments. Evidence of neurodgeneration and gliosis in the cortex.	✔	✔	?	✔	✔	✔	?	?
O'Rourke et al. (2015)	C57BL/6J	Human C9orf72 BAC transgenic. Full C9orf72 gene region plus ~ 110 kb upstream and ~20 kb downstream.	~ 100 - 1000 repeats; unknown copy number	No behavioural, social or motor abnormalities. No evidence of neuromusculas innervation dysfunction. No evidence of neuronal loss or gliosis.	✗	✔	✔	✔	✗	✗	✗	✗
Peters et al. (2015)	SJL/B6	Human C9orf72 BAC transgenic. Exons 1 - 6 of the C9orf72 gene plus 140.5 kb upstream.	300/500 repeats; 2 copies	No behavioural, social or motor abnormalities. No evidence of neuronal loss or gliosis. No alterations to gene expression.	✗	✔	✔	✔	✗	✗	✗	✗
Jiang et al. (2016)	C57BL6/CH3 injected, backcrossed into C57BL/6J	Human C9orf72 BAC transgenic. Exons 1 -5 of C9orf72 gene plus 140 kb upstream.	450 repeats; 1 copy; high expression	No motor abnormalities in any line tested. Males of the medium and high repeat expressing lines displayed cognitive and behavioural impairments. Mild age-dependent neuronal loss in medium and high expressors.	✗	✔	✔	✔	✔	✔	✗	✗
			450 repeats; 1 copy; medium expression		?	✔	✔	✔	?	?	✗	✗
			450 repeats; 1 copy; low expression		?	✔	✔	✔	?	?	✗	✗
			110 repeats; 1 copy; medium expression		?	✗	✗	✗	?	?	✗	✗
Liu et al. (2016)	FVB/NJ	Human C9orf72 BAC transgenic. Full C9orf72 gene region plus ~ 52 kb upstream and ~ 19.4 kb downstream.	500 repeats; 1 copy	500 and 550/32 mice either develop acute or slower progressive phenotype, or are asymptomatic. Reduced survival. Symptomatic mice show behavioural and motor impairments.Evidence of neurodegeneration in acute end-stage and progressive mice. No neurodegeneration in asymptomatic mice.	✔	✔	✔	✔	✔	?	?	?
			500/32 repeats; 2 copies		✔	✔	✔	✔	✔	?	?	?
			36/29 repeats; 4 copies		✔	✗	✗	?	✔	?	?	?
Herranz-Martin et al. (2017)	C57BL/6J	AAV based overexpression with Cytomegalovirus promoter. Cerebrospinal fluid injection at P1.	10 repeats	Neither repeat length showed evidence of gliosis or neurodegeneration. Neuromuscular junction defects and gait abnormalities in 102 repeat mice.	✔	✔	?	?	✗	?	?	?
			102 repeats interrupted every 10 -17 repeats by TCGAG sequence		✔	✔	?	?	✔	?	?	?

1.7 Thesis aims

Although significant advances in the modelling and understanding of the pathobiology of *C9orf72*-linked FTD and ALS have been made in recent years, the primary cause of disease remains unknown. The majority of studies indicate that loss of function of the C9orf72 protein is unlikely to be insufficient to cause disease alone whilst the arginine-containing DPR proteins have consistently been shown to induce neurodegenerative phenotypes. However, the relative contribution of toxicity from repeat containing RNA or other DPR protein species is still debated. Recapitulating the disease phenotype in *in vivo* models has led to mixed results, with the majority not modelling the disease when haploinsufficiency, RNA toxicity and DPR toxicity are studied independently, or in combination. The models that show a neurodegenerative phenotype which best mimics human *C9orf72* disease are ones that exhibit multiple aspects of *C9orf72* pathology, suggesting that the disease is likely a result of a combination of pathogenic factors. Nevertheless, understanding the independent contribution of DPR proteins, repeat RNA and the C9orf72 protein itself, are crucial to elucidating the role they each play in disease pathogenesis. Understanding the mechanism by which they each mediate toxicity will provide valuable information that may assist in the development of future treatments targeting *C9orf72* FTD and ALS.

The work in this thesis primarily focuses on DPR protein induced toxicity. By using post-mortem *C9orf72* patient brain tissue and cellular systems, this work explores a possible method of mitigating DPR protein aggregation, characterisation of potential post-translational modifications of these proteins, and the generation of a novel patient-derived iPSC model that can be utilised to study the endogenous production and toxicity of DPR proteins *in vivo*.

The primary aims of work in this thesis related to *C9orf72* FTD/ALS are;

1. To determine whether the aggregation of DPR proteins can be reduced using the molecular chaperone, HSJ1a (Chapter 3).

2. To use CRISPR-Cas9 genome engineering to generate a *C9orf72* patient-derived iPSC line where endogenously produced DPR proteins are RAN translated with the addition of a C-terminal eGFP tag (Chapter 4).

3. To characterise whether poly-GR inclusions in *C9orf72* patient post-mortem brain tissue is post-translationally modified by arginine methylation using novel methylated-GR antibodies (Chapter 5).

Additionally, this thesis contains a chapter un-related to *C9orf72*-linked FTD/ALS. Instead, this chapter explores the pathology of other subtypes of FTLD. The primary aim of this work was;

4. To characterise the abundance, distribution and morphology of hnRNP R and hnRNP Q inclusions in post-mortem brain tissue of different FTLD subtypes (Chapter 6).

Chapter 2. Materials and methods

2.1 Molecular techniques used in Chapter 3

2.1.1 Generation of GFP tagged constructs

eGFP tagged DPR pcDNA3.1 (pcDNA3.1-DPR) constructs were generated by subcloning the eGFP coding sequence from a pEGFP-C3 construct into pcDNA3.1 constructs containing the coding sequence for each DPR protein previously generated by Mizielinska et al., 2014. pcDNA3.1-DPR was digested with *BamH1* and *Bmt1*, and pEGFP-C3 was digested with *BglII* and *Bmt1*, both at 37 °C for 2 hours in NEBuffer 3.1. The eGFP insert was ligated into the linearised pcDNA3.1-DPR construct at a molar ratio of 3:1 with T7 DNA Ligase and 2x T7 DNA Ligase Buffer at 25 °C for 30 minutes (Alberts et al., 2002; Green and Sambrook, 2012). All enzymes and buffers used were purchased from New England Biolabs.

2.1.2 Generation of mCherry and BFP tagged constructs

mCherry and blue fluorescent protein (BFP) tagged pcDNA3.1-DPR constructs were generated using a combined digestion and ligation method. mCherry and BFP inserts were generated by digesting pmCherry-C3-SOD1 and pBAD-mtagBFP2 with *BamH1* and *Bmt1* in NEBuffer 3.1 at 37 °C for 2 hours. The mCherry or BFP inserts were combined with a non-linearised pcDNA3.1-DPR construct, 10mM ATP, *Bmt1*, *BamH1*, NEBuffer 3.1 and T4 DNA Ligase. The reaction was incubated in an Mastercycler Gradient (Eppendorf) where it cycled through 6 rounds of 37 °C for 5 minutes (digestion) followed by 21 °C for 5 minutes (ligation). All enzymes and buffers used were purchased from New England Biolabs.

2.1.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments and polymerase chain reaction (PCR) products by size. Between 1 % and 1.5 % UltraPure™ Agarose

powder (Life technologies) was added to 100 ml of 1X TBE (tris-boric acid-ethylenediaminetetraacetic acid) depending on the predicted size of the product of interest. The gel mixture was heated in a microwave for 2 minutes and left to cool to approximately 50 °C before the addition of 0.005 % SafeView Nucleic Acid Stain (NBS Biologicals) or 10 µL Ethidium Bromide solution (500 µg/ml in H2O; Sigma-Aldrich) per 100 mL gel, inside a chemical fume hood. The gel solution was poured into a holder containing a lane comb and left to set at room temperature. Samples were diluted in 6X Gel Loading Dye, Purple (New England Biolabs) to a final concentration of 1X. Samples were loaded onto agarose gels and electrophoresis was carried out in 1 x TBE buffer at 100 V for between 30 - 90 minutes. DNA bands of interest were visualised using a Gel Doc XR system with Quantity One software (BioRad). DNA band sizes were determined by comparing the position of the DNA band on the gel relative to the position of the Quick Load Purple 1 kb DNA ladder (New England Biolabs) or Hyperladder I (Bioline) for bands of interest up to 1 kb and Hyperladder IV (Bioline) for bands of interest larger than 1 kb.

2.1.4 Gel extraction

Following restriction enzyme digest, DNA bands of interest were visualised using a UV transilluminator and were excised using a scalpel, with care taken to limit exposure to the UV radiation. DNA was extracted using the QIAquick Gel extraction kit (QIAGEN) according to the manufacturer's instructions. DNA was eluted in 50 µL double-distilled water (ddH$_2$O).

2.1.5 Preparation of LB broth

To make 500 mL of Luria-Broth (LB), 10g LB powder (Sigma-Aldrich) was added to 500 mL ddH$_2$O and sterilised by autoclaving. The solution was left to cool, sealed and stored at room temperature until use. Immediately prior to use either 50 µg/mL

ampicillin or 10 µg/mL kanamyacin (depending on the resistance gene in the plasmid) was added to the solution.

2.1.6 Preparation of LB agar plates

To make 500 mL of LB agar, 7.5 g Agar powder (Sigma-Aldrich) was added to 500 mL LB broth (Section 2.1.5) and was sterilised by autoclaving. The solution was left to cool to approximately 40 °C before the addition of either 50 µg/mL ampicillin or 10 µg/mL kanamyacin (depending on the resistance gene in the plasmid). LB agar was poured into 10 cm Petri dishes with approximately 20 ml added to each dish. The agar was left to set at room temperature before covering and storing the dishes at 4 °C until use.

2.1.7 Transformation of chemically competent *E.coli*

Stable Competent *E.coli* C3040H (New England Biolabs) were thawed on ice for 10 minutes. 1 - 2 µl of DNA containing ~ 100ng was added to 25 µl of *E.coli.* Cells were incubated on ice for 30 minutes before heat shock at 42 °C for 45 seconds, followed by a further incubation on ice for 2 minutes. 950 µl of room temperature SOC growth media was added to the cells and then incubated at 37 °C shaking at 250 rpm for 1 hour. 100 µl of the cultures was spread on LB agar plates containing the appropriate antibiotic. Agar plates were incubated overnight at 37 °C.

2.1.8 Colony polymerase chain reaction

To screen for colonies that had been transformed by the desired DNA plasmid, colony PCR was performed before selecting colonies to grow up. Colonies of transformed E.coli on LB agar plates were individually picked using a pipet tip, streaked onto a fresh LB agar plate and then placed into a 0.5 ml tube containing PCR reaction mixtures set up according to the volumes outlined in Table 2.1. Forward and reverse primers used are listed in Table 2.2 , and the cycling conditions of the PCR are listed

in Table 2.3. DNA was visualised on a UV transillumintor on the ChemiDoc Imaging system (Bio-Rad). As the primers were designed to recognise a sequence in the mCherry or BFP tags, only colonies that contained a plasmid expressing tagged constructs produced a PCR product.

Table 2.1 Reagents used for colony PCR

Reagent	Supplier	Volume per reaction (µl)
GoTaq Green Master Mix	Promega	6.25
10 µM Forward Primer	Sigma-Aldrich	1
10 µM Reverse Primer	Sigma-Aldrich	1
ddH$_2$O	-	4.25
DNA	-	1 colony

Table 2.2 Primers used for colony PCR and plasmid sequencing

Primer	Sequence (5' to 3')
eGFP Forward	ACCACTACCAGCAGAACAC
mCherry Forward	AACGGCCACGAGTTCGAGAT
mCherry Reverse	GTGTAGTCCTCGTTGTGGGA
BFP Forward	CACCGTGGACAACCATCACT
BFP Reverse	CCCAGTTTGCTAGGGAGGTC

Table 2.3 Cycling conditions used for colony PCR

Step	Temperature (ºC)	Time	Cycles
Initial Denaturation	94	2 minutes	1
	94	30 seconds	
	60	1 minute	34
	72	1 minute	
Final Extension	72	2 minutes	1
Hold	4	∞	N/A

2.1.9 Plasmid isolation and purification

Transformed *E.coli* C3040H colonies grown on agar plates were picked into 5 ml LB broth containing the appropriate antibiotic (100 µg/ml ampicillin or 50 µg/ml kanamyacin) and incubated shaking overnight at 37 °C. DNA was isolated from the bacteria the following day using the PureLink® HiPure Plasmid Filter Purification Kit for Mini preparation (Invitrogen) according to the manufacturer's instructions. Briefly, bacterial cells were lysed in an alkaline lysis solution then neutralized by the addition of a neutralization buffer containing acid. The lysate was centrifuged for 10 minutes at room temperature. Following centrifugation, the lysate was transferred to a tube containing a DNA binding membrane. Impurities (RNA, proteins, dyes) were removed by a series of wash and centrifugation steps using a wash solution containing ethanol. To elute the plasmid DNA, 100 µl of elution buffer was added to the column membrane and left to stand for 1 minute before final centrifugation.

To prepare higher concentrations of plasmid stocks, plasmid Midipreps were performed. First minipreps were carried out, however 50 µl of the LB culture was retained prior DNA extraction and added to 50 ml LB broth containing the appropriate antibiotic, and incubated shaking overnight at 37 °C. Midipreps were performed using the ZymoPURE™ Plasmid Midiprep kit (Zymo Research) according to the manufacturer's instructions. The process and principle is similar to that of the miniprep process explained above, except larger volumes of the buffers are used to deal with the greater amount of bacteria being lysed. The plasmid DNA was eluted in a volume of 200 µl elution buffer.

2.1.10 Bacterial glycerol stocks

For long-term storage of the plasmids, bacterial stocks were made by diluting 500 µl of the bacterial starter culture grown for a miniprep in 500 µl of 50 % glycerol. Bacterial glycerol stocks were stored at - 80 °C.

2.1.11 DNA UV spectroscopy

The concentration of DNA was determined by loading 1 µl of the DNA sample onto a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) and measuring the absorbance at 260 nm. The spectrophotometer was calibrated prior to measuring the sample DNA concentration with 1 µl of a reference solution that had been used to elute the DNA. The purity of the DNA sample was checked using the 260 nm to 280 nm ratio.

2.1.12 Plasmid sequencing

DNA sequencing was performed by Source Bioscience. Plasmid DNA samples were sent at a concentration of 100 ng/µl and primers were sent at a concentration of 3.2 pmol/ µl. Primers used for plasmid sequencing are listed in Table 2.2.

2.2 Molecular techniques used in Chapter 4

2.2.1 Guide RNA design

Guide RNAs targeting the genomic region downstream of the C9orf72 hexanucleotide repeat expansion were designed using multiple online Guide RNA design tools: MIT CRISPR Design Tool (http://crispr.mit.edu), CRISPOR (http://crispor.tefor.net) and gRNA Designer Tool (Atum; https://www.atum.bio/eCommerce/cas9/input). Standard default settings were used for all online tools used. A sequence downstream of the GGGGCC hexanucleotide repeat expansion, shown in Table 2.4, was inputted into these online algorithms and potential Guide RNA sequences were generated along with various scores of specificity, efficiency and quality. Based on these scores, four guide RNA sequences were chosen for evaluation of targeting efficiency in iPSCs. Two were chosen to target the forward DNA strand and two were chosen to target the reverse DNA strand. Guide RNA sequences are listed in Table 2.5 with their protospacer adjacent motif (PAM) in brackets. Throughout the Guide RNA

optimisation process, experiments were run in parallel with an Alt-R® CRISPR-Cas9 Control Kit, Human (Integrated DNA technologies) containing positive and negative control guide RNAs that target the human hypoxanthine-guanine phosphoribosyltransferase (HPRT) gene.

Table 2.4 Nucleotide sequence inputted into online Guide RNA design algorithms

Nucleotide Sequence Inputted (5' to 3')
TGGCGAGTGGGTGAGTGAGGAGGCGGCATCCTGGCGGGTGGCTGTTTGGGGTTCGG CTGCCGGGAAGAGGCGCGGGTAGAAGCGGGGGGCTCTCCTCAGAGCTCGACGCATTTT TACTTTCCCTCTCATTTCTCTGACCGAAGCTGGGTGTCGGGCTTTCGCCTCTAGCGACT GGTGG

Table 2.5 CRISPR-Cas9 guide RNA sequences

Guide RNA Name	Guide RNA Sequence (5' to 3')	Target Strand
Guide 1	GGGCTTTCGCCTCTAGCGAC (TGG)	Forward
Guide 5	TAAAAATGCGTCGAGCTCTG (AGG)	Reverse
Guide 6	AAGAGGCGCGGGTAGAAGCG (GGG)	Forward
Guide 12	GAAAGCCCGACACCCAGCTT (CGG)	Reverse

2.2.2 PCR to amplify targeted genomic region

PCR using Q5 Hot Start High Fidelity Polymerase (New England Biolabs) was used to amplify specific regions of DNA targeted by Guide RNAs. PCR reaction mixtures were set up according to the volumes outlined in Table 2.6. Forward and reverse primers used are listed in Table 2.7, and the cycling conditions of the PCR are listed in Table 2.8. To confirm the presence and determine the size of a PCR product, a small aliquot of PCR product (2 µl – 10 µl) was separated by agarose gel electrophoresis as outlined in Section 2.1.3.

Table 2.6 Reagents used for PCR to assess Guide RNA targeting

Reagent	Supplier	Volume per reaction (µl)
Q5 Hot Start High Fidelity Master Mix	New England Biolabs	25
10 µM Forward Primer	Sigma-Aldrich	2.5
10 µM Reverse Primer	Sigma-Aldrich	2.5
H_2O	-	Up to 50
100 ng DNA	-	-

Table 2.7 Primers used for PCR to assess Guide RNA targeting

Primer Name	Primer Sequence (5' to 3')	Annealing Temperature (ºC)
C9.lg1_Forward	TGGCTGTTTGGGGTTCGG	67
C9.lg1_Reverse	GTACCCGAGGCTCCCTTTTC	
HPRT Forward	GGGATTGTATTTCCAAGGTTTCTAG	61
HPRT Reverse	TCAAATCCCTGAAGTATTCATTA	

Table 2.8 Cycling conditions used for PCR to assess Guide RNA targeting

Step	Temperature (ºC)	Time	Cycles
Initial Denaturation	98	30 seconds	X 1
Denaturation	98	5 seconds	X 35
Annealing	67 / 61	10 seconds	
Extension	72	20 seconds	
Final Extension	72	2 minutes	X 1
Hold	10	∞	N/A

2.2.3 Heteroduplex mobility assay

PCR products were annealed in a heteroduplex mobility assay. 200 ng of PCR products were mixed with 2 µl 10X NEB Buffer 2 (New England Biolabs) and ddH_2O up to 19 µl. The reaction was incubated according to the annealing conditions outlined in Table 2.9. Following annealing, samples were run by electrophoresis on a pre-cast

6 % Novex™ TBE gel in 1X TBE buffer for 40 minutes at 140 V. Quick-Load Purple Low Molecular Weight DNA Ladder (New England Biolabs) was loaded to allow approximation of product band size. Following electrophoresis, the gel was incubated in SYBR™ Green I Nucleic Acid Gel Stain (Thermo Fisher Scientific) in 1X TBE for 10 minutes at room temperature and visualised using the Gel Doc XR system with Quantity One software (BioRad).

Table 2.9 Cycling conditions used for heteroduplex mobility assay

Step	Temperature (^{o}C)	Ramp Rate	Time
Initial Denaturation	95	-	5 minutes
Annealing	95 – 85	- 2 ^{o}C / sec	-
	85 – 25	- 0.1 ^{o}C / sec	-
Hold	10	-	∞

2.2.4 T7 endonuclease assay

19 µl of annealed PCR product (Section 2.2.3) was incubated with 1 µl T7 Endonuclease (New England Biolabs) at 37 ^{o}C for 30 minutes. 1.5 µl of 0.25 M ethylenediaminetetraacetic acid (EDTA) was added to stop the reaction. Samples were mixed with Novex™ Hi-Density TBE Sample Buffer (5X) and run by electrophoresis on a pre-cast 6 % Novex™ TBE gel (Thermo Fisher Scientific) in 1X TBE buffer for 40 minutes at 140 V. Quick-Load Purple Low Molecular Weight DNA Ladder was loaded to allow approximation of product band size. Following electrophoresis, the gel was incubated in SYBR™ Green I Nucleic Acid Gel Stain (Thermo Fisher Scientific) in 1X TBE for 10 minutes at room temperature and visualised using the Gel Doc XR system with Quantity One software (BioRad).

2.2.5 Sequencing

DNA sequencing was performed by Source Bioscience using dGTP chemistry to aid read-through of GC-rich repeats. PCR products were sent at a concentration of 10

ng/μl and primers sent at a concentration of 3.2 pmol/μl. Primers used for sequencing are listed in Table 2.10.

Table 2.10 Primers used for sequencing to test Guide RNAs

Primer Name	Sequence (5' to 3')
C9.lg_Forward	TGGCTGTTTGGGGTTCGG
C9.lg1_Reverse	GTACCCGAGGCTCCCTTTTC

2.2.6 Single stranded donor DNA design

Three single stranded donor DNA oligonucleotides (ssODN) were designed. Each ssODN was designed to insert monomeric eGFP in frame with either poly-GP, poly-GR or poly-GA only. The ssODN were designed to contain the eGFP sequence flanked by two 100 base pair homology arms targeting the area where *Streptococcus pyogenes* Cas9 (spCas9) and Guide 6 are predicted to introduce a double-strand break site in the genomic DNA. Single base changes were introduced into the 5' homology arms such that stop codons were removed between the hexanucleotide repeat and eGFP in frame with the DPR protein being tagged, and stop codons were introduced between the repeat sequence and eGFP in frame with the two DPR proteins not being tagged. The ATG start codon of eGFP was also mutated to allow for RAN translation of eGFP. In the 3' homology arm, a single base change was introduced to mutate the PAM site. The nucleotide sequences of the single stranded donor DNA oligonucleotides are listed in Table 2.11.

Table 2.11 Nucleotide sequences of ssDNA used as donor DNA for homology directed repair

	Nucleotide Sequence (5' to 3')	Base Pairs
Donor DNA in frame with poly-GP	CGGCGGAGGCGCAGGCGGTGGCGAGTGGGTGAGCGAG GAGGCGGCATCCTGGCGGGTGGCTGTTTGGGGTTCGGC TGCCGGGAAGAGGCGCGGGTAGAAGCGGTGAGCAAGGG CGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGC CCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGC AGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACG ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGG AGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA CCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC GGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC GGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAG GACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAC ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAA CCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCC CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT GACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA CAAGTAAGTGGCTCTCCTCAGAGCTCGACGCATTTTTACT TTCCCTCTCATTTCTCTGACCGAAGCTGGGTGTCGGGCTT TCGCCTCTAGCGACTGGTGGAATTGC	918
Donor DNA in frame with poly-GR	GGCGGAGGCGCAGGCGGTGGCGAGTGGGTGAGTGAGG AGGCGGCATCCTGGCGGGTGGCTGTTTGGGGTTCGGCT GCCGGGAAGAGGCGCGGGCAGAAGCGGGTGAGCAAGG GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTG TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCT GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT GCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGT GCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCA CGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAA GACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACA ACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAA CGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGA GGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGA ACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC AACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGAC CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC GTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT GTACAAGTAAGTGGCTCTCCTCAGAGCTCGACGCATTTTT ACTTTCCCTCTCATTTCTCTGACCGAAGCTGGGTGTCGGG CTTTCGCCTCTAGCGACTGGTGGAATTGCC	919

Donor DNA in frame with poly-GA	GCGGCGGAGGCGCAGGCGGTGGCGAGTGGGCGAGTGA GGAGGCGGCATCCTGGCGGGTGGCTGTTTGGGGTTCGG CTGCCGGGAAGAGGCGCGGGTAGAAGCGTGAGCAAGGG CGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGC CCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGC AGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACG ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGG AGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA CCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC GGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC GGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAG GACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAC ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAA CCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCC CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT GACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA CAAGTAAGTGGCTCTCCTCAGAGCTCGACGCATTTTTACT TTCCCTCTCATTTCTCTGACCGAAGCTGGGTGTCGGGCTT TCGCCTCTAGCGACTGGTGGAATTGCC	919

2.2.7 *In vitro* transcription

Synthetic double stranded DNA oligonucleotides were obtained from Integrated DNA Technologies and resuspended in TE buffer (Integrated DNA Technologies) to a stock concentration of 50 ng/µl. The DNA oligonucleotides were designed to have a T7 promotor sequence followed by the reverse complement sequence of the required donor DNA sequence (Table 2.12). *In vitro* transcription of the DNA oligonucleotides was performed using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (New England Biolabs) according to manufacturer's instructions. FLuc control template DNA (supplied in the RNA synthesis kit) containing the sequence for the firefly luciferase gene under the transcriptional control of the T7 promoter was used as a positive control. RNA synthesis was performed as follows; 150 ng of template DNA was mixed and incubated with 10 µl 2X NTP buffer mix, 5 µl ddH$_2$O and 2 µl T7 RNA polymerase for 2 hours at 37 ^{o}C. To remove the DNA template, the mixture was incubated with 2 µl of DNase (NEB) for a further 15 minutes at 37 ^{o}C. RNA was purified using the MEGAclear™ Transcription Clean-Up Kit (AMBION) and eluted in 45 µl of RNase free ddH$_2$O. 1.5 µl of each RNA product was run on a 1 % agarose gel by electrophoresis, as outlined in Section 2.1.3, to confirm the size of the RNA products. RNA products were stored at – 80 ^{o}C.

Table 2.12 Nucleotide sequences of DNA oligonucleotides used for *in vitro transcription*

	Sequence (5' to 3')
T7_GP_donor	GAAATTAATACGACTCACTATAGGGGGCAATTCCACCAGTCGCTA GAGGCGAAAGCCCGACACCCAGCTTCGGTCAGAGAAATGAGAGG GAAAGTAAAAATGCGTCGAGCTCTGAGGAGAGCCACCTATTTGTA TAGTTCATCCATGCCATGTGTAATCCCAGCAGCTGTTACAAACTCA AGAAGGACCATGTGGTCTCTCTTTTCGTTGGGATCTTTCGAAAGG GCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGG GCCATCGCCAATTGGAGTATTTTGTTGATAATGGTCTGCTAGTTGA ACGCTTCCATCTTCAATGTTGTGTCTAATTTTGAAGTTAACTTTGAT TCCATTCTTTTGTTTGTCTGCCATGATGTATACATTGTGTGAGTTAT AGTTGTATTCCAATTTGTGTCCAAGAATGTTTCCATCTTCTTTAAAA TCAATACCTTTTAACTCGATTCTATTAACAAGGGTATCACCTTCAAA CTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCATCTTTGAAAAAT ATAGTTCTTTCCTGTACATAACCTTCGGGCATGGCACTCTTGAAAA AGTCATGCCGTTTCATATGATCTGGGTATCTTGAAAAGCATTGAAC ACCATAAGAGAAAGTAGTGACAAGTGTTGGCCATGGAACAGGTAG TTTTCCAGTAGTGCAAATAAATTTAAGGGTAAGTTTTCCGTATGTT GCATCACCTTCACCCTCTCCACTGACAGAAAATTTGTGCCCATTAA CATCACCATCTAATTCAACAAGAATTGGGACAACTCCAGTGAAAA GTTCTTCTCCTTTACTCGCTTCTACCCGCGCCTCTTCCCGGCAGC CGAACCCCAAACAGCCACCCGCCAGGATGCCGCCTCCTCGCTCA CCCACTCGCCACCGCCTGCGCCTCCGCCG
T7_GR_donor	GAAATTAATACGACTCACTATAGGGGGCAATTCCACCAGTCGCTA GAGGCGAAAGCCCGACACCCAGCTTCGGTCAGAGAAATGAGAGG GAAAGTAAAAATGCGTCGAGCTCTGAGGAGAGCCACCTATTTGTA TAGTTCATCCATGCCATGTGTAATCCCAGCAGCTGTTACAAACTCA AGAAGGACCATGTGGTCTCTCTTTTCGTTGGGATCTTTCGAAAGG GCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGG GCCATCGCCAATTGGAGTATTTTGTTGATAATGGTCTGCTAGTTGA ACGCTTCCATCTTCAATGTTGTGTCTAATTTTGAAGTTAACTTTGAT TCCATTCTTTTGTTTGTCTGCCATGATGTATACATTGTGTGAGTTAT AGTTGTATTCCAATTTGTGTCCAAGAATGTTTCCATCTTCTTTAAAA TCAATACCTTTTAACTCGATTCTATTAACAAGGGTATCACCTTCAAA CTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCATCTTTGAAAAAT ATAGTTCTTTCCTGTACATAACCTTCGGGCATGGCACTCTTGAAAA AGTCATGCCGTTTCATATGATCTGGGTATCTTGAAAAGCATTGAAC ACCATAAGAGAAAGTAGTGACAAGTGTTGGCCATGGAACAGGTAG TTTTCCAGTAGTGCAAATAAATTTAAGGGTAAGTTTTCCGTATGTT GCATCACCTTCACCCTCTCCACTGACAGAAAATTTGTGCCCATTAA CATCACCATCTAATTCAACAAGAATTGGGACAACTCCAGTGAAAA GTTCTTCTCCTTTACTCCGCTTCTGCCCGCGCCTCTTCCCGGCAG CCGAACCCCAAACAGCCACCCGCCAGGATGCCGCCTCCTCACTC ACCCACTCGCCACCGCCTGCGCCTCCGCC

2.2.8 Reverse transcription

Reverse transcription was performed using SuperScript III Reverse Transcriptase (Invitrogen) with gene-specific primers (Table 2.13) resuspended at 10 pmol/ µl. First-strand cDNA synthesis was performed as follows; 20 µg of RNA was incubated at 65 °C for 5 minutes with 16 µl of gene-specific primer, 40 µl of 10 µM dNTP mix and RNase-free ddH$_2$O up to 520 µl, followed by a 1 minute incubation on ice. The following reagents were then added to the reaction mixture and incubated at 55 °C for 1 hour 30 minutes; 160 µl of 5X First-Strand buffer, 40 µl of 0.1 M dithiothreitol (DTT), 20 µl of RNaseOUT Recombinant RNase Inhibitor (Invitrogen) and 40 µl of SuperScript III reverse transcriptase. The reaction was inactivated by heating to 70 °C for 15 minutes. To remove the RNA template, the mixture was incubated with 20 µl RNase H (NEB) for 20 minutes at 37 °C. cDNA was purified using QIAquick Gel Extraction kit (Qiagen) as per the manufacturer's instructions and eluted in 30 µl nuclease-free ddH$_2$O. The eluted single stranded DNA (ssDNA) was run on a 1 % agarose gel by electrophoresis and visualised using the Gel Doc XR system with Quantity One software (BioRad). The band at approximately 1 kb (the estimated size of ssDNA donor) was excised from the gel. The ssDNA was extracted using the QIAquick Gel Extraction kit (Qiagen) as outlined in Section 2.1.4. ssDNA was eluted in 30 µl ddH$_2$O and stored at – 80 °C.

Table 2.13 Primers used for reverse transcription reactions

Primer Name	Primer Sequence (5' to 3')
RT Primer_GP	CGGCGGAGGCGCAGGCGG
RT Primer_GR	GGCGGAGGCGCAGGCGGT

2.3 Cell culture of SK-N-SH cells

2.3.1 Maintenance of SK-N-SH cells

SK-N-SH human neuroblastoma cells were maintained in Dulbecco's Modified Eagle Medium F-12 (D-MEM/F-12 GlutaMAX™; Thermo Fisher Scientific) supplemented with 10 % fetal calf serum (FCS; Invitrogen), 100 U/ml penicillin and 100 µg/ml streptomycin (Pen/Strep; Invitrogen). Cells were incubated at 5 % CO_2 and 37 ºC in T-75 flasks (VWR) and passaged every 2 - 3 days (when the cells were 80 – 90 % confluent) up to passage number 20. All experiments performed with SK-N-SH cells were performed on cells under passage number 20 and were repeated in triplicate.

2.3.2 Passage of SK-N-SH cells

To passage the SK-N-SH cells, media was removed and the cells were washed with phosphate buffered saline (PBS) then incubated with 1 ml trypsin with EDTA (Invitrogen) for 2 minutes at 37 ºC. Cells were resuspended in 9 ml DMEM F-12 and 1 ml was transferred into a new T-75 flask containing 10 ml DMEM F-12 into order to split the cells at a ratio of 1:10.

2.3.3 Cryopreservation of SK-N-SH cells

Near confluent cells with a low passage number were trypsinised and resuspended in D-MEM F-12 (containing 10 % FCS and Pen/Strep). Cells density was counted using a hemocytometer slide and the cells were then centrifuged at 1000 rpm for 5 minutes. Following removal of the supernatant, the cell pellet was resuspended in the appropriate volume of freezing media (90 % FCS + 10 % dimethyl sulphoxide (DMSO)) to create a density of 1 x 10^6 cells/ml. The resuspended cells were then aliquoted in 1 ml volumes into cryovials. Cryovials were placed into a Mr. Frosty™ Freezing container (Thermo Fisher Scientific) filled with isopropanol for gradual

cooling and were placed at -80 ºC overnight. Once frozen, the cryovials were moved into a liquid nitrogen tank for long-term storage.

2.3.4 Revival of cryopreserved SK-N-SH cells

To defrost cells from cryovials preserved in liquid nitrogen, individual cryovials were rapidly thawed and resuspended in 1 ml DMEM F-12 before being transferred to a T-75 flask containing 10 ml DMEM F-12 (containing 10 % FCS and Pen/Strep) and incubated at 37 ºC and 5 % CO_2. Defrosted cells were grown and passaged as described in Section 2.3.2 for approximately 1 week before being used in experiments.

2.3.5 Transient transfections

SK-N-SH cells were seeded into either 8-well chamberslides (VWR) for immunocytochemistry experiments or 6-well dishes (Starlab) for biochemical experiments at densities of either 3×10^4 cells/well or 50×10^4 cells/well, respectively 24 hours prior to transfection. Cells were transfected using Lipofectamine™ and Plus™ reagents (Invitrogen) according to the manufacturer's instruction. The volumes of reagents used in the transfections are listed in Table 2.14. Briefly, total DNA was mixed with appropriate volumes of DMEM F-12 serum free media (SFM) and Plus reagent and incubated for 15 minutes at room temperature. Following incubation, the appropriate volume of SFM mixed with Lipofectamine reagent was added to the DNA/Plus mixture and incubated for a further 15 minutes at room temperature. During this incubation, the cells were washed twice with PBS. The appropriate volume of SFM was added to the DNA/Plus/Lipofectamine mix following incubation, and then the appropriate volume of the mixture was added to the cells before incubation at 37 ºC and 5 % CO_2. Following a 3 hour incubation, the media was removed from the cells and replaced with DMEM F-12 containing 10 % FCS and Pen/Strep. Cells were

grown for a further 24 hours prior to fixation for immunocytochemical experiments, or lysis for biochemical experiments.

Table 2.14 Volumes of reagents used for transient transfection of SK-N-SH cells

	6-Well Plate	**8-Well Chamberslide**
Total DNA	2000 ng	200 ng
Plus Reagent	8 µl	50 µl
SFM	100 µl	1 µl
Lipofectamine Reagent	50 µl	0.5 µl
SFM	4 µl	12.5 µl
Final Volume	750 µl	150 µl

2.3.6 Plasmids

All plasmid constructs used in this thesis are listed in Table 2.15.

2.3.7 Immunocytochemistry

SK-N-SH cells were fixed (4 % paraformaldehyde; 10 minutes), permeabilised (0.1 % Triton X-100; 10 minutes) and non-specific binding was blocked (10 % donkey serum and 3 % bovine serum albumin; 1 hour) prior to incubation with primary antibodies for 1 hour. Following PBS washes, cells were incubated with secondary antibodies. In most instances, cell nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich), however if the cells contained a BFP tagged construct, nuclear staining was performed using propidium iodide (PI) (Invitrogen). Slides were mounted on glass coverslips with Fluorescent Mounting Medium (DAKO) and stored at 4 °C. Antibodies and dilutions are listed in Table 2.16 and Table 2.17.

.

2.3.8 Immunoblotting

SK-N-SH cells were lysed in radioimmunoprecipitation (RIPA) buffer (150 mM NaCl, 50 mM Tris.HCl (pH 7.6), 1 % NaDOC, 0.1 % sodium dodecyl sulfate (SDS) and 1 % IgePal) containing 2 % protease inhibitor cocktail on ice for 10 minutes. Lysates were collected and sonicated for 10 seconds. 2x SDS loading buffer (0.5M Tris (pH 6.8), 20 % glycerol, 4 % SDS, 2.5 % β-mercaptoethanol, 0.004 % bromophenol blue) was added to the lysates and samples were heated at 98 ºC for 5 minutes. Proteins in the lysates were separated via electrophoresis on 10 % SDS-PAGE gels and electrophoretically transferred onto nitrocellulose membranes. Membranes were blocked in 5 % milk in phosphate buffered saline with 0.05 % Tween20 (PBS-T) for 1 hour. Membranes were incubated overnight at 4 ºC with primary antibody diluted in 5 % milk in PBS-T, then washed 3-times in PBS-T prior to incubation for 1 hour with secondary antibodies diluted in 5 % milk in PBS-T. Protein bands were detected using the ChemiDoc imaging system (BioRad) following treatment with ECL Prime Western blotting Detection Reagent (GE). Antibodies and dilutions are listed in Table 2.16 and Table 2.17.

Table 2.15 Plasmids used in cloning and transient transfections

Vector	Insert	Tag	Resistance	Source
pcDNA3.1(+)	GA36	Untagged	Ampicillin	Prof. A. Isaacs (UCL)
	GA100			
	PR36			
	PR100			
	GR36			
	GR100			
	GA36	N-Terminal GFP		L. Gittings (This Study)
	GA100			
	PR36			
	PR100			
	GR36			
	GR100			
	GA36	N-Terminal mCherry		
	GA100			
	PR36			
	PR100			
	GA36	N-Terminal BFP		
	GA100			
	PR36			
	PR100			
pCMVtag3a	HSJ1a	N-Terminal myc	Kanamycin	Prof. M. Cheetham (UCL)
pEGFP-N1	HSJ1a	GFP	Kanamycin	Prof. M. Cheetham (UCL)
pCMVtag3b	Empty	Untagged	Kanamycin	Sigma-Adlrich
pCI-neo	TDP-43	N-Terminal myc	Ampicillin	Dr. C. Vance (KCL)
pEGFP-N1	Empty	N-Terminal GFP	Kanamycin	Prof. M. Cheetham (UCL)
pEGFP-C3	LC3	N-Terminal GFP	Kanamycin	Addgene
pmCherry-c3	SOD1	Untagged	Kanamycin	Prof. M. Cheetham (UCL)
pBAD-mtagBFP2	Empty	BFP	Ampicillin	Addgene

Table 2.16 Primary antibodies used for immunocytochemistry and immunoblotting in Chapter 3

Target	Host	Source	Product Code	Dilution	Application
Myc	Mouse	Sigma-Aldrich	M4439	1:1000	ICC
GA	Rabbit	Proteintech	24492-1-AP	1:400	ICC
PR	Rabbit	Adrian Isaacs	PR30	1:500	ICC
GR	Rat	Kind gift from Friedrich Grässer	GR-5H9	1:50	ICC
P62	Rabbit	Abcam	ab56416	1:100	ICC
GFP	Mouse	Sigma-Aldrich	11814460001	1:1000	WB
HSJ1	Rabbit	Proteintech	10838-1-AP	1:500	WB
β-tubulin	Mouse	Sigma	T4026	1:3000	WB

Table 2.17 Secondary antibodies used for immunocytochemistry and immunoblotting in Chapter 3

Target	Host	Conjugate	Source	Product Code	Dilution	Application
Mouse	Donkey	Alexa Fluor 488	Invitrogen	A-21202	1:1000	ICC
Rabbit		Alexa Fluor 594		A-21207		ICC
Mouse		Alexa Fluor 594		A-21203		ICC
Rabbit		Alexa Fluor 488		A-21206		ICC
Mouse	Goat	Horseradish Peroxidase	Pierce	31430	1:30,000	WB
Rabbit				31460		WB

2.4 Cell culture of iPSC

2.4.1 Generation and characteristics of iPSC lines

Human iPSCs were obtained from Dr Selina Wray who derived the cell lines from dermal fibroblasts grown from patient skin biopsies. The characteristics of the iPSC line used in this study is outlined in Table 2.18 and has been previously described by Simone et al., 2018.

Table 2.18 Characteristics of iPSC line used in this study

ID	Gender	Clinical Diagnosis	Age at disease onset (years)	Age at biopsy (years)	*C9orf72* repeat expansion size
DN19	Male	ALS	52	58	~ 638

2.4.2 Maintenance of iPSC

Human iPSCs were maintained on 6-well plates pre-coated with a layer of Geltrex (150 µg/ml; Thermo Fisher Scientific) basement membrane matrix in Essential 8 ™ medium (containing DMEM/F-12, L-ascorbic acid, selenium, transferrin, $NaHCO_3$, insulin, fibroblast growth factor-2 and transforming growth factor β1; Thermo Fisher Scientific). Cells were incubated at 37 °C and 5 % CO_2. Essential 8 ™ medium was replaced daily.

2.4.3 Passage of iPSC

iPSCs were passaged when cells reached approximately 70 % confluency (approximately every 2 – 3 days) by incubating the cells at 37 °C for 5 minutes with 0.5 mM EDTA in PBS. Following removal of EDTA, cells were resuspended in 1 ml Essential 8 ™ medium and split at the desired ratio (typically 1:6) into new Geltrex coated plates containing Essential 8™ medium.

2.4.4 Cryopreservation of iPSC

iPSCs of approximately 70 % confluency were dissociated by incubating with 0.5 mM EDTA at 37 $^{\circ}$C for 5 minutes. Following removal of EDTA, cells were resuspended in 1 ml of freezing medium containing 90 % Essential 8^{TM} Medium and 10 % DMSO and transferred into a cryovial. Cryovials were placed into a cryopreserving Mr. Frosty™ Freezing Container, filled with isopropanol for gradual cooling, and were transferred into a - 80 $^{\circ}$C freezer overnight. Once thoroughly frozen, the cryovials were transferred into a liquid nitrogen tank for long term storage.

2.4.5 Revival of cryopreserved iPSC

To thaw cryopreserved iPSCs, cryovials were rapidly thawed and resuspended in 1 ml of Essential 8^{TM} medium. Cells were transferred to a 15 ml Falcon tube containing 5 ml Essential 8^{TM} medium and centrifuged at 300 g for 5 minutes to pellet the cells. The supernatant was removed, and the cell pellet was resuspended in 1 ml of Essential 8^{TM} medium before being transferred into one well of a new Geltrex coated 6-well plate. Cells were incubated at 37 $^{\circ}$C and 5 % CO_2 and were maintained and passaged for at least one week before use in any experiments.

2.4.6 Nucleofection of iPSC

Nucleofection of CRISPR-Cas9 system components into iPSCs was performed using the Amaxa 4D-Nucleofector^{TM} System (Lonza) in conjunction with the P3 Primary Cell 4D-Nucleofector® X Kit (Lonza). Alt-R *Streptococcus pyogenes* Cas9 protein (SpCas9), crRNA, tracrRNA and Electroporation Enhancer were obtained from Integrated DNA Technologies. Approximately 6 hours prior to nucleofection, the iPSC media was changed to Essential 8^{TM} medium containing 10 µM ROCK inhibitor (Sigma-Aldrich). Guide RNA and SpCas9 protein were assembled into ribonucleoprotein complexes prior to nucleofection. Guide RNA was formed by mixing 200 µM tracrRNA with 200 µM crRNA in a total volume of 5 µl IDT duplex buffer

(Integrated DNA Technologies), heating to 95 °C for 5 minutes and allowed to cool at room temperature for at least 45 minutes. Once cooled, ribonucleoprotein complexes were formed by adding Alt-R SpCas9 nuclease to the tubes containing the guide RNA complexes and incubated at room temperature for 20 minutes. In nucleofections where single stranded donor DNA (ssDNA) was provided for homology directed repair, 1 µg of donor ssDNA and 4 µM Electroporation Enhancer was added to the ribonucleoprotein complex mixture. During Guide RNA optimisation experiments, no ssDNA was provided. iPSCs were prepared for nucleofection by a 5 minute treatment with 0.5 µM EDTA, resuspended in Essential 8™ media with 10 µM rho-associated protein kinase (ROCK) inhibitor and counted using a haemocytometer. 8×10^5 cells per nucleofection were centrifuged at 1300 rpm for 3 minutes and then resuspended in 100 µl of Amaxa P3 Primary Cell 4D-Nucleofector solution per nucleofection. 100 µl of resuspended iPSC was transferred to the tube containing the preassembled SpCas9 ribonucleoprotein complex, mixed and then transferred to an Amaxa 4D Nucleofector cuvette. Cuvettes were inserted into the Amaxa 4D Nucleofector device and nucleofected in solution Primary Cell P3 using the custom program pulse CA137. Following nucleofection, cells were resuspended in 4 ml Essential 8™ media containing 10 µM ROCK inhibitor and plated into one well of a Geltrex coated 6-well plate then incubated at 37 °C, 5 % CO_2. 24 hours post nucleofection media was changed to Essential 8™ media without ROCK inhibitor.

2.4.7 DNA extraction from iPSC

DNA was extracted from iPSCs using the DNeasy Blood & Tissue Kit (Qiagen), as per the manufacturer's instructions. DNA was eluted in 200 µl ddH_2O. The concentration of DNA samples was determined using the NanoDrop ND-1000 spectrophotometer as outlined in Section 2.1.11.

2.4.8 Single cell colony isolation

When iPSCs reached 70 – 80 % confluency following nucleofection, the cells were re-plated into one well of a Geltrex coated 6-well plate at a density of 500 cells/well in Essential 8™ media supplemented with 1X RevitaCell™ (Thermo Fisher Scientific). Essential 8™ media was replaced daily for approximately 6 – 7 days until distinct iPSC colonies were visible under the microscope. A fine liner marker pen was used to mark the position of individual colonies on the underneath of the plate. iPSC colonies that had contacted another iPSC colony were not marked. The plate was then returned to the laminar flow hood and each marked colony was individually picked from the plate using a P1000 pipette. The colony was placed into one well of a Geltrex coated 24-well plate well in Essential 8™ media supplemented with 1X RevitaCell™. A 50 % media change (Essential 8™) was performed 2 days after colony plating and daily Essential 8™ media changes were performed from day 3 onwards. iPSC colonies were passaged and maintained in 24- and 96-well Geltrex coated plates when the cells reached 70 – 80 % confluency.

2.4.9 DNA extraction from iPSC colonies and colony PCR

Confluent iPSC colonies in 96-well plates were lysed for DNA to be used in a colony PCR screen. 40 µl QuickExtract™ DNA Extraction Solution (Cambio) was added to each well and the plate was incubated at 65 °C for 15 minutes. The plate was then agitated for 15 seconds and incubated for a further 15 minutes at 98 °C. 1 µl of lysed product per well was used for the colony PCR reaction. PCR reactions were set up using the reagents and volumes outlined in Table 2.6, with the exception of 1 µl of lysed product rather than 100 ng DNA. Forward and reverse primers used for the colony PCR are listed in Table 2.19, and the cycling conditions of the PCR are listed in Table 2.20. Primers 1 and 2 were used to confirm the presence of DNA. Primers 3 and 4 were used to confirm the presence of eGFP. To confirm the presence and

determine the size of the PCR product, the PCR product was separated by agarose gel electrophoresis as outlined in Section 2.1.3.

Table 2.19 Primers used for iPSC colony PCR

	Primer Name	Primer Sequence (5' to 3')	Annealing Temperature (ºC)
1	OCT4_Forward	TTCTGGCGCCGGTTACAGAACCA	60
2	OCT4_Reverse	GACAACAATGAAAATCTTCAGGAGA	
3	mEGFP3_Forward	TATATCATGGCCGACAAGCA	67
4	C9.lg1_Reverse	GTACCCGAGGCTCCCTTTTC	

Table 2.20 Cycling conditions used for colony iPSC PCR reactions

Step	Temperature (ºC)	Time	Cycles
Initial Denaturation	98	30 seconds	X 1
Denaturation	98	5 seconds	X 35
Annealing	67 / 60	20 seconds	
Extension	72	30 seconds	
Final Extension	72	2 minutes	X 1
Hold	10	∞	N/A

2.4.10 PCR to validate eGFP insertion in iPSC CRISPR clones

DNA was extracted from iPSCs as described in Section 2.4.7. Two PCRs using Q5 Hot Start High Fidelity Polymerase (New England Biolabs) was used to confirm the presence of eGFP in the CRIPSR clones. PCR reaction mixtures were set up according to the volumes outlined in Table 2.21. Forward and reverse primers used are listed in Table 2.22, and the cycling conditions of the PCR are listed in Table 2.23. To confirm the presence and determine the size of a PCR product, 10 µl was separated by agarose gel electrophoresis as outlined in Section 2.1.3.

Table 2.21 Reagents used for PCR to screen for eGFP

Reagent	Supplier	Volume per reaction (µl)
Q5 Hot Start High Fidelity Master Mix	New England Biolabs	6
10 µM Forward Primer	Sigma-Aldrich	1
10 µM Reverse Primer	Sigma-Aldrich	1
5 % DMSO	Sigma-Aldrich	0.6
H_2O	-	Up to 12
100 ng DNA	-	-

Table 2.22 Primers used for PCR to screen for eGFP

	Primer Name	Primer Sequence (5' to 3')	Annealing Temperature (ºC)
1	mEGFP4_Forward	GTCCAAGCTGAGCAAAGACC	
2	mEGFP5_Forward	ACATGGTCCTGCTGGAGTTC	63
3	C9.lg2_Reverse	ATCTCATCCCGCATGATCTC	

Table 2.23 Cycling conditions used for PCR to screen for eGFP

Step	Temperature (ºC)	Time	Cycles
Initial Denaturation	98	30 seconds	1
Denaturation	98	5 seconds	
Annealing	63	30 seconds	30
Extension	72	40 seconds	
Final Extension	72	2 minutes	1
Hold	10	∞	N/A

2.5 Human Pathology

2.5.1 Cases

The brains used in this study were donated to the Queen Square Brain Bank for Neurological Disorders (QSBB; Institute of Neurology, University College London), the Manchester Brain Bank (University of Manchester) and the MRC London Neurodegenerative Diseases Brain Bank (Institute of Psychiatry, Psychology & Neuroscience (IOP), King's College London). All cases used were grouped according to the pathological diagnosis. Ethical approval for the study was obtained from the Local Research Ethics Committee of the National Hospital for Neurology and Neurosurgery.

In Chapter 5, 38 genetically and pathologically confirmed *C9orf72* cases were analysed (FTLD-TDP A = 18, FTLD-TDP-B = 5, MND = 15) from the three different brain banks (QSBB = 18, IOP = 8, Man = 12). Where applicable, data was compared to 5 neurologically normal controls and 5 non-*C9orf72* FTLD-TDP A disease controls.

In Chapter 6, 12 pathologically confirmed FTLD-FUS cases were analysed (NIFID = 5, aFTLD-U = 7). These were compared with 5 neurologically normal controls. All cases in Chapter 6 were obtained from QSBB.

The pathological and genetic details of the FTLD, MND and control cases used in this study are summarised in Table 2.25. Full case details are described in the appendices.

2.5.2 Tissue processing

For cases processed at QSBB, formalin fixed brain tissue was embedded in paraffin wax by dehydration in a series of ethanol baths, prior to clearing with chloroform, and infiltration with paraffin wax as detailed in Table 2.24. Embedded tissue was stored

until required. 8 μm sections were cut using a Leica microtome. Sections were then transferred into 30 % ethanol, floated onto warm water and mounted onto glass microscope slides (Solmedia). The slides containing the sections were dried at 37 °C for several hours and stored until required. For cases obtained from the Manchester Brain Bank and the MRC London Neurodegenerative Diseases Brain Bank, these were processed according to the standard protocol used at these brain banks in accordance to guidelines set out by the UK Brain Banks Network and in a similar manner to the protocol used at QSBB, described in Table 2.24.

Table 2.24 Paraffin embedding process of post mortem brains at QSBB

Step	Number of Washes	Reagent	Duration (hours)
1	1	70 % Ethanol	6
2	2	90 % Ethanol	6
3	4	100 % Ethanol	6
4	2	Chloroform	6
5	3	Paraffin Wax	6

2.5.3 Immunohistochemistry

8 μm thick paraffin embedded frontal cortex and hippocampal sections to be used for immunohistochemistry were incubated at 60 °C overnight prior to starting the immunohistochemistry protocol. Positive and negative controls samples were included in every immunochemical experiment to confirm the experimental procedure had worked correctly. Positive controls were samples known to previously confirmed to contain the protein of interest by immunohistochemical methods. Negative controls were samples previously confirmed to not contain the protein of interest by immunohistochemical methods. Slides containing tissue sections were dewaxed in a series of three 5 minute xylene washes, followed by rehydration in 100 %, 95 % and 70 % ethanol. Slides were incubated in 0.3 % H_2O_2 in methanol for 10 minutes to block endogenous peroxidase activity followed by a 5 minute wash in ddH_2O. The slides were then boiled in 0.1 M citrate buffer (pH 6.0) in a pressure cooker for 10

minutes at maximum pressure for antigen retrieval. Non-specific binding was blocked by incubating the slides in 10 % non-fat milk (Marvel) in PBS for 30 minutes at room temperature. Slides were then incubated in the relevant primary antibody (Table 2.26) for 1 hour at room temperature or 4 °C overnight and then washed twice in PBS. Slides were incubated in the relevant secondary antibody (Table 2.27) for 30 minutes at room temperature, followed by two more PBS washes. Slides were then incubated with avidin-biotin complex (ABC; DAKO) for 30 minutes at room temperature and washed twice with PBS. Antibody binding was visualised by developing the slides in diaminobenzidine (DAB; Sigma-Aldrich) activated by H_2O_2. Slides were incubated in a solution of 500 µg DAB in 100 ml PBS, activated with 32 µl 30 % H_2O_2 for 3 minutes, after which the colour intensity was checked using a light microscope. Slides were counterstained in Mayers haemotoxylin (BDH) for 10 seconds, before washing with ddH$_2$O. Sections were dehydrated by washing in 70 %, 95 % and 100 % ethanol, cleared in two washes of xylene and then mounted with Depex mounting media (BDH).

2.5.4 Immunofluorescence

Double immunohistochemistry was performed on 8 µm thick paraffin embedded frontal cortex and hippocampal sections. Dewaxing, pre-treatment and blocking of sections were performed as described in Section 2.5.3. Antibodies used for double immunohistochemistry staining are listed in Table 2.26 and Table 2.27.

For double immunofluorescence labelling of poly-GR inclusions with ADMA-GR or SDMA-GR in Chapter 5, tissue sections were incubated for 1 hour at room temperature with both primary antibodies. Sections were then washed with PBS-T and incubated with species-appropriate Alexa Fluor 488 and 568 secondary antibodies (Invitrogen, 1:1000) for 1 hour at room temperature. Tissue sections were

washed with PBS-T and mounted using Vectashield anti-fade mounting medium containing DAPI (Vector Laboratories) for nuclear counterstaining.

For double immunofluorescence labelling of FUS and hnRNP R in Chapter 6, tissue sections were incubated with the first primary antibody overnight at 4 °C. Sections were then washed with tris-buffered saline with tween (TBS-T) and incubated with the species-appropriate secondary antibody (DAKO) for 30 minutes at room temperature, prior to incubation with ABC for 30 minutes at room temperature. Binding of the first antibody was visualized using TSA fluorescein amplification kit (Perkin-Elmer) which was applied to the section for 15 minutes at room temperature. Tissue sections were then incubated with the second primary antibody overnight at 4 °C. Sections were washed with TBS-T and incubated with the species-appropriate secondary antibody (DAKO) for 30 minutes at room temperature, prior to incubation with ABC for 30 minutes at room temperature. Binding of the second antibody was visualized using TSA Cyanine 3 amplification kit (Perkin-Elmer). Cross reactivity of antibodies was controlled for by omitting the primary antibodies from sections that were subsequently incubated with secondary antibodies and TSA. Tissue sections were washed and mounted using Vectashield anti-fade mounting medium containing DAPI (Vector Laboratories) for nuclear counterstaining.

Fluorescent images were captured at 40x or 63x magnification using a Leica DM5500B fluorescence microscope. Antibody staining was identified and imaged using the appropriate fluorescent channels, and co-localisation was confirmed on the combined images. All fluorescent images of human brain sections presented in this thesis are maximum projection of z-stack images following blind 3D deconvolution.

2.5.5 Nanostring mRNA expression analysis

RNA extractions prior to Nanostring analysis was performed from frozen brain tissue by Ms Priya Gami-Patel (Research Technician at QSBB). Total RNA was extracted from grey matter within the frontal and temporal cortices of FTLD-FUS (n = 5), FTLD-TDP A (n= 19), FTLD-TDP B (n = 3), FTLD-TDP C (n = 7) and normal controls (n = 6) using the Qiagen Rneasy kit, as per the manufacturer's instructions. 100 ng of total RNA from each sample was analysed using the NanoString nCounter analysis system (Nanostring Technologies, Seattle, WA) using a pre-designed codeset, which has been previously reported (Gami-Patel et al., 2016). The codeset contained probes for detection of the gene of interest; hnRNP R. Probes were designed according to the manufacturer's design principles (Geiss et al., 2008), including screening for inter-reporter, intra-reporter and capture probe interactions, and selection for probes with optimal melting temperatures. The laboratory running the assay was blinded to case diagnoses, and samples of cases or controls were randomly assigned to plates to avoid run-order bias. Raw counts were subjected to a technical normalisation and normalised to the geometric mean using nSolver Analysis Software v2.0 (NanoString). Biological normalisation was performed using reference genes (*CLTC, GAPDH, GUSB, HPRT1, PGK1, and TUBB*) included in the codeset.

Table 2.25 *Details of post-mortem cases used for pathology analysis*

Pathological Diagnosis	Gender	AAO (years ± SD)	AAD (years ± SD)	Disease Duration (years ± SD)	Mutation	Source Brain Bank
FTLD-TDP A	18 M: 12 F	59 ± 8	67 ± 9	7 ± 3	*C9orf72*: 18 *GRN*: 1 Unknown: 11	QSBB: 23 IOP: 0 Man: 7
FTLD-TDP B	2 M: 6 F	63 ± 6	69 ± 6	7 ± 8	*C9orf72*: 5 Unknown: 3	QSBB: 4 IOP: 4 Man: 0
FTLD-TDP C	5 M: 2 F	57 ± 7	71 ± 5	14 ± 5	None known	QSBB: 7 IOP: 0 Man: 0
MND	13 M: 8 F	60 ± 10	62 ± 10	3 ± 2	*C9orf72*: 15 Unknown: 6	QSBB: 8 IOP: 8 Man: 5
NIFID	1 M: 4 F	52 ± 13	53 ± 2	3 ± 2	None known	QSBB: 5 IOP: 0 Man: 0
aFTLD-U	4 M: 3 F	47 ± 5	54 ± 4	7 ± 3	None known	QSBB: 7 IOP: 0 Man: 0
Control	4 M: 6 F	N/A	79 ± 6	N/A	None known	QSBB: 10 IOP: 0 Man: 0

M = male, F = female, AAO = age at disease onset, AAD = age at death, QSBB = Queen Square Brain Bank, IOP = King's College Institute of Psychiatry Brain Bank, Man = Manchester Brain Bank, FTLD = frontotemporal lobar degeneration, MND = motor neuron disease, NIFID = neuronal intermediate   filament inclusion disease, aFTLD-U = atypical frontotemporal lobar degeneration.

Table 2.26 Primary antibodies used for immunohistochemistry and immunofluorescence

Target	Host	Manufacturer	Product code	Dilution	Application
Unmethylated-GR	Rabbit	UCB	4021	1:50	IHC
Asymmetric-Dimethylarginine (ADMA)-GR	Rabbit		4022	1:2000	IHC, IF
Symmetric-Dimethylarginine (SDMA)-GR	Rabbit		4023	1:50	IHC, IF
Pan-GR	Rat	Kind gift from Friedrich Grässer	GR5-H9	1:25	IHC, IF
p62	Mouse	BD Biosciences	610833	1:200	IHC
hnRNP R	Rabbit	Abcam	ab30930	1:200	IHC
hnRNP R	Goat	Santa Cruz Biotechnology	sc-16541	1:200	IF
hnRNP Q	Rabbit	Thermo Fisher Scientific	PA5-15009	1:200	IHC
FUS	Rabbit	Novus Biologicals	NB100-565	1:200	IHC, IF
TRN1	Mouse	Abcam	ab10303	1:500	IHC

IHC = Immunohistochemistry, IF = Immunofluorescence

Table 2.27 Secondary antibodies used for immunohistochemistry and immunofluorescence

Target	Host	Conjugate	Manufacturer	Product Code	Dilution
Rabbit	Swine	Biotin	DAKO	E0353	1:200
Goat	Rabbit		DAKO	E0466	
Mouse	Goat		Vector Laboratories	BA-2020	
Rabbit	Donkey	Alexa Fluor 568	Invitrogen	A10042	1:1000
Goat	Donkey			A11057	
Mouse	Goat			A11031	
Rabbit	Goat	Alexa Fluor 488		A11008	
Goat	Chicken			A21467	
Mouse	Chicken			A21200	
Rat	Chicken			A-21470	

2.6 Microscopy and Image Analysis

2.6.1 Bright field microscopy

Bright field images of antibody staining in human post-mortem tissue in Chapters 5 and 6 were captured using a Nikon Digital Sight DS-Fi2 camera on a Nikon Eclipse Ni-U microscope. Images were taken at 200x, 400x or 600x magnification.

2.6.2 Fluorescence microscopy

Fluorescent images in Chapter 3 were obtained using the Carl Zeiss LSM700 laser-scanning confocal microscope using the 40x oil immersion objective and Zeiss Zen 2009 software. The excitation/emission spectrum of the channels used was as follows; DAPI: 364/475-525nm, FITC: 488/505-530nm, Cy3: 543/560nm. Fluorescent images in Chapters 5 and 6 were obtained using Leica DM5500B fluorescence microscope using the 40x or 63x objectives and Leica Application Suite X software.

2.6.3 Manual counting of inclusions using epifluorescence microscopy

For the manual scoring of inclusion incidence in transfected cells in Chapter 3, slides were visualised on a Nikon 80i fluorescent microscope using the 20x objective. DAPI staining was used to manually count total cell number and inclusions were counted manually. Transfection efficiency was calculated as the percentage of GFP positive cells out of the total DAPI positive cells. Inclusion incidence was expressed as either the percentage of transfected cells with inclusions, or the percentage of total cells with inclusions.

2.6.4 Axioscan microscopy

Regions of interest corresponding to the wells of the chamberslides were selected and imaged at 200x on the AxioScan Z1 (Zeiss) slide scanner in three fluorescent channels. The excitation/emission spectrum of the channels used was as follows;

DAPI: 365/445-450nm, FITC: 470/525-550nm, Texas Red: 550/605-670nm. The images were tiled together by the AxioScan and regions of interest were then manually selected and exported using the Zen Lite software (Zeiss).

2.6.5 Volocity analysis of GFP-GA100 inclusions

Analysis of confocal and AxioScan images was performed using Volocity image analysis software (Perkin Elmer). The analysis protocol designed to quantify the number of GFP-GA100 inclusions is outlined in Table 2.28.

Table 2.28Volocity protocol for the quantification of GFP-GA100 inclusions.

Task	Steps in Volocity Protocol
Identify DAPI stained nuclei to give population 'DAPI'	Find objects using automatic thresholds in channel 405
	Exclude objects touching edge of image
	Separate touching objects: size guide 50 μM^2
	Exclude objects greater than 400 μM^2
Identify GFP-GA100 inclusions to give population 'GFP-GA100'	Find objects using automatic thresholds channel 488
	Fill in holes in objects
	Exclude objects touching edge of image
	Separate touching objects: size guide 10 μM^2
	Exclude objects smaller than 5 μM^2

2.6.6 Quantification of unmethylated-GR 'speckled-nuclei'

Ten 20x magnification bright field images of frontal cortex sections per case were taken of each slide stained with unmethylated-GR using a Nikon Digital Sight DS-Fi2 camera on a Nikon Eclipse Ni-U microscope. Using these images, the total number of neurons and the number of neurons containing nuclear speckles were manually counted by eye and these values were used to calculate the percentage of neurons containing nuclear speckles.

2.6.7 Quantification of methyl-GR inclusions

Slides stained with ADMA-GR and SDMA-GR antibodies were scanned at 40x magnification using a Leica Slide Scanner SCN400 and viewed on Digital Image Hub (Leica Biosystems). Regions of interest were manually selected and PicPick (NGWIN) software was used to extract 10 images per region per case at 40x magnification. Extracted images had a size of 1280 x 960 pixels. Extracted images were loaded into ImageJ (National institute of health) and the Cell Counter tool was used to manually quantify the number of ADMA-GR and SDMA-GR inclusions.

2.6.8 Quantification of FUS, hnRNP R, hnRNP Q and TRN1 inclusions

FUS, TRN1, hnRNP R and hnRNP Q stained slides were scanned at 400x magnification using a Leica Slide Scanner SCN400. Aperio Imagescope was used to digitally marked and extract regions of interest from the image to decrease file size. Extracted images were loaded into ImageJ and a macro, developed by a collaborator (Yau Lim, Kings College London), was used to select a region of interest on the image. From within this region 10 random squares per image were generated based on coordinates of random squares generated by a Python script. Random squares were selected at 1000 x 1000 pixels square, which represented 500 μm^2, for images of the frontal cortex, and 500 x 500 pixels square, which represented 250 μm^2, for images of the granule cell layer of the dentate gyrus. The number of FUS, TRN1,

hnRNP R and hnRNP Q positive inclusions in each image was manually quantified using the Cell Counter tool in ImageJ.

2.7 Statistical analysis

Statistical analyses were performed using Microsoft Excel or GraphPad Prism. A significance level of $p < 0.05$ was used for all statistical tests. The level of significance is demonstrated in figures as * for $p < 0.05$, ** for $p < 0.005$, *** for $p < 0.0005$. Where statistical tests have been used in this thesis, the type of statistical test, n number, p value and r value (where appropriate) have been described in the corresponding figure legend.

2.8 Software list

Adobe Illustrator CC 2015.3 (Adobe)

Adobe Photoshop elements 6 (Adobe)

Aperio Imagescope (Leica Biosystems)

Digital Image Hub (Leica Biosystems)

Geneious (Biomatters)

Graphpad Prism v7 (Graphpad Software Inc.)

Illustrator for Biological Sequences (GPS)

Image Lab (Bio-Rad)

ImageJ v1.50i (National institute of health)

Leica Application Suite X (Leica Biosystems)

Microsoft Excel (Microsoft)

ND-1000 Nanodrop software

PicPick (NGWIN)

Quantity One (BioRad)

SnapGene Viewer Software (GSL Biotech)

Volocity Image analysis software (Perkin Elmer)

Zen Black 2011 (Zeiss)

Zen Blue 2011 (Zeiss)

Chapter 3. Characterisation and manipulation of *C9orf72* dipeptide repeat protein aggregation using the molecular chaperone, HSJ1

3.1 Introduction

3.1.1 Protein aggregation and molecular chaperones in neurodegeneration

Intracellular or extracellular proteinaceous inclusions in specific brain regions are a pathological hallmark of many neurodegenerative diseases (Ross and Poirier, 2005; Soto, 2003; Yerbury et al., 2016). These inclusions are primarily composed of misfolded, aggregated and potentially toxic forms of one or more specific proteins. One proposed disease mechanism is that pathological protein aggregation contributes towards toxicity of neurons and subsequent neurodegeneration by physically obstructing axonal transport, sequestering essential proteins and disrupting overall protein homeostasis of the cell (Lee et al., 2011). Neurons are particularly vulnerable to this toxicity as they rely heavily on axonal transport between the cell body and synaptic terminals, and being a postmitotic cell type, they do not have an ability to disperse protein aggregates via cell division or be readily replaced (Yerbury et al., 2016). Neurons, therefore, depend heavily on an intrinsic network of protein quality control mechanisms designed to maintain proteostasis; a state in which all proteins in the proteome are in a conformation, concentration and location that is required for correct functioning of the cell (Balch et al., 2008).

Cells utilise several pathways to regulate the biogenesis, folding, trafficking and degradation of proteins to ensure that proteostasis is maintained. Any disruptions to these processes, caused by genetic or environmental factors, or an imbalance in misfolded or mutated proteins in the cell can often lead to disease (Zarouchlioti et al.,

2018). Cells employ intrinsic degradation systems, such as autophagy and the ubiquitin-proteasome system, to clear misfolded or aggregated proteins from the cell. These systems involve the compartmentalisation, degradation and recycling of misfolded or unfolded proteins by lysosomes or proteasomes, respectively (Nandi et al., 2006; Ravikumar et al., 2010; Tanaka and Matsuda, 2014). Cells are also equipped with a variety of pathways that aim to maintain proteostasis via transcriptional mechanisms, often in response to cellular stress. Examples include the heat shock response (HSR) and the unfolded protein response (UPR). The HSR is utilised to respond to stress throughout the cytoplasm and nucleus via transcriptional up-regulation of genes encoding heat shock factors (e.g. HSF1). In contrast, the UPR is activated in response to an accumulation of unfolded or misfolded proteins in the endoplasmic reticulum in response to stress (Hetz et al., 2015; Richter et al., 2010). Both the HSR and UPR pathways serve to restore protein homeostasis by reducing protein translation, degrading misfolded proteins, and activating signalling pathways that lead to the increased production of molecular chaperones involved in protein folding (Diaz-Villanueva et al., 2015).

Molecular chaperones are proteins that interact with, stabilise and / or aid in folding another protein, without becoming part of the final conformation of the protein (Ellis, 1987). They are a heterogenous and functionally diverse family of proteins which play an essential role in many cellular processes, such as protein folding, trafficking, quality control and degradation. Often described as the 'guardians of the proteome', molecular chaperones are usually the first line of defence against protein misfolding and aggregation, and their diverse range of functions serves to ensure the integrity of the intracellular environment (Jeng et al., 2015; Yerbury et al., 2016).

The term molecular chaperone encompasses many proteins that can be sub-classified into several categories based on their monomeric molecular weight (Hartl

et al., 2011; Jeng et al., 2015; Smith et al., 2015). The major families are Hsp90, Hsp70, Hsp60, Hsp40 and the small Hsps (sHSP). Each family of molecular chaperones has a unique molecular architecture and domain organisation that allows them to carry out a large variety of functions to quality control the intracellular protein environment effectively (Saibil, 2013). Many molecular chaperones are ubiquitously expressed to carry out 'house-keeping' roles within the cell, however under conditions of cellular stress, many molecular chaperones are upregulated to provide the cell with the resources to cope with potential threats to protein homeostasis and cellular function. Through their diverse range of functions and protein interactions, molecular chaperones quality control the intracellular protein environment, however, a failure in this system can lead to protein aggregation and disease.

Given their ability to refold, stabilise and prevent the aggregation of proteins, molecular chaperones are of particular interest in neurodegenerative diseases, where protein misfolding and aggregation is the primary pathological hallmark of disease. Indeed, evidence that chaperones may be attempting to resolve protein aggregation in these diseases comes from pathological studies that show molecular chaperones co-localising with protein inclusions (Barral et al., 2004; Broadley and Hartl, 2009; Jana et al., 2000; Uryu et al., 2006). While this is often cited as evidence that the molecular chaperones are attempting to circumvent the accumulation of misfolded or mutant protein, the very presence of an inclusion represents a failure of the cellular proteostasis network, as it was unable to prevent the misfolding and aggregation of the protein in the first place. Additionally, it has been suggested that the aggregates of protein may have sequestered the chaperone proteins into the inclusions, thereby further exacerbating the protein aggregation, as it reduces the available pool of chaperones to maintain proteostasis (Yang and Hu, 2016).

In any case, it is clear that in neurodegenerative disease, impairment of the chaperone and protein quality control machinery occurs, and this is likely to be a significant contributing factor to the formation neurotoxic protein inclusions and subsequent disease. Evidence to further support this notion comes from the identification of mutations in several molecular chaperones as the cause of several familial neurodegenerative diseases (Smith et al., 2015), thereby highlighting the importance of molecular chaperones in maintaining neuronal proteostasis.

Numerous studies have been conducted to determine whether manipulation of molecular chaperones could reduce protein aggregation and modulate toxicity, with many demonstrating beneficial effects of over-expression of many of chaperones in several neurodegenerative diseases (Smith et al., 2015). For example, in both cellular and transgenic mouse models of Alzheimer's disease (AD) overexpression of Hsp70 chaperones has been shown to reduce amyloid-beta aggregation and suppress disease-related phenotypes (Evans et al., 2006; Hoshino et al., 2011). Similar neuroprotective effects of various chaperone overexpression have also been demonstrated in models of Parkinson's disease (PD), ALS and Huntington's disease (HD) (Auluck et al., 2002; Cox et al., 2014; Jiang et al., 2012; Koyama et al., 2006; Patel et al., 2005; Wyttenbach et al., 2002). Additionally, pharmacological upregulation of chaperones, such as those in the Hsp70 system via Hsp90 inhibition, has been shown to be efficacious in reducing α-synuclein and amyloid beta cytotoxicity and aggregation in PD and AD models, respectively (Ansar et al., 2007; McFarland et al., 2014; McLean et al., 2002; Putcha et al., 2010), suggesting that manipulation of molecular chaperones could represent a potential therapeutic target for neurodegenerative disorders.

Given that the DPR proteins produced by RAN translation of the *C9orf72* repeat expansion form pathological protein aggregates, this study aims to address whether

molecular chaperones could also be used to reduce protein aggregation and modulate toxicity in *C9orf72* FTD/ALS. It should be noted that, after this study was completed, another study investigating the beneficial effects of a molecular chaperone on the DPR proteins was published. In 2017 Cristofani *et al.* demonstrated that the small heat shock protein B8 (HSPB8) was able to reduce accumulation of insoluble species of all DPR proteins formed of 100 repeats when overexpressed in the motor neuron-like cell type, NSC34.

3.1.2 HSJ1

This study will focus specifically on the molecular chaperone, HSJ1a, because its overexpression has been shown to be particularly effective in reducing inclusion formation of several proteins associated with neurodegenerative disorders in both cell and animal models (Chen et al., 2016; Howarth et al., 2007; Labbadia et al., 2012; Novoselov et al., 2013; Rose et al., 2011; Westhoff et al., 2005). HSJ1 is a member of the Hsp40 molecular co-chaperone family (Cheetham and Caplan, 1998). The Hsp40 (or DNAJ) family is a group of molecular co-chaperones, which act via a highly conserved 70 amino acid J-domain to stimulate the ATPase activity and substrate binding of Hsp70 proteins (Cheetham and Caplan, 1998; Kampinga and Craig, 2010). This interaction allows the Hsp70 chaperone machinery to carry out a diverse range of cellular functions, including modulating protein folding, degradation and protein-protein interactions (Smith et al., 2015). In addition to this general Hsp40 function, HSJ1 also has an additional function in that it targets polyubiquitinated proteins from the Hsp70 machinery to the proteasome via its two unique ubiquitin interacting motifs (UIMs) (Chapple et al., 2004; Howarth et al., 2007; Westhoff et al., 2005).

HSJ1 is encoded by the gene *DNAJB2* and is alternatively spliced into two isoforms both of which are preferentially expressed in neuronal tissue but with distinct subcellular localisations (Cheetham et al., 1992). The smaller 36 kDa isoform, HSJ1a,

exhibits cytoplasmic and nuclear localisation while the larger 42 kDa isoform, HSJ1b, is localised to the cytoplasmic face of the endoplasmic reticulum via C terminal geranylgeranylation (Chapple et al., 2004). Both isoforms contain an N-terminal J-domain with an adjacent glycine/phenylalanine (G/F) rich region and two unique UIMs. This study focuses only on the HSJ1a isoform.

Several studies have demonstrated a protective role for HSJ1a in cellular and animal models of neurodegenerative disease. In neuronal cell models, HSJ1a has been shown to decrease aggregation of several proteins known to aggregate and form inclusions in neurodegenerative disorders, including expanded polyglutamine containing proteins, TDP-43, mutant SOD1 and mutant Parkin (Chen et al., 2016; Gao et al., 2011; Howarth et al., 2007; Labbadia et al., 2012; Rose et al., 2011). Studies using animal models have also indicated a protective role for HSJ1a. Transgenic overexpression of HSJ1a has been shown to reduce mutant huntingtin aggregation and improve neurological performance in the R6/2 model of Huntington's disease (Labbadia et al., 2012), and in a mouse model of SOD1 ALS, overexpression of HSJ1a was able to reduce mutant SOD1 aggregation and enhance motor neuron survival (Novoselov et al., 2013). No other individual chaperone has been shown to have such a wide range of activity, or efficacy in the mammalian CNS, with no negative effects of overexpression.

Evidence suggests that HSJ1a is effective at reducing the aggregation of several other proteins associated with neurodegenerative disease however, it's effect on DPR proteins in C9orf72 FTD/ALS has not been reported. Preliminary data generated in the Cheetham laboratory to investigate the effect of HSJ1a on DPR proteins has indicated that co-expression of HSJ1a and poly-GA in a cellular model is able to reduce the number of poly-GA inclusions. The ability of HSJ1a to reduce inclusion incidence was found to increase with longer co-expression time periods and

increasing concentrations of HSJ1a transfected. Conversely, the disaggregating ability of HSJ1a decreased as the length of the poly-GA repeats (Christina Zarouchlioti, unpublished data).

3.1.3 Chapter 3 aims

Given the broad range of activity of HSJ1a against various pathological protein inclusions, the current study aimed to determine whether HSJ1a could have similar beneficial effects on reducing aggregation of *C9orf72* DPR proteins overexpressed in SK-N-SH cells. Preliminary data generated in the Cheetham laboratory suggested that co-expression of HSJ1a and poly-GA could reduce inclusion incidence in a time, dose and repeat-size dependent manner (Christina Zarouchlioti, unpublished data). However, this study relied on use of an anti-GA antibody to detect poly-GA, which was thought to be unable to detect all forms of poly-GA protein within the cell. To generate a more consistent and representative method of DPR protein detection without the need for specific antibodies, fluorescent tags were added to the DPR protein overexpression constructs. In addition to aiding their detection, the addition of fluorescent tags will also broaden the spectrum of experimental applications that the constructs can be utilised for. This study focused specifically on three of the DPR protein species, poly-GA, poly-PR and poly-GR, as these proteins have been demonstrated to be the most toxic in several models (Kwon et al., 2014; Mizielinska et al., 2014; Yamakawa et al., 2015; Zhang et al., 2016). Finally, this project additionally involved optimisation of methods used to quantitively assess protein inclusion incidence in cells.

The principle aims of Chapter 3 were to;

- Generate a library of fluorescently tagged DPR protein constructs that can be used to investigate DPR protein overexpression in a cell model.

- Investigate the effect of HSJ1a on inclusions formed by DPR proteins in an overexpression cell model.

3.2 Results

3.2.1 Generation of fluorescently tagged DPR proteins

To study DPR protein toxicity independently from RNA toxicity, 'protein-only' DNA constructs that produce each individual DPR protein without the repetitive GGGGCC RNA were used. These constructs were designed using alternative codons to code for each DPR protein (Mizielinska et al., 2014). Previous immunofluorescence studies using these constructs showed that antibodies raised against the DPR proteins were able to detect protein inclusions, but the detection of soluble, diffuse forms was poor (Ridler, 2016). To determine whether soluble forms of the DPR proteins could be detected, and to better visualise the proteins in cells, fluorescently tagged DPR constructs were generated using a restriction enzyme-based sub-cloning approach (Figure 3.1). Using this strategy, a library of fluorescently tagged poly-GA, poly-GR and poly-PR constructs, of both 36 and 100 repeat lengths, were generated by the addition of eGFP, mCherry or BFP at the N-terminus of the protein (Table 3.1). To generate the N-terminally tagged constructs, the eGFP, mCherry or BFP coding sequence were excised from pEGFP-C3, pMCherry-C3-SOD1 and mTagBFP2-C3, respectively, and sub-cloned into a linearised pcDNA3.1-DPR. The plasmids were subsequently sequenced to confirm the presence of the fluorescent tag in frame with the DPR protein open reading frame. The constructs generated in this study are shown in Table 3.1. mCherry and BFP tagged poly-GR constructs were not made due to technical difficulties with cloning and time restraints.

Table 3.1 Fluorescently tagged 'protein-only' DPR protein constructs generated in this study.

	No. of repeats	eGFP	mCherry	BFP
Poly-GA	36	✓	✓	✓
	100	✓	✓	✓
Poly-PR	36	✓	✓	✓
	100	✓	✓	✓
Poly-GR	36	✓	✗	✗
	100	✓	✗	✗
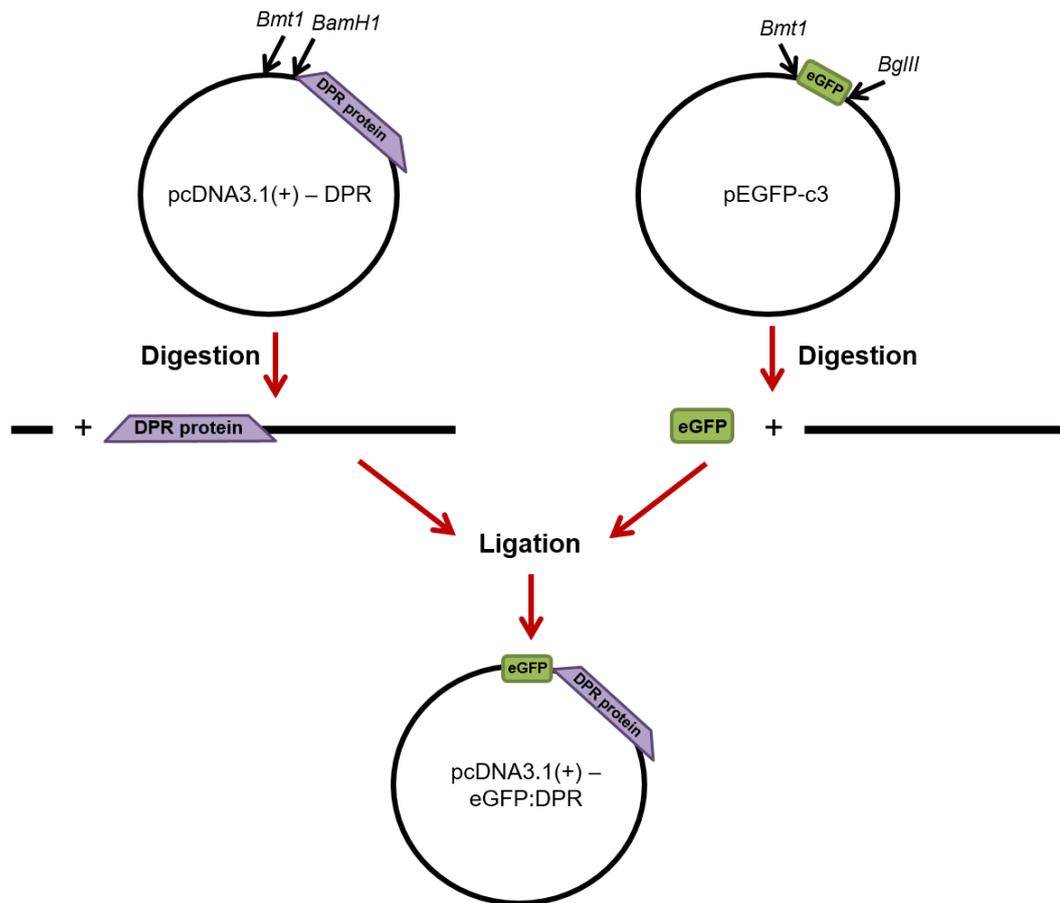
Figure 3.1. Schematic diagram of the cloning strategy used to generate GFP tagged DPR protein constructs

The eGFP coding sequence (green) was excised from the pEGFP-C3 plasmid using BglII and Bmt1 restriction enzymes and ligated into a linearised pcDNA3.1(+) construct containing the coding sequence for a DPR protein of either 36 or 100 repeats (purple).

3.2.2 Characterisation of fluorescently tagged DPR proteins in SK-N-SH cells

To characterise the subcellular localisation of the fluorescently tagged DPR proteins, the constructs generated were independently, transiently transfected into SK-N-SH cells, fixed 24 hours later and examined by fluorescence microscopy. These were compared to untagged DPR protein constructs, which were also independently transfected into SK-N-SH cells, stained with appropriate antibodies and visualised by fluorescence microscopy. DAPI was used as a nuclear stain for the untagged, eGFP and mCherry expressing constructs, and PI was used for the BFP tagged constructs.

All three of the fluorescently tagged poly-GA36 proteins showed diffuse cytoplasmic and nuclear localisation, in contrast to the untagged poly-GA36 detected with the anti-poly-GA antibody, which was only rarely detected in inclusions and never showed a diffuse staining pattern (Figure 3.2 A). Interestingly, eGFP-poly-GA36 occasionally formed bright, star-like inclusions, similar in shape to the characteristic pathological inclusions seen in *C9orf72* FTD/ALS patient post-mortem tissue (Figure 3.2 A).

Both the untagged and all three of the fluorescently tagged poly-GA100 proteins formed cytoplasmic inclusions (Figure 3.2 B). Due to the bright intensity of these inclusions, it was difficult to simultaneously capture an image of the poly-GA100 inclusions and the diffuse poly-GA100 staining. However, when the exposure time was increased, it was also possible to detect cells with a diffuse cytoplasmic distribution of GFP-GA100. This was also observed with mCherry-GA100 and BFP-GA100 proteins (data not shown). Diffuse cytoplasmic distribution of untagged-GA100, stained with an anti-poly-GA antibody, was not detected.

All the poly-PR36 and poly-PR100 constructs, both tagged and untagged, showed clear localisation to the nucleus, specifically to a sub-nuclear compartment, that is

likely to be the nucleolus (Figure 3.3). Further staining with nucleolar markers, such as nucelolin and nucleophosmin, is required to confirm this.

In contrast, poly-GR36 and poly-GR100 showed very different localisation patterns from each other. GFP-GR36 was localised to the nucleus in the subnuclear compartment, likely to be the nucleolus (Figure 3.4 A). In contrast, GFP-GR100 was completely excluded from the nucleus and showed a diffuse, cytoplasmic distribution (Figure 3.4 B). A benefit of having the fluorescently tagged constructs is highlighted with poly-GR, as the anti-GR antibody was unable to detect the untagged proteins, but the eGFP tag allowed the localisation of the protein to be visualised.
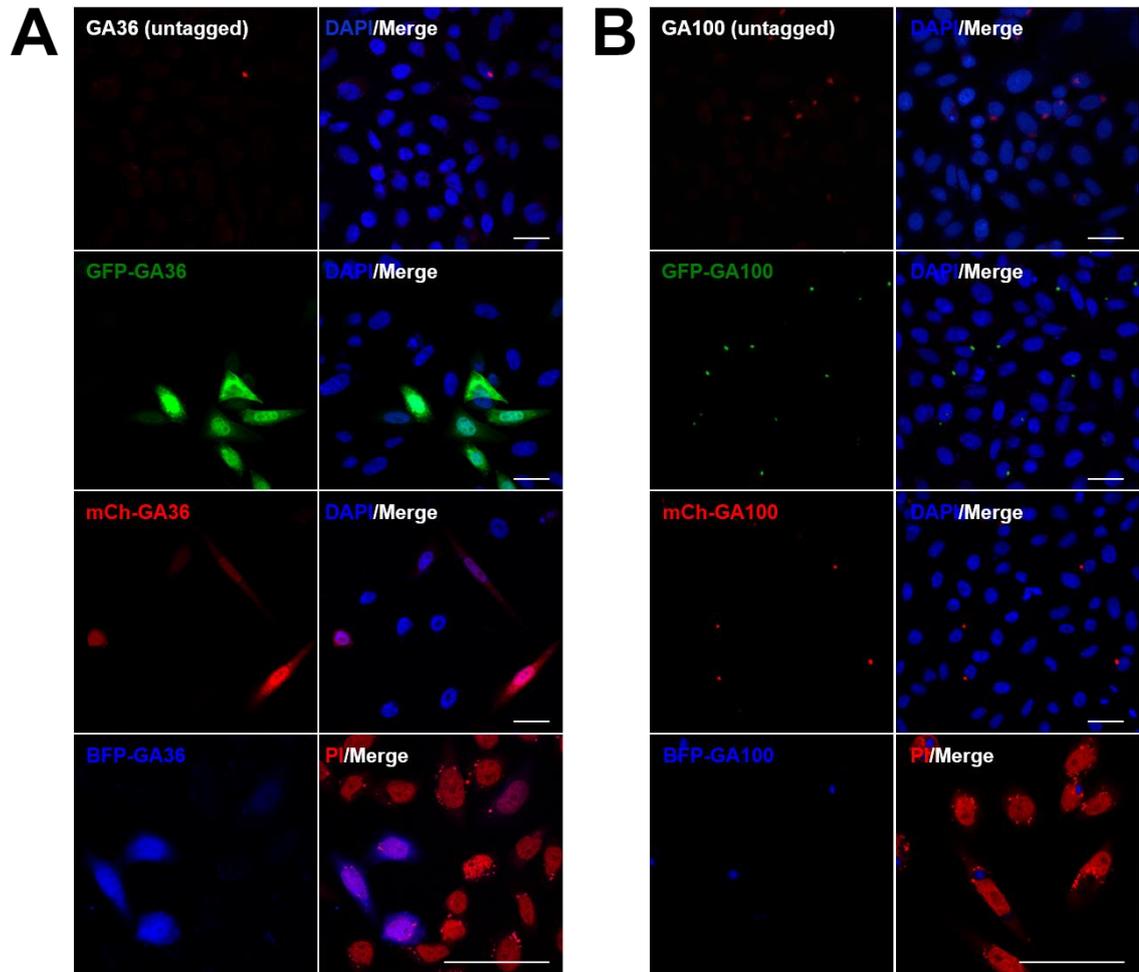
Figure 3.2. Localisation of untagged and fluorescently tagged poly-GA36 and poly-GA100 in SK-N-SH cells.

A) Anti-poly-GA antibody detection of untagged-GA36 shows rare cytoplasmic poly-GA inclusions but no diffuse protein. GFP-GA36, mCh-GA36 and BFP-GA36 all predominantly show diffuse cytoplasmic and nuclear localisation. **B**) Anti-poly-GA antibody detection of untagged-GA100 shows cytoplasmic poly-GA inclusions. GFP-GA100, mCh-GA100 and BFP-GA100 all also show cytoplasmic inclusions. In both A and B, either DAPI or PI was used to stain cell nuclei as specified in each panel. Scale bars represent 20 µM. Images are representative of 3 individual experiments.
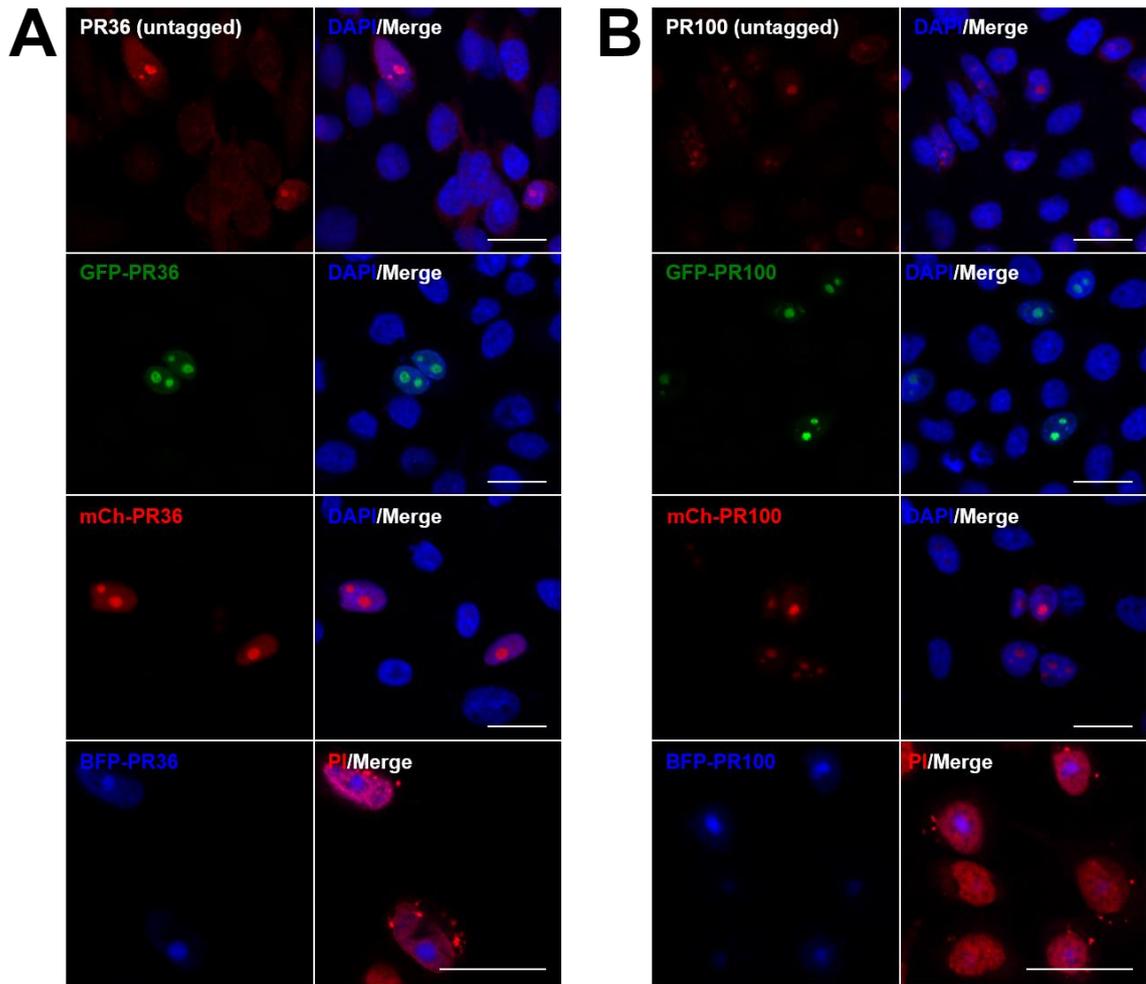
Figure 3.3. Localisation of untagged and fluorescently tagged poly-PR36 and poly-PR100 in SK-N-SH cells.

A) Anti-poly-PR antibody detection of untagged-PR36 and all three fluorescently tagged PR36 constructs show localisation to a sub-compartment of the nucleus. **B**) Anti-poly-PR antibody detection of untagged-PR100 and all three fluorescently tagged PR100 constructs show localisation to a sub-compartment of the nucleus. In both A and B, either DAPI or PI was used to stain cell nuclei, as specified in each panel. Scale bars represent 20 µM. Images are representative of 3 individual experiments.
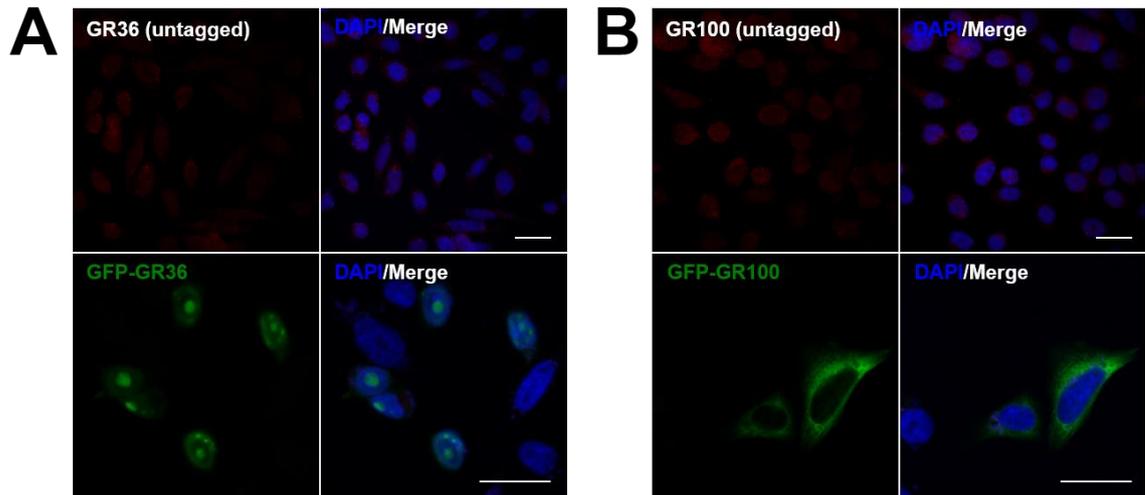
Figure 3.4 Localisation of untagged and GFP tagged poly-GR36 and poly-GR100 in SK-N-SH cells.

A) Anti-poly-GR antibody was unable to detect untagged-GR36. GFP-GR36 localised to a sub-compartment of the nucleus. **B**) Anti-poly-GR antibody was unable to detect untagged-GR100. GFP-GR100 showed diffuse localisation throughout the cytoplasm. In both A and B, DAPI was used to stain cell nuclei. Scale bars represent 20 µM. Images are representative of 3 individual experiments.

DPR protein inclusions in *C9orf72* FTD/ALS patient post mortem tissue have the characteristic feature of staining negatively for TDP-43, but positively for p62. As poly-GA100 (tagged or untagged) was the DPR protein that most frequently formed cytoplasmic protein inclusions in the SK-N-SH cell model, it was next determined whether these inclusions were also TDP-43 negative and p62 positive. SK-N-SH cells expressing GFP-GA100 were stained with a p62 antibody, which demonstrated that the poly-GA100 inclusions were p62-positive (Figure 3.5 B) In contrast, when GFP-GA100 was co-expressed with a TDP-43-myc construct, TDP-43 showed its usual nuclear localisation and was excluded from the GFP-GA100 inclusions (Figure 3.5 C). The GFP-GA100 protein, therefore, mimicked the TDP-43 negative and p62 positive characteristic of pathological DPR protein inclusions observed in *C9orf72* FTD/ALS patient tissue.
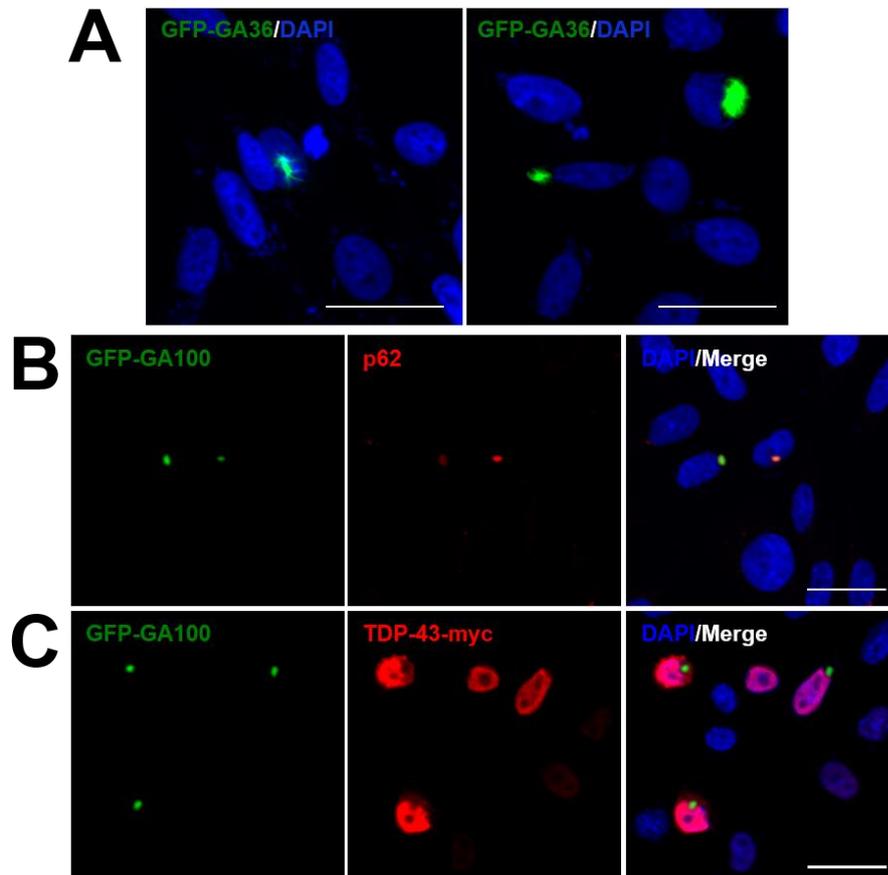
Figure 3.5. GFP-GA100 inclusions have similar characteristics to pathological poly-GA inclusions.

A) Star-like inclusions formed by GFP-GA36. **B**) p62 co-localises with GFP-GA100 inclusions. **C**) TDP-43-myc does not co-localise with GFP-GA100 inclusions. DAPI was used to stain cell nuclei. Scale bar represents 20 µM. Images are representative of 3 individual experiments.

An additional benefit of the GFP-tagged DPR protein constructs is that they can be visualised by immunoblotting. None of the antibodies raised against the DPR proteins used in this study were able to detect the untagged DPR proteins by immunoblotting (data not shown); however, some of the GFP tagged DPR proteins were detectable using an anti-GFP antibody (

Figure 3.6). The smaller GFP-tagged DPR proteins, consisting of 36 repeats were detected, at approximately 40 kDa. GFP alone was detected at 27 kDa for comparison. In contrast, it was difficult to detect the larger 100 repeat GFP-tagged DPR proteins at the expected weight of approximately 70 kDa; Occasionally, an

immunoreactive smear at the top of the blot, thought to represent aggregation and formation of high-molecular-weight complexes of the DPR proteins in the wells of the stacking gel, was detected. However, this effect was not consistently reproducible.
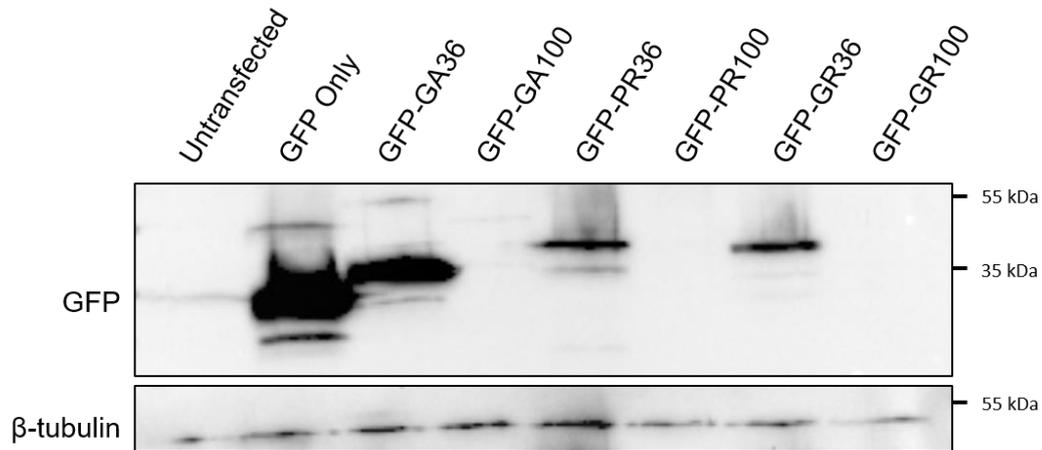
Figure 3.6. Detection of GFP tagged DPR proteins by immunoblotting.
Representative western blot of GFP tagged DPR proteins in transfected SK-N-SH cell lysates. GFP tagged proteins were detected using an anti-GFP antibody. GFP is detected at its predicted weight of 27 kDa. DPR proteins consisting of 36 repeats are detected at 40 kDa, however the DPR proteins with 100 repeats were not detected at the expected position. β-tubulin was used as a loading control.

3.2.3 Recruitment of HSJ1a to GFP-GA100 inclusions

As poly-GA was the only protein of the three DPR proteins investigated that clearly formed cytoplasmic aggregates, which recapitulated *C9orf72* FTD/ALS patient pathology, after characterisation of the fluorescently tagged DPR proteins, the effect of HSJ1a expression on the incidence of cytoplasmic poly-GA inclusions was investigated. Preliminary data generated in the Cheetham laboratory indicated that HSJ1a can significantly reduce the number of untagged poly-GA inclusions (Christina Zarouchlioti, unpublished data), however the addition of a tag to a protein is well-known to potentially alter the properties of the protein and its interactors (Jensen, 2012; Snapp, 2005). For this reason, it was important to assess whether the addition of fluorescent tags to the poly-GA protein could alter its response to the chaperone function of HSJ1a. For the remainder of this study, only the GFP tagged constructs were used.

Firstly, to determine whether HSJ1a is recruited to GFP-GA100 inclusions, myc-HSJ1a and GFP-GA100 were co-transfected into SK-N-SH cells. As expected, in cells expressing myc-HSJ1a alone, HSJ1a showed a diffuse pattern of localisation throughout both the cytoplasm and nucleus (Figure 3.7 A). In contrast, in cells co-expressing myc-HSJ1a with GFP-GA100, HSJ1a was recruited to GFP-GA100 and formed a distinct ring around the inclusion (Figure 3.7 B). To determine whether this ring may have structural importance to the interaction with the inclusion or was an artefact of the anti-myc antibody being unable to penetrate the inclusion, GFP-GA100 expressing cells were stained with an anti-GFP antibody (Figure 3.7 C). This antibody showed the same ring-like structure around the GFP-GA100 inclusion, indicating that the ring formed by HSJ1a is likely to be a result of the antibody being unable to penetrate to the centre of the inclusion.
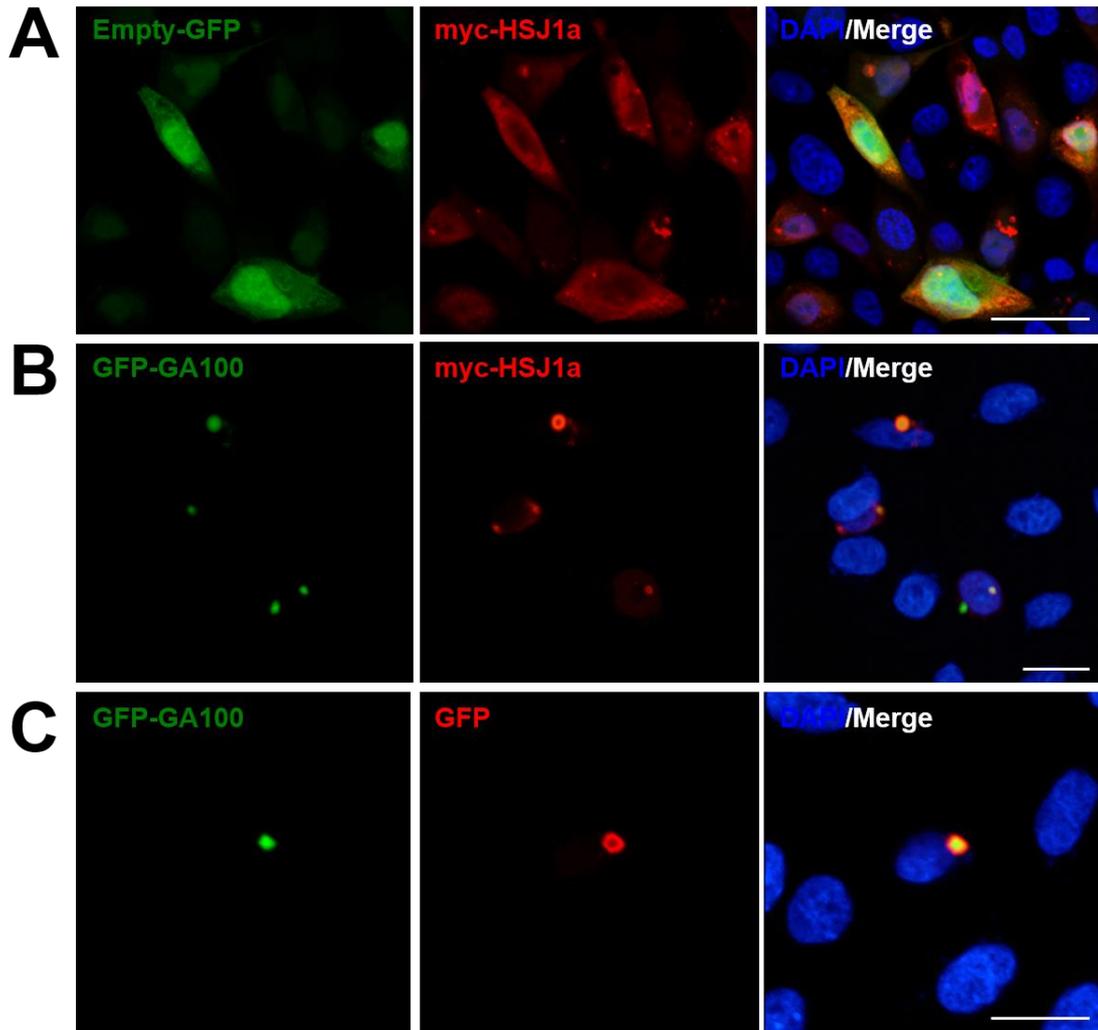
Figure 3.7 HSJ1a is recruited to GFP-GA100 inclusions

A) myc-HSJ1a (red) shows a diffuse pattern of localisation throughout the cytoplasm and nucleus when co-transfected with an empty-GFP plasmid (green). **B**) myc-HSJ1a (red) is recruited to and forms a ring around GFP-GA100 inclusions (red). **C**) Anti-GFP antibody (red) also forms a ring around GFP-GA100 inclusions (green). In all panels, cell nuclei are stained with DAPI. Scale bars represent 20 µM. Images are representative of 3 individual experiments. Images in panel A were acquired by Christina Zarouchlioti.

3.2.4 Effect of HSJ1a on GFP-GA inclusions

To investigate whether HSJ1a influenced the incidence of GFP-GA inclusions, SK-N-SH cells were transiently transfected for 24 hours with GFP-GA36 or GFP-GA100, and either myc-HSJ1a, or an empty pCMV-tag3b construct as a control, at a 1:1 ratio. Using fluorescence microscopy, the number of cells with GFP-GA inclusions were manually counted. Transfection efficiency was assessed as the percentage of GFP positive cells and was approximately 30 % across all conditions. The expression of HSJ1a appeared to reduce the percentage cells with GA inclusions for the GFP-GA36, but not the GFP-GA100, however this reduction was not statistically significant ($n = 3$, $p = 0.14$ and $p = 0.98$ respectively) (Figure 3.8 A-B). The percentage of transfected cells with GFP-GA36 inclusions was much lower than those with GFP-GA100 inclusions, suggesting that the longer poly-GA repeats are more likely to form inclusions.

To determine whether the load of GFP-GA100 may be overwhelming the chaperone function of HSJ1a, the ratio of GFP-GA100 and HSJ1a was altered so that SK-N-SH cells were transfected with 10 times more myc-HSJ1a plasmid than GFP-GA100 (Figure 3.8 C). This increase in HSJ1a reduced the incidence of GFP-GA100 inclusions, however, once again, this did not reach statistical significance ($n = 3$, $p = 0.15$).

Immunoblotting of GFP-GA also indicated that HSJ1a did not alter the level of total GFP-GA protein (Figure 3.8 D). In both the presence and absence of HSJ1a, the protein levels of GFP-GA36 and GFP-GA100 were unaffected. It should be noted that, once again, GFP-GA100 was more difficult to detect by western blot than GFP-GA36, possibly because of aggregation of the poly-GA protein meant that the protein remained in the wells of the stacking gel, or that the epitope was inaccessible to the GFP antibody.
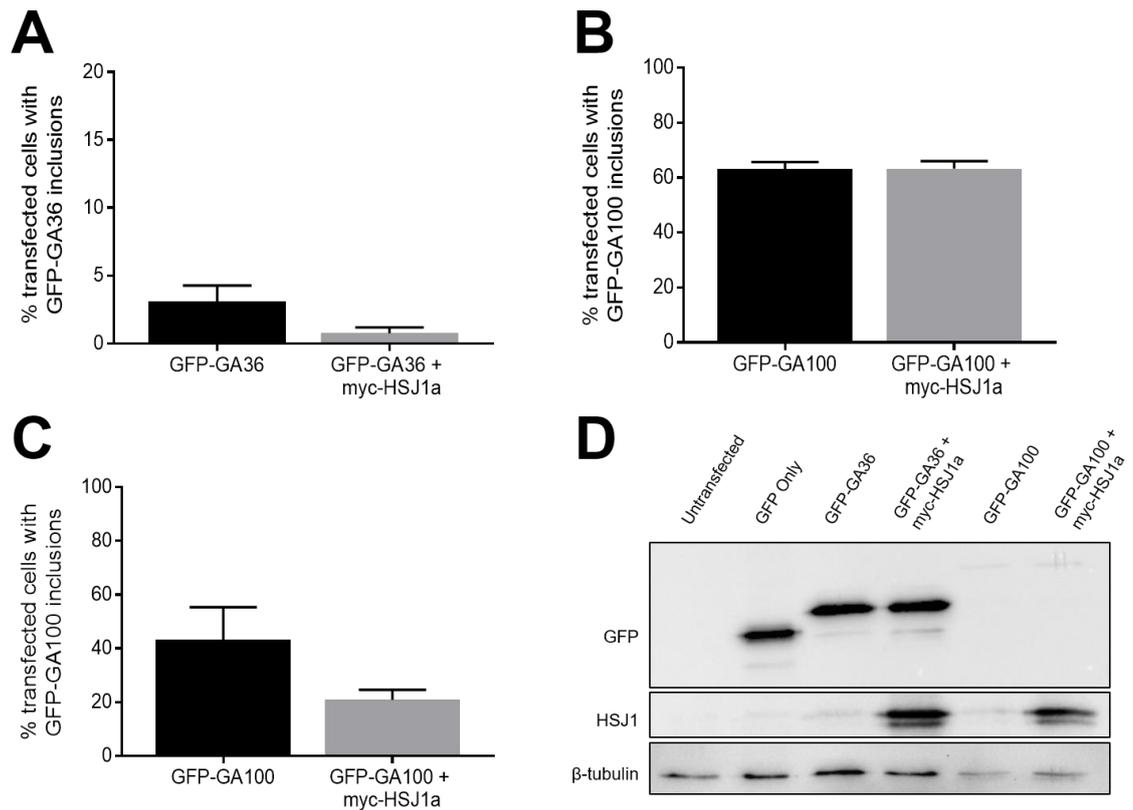
Figure 3.8 Effect of myc-HSJ1a on GFP-GA36 and GFP-GA100 inclusions co-transfected SK-N-SH cells

A) Percentage of transfected cells with GFP-GA36 inclusions in the absence and presence of myc-HSJ1a, where GFP-GA36 and myc-HSJ1a are co-expressed at a 1:1 ratio. p = 0.14.
B) Percentage of transfected cells with GFP-GA100 inclusions in the absence and presence of myc-HSJ1a where GFP-GA100 and myc-HSJ1a are co-expressed at a 1:1 ratio. p = 0.98.
C) Percentage of transfected cells with GFP-GA100 inclusions in the absence and presence of myc-HSJ1a where GFP-GA100 and myc-HSJ1a are co-expressed at a 1:10 ratio. p = 0.15.
In A – C, Student's un-paired T-test were used to test statistical significance, error bars represent standard error of the mean and n = 3 individual experiments. **D**) Western blot detection of total GFP-GA36 or GFP-GA100 in the absence and presence of myc-HSJ1a in transfected SK-N-SH cell lysates using an anti-GFP antibody. β-tubulin was used as a loading control.

3.2.5 Effect of HSJ1a on GFP-PR and GFP-GR

The sub-cellular localisation studies of the fluorescently tagged DPR proteins showed that the GFP-tagged poly-PR and poly-GR are mostly localised to a sub-nuclear compartment, with the exception of GFP-GR100. Although it is unlikely that this intra-nuclear accumulation of the proteins is an inclusion, as previous work from other groups suggests these proteins most likely localise to the nucleolus, it was next investigated whether HSJ1a could alter the number of cells with nuclear accumulation of poly-PR or poly-GR localisation. SK-N-SH cells were transfected with GFP-PR36, GFP-PR100, GFP-GR36 or GFP-GR100 and myc-HSJ1a or an empty vector. 24 hours post transfection the number of transfected cells containing GFP-PR or GFP-GR in a sub-nuclear compartment was counted. In all cases, HSJ1a had no significant effect on the percentage of transfected cells containing either poly-PR or poly-GR in the sub-nuclear compartment (Figure 3.9) (n = 3, GFP-PR36 $p = 0.17$, GFP-PR100 $p = 0.39$, GFP-GR36 $p = 0.49$, GFP-GR100 $p = 0.46$).
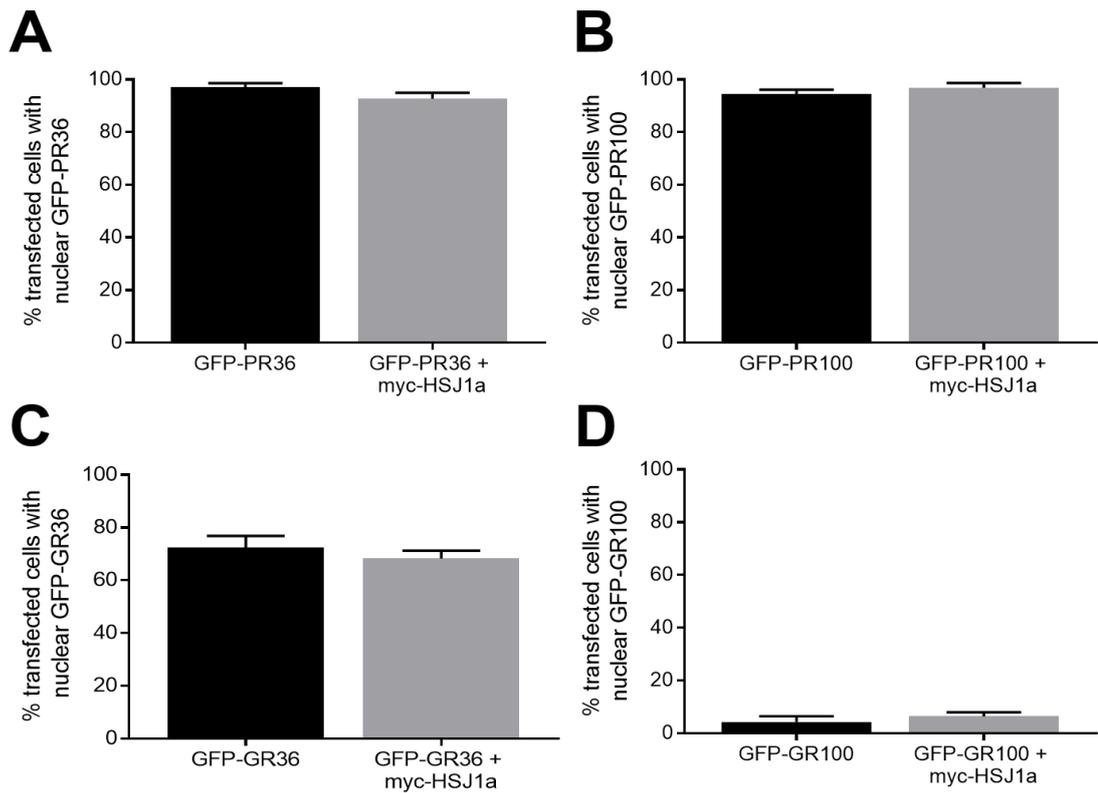
Figure 3.9 Effect of myc-HSJ1a on the nuclear localisation of GFP-PR and GFP-GR

Percentage of transfected SK-N-SH cells with nuclear **A**) GFP-PR36, **B**) GFP-PR100, **C**) GFP-GR36 and **D**) GFP-GR100 in the presence and absence of myc-HSJ1a, where the GFP-tagged DPR protein and myc-HSJ1a are co-expressed at a 1:1 ratio. Myc-HSJ1a has no significant effect on expression of nuclear GFP-PR and GFP-GR of either repeat length. Student's un-paired T-test were used to test statistical significance, error bars represent standard error of the mean and n = 3 individual experiments.

3.2.6 Optimisation of DPR protein inclusion quantification

To account for potential inconsistencies with the manual method of inclusion quantification, such as human error or investigator bias, alternative methods of quantification were investigated. As described previously, GFP-GA100 was transiently transfected with myc-HSJ1a or an empty plasmid in SK-N-SH cells at either a 1:1 or 1:10 ratio. The cells were fixed 24 hours post-transfection and the number of GFP-GA100 inclusions was assessed by three different, independent methods. The first method was manual assessment using an epifluorescent

microscope, scoring the presence of inclusions; the second method involved taking confocal images of multiple fields of view of each condition followed by using the image analysis software, Volocity, to count the inclusions (See Section 2.6.5 for Volocity protocol); and the third method involved using the Axioscan slide scanner to image the entire slide that the cells were plated on, and then, again, using Volocity to calculate inclusion incidence. In this experiment it was not possible to assess the percentage of transfected cells because the cells were not transfected with a transfection marker and the image analysis software could not score diffuse GFP-GA100 expression in addition to inclusions, because the intensity of diffuse GFP-GA100 was too low compared to the bright intensity of the inclusions. Instead, in all three methods, the number of GFP-GA100 inclusion was given as a percentage of total cell number, which was assessed by DAPI staining.

Figure 3.10 shows a comparison of the three methods of inclusion quantification as assessed by the percentage of total DAPI positive cells with GFP-GA100 inclusions in the absence and presence of myc-HSJ1a at a 1:1 (A – C) and 1:10 (D – E) ratio. In all instances, HSJ1a had no significant effect on the number of GFP-GA100 inclusions. P values obtained from a Student's unpaired t-test for each comparison are listed in Table 3.2. For both ratios of GFP-GA100 and myc-HSJ1a co-expression, a one-way analysis of variance (ANOVA) was performed to determine whether the average percentage change in inclusion number detected by each quantification method differed significantly from each other. No significant differences were found between methods, for both expression ratios, indicating that there is a good level of consistency between methods (Figure 3.11) (1:1 ratio $p = 0.8429$, 1:10 ratio $p = 0.6345$). In terms of efficiency, however, use of the Axioscan slide scanner followed by Volocity software analysis was the least time-consuming method and was also likely to be the least biased because both image acquisition and analysis were automated.
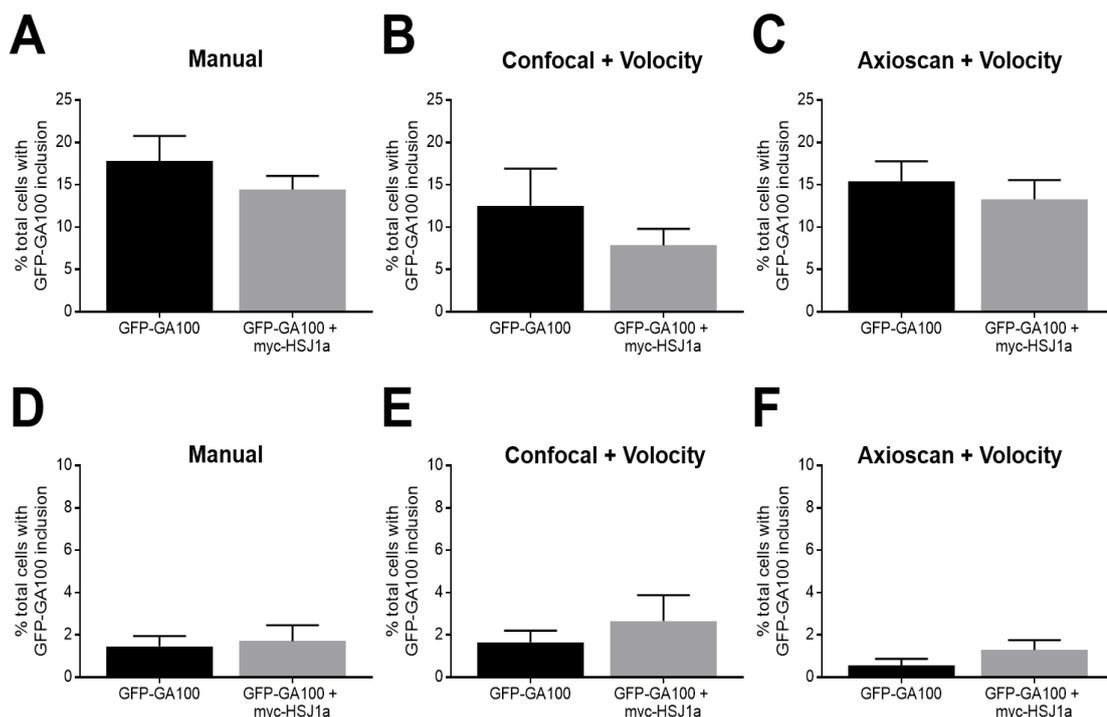
Figure 3.10 Quantification of GFP-GA100 inclusions assessed by three different methods.

Percentage of SK-N-SH cells with GFP-GA100 inclusions in the absence and presence of myc-HSJ1a at with a 1:1 (**A** – **C**) or 1:10 (**D** – **F**) ratio, as assessed by manual counting (**A** and **D**), confocal imaging followed by Volocity analysis (**B** and **E**) and Axioscan imaging followed by Volocity analysis (**C** and **F**). Student's un-paired T-test were used to test statistical significance, error bars represent standard error of the mean and n = 3 individual experiments.

Table 3.2 P values obtained from Student's unpaired T-Tests to compare inclusions quantification by three different methods.

Quantification Method	GFP-GA100 to myc-HSJ1a Expression Ratio	Corresponding Figure	p value	Significantly Different? (p < 0.05)
Manual	1:1	3.10 A	0.38	No
Confocal + Volocity	1:1	3.10 B	0.35	No
Axioscan + Volocity	1:1	3.10 C	0.55	No
Manual	1:10	3.10 D	0.80	No
Confocal + Volocity	1:10	3.10 E	0.49	No
Axioscan + Volocity	1:10	3.10 F	0.25	No
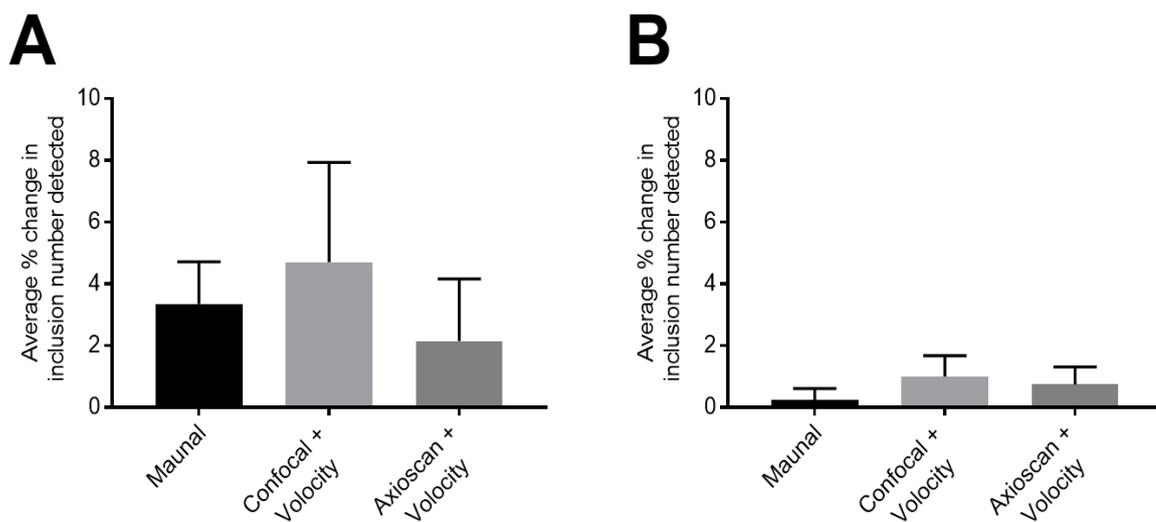
Figure 3.11 Comparison of the three different methods used for GFP-GA100 inclusion quantification.

A) Average percentage change in inclusion number detected for each quantification method at the 1:1 co-expression ratio. **B**) Average percentage change in inclusion number detected for each quantification method at the 1:10 co-expression ratio. No significant difference between the three quantification methods (manual, confocal + Volocity and Axioscan + Volocity) was observed for either ratio of GFP-GA100 to myc-HSJ1a co-expression (One-way ANOVA, 1:1 ratio p = 0.8429, 1:10 ratio p = 0.6345)

3.3 Discussion

3.3.1 Characterisation of tagged DPR protein constructs

The generation of fluorescently tagged poly-GA, poly-PR and poly-GR constructs enabled better detection of DPR proteins in transfected cells, compared with detection using antibodies. This difference is most clearly illustrated with the GA36 protein. All three of the fluorescently tagged GA36 proteins were clearly diffusely localised throughout the cytoplasm of transfected cells, with cytoplasmic inclusions occasionally detected, however, the anti-GA antibody was only able to detect inclusions and did not show any diffuse poly-GA. Similarly, the anti-GR antibody used in this study was unable to detect GR36 or GR100, but the GFP tag enabled detection of GFP-GR36 in the nucleus and GFP-GR100 in the cytoplasm.

The sub-cellular localisation of the fluorescently-tagged DPR proteins in this study recapitulates what has been reported in other cellular models. Several groups have reported that poly-GA forms cytoplasmic inclusions in cellular models (May et al., 2014; Wen et al., 2014; Zhang et al., 2014), and poly-PR has been frequently reported to localise to the nucleolus (Cristofani et al., 2017; Kwon et al., 2014; Lee et al., 2016; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2015). This is unsurprising given that recent work has demonstrated that poly-PR is able to undergo liquid-liquid phase transition; the physiological phenomenon which allows cells to form membrane-less organelles, such as nucleoli (Boeynaems et al., 2017; Lee et al., 2016; Weber and Brangwynne, 2015). More specifically, super-resolution fluorescent imaging of the ultrastructure of the nucleolus has revealed that both poly-PR and poly-GR localise to the outermost region of the nucleolus; the granular component, where final ribosomal biogenesis occurs (Boisvert et al., 2007; Lee et al., 2016). Although no nucleolar markers were used in this study, so definitive conclusions cannot be drawn, based on the existing literature and the shape and size of the intra-nuclear

accumulation of poly-PR seen in this study, the nucleolus is likely to be where poly-PR is localising to in this study also. The difference in localisation between the longer (GR100) and shorter (GR36) forms of poly-GR seen in this study has also been reported by other groups. Short forms of poly-GR consisting of 30 repeats have been shown to accumulate in the nucleus (Tao et al., 2015), while longer forms of poly-GR consisting of 80 or more repeats are reported to be diffusely localised throughout the cytosol (May et al., 2014; Yamakawa et al., 2015; Yang et al., 2015). The localisation of the poly-GR protein is clearly altered by the length of the GR repeat, but it is currently unclear what causes this localisation difference. One possible explanation could be that as the length of the poly-GR protein increases, it's large physical size or highly positive charge prevent it from crossing the nuclear envelope. It is interesting, however, that poly-PR and poly-GR possess similar structural and charge properties and yet larger forms of poly-PR are clearly able to enter the nucleus, perhaps suggesting that there is a unique feature of larger poly-GR proteins which specifically prevent it from doing the same.

The inclusions formed by the GFP tagged poly-GA constructs successfully modelled the pathology of DPR protein inclusions in *C9orf72* FTD/ALS patient brain tissue. Overexpression of GFP-GA36 in SK-N-SH cells led to the occasional formation of inclusions which have the characteristic 'star-like' morphology, a feature of inclusions that is frequently observed in *C9orf72* pathology (Ash et al., 2013; Mori et al., 2013a). Interestingly, it was only the poly-GA proteins which formed cytoplasmic inclusions in this cellular model, however in *C9orf72* FTD/ALS patient tissue all the DPR proteins are detected in cytoplasmic inclusions. This, perhaps indicates, the need for co-expression of DPR proteins or other factors to drive cytoplasmic inclusion formation of the poly-PR and poly-GR proteins. Although it was not performed in this study, generation of the different coloured tagged DPR protein constructs could facilitate future investigation of co-expression and the interaction between multiple DPR

proteins in a cell model because it is evident from detection of multiple DPR proteins in inclusions in human brain tissue that these proteins do not exist in isolation (Mori et al., 2013c). Nevertheless, the cytoplasmic inclusions formed by poly-GA were demonstrated to co-localise with p62 but were TDP-43 negative, which mimics another key characteristic hallmark of *C9orf72* pathological inclusions (Al-Sarraj et al., 2011).

As well as enhancing their detection by microscopy, the addition of a GFP tag to the DPR proteins also facilitated their detection by immunoblotting via anti-GFP antibodies. This could be beneficial for future biochemical studies of these proteins, given that detection of DPR proteins by western blot using DPR-specific antibodies was not otherwise possible in the Cheetham laboratory. GFP-GA36, GFP-PR36 and GFP-GR36 were all visible at their predicted size of 40 kDa. However, optimisation of the DPR protein extraction or immunoblotting protocol is required to enable detection of the larger 100 repeat DPR proteins by western blot. Occasionally, GFP-GA100, GFP-PR100 and GFP-GR100 can be detected as an immunoreactive smear at the top of the blots, presumed to represent the accumulation of the proteins in the wells. However, this was not consistently reproducible. This phenomenon has been reported in other biochemical studies of DPR proteins (Yamakawa et al., 2015). For the poly-GA protein, the accumulation in the stacking gel is likely to be due to the protein forming large insoluble aggregates which are unable to enter the SDS-PAGE gel. It is well known that proteins which have a tendency to form aggregates become insoluble and cannot be detected by conventional western blot (Tebbenkamp and Borchelt, 2009). It is, therefore, unsurprising that large poly-GA proteins are difficult to detect given the biophysical properties of a polymer formed from poly-GA, which has been shown to have a strong propensity to aggregate, forming amyloidogenic fibrils that stain positive with Congo red or thioflavin T (Chang et al., 2016; May et al., 2014). In contrast, the accumulation of poly-PR and poly-GR in the stacking gel is

likely to reflect their highly positive charge, which may not be able to be neutralised by SDS detergent, thus preventing the protein from moving through the polyacrylamide towards the positively charged anode. Optimisation of the protein extraction with different lysis buffers could be carried out to improve the protein's detection by western blot, for example increasing the concentration of SDS or including urea in the lysis buffer may help to increase the solubility of the DPR proteins. Several groups investigating DPR proteins have also encountered this problem with detection of larger DPR proteins by western blot (May et al., 2014; Yamakawa et al., 2015; Zhang et al., 2016; Zhang et al., 2014). Many studies, therefore, now use alternative biochemical assays, such as dot blots, to measure total DPR protein expression as this does not require separation of the protein by size (Lopez-Gonzalez et al., 2016; Mizielinska et al., 2014; Wen et al., 2014). Conversely, other groups have exploited the aggregation properties, particularly of the poly-GA protein, and used filter retardation assays to assess the solubility of these proteins (Chang et al., 2016; Cristofani et al., 2017).

3.3.2 Effect of HSJ1a on GFP-tagged DPR proteins

Following the generation and characterisation of the fluorescently tagged DPR protein constructs, the GFP tagged constructs were used to investigate the effect HSJ1a on GFP tagged poly-GA, poly-PR and poly-GR. Preliminary data generated in the Cheetham laboratory suggested that HSJ1a was able to reduce the incidence of inclusions formed by untagged poly-GA (Christina Zarouchlioti, unpublished data), however, the addition of a GFP tag to the protein may alter the behaviour of poly-GA and its interaction with of HSJ1a; an important factor to consider when planning future experiments utilising cell models to investigate mechanisms of reducing DPR protein aggregation. Thus, the effect of co-expression of HSJ1a was investigated on the newly generated GFP-tagged DPR proteins.

HSJ1a had no effect on the nucleolar localisation of GFP-PR or GFP-GR of either repeat length. This was unsurprising given that these proteins are not thought to be aggregated in an inclusion, but rather they are thought to accumulate within a sub-compartment of the nucleus, most likely the nucleolus, due to their ability to phase separate (Boeynaems et al., 2017; Lee et al., 2016). Thus, although HSJ1a is expressed in the nucleus (Chapple et al., 2004), it had no significant effect on the number of cells containing nuclear GFP-PR or GFP-GR, possibly because these proteins were not pathologically aggregating or disrupting proteostasis, thus were unlikely to have generated signals to employ the chaperoning abilities of HSJ1a; poly-ubiquitination or Hsp70 recruitment (Cheetham et al., 1994; Westhoff et al., 2005).

In contrast, GFP-GA100 formed dense, cytoplasmic protein inclusions that clearly recruited HSJ1a. It is possible that the chaperone could have been recruited there because it had recognised a misfolded protein and is binding in an attempt to restore proteostasis. This effect of HSJ1a has been demonstrated in many other cellular studies of aggregated proteins associated with neurodegenerative diseases (Chen et al., 2016; Howarth et al., 2007; Labbadia et al., 2012; Novoselov et al., 2013; Rose et al., 2011). In these studies, HSJ1a was found to have a beneficial effect in reducing inclusion incidence, either via interactions with client proteins through its UIM domains and stimulating proteasomal degradation (Howarth et al., 2007; Labbadia et al., 2012) or by stimulating the interaction of the client protein with Hsp70 through its J-domain (Chen et al., 2016; Novoselov et al., 2013; Rose et al., 2011). The same beneficial effect was not seen in this study, as co-expression of myc-HSJ1a was unable to significantly reduce inclusion incidence or protein levels of either GFP-GA36 or GFP-GA100. However, when co-expressed at a 1:1 ratio, there was a consistent trend for HSJ1a to decrease the percentage of transfected cells with poly-GA inclusions, perhaps indicating that HSJ1a is reducing poly-GA aggregation to some extent. It is possible that inter-experiment variability, such as differences in transfection

efficiency, may account for the fact this trend did not reach statistical significance. This is unlikely, however, as the trend to reduce inclusions incidence was also not replicated when HSJ1a expression was 10 times higher than that of poly-GA (1:10 ratio), suggesting that, even in excess, HSJ1a alone may be insufficient to significantly reduce the number of inclusions formed by GFP tagged poly-GA. It is important to note, however, that, although no reduction in inclusion incidence was observed after the addition of myc-HSJ1a, the chaperone may be making smaller, subtler changes to the poly-GA inclusions, such as altering inclusion size, morphology or the number of inclusions within a cell. These parameters were not assessed during this study, so the effect of HSJ1a on them cannot yet be deduced, however future work should include these alternative assays to the assess effect of HSJ1a as this may potentially reveal more subtle effects that inclusion incidence alone is not sensitive enough to detect.

Given that preliminary data had indicated HSJ1a was able to reduce inclusion incidence of untagged poly-GA, the current data suggests that the addition of the GFP tag alters the aggregation propensity of the poly-GA protein, making it more difficult to disaggregate, or the tag hinders the chaperone function of HSJ1a. GFP is well-known to dimerise and self-aggregate (Day and Davidson, 2009; Tsien, 1998; Yang et al., 1996), particularly at high concentrations as in an over-expression model. Therefore, it is possible that tagging the poly-GA protein with GFP may have caused an enhancement in the poly-GA proteins ability to form inclusions. This may then have made it more difficult for HSJ1a to reduce aggregation of the DPR protein. To try to reduce self-aggregation of the GFP tag, the enhanced GFP variant (eGFP) was used in this study, as this variant has been shown to be less aggregation prone (Krasowska et al., 2010). Nevertheless, weak dimerization of eGFP is still possible (Shaner et al., 2005), and thus this may still have increased protein aggregation. It is also possible that the GFP tag may have physically blocked or interfered with HSJ1a binding or

activity. At 27 kDa, GFP is relatively large in comparison to the size of the GA100 protein (predicted size = 70 kDa), so it is possible that the tag may physically hinder HSJ1a interaction, thus preventing it from performing as a molecular chaperone. It is also worth noting that some proteins are sensitive to the position of GFP attachment (Crivat and Taraska, 2012). In this study, the DPR proteins were only N-terminally tagged with GFP, however when generating GFP-fusion proteins it is recommended that both sides of GFP fusion proteins be tagged in order to test for the effect of the tag on the functionality and sub-cellular localisation of proteins (Crivat and Taraska, 2012). C-terminally tagging the DPR proteins may have a different effect on their localisation or propensity to form inclusions, and thus a different effect on the ability of HSJ1a to reduce aggregation.

An alternative explanation for the ineffectiveness of HSJ1a at reducing the GFP tagged poly-GA inclusions, is that the aggregating poly-GA protein has sequestered the chaperone into the inclusion, thus rendering it unable to carry out mechanisms to restore proteostasis. This would explain why HSJ1a is localised to the GFP-GA100 inclusions but does not significantly reduce them. Other molecular chaperones, including members of the Hsp70 family, have been found to associate with DPR proteins in mass spectrometry studies (Lee et al., 2016; Lin et al., 2016). It is unclear whether this association is a result of the molecular chaperone machinery attempting to restore proteostasis, or whether the proteins have been aberrantly sequestered into the inclusion. If the former is true, overexpression of these other chaperones could be more effective than HSJ1a at reducing GFP-GA inclusions.

Since performing this study, another molecular chaperone has been demonstrated to have a significant effect on the accumulation of insoluble forms of all five of the DPR proteins (Cristofani et al., 2017). The small heat shock protein B8 (HSPB8) was able to reduce the accumulation of insoluble species formed by all DPR proteins formed

of 100 repeats, when overexpressed in the motor neuron-like cell type, NSC34. Interestingly, HSPB8 was found to have a particularly pronounced effect on reducing insoluble forms of the DPR proteins investigated here, poly-GA, poly-GR and poly-PR, however it is important to note that the proteins expressed in this study were tagged with the FLAG epitope, which is significantly smaller than the GFP tag used in this study and thus much less likely to hinder any interaction between DPR protein and chaperone. HSPB8 has previously been shown to induce the removal of several neurodegeneration-related aggregation prone proteins via autophagy mediated disposal (Carra et al., 2008; Crippa et al., 2010; Wilhelmus et al., 2006), and the authors demonstrate that overexpression of HSPB8 is able to reduce insoluble forms of all DPR proteins, except poly-PR, via the same autophagic mechanism in NSC34 cells. It is highlighted, however, that different mammalian cell types may utilise different protein clearance pathways, thus it is important to note that other clearance systems may act, either independently or simultaneously, to reduce DPR protein aggregation in other models. This provides another possible explanation for the ineffectiveness of HSJ1a to reduce GFP-GA inclusions; perhaps other proteostasis pathways are preferred over proteasomal degradation or Hsp70/Hsp40 mediated protein refolding in the SK-N-SH cell type.

Regardless of the pathway utilised in these relatively simple cell models, investigation into the removal or reduction of aggregation of DPR proteins in a more disease-relevant cell model, such as neurons derived from *C9orf72* patient iPSCs, is required to assess endogenous DPR protein removal mechanisms in cells that are affected in the disease process. This has been problematic, however, due to the difficulties in detecting endogenous DPR proteins in patient-derived cells.

Despite the effectiveness of HSPB8 to reduce the insoluble forms of the DPR proteins, and potentially HSJ1a when the DPR protein is untagged, it should be noted

that overexpression of molecular chaperones to reduce aberrant protein aggregation, is probably not an ideal therapeutic option. Molecular chaperones form an essential component of the cellular proteostasis machinery in every cell, thus a generalised increase in levels of these proteins to combat a specific aberrant protein aggregation in one cell type could have detrimental off-target effects on essential proteins in cells throughout the body. For this reason, pharmacological compounds that are more specifically targeted to a protein prone to aggregation or a particular cell type would be a more preferable method of manipulating proteostasis as these are likely to have less off-target effects and would be more therapeutically viable than overexpressing endogenous molecular chaperones.

3.3.3 Optimisation of inclusions counting

In terms of assessing which method was best for quantifying protein inclusions, all methods used showed similar results. There were no significant differences between methods in their ability to detect the average percentage change in GFP-GA100 inclusion number. This indicates consistency between the three methods and suggests that any could be used in future quantifications of protein inclusions. However, given that the Axioscan scanner can scan multiple slides at once, and the image analysis software is considerably faster than manual counting, using the Axioscan in combination with an image analysis software is likely to be the most efficient method of quantifying protein inclusions. Additionally, use of image analysis methods, generated either from confocal or Axioscan imaging, allows for assessment of other parameters in addition to inclusions incidence, for example inclusion size, morphology or intensity. This may help to detect effects of HSJ1a that are not linked to inclusion incidence. However, it is important to note that problems still exist with using the Axioscan method, for example, the images are at a lower resolution and this method was unable to assess diffuse GFP-GA100 simultaneously with GFP-GA100 inclusions because the signal was too low compared to the intensity of the

inclusions. These factors could mean that smaller, subtler effects could be more easily missed with the Axioscan and Volocity method, compared to assessment by eye.

3.3.4 Conclusions

In summary, this project involved the generation of fluorescently tagged DPR protein constructs that have allowed for easier detection of DPR proteins by immunofluorescence and by immunoblotting. The tagged proteins show similar localisation to what has been reported by other groups and inclusions formed by these DPR proteins recapitulate several aspects of pathological inclusions seen in *C9orf72* patient brain tissue. Unlike current antibody detection, the fluorescent tags allow both diffuse and aggregated DPR proteins to be visualised and, by having each DPR protein tagged with three different colours, co-expression of the DPR proteins can be performed to investigate interactions between the proteins. Some of the fluorescently tagged constructs generated were then used to investigate the effect of overexpression of HSJ1a on DPR protein inclusions. Several methods were used to quantify GFP-GA inclusions and found HSJ1a to have no significant effect on the incidence of GFP-GA inclusions, despite the chaperone being recruited to the inclusions. Similarly, no effect of HSJ1a was found on GFP-PR or GFP-GR presence in the nucleolus. Overall, the results suggest that HSJ1a does not significantly affect the GFP-tagged *C9orf72* DPR proteins *in vitro*.

Chapter 4.   Generation of fluorescently tagged DPR proteins in *C9orf72* patient iPSCs using CRISPR-Cas9 technology

4.1 Introduction

Whilst the fluorescently tagged DPR protein constructs generated in Chapter 3 are a useful tool for studying these proteins in mammalian in cell lines, this method relies on the overexpression of DPR proteins to levels that are unlikely to be physiologically relevant. Additionally, these overexpression experiments typically utilise immortalised cell lines that arise from cell types unaffected in *C9orf72* FTD/ALS thus may not fully recapitulate some aspects of disease pathology. To gain a better understanding of how DPR proteins may cause toxicity at physiologically relevant levels in cells affected in *C9orf72* FTD/ALS, the next part of this project involved the generation of an iPSC line derived from a *C9orf72* patient, where CRISPR-Cas9 genome editing was used to insert the coding sequence of a fluorescent tag downstream of the repeat expansion. This would cause endogenous RAN translation to produce DPR proteins with a C-terminal fluorescent tag. The generation of these cells will provide a useful tool for future experiments to be performed on endogenously produced *C9orf72* DPR proteins in disease-relevant cells types derived from iPSCs, such as motor or cortical neurons.

4.1.1   Justification for addition of eGFP to endogenous DPR proteins

Although the addition of a GFP tag to the DPR proteins in Chapter 3 proved to be a hinderance in studying of the effect of molecular chaperones on DPR proteins, monomeric eGFP was again chosen as the tag to insert downstream of the *C9orf72* repeat expansion in C9-iPSCs for several reasons outlined below.

Firstly, expression of DPR proteins in iPSCs and iPSC-derived cells is low, with some DPR proteins being undetectable. Given that the GFP tag added to poly-GA in Chapter 3 made the protein more resistant to the effects of a molecular chaperone, it could be hypothesised that the GFP tag was stabilising the DPR protein, therefore increasing the stability of DPR proteins would enhance their detection and make them easier to study. The addition of a GFP is not predicted to reduce the relative toxicity of DPR proteins as several groups have previously demonstrated that addition of a fluorescent tag to DPR proteins does not alter their toxicity when overexpressed *in vitro* or *in vivo* models (Schludi et al., 2017; Tao et al., 2015; Zhang et al., 2016). Another reason for choosing an eGFP tag was that it could be utilised during the process of generating CRISPR-Cas9 editing C9-iPSC lines to isolate GFP-positive cells by fluorescence activated cell sorting (FACS). Finally, addition of a fluorescent tag to DPR proteins will enable them to be studied *in vivo* over time using fluorescence longitudinal imaging. This is a novel method of studying these proteins, as previous studies of DPR proteins in C9-iPSCs have relied on their detection in fixed or lysed cells (Almeida et al., 2013; Dafinca et al., 2016; Donnelly et al., 2013; Lopez-Gonzalez et al., 2016; Simone et al., 2018; Su et al., 2014a; Westergard et al., 2016).

4.1.2 Detection of DPR proteins in *C9orf72* iPSC-derived neurons

Several studies have demonstrated that motor and cortical neurons derived from C9-iPSCs endogenously produce DPR proteins by RAN translation. Poly-GP is the most abundant, but poly-GA and poly-GR have also been detected in some studies (Almeida et al., 2013; Dafinca et al., 2016; Donnelly et al., 2013; Simone et al., 2018; Su et al., 2014a; Westergard et al., 2016). As previously mentioned, these studies all rely on detection of DPR proteins in fixed or lysed cells by immunofluorescence (Donnelly et al., 2013; Su et al., 2014a; Westergard et al., 2016), dot blots (Almeida et al., 2013; Dafinca et al., 2016; Lopez-Gonzalez et al., 2016), or Meso Scale Discovery (MSD) immunoassay (Simone et al., 2018). However, as demonstrated in

Chapter 3, antibody-based methods do not necessarily provide detection of all forms of the proteins within the cell. Additionally, they only provide information about one-time point within the life of a neuron, making it more difficult to deduce the long-term effects of these proteins. The generation of a C9-iPSC line with fluorescently tagged DPR proteins would allow for a completely novel method of studying these proteins. Using the fluorescent tag, the production, transport, localisation and toxicity of these proteins could be studied *in vivo* for extended periods of time using live fluorescence longitudinal imaging.

The iPSC line used in this study was chosen because it contains a large number of GGGCC repeats (~ 638) within the *C9orf72* gene and has previously been demonstrated to have the highest level of expression of DPR proteins detectable by MSD immunoassay of all the *C9orf72* iPSC lines available in the Isaacs laboratory (Simone et al., 2018).

4.1.3 Principles of CRISPR-Cas9 genome editing

To insert the eGFP sequence into the genome of a C9-iPSC, CRISPR-Cas9 genome engineering technology was used. The CRISPR-Cas9 system is a prokaryotic defence mechanism that has been adapted in recent years to become a versatile genome-editing system (Cong et al., 2013; Jinek et al., 2012; Ran et al., 2013; Sander and Joung, 2014; Wright et al., 2016). The process involves inducing a single or double strand break at a specified cut site adjacent to a protospacer adjacent motif (PAM) using a Cas9 endonuclease, which is guided there by specifically designed small guide RNAs (gRNA). Once a break has been induced in the DNA, cellular DNA repair mechanisms are triggered to repair the break via two different pathways; the efficient but error-prone non-homologous end joining (NHEJ) pathway, or the less efficient but higher fidelity homology directed repair (HDR) pathway. The NHEJ pathway is the most active repair mechanism, capable of rapidly repairing DNA

damage, however this process can result in small nucleotide insertions or deletions at the site of the DNA break as there is no DNA template to guide the repair. This can be useful in genome engineering to create knock-out models. In contrast, the HDR pathway is much slower and requires a template DNA sequence that has some sequence homology to the damaged DNA strand. This template is typically referred to as the 'repair template' and can be from a sister chromatid, homologue or exogenously provided DNA. The process is error-free if the DNA template used for repair is identical to the original DNA sequence, but it can also be used to create specific mutations or insertions if an exogenously provided DNA repair template has enough homology to the sequence surrounding the target insert site.

4.1.4 Strategy of CRISPR-Cas9 mediated eGFP insertion

In this project, the HDR process was utilised to insert the protein-coding monomeric eGFP sequence downstream of the *C9orf72* repeat expansion in a patient-derived iPSC line. The strategy designed to do this involved inducing a double strand break downstream of the *C9orf72* repeat by nucleofecting C9-iPSCs with (1) recombinant *Streptococcus pyogenes* Cas9 (SpCas9) nuclease protein, (2) a gRNA designed to target the genomic region downstream of the *C9orf72* repeat, and (3) a single stranded DNA oligonucleotide (ssDNA) containing the eGFP sequence flanked by two homology arms to act as a repair template for HDR. Several methods of inserting these essential components of the CRISPR-Cas9 system are available and these were considered when designing the experiment (Li et al., 2016; Yang et al., 2014).

The method used to deliver the Cas9 protein and gRNAs into cells is known to influence the frequency of off-target mutations, with fewer off-target effects reported following the delivery of recombinant Cas9 protein compared to Cas9 expression from a plasmid (Kim et al., 2014; Liang et al., 2015; Zischewski et al., 2017). This is due to the Cas9 protein being subjected to the endogenous cellular degradation processes,

and thus has a limited window of opportunity to induce double stranded breaks in the genome compared to Cas9 being constantly expressed from an expression plasmid (Kim et al., 2014). Therefore, to reduce the chances of off-target mutations in this project recombinant SpCas9 was chosen as the method of delivery of the nuclease instead of a plasmid expressing the SpCas9 nuclease.

The method of delivery of the donor DNA repair template using ssDNA or double stranded DNA (dsDNA) was also a consideration. A ssDNA was chosen as the exogenous DNA repair template because single stranded DNA is typically inserted at higher efficiency and is less toxic to cells than dsDNA repair templates (Ran et al., 2013). Previously, use of ssDNA as repair templates in CRISPR genome editing has been limited by the length of the ssDNA because the limit of commercially synthesisable ssDNA was up to 200 bases. However, recent studies have demonstrated the successful generation of longer ssODN molecules by '*in vitro* Transcription and Reverse Transcription (*iv*TRT)' that efficiently serve as repair templates in CRISPR-Cas9 mediated sequence insertion (Miura et al., 2015; Yoshimi et al., 2016). Given that the ssDNA required to insert the eGFP sequence in this study was approximately 1 kb long, the principles of *iv*TRT were used to generate the ssDNA repair template.

4.1.5 Chapter 4 aims

This chapter describes the generation of CRISPR-Cas9 edited C9-iPSCs where eGFP was inserted downstream of the hexanucleotide repeat expansion to generate a model where fluorescently tagged DPR proteins can be studied at endogenous levels in disease-relevant cell types. To achieve CRISPR-Cas9 genome editing, several steps were involved to design and synthesise the necessary components of the system. The principle aims of Chapter 4 were;

- Design gRNA to target the genomic region downstream of the *C9orf72* hexanucleotide repeat expansion.

- Test the efficiency of designed gRNA at targeting this region.

- Design and synthesise single stranded DNA oligonucleotides to act as a repair template in homology directed repair.

- Utilise the designed gRNA and ssDNA template in CRISPR-Cas9 genome engineering of *C9orf72* patient-derived iPSCs to C-terminally tag an endogenous DPR protein with eGFP.

4.2 Results

4.2.1 Guide RNA design

Several online gRNA design tools were used to design gRNAs targeting the region downstream of the *C9orf72* repeat expansion. The online tools used were; CRISPOR, CRISPR.MIT.EDU design tool and ATUM crispr grna design tool (available online from http://crispor.tefor.net/, https://www.atum.bio/eCommerce/cas9/input, and http://crispr.mit.edu). The region immediately downstream of the *C9orf72* repeat expansion has an extremely high GC-content (89.8 %), making it difficult to design effective gRNAs that will target this region. For this reason, gRNAs were designed to target the region further downstream of the *C9orf72* repeat. At the translational level, this also provides the additional benefit of creating a 'linker-region' between the DPR protein and the eGFP tag. To account for this, the genomic sequence of *C9orf72* that was 88 bp downstream of the *C9orf72* repeat was inputted into the online gRNA design tools. The sequence inputted was 177 bp long. Where required, gRNA length was selected as 20 bp and the PAM site for the gRNAs to target was selected as 5'-NGG-3'. Each online gRNA design tool provided a list of possible gRNA sequences, with various efficiency, quality or specificity scores, and the number of predicted off-target regions, depending on the algorithms used to generate the gRNA sequences. Based on the efficiency, quality and specificity scores, ten gRNAs were selected for further analysis (Table 4.1). These gRNAs were chosen based on whether they were predicted by more than one of the three gRNA design tools, and whether they had high quality, efficiency or specificity scores. A mixture of gRNAs targeting the forward and reverse strands were chosen, in case one of the strands is more efficiently targeted than then other.

Table 4.1 Guide RNA sequences designed to target downstream of C9orf72 repeat expansion

	Guide Name	Sequence	Target Strand	MIT tool		CRISPOR Tool			Atum Tool	Chosen Guides
				Quality Score	No. of off-targets	Specificity Score	Efficiency Score	No. of off-target	Quality Score	
1	Guide 1	GGGCTTTCGCCTCTAGCGAC (TGG)	Fwd	96	34 (9 in genes)	97	51	22	56.78	✓
2	Guide 2	CTTTCGCCTCTAGCGACTGG (TGG)	Fwd	94	48 (12 in genes)	94	54	41	/	-
3	Guide 5	TAAAAATGCGTCGAGCTCTG (AGG)	Rev	82	82 (7 in genes)	84	67	77	100	✓
4	Guide 6	AAGAGGCGCGGGTAGAAGCG (GGG)	Fwd	82	102 (19 in genes)	78	45	137	/	✓
5	Guide 12	GAAAGCCCGACACCCAGCTT (CGG)	Rev	74	126 (22 in genes)	76	52	101	/	✓
6	Guide 4	GCTGTTTGGGGTTCGGCTGC (CGG)	Fwd	84	148 (33 in genes)	/	/	/	81.94	-
7	Guide 8	CTGTTTGGGGTTCGGCTGCC (GGG)	Fwd	80	146 (27 in genes)	/	/	/	85.56	-
8	Guide 7	CTCATTTCTCTGACCGAAGC (TGG)	Fwd	81	157 (15 in genes)	87	53	117	/	-
9	Guide 121	GCTTCGGTCAGAGAAATGAG (AGG)	Rev	/	/	65	69	131	/	-
10	Guide 66	CTTCTACCCGCGCCTCTTCC (CGG)	Rev	/	/	/	/	/	74.91	-

Selected guide RNA sequences with corresponding predicted quality, efficiency and specify scores generated by the three online gRNA design tools (MIT, CRISPOR and Atum).

The ten gRNAs were then checked for off-target effects using additional online tools: CCTop CRISPR-Cas9 target online predictor and Off-Spotter (available online from http://crispr.cos.uni-heidelberg.de/ and https://cm.jefferson.edu/Off-Spotter/). To check for the number of off-target effects, the 177 bp sequence of *C9orf72* previously used to design the gRNAs, was inputted into these online tools with each of the 10 gRNA sequences individually. These tools returned alternative genomic sequences that the gRNAs could potentially bind to, along with the number of base pairs that were mismatched between these sequences and the gRNA. gRNAs that had one or two mismatches with other genomic sites were disregarded on the basis that these were more likely to have off-target effects as SpCas9 is known to have high tolerance for single or double base pair mismatches (Cho et al., 2014; Hsu et al., 2013). In contrast, those with more than two mismatches with other genomic sites were considered to be more specific gRNAs as research suggests that less than 5 % gRNAs with more than two mismatches will induce off-target effects (Anderson et al., 2015). After off-target analysis, four gRNAs were chosen for *in vivo* assessment based on their high efficiency scores and low predicted off-target binding. Two of the gRNAs targeted the forward strand (Guide 1 and Guide 6) and two targeted the reverse strand (Guide 5 and Guide 12).

4.2.2 Guide RNA testing

The ability of the four chosen gRNAs to target the desired genomic region downstream of the *C9orf72* repeat was next tested in iPSC. Separate populations of *C9orf72* patient-derived iPSCs were nucleofected with the SpCas9 protein and the different gRNA independently. One population of cells was only nucleofected with the SpCas9 protein to act as a 'no gRNA control'. Additionally, a commercially available positive and negative control were included in the gRNA testing process. The positive control gRNA targeted the hypoxanthine phosphoribosyltransferase (HPRT) gene

and the negative control gRNA was a computationally validated negative control designed to not target the HPRT gene.

The process for gRNA testing involved extracting DNA from the nucleofected cells and amplifying the targeted genomic regions by PCR. The PCR products were then used in a heteroduplex mobility assay, T7 endonuclease assay and sequenced, to verify correct targeting and efficient cutting at the targeted genomic site by the gRNAs and SpCas9 protein.

4.2.2.1 PCR amplification of targeted region

48 hours post nucleofection, cells were lysed, and DNA was extracted from each nucleofected iPSC population and the genomic region targeted by the gRNAs were amplified by PCR. For the gRNAs targeting the region downstream of the *C9orf72* repeat, the targeted region was amplified using a forward primer located between the repeat expansion and the region the gRNAs targeted, and a reverse primer located downstream of the region the gRNAs targeted.

For the positive and negative control gRNAs, the targeted region in the HPRT gene was amplified using primers provided in the Alt-R CRISPR-Cas9 human control kit (Integrated DNA Technologies). PCR products were confirmed to be the expected size – 669 bp for the HPRT targeted regions and 321 bp for the *C9orf72* targeted region by electrophoresis (Figure 4.1).
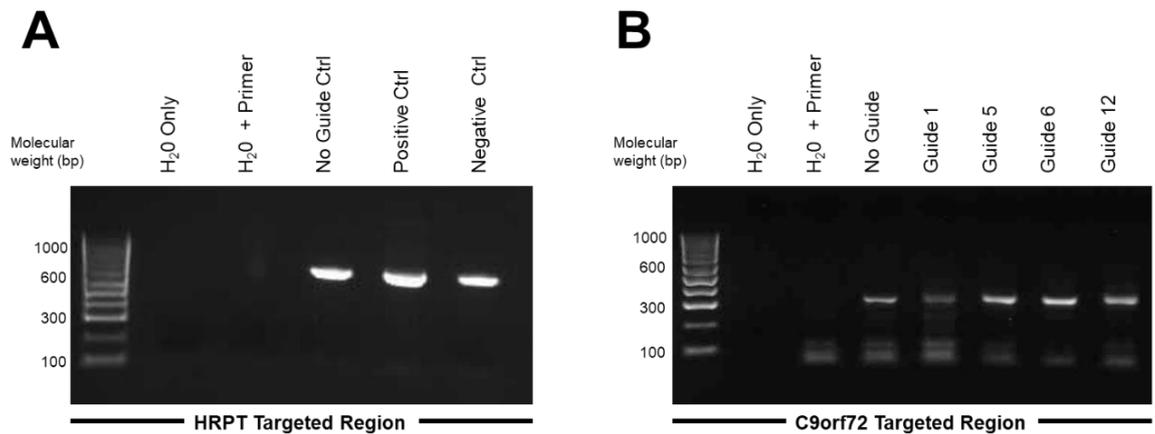
Figure 4.1 PCR amplification products of SpCas9 targeted genomic region

A) PCR amplification products of the SpCas9 targeted region in the HRPT control gene. Predicted product size = 669 bps. **B**) PCR amplification products of the SpCas9 targeted region in the *C9orf72* gene. Predicted product size = 321 bps.

4.2.2.2 Heteroduplex mobility assay

Following PCR amplification, PCR products were denatured and slowly reannealed, allowing for the formation of homoduplex and heteroduplex DNA products which were then exploited in the heteroduplex mobility assay to determine whether the gRNAs had targeted the relevant genomic regions correctly.

Homoduplex DNA is created when the two strands of DNA base-pair match perfectly, whilst heteroduplexes are created when there is a mismatch between the two strands of DNA. Following gRNA guidance and Cas9 cleavage of DNA, the DNA is rapidly repaired and, in some instances, insertions or deletions (indels) of DNA sequences can occur. This means that, following Cas9 cleavage and DNA repair, not all strands of DNA in the PCR products will perfectly match, creating heteroduplex DNA.

Homoduplex and heteroduplex DNA have different physical properties, such as melting temperatures and mobility in PAGE gels, which can be exploited to distinguish between the two DNA species. In the heteroduplex mobility assay, the angle between

matched and mismatched DNA strands caused by an indel means that heteroduplex DNA migrates at a significantly lower rate than homoduplex DNA under native conditions (Zischewski et al., 2017). The two forms can therefore be easily distinguished based on their mobility in a PAGE gel. The homoduplex DNA will run at the predicted size of the PCR product, whilst the heteroduplex DNA will run at a higher molecular weight.

In the heteroduplex mobility assay performed on the HPRT positive and negative controls, homoduplex DNA can be detected at the predicted size of 669 bp in all conditions. Higher molecular weight heteroduplex DNA is only detected in the positive control, indicating that indels have occurred (Figure 4.2 A). As expected, no higher molecular weight complexes were observed above the 669 bp band in the negative control or the no guide control. The absence of the higher molecular weight complexes indicates no heteroduplexes were formed and suggests no insertions or deletions occurred in these conditions.

Similarly, in the heteroduplex mobility assay performed on the PCR product of the *C9orf72* targeted region, homoduplex DNA is observed at the expected molecular weight of 321 bp across all conditions, but heteroduplex DNA, running at higher molecular weights, is only observed in the conditions that included gRNAs (Figure 4.2 B). This suggests that all four gRNAs are targeting the SpCas9 protein to the correct genomic region, resulting in indels in the repaired DNA. The lack of heteroduplex DNA in the no guide control confirms that the SpCas9 protein is actively cutting the DNA in the targeted region only in the presence of a gRNA. Products at lower molecular weights are thought to be primer dimers, as they also appear in the 'H$_2$O + Primer' control, which does not contain any DNA.
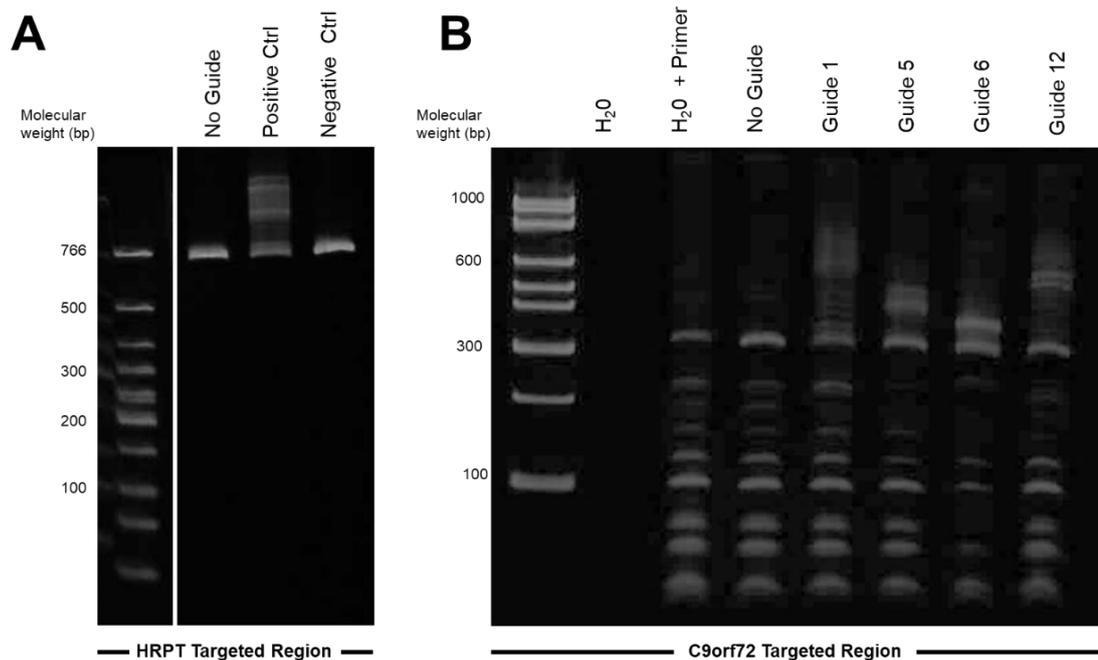
Figure 4.2 Heteroduplex mobility assay on PCR products from SpCas9 targeted genomic region

Homoduplex and heteroduplex PCR products of the **A**) HRPT targeted region and **B**) *C9orf72* targeted region. The homoduplex PCR products run at the predicted weight of (**A**) 669 bps or (**B**) 321 bps. Where SpCas9 has successfully cleaved the targeted region, heteroduplex PCR products run at a higher molecular weight than the homoduplex products.

4.2.2.3 T7 Endonuclease I assay

A second assay performed to check that the gRNAs are targeting the desired genomic region was the T7 endonuclease I assay. PCR products were again denatured and slowly reannealed, allowing for the formation of homoduplex and heteroduplex DNA products. The products were then incubated with T7 endonuclease I; an enzyme that recognises and cleaves non-perfectly matched DNA. Following T7 endonuclease I incubation, the products are run on a PAGE gel and any cleaved products will run at a lower molecular weight than the predicted size of the PCR product. Assuming the gRNA only allows the SpCas9 protein to produce a double strand break in the DNA once, the predicted size of the two T7 endonuclease I cleavage products can be calculated based on where the gRNA would target in the PCR product.

In the T7 endonuclease I assay performed on the HPRT positive and negative controls, two cleavage products that correspond to the predicted cleavage product sizes (198 bp and 471 bp) can be seen in the positive control only. These products had a lower molecular weight than the homoduplex PCR product (669 bp), and the higher molecular weight heteroduplex PCR products (> 669 bp). In contrast, no cleavage products were seen in the no guide control or the negative control, indicating that the SpCas9 protein did not cut the DNA and induce indels in these conditions (Figure 4.3).

In the T7 endonuclease I assay performed on the PCR product of the *C9orf72* targeted region, the no guide control showed the predicted PCR product at 321 bp and a few smaller products at lower molecular weights. These are likely a result of primer-dimer formation and does not indicate T7 endonuclease I cleavage. The presence of the primer-dimers on the gel makes it difficult to interpret where the correct cleavage products are for each of the gRNAs. However, Guide 6 and Guide 12 both had additional lower molecular weight bands corresponding to the correct size of their predicted cleavage products (Figure 4.3). Guide 1 and Guide 5 had multiple bands in addition to the correct size of their cleavage products, making it difficult to deduce the specificity of SpCas9 cleavage (Figure 4.3), therefore these guides were disregarded as potential gRNAs in future experiments. Of the two remaining gRNAs, Guide 6 was chosen for further analysis because it targets a region closer to the end of the *C9orf72* repeat expansion sequence than Guide 12.

A

Molecular weight (bp): 766, 500, 300, 200, 100, 50

No Guide | Positive Ctrl | Negative Ctrl

━ **HRPT Targeted Region** ━

Guide	Product 1	Product 2
Ctrl Guide	198 bp	471 bp

B

Molecular weight (bp): 766, 500, 300, 200, 100, 50

No Guide | Guide 1 | Guide 5 | Guide 6 | Guide 12

━━ **C9orf72 Targeted Region** ━━

Guide	Product 1	Product 2
Guide 1	133 bp	188 bp
Guide 5	57 bp	264 bp
Guide 6	46 bp	275 bp
Guide 12	101 bp	201 bp
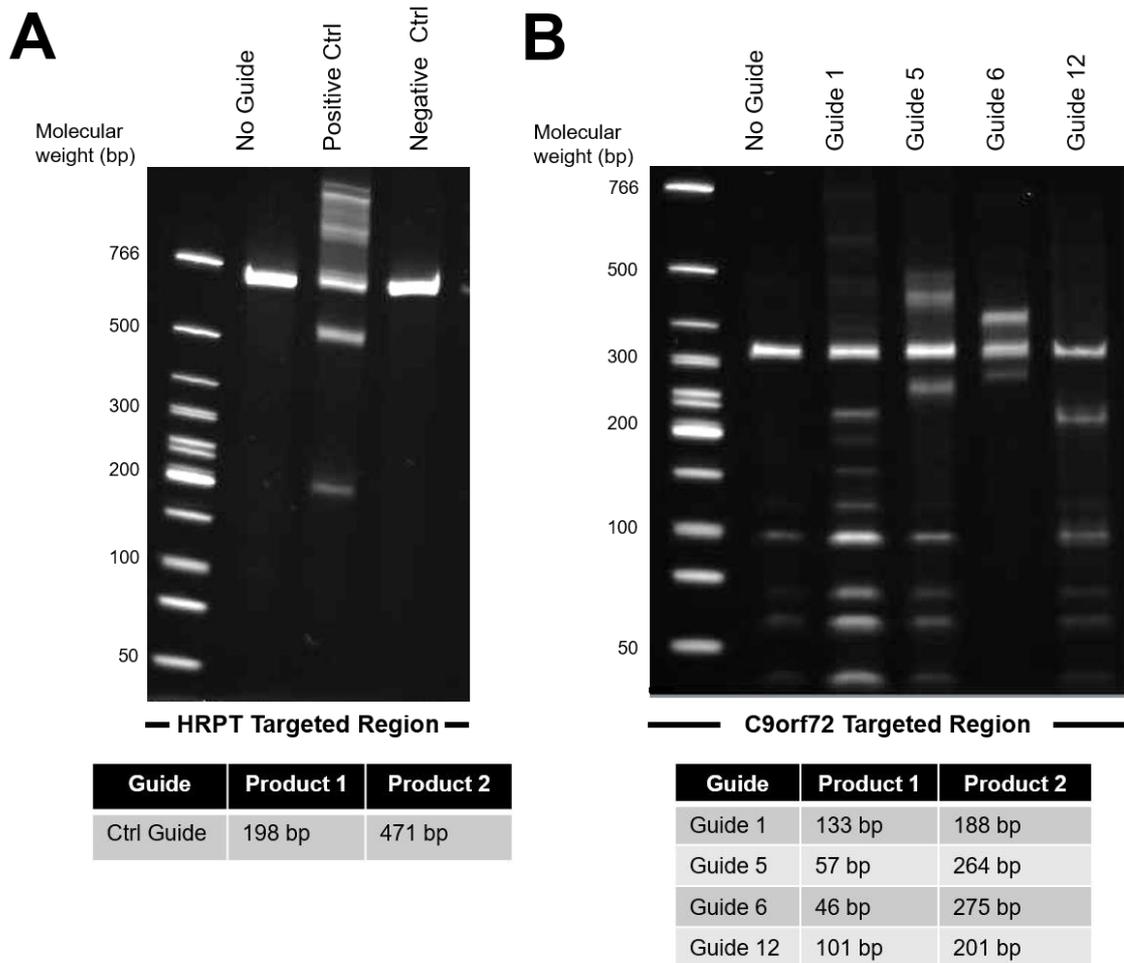
Figure 4.3 T7 endonuclease I assay on PCR products from SpCas9 targeted genomic region

PCR products of the **A**) HRPT targeted region and **B**) *C9orf72* targeted region, following cleavage by T7 endonuclease I. PCR products run at the predicted weights of (**A**) 669 bps or (**B**) 321 bps. T7 endonuclease I cleaved products run at lower molecular weights than the predicted PCR products. Heteroduplex run at a higher predicted weight than the predicted PCR products. The predicted size of the two cleavage products for each gRNA are listed in the tables below the gel images.

4.2.2.4 Sequencing of targeted region

To confirm correct targeting of the Guide 6, the PCR product was sequenced and compared to the sequencing of the no guide control product (Figure 4.4). For the no guide control, the sequencing matched the predicted sequence of the PCR product and the sequencing produced clean peaks on the chromatogram. In contrast, the sequencing of the Guide 6 product only matched the predicted sequence of the PCR product up until the predicted PAM site for Guide 6 (CGG). After this point, the sequencing reaction failed, as shown by the drop off in peaks on the chromatogram beyond the region that is targeted by Guide 6. The sequencing failure represents an inability to accurately distinguish bases within the sequence due to the multiple, random insertions or deletions that have occurred DNA following SpCas9 cleavage and DNA repair. The sequencing result confirmed that Guide 6 was targeting the SpCas9 protein to the desired region within the *C9orf72* gene and therefore, this gRNA was used for all future experiments.
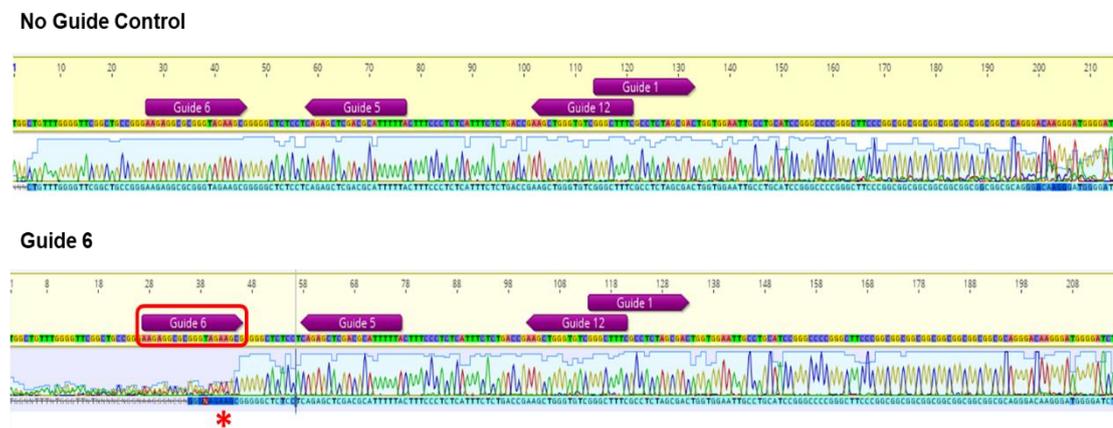
Figure 4.4 Sequencing chromatograms of PCR products from SpCas9 targeted C9orf72 genomic region in the presence of guide 6 or no guide RNA control

Screenshots of the chromatograms produced from sequencing the PCR amplification products of the targeted *C9orf72* genomic region where SpCas9 was not guided or guided by guide 6. The binding locations of each gRNA are shown in the purple arrows. Sequencing of the no guide control product showed an exact match to the predicted PCR product sequence. Sequencing of the guide 6 product failed beyond the predicted PAM site for guide 6 (CGG) as shown by the drop off in peaks on the chromatogram beyond the region that is targeted by Guide 6.

4.2.3 Donor DNA design

Three different ssDNA sequences were designed to act as repair templates during HDR. Each sequence was designed to insert monomeric eGFP in frame with a different DPR protein that is translated from the sense strand (poly-GA, poly-GR or poly-GP), whilst preventing the translation of the other two DPR proteins. The ssDNAs were named eGFP-GA, eGFP-GR and eGFP-GP, depending on which DPR protein they were designed to tag.

Each ssDNA designed included approximately 100 bp of 5' homology arm, 714 bp of eGFP coding sequence, and 100 bp of 3' homology arm. The sequences of the ssDNAs can be found in Section 0. When designing the ssDNAs, various factors had to be considered.

In the 5' homology arm of the ssDNA, any stop codons that were in frame with the targeted DPR protein between the end of the C9orf72 repeat expansion and the beginning of the eGFP sequence had to be mutated, to allow protein translation to continue beyond the DPR protein and into the eGFP sequence. Conversely, if they were not already present, stop codons were inserted in the two reading frames that were not being targeted. This was to ensure that only one DPR protein product would be tagged with eGFP.

To ensure that the eGFP tag is translated by RAN translation, the ATG start codon of the eGFP was removed. This was to ensure that any eGFP signal seen in CRISPR-engineered cells will not be a result of canonical ATG-initiated translation.

In the 3' homology arm of the ssDNA, the PAM site recognised by SpCas9 (5'-NGG-3') was mutated so that it can no longer by recognised by the SpCas9 protein. This means that, once the ssDNA has been inserted by HDR, any SpCas9 protein

remaining in the cell is unable to create a double-stranded break in the DNA at the targeted site again, preventing further DNA insertion or deletions.

4.2.4 Donor DNA synthesis

At the time of starting this project, commercial synthesis of ssDNA oligonucleotides longer than 200 bps was unavailable. Therefore, to create the designed ssDNA sequences of over 900 bps, these oligonucleotides were synthesised in-house using *in vitro* transcription reverse transcription (ivTRT) (Miura et al., 2015). During this process, the sequence of the desired ssDNA was synthesised commercially as a double-stranded block of DNA (Integrated DNA Technologies). This was then *in vitro* transcribed to create an RNA product, which was used as the template for reverse transcription. Following synthesis of cDNA from reverse transcription, the RNA was degraded using RNaseH, leaving behind a single strand of DNA which can be used as a ssDNA template in HDR (Figure 4.5).
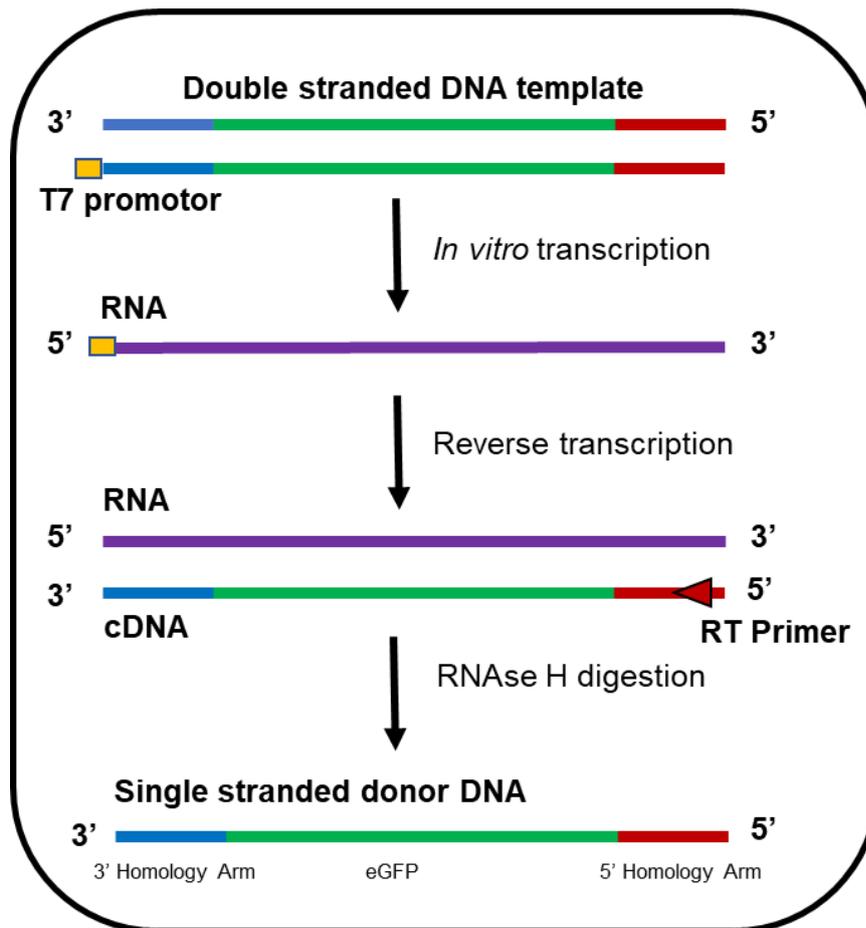
Figure 4.5 ssDNA synthesis by ivTRT strategy

Schematic of ssDNA synthesis by ivTRT. Synthetic double stranded DNA containing the desired ssDNA sequence and a T7 promoter (yellow box) is used as a template for in vitro transcription of RNA. cDNA (ssDNA) is then synthesized by reverse transcription of the RNA using gene-specific primers. The RNA template is degraded by RNaseH.

4.2.4.1 In vitro transcription

In vitro transcription using T7 RNA polymerase was performed on commercially synthesised double-stranded DNA blocks containing the sequence designed to insert eGFP in frame with poly-GP and poly-GR under the transcriptional control of the T7 promoter. As a control, FLuc control template DNA containing the firefly luciferase gene, also under the transcriptional control of the T7 promoter, was *in vitro* transcribed.

To evaluate transcript integrity and confirm the transcripts were the correct length, one tenth of each transcription reaction was analysed by gel electrophoresis (Figure 4.6). Both the eGFP-GP and eGFP-GR transcription products appeared at the predicted size of ~ 1 Kb. The FLuc control transcription products was smaller than the predicted size of 1.8 Kb, however the presence of a product indicated the *in vitro* transcription reaction was working correctly. All three transcription products had a slight smear below the expected transcript length. This smear likely represents contamination with RNase. To reduce this contamination, the RNA transcripts were purified using an RNA clean-up kit (Qiagen), prior to use in reverse transcription reactions.
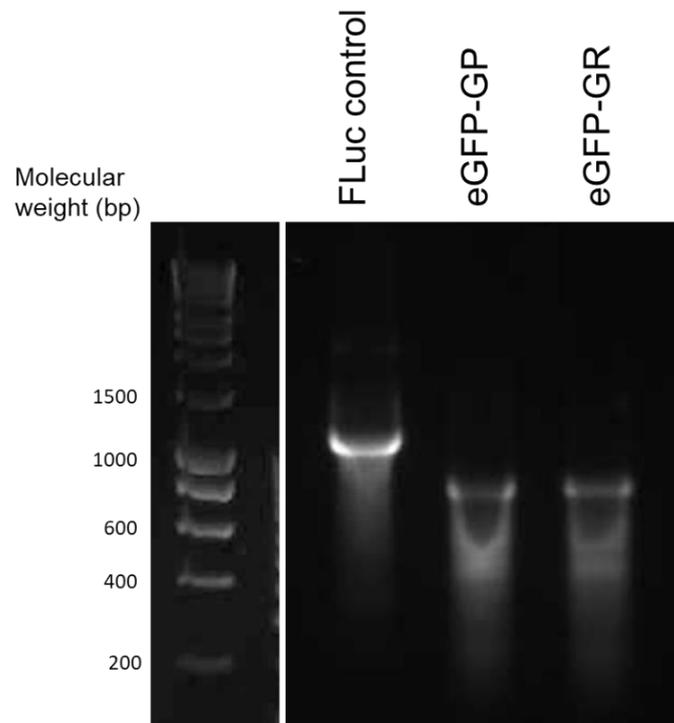
Figure 4.6 in vitro transcription products

RNA products from in vitro transcription of the synthetic double stranded DNA sequences encoding the firefly luciferase gene (FLuc control), eGFP-GP donor DNA or eGFP-GR donor DNA.

4.2.4.2 Reverse transcription

Following production of RNA transcripts containing the eGFP-GP and eGFP-GR ssDNA sequences, these products were used as a template for cDNA synthesis by reverse transcription. Reverse transcription was performed using Superscript III reverse transcriptase with primers specific for the first 18 nucleotides of the eGFP-GP or eGFP-GR donor template RNA. The *in vitro* transcription product of the FLuc control template was used as a control. Following reverse transcription, the cDNA products were incubated with RNase H to remove any remaining RNA template.

The reverse transcription reaction was analysed by gel electrophoresis (Figure 4.7). In the lanes containing eGFP-GP and eGFP-GR a band of approximately 1 kb was detected, which corresponds to the predicted size of the eGFP-GP and eGFP-GR ssDNA donor templates. No products were observed in the FLuc controls, indicating specific reverse transcription of the desired products only in the eGFP-GP and eGFP-GR reactions. The smear in each lane likely reflects template RNA remaining in the sample that has not been degraded by RNase H.

eGFP-GP and eGFP-GR ssDNA donor templates were purified from the gel. The products were quantified and sequenced. Unfortunately, the yield of the ssDNA donor templates was much lower than what was required for a subsequent nucleofection into iPSCs. Additionally, all sequencing reaction attempts failed, so the sequence of the ssDNA donors was unable to be verified. This meant that an alternative method was required to generate the desired ssDNA donor templates.
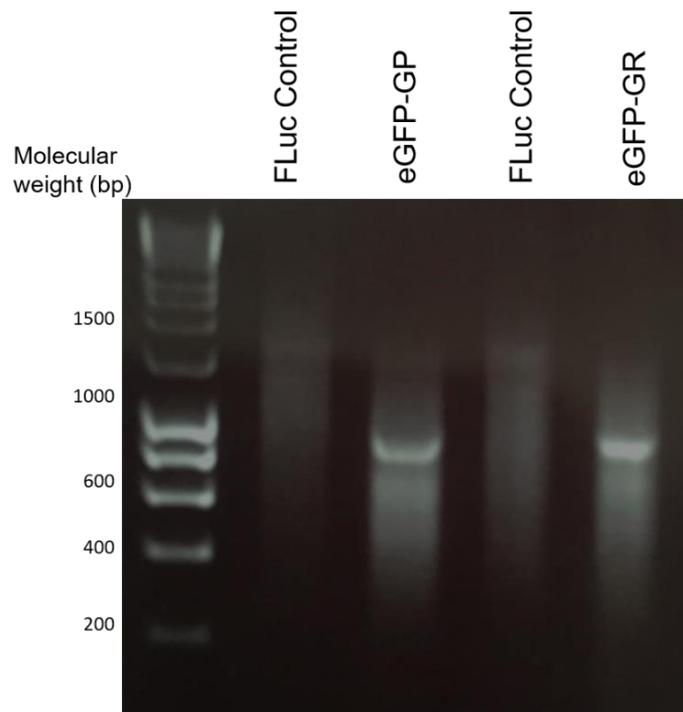
Figure 4.7 Reverse transcription products

cDNA products from reverse transcription of RNA encoding the the firefly luciferase gene (FLuc control), eGFP-GP donor DNA or eGFP-GR donor DNA.

4.2.5 Commercial synthesis of donor DNA

During the process of attempting to synthesise the ssDNA oligonucleotides by *iv*TRT, Integrated DNA Technologies introduced a service to commercially synthesise long ssDNA sequences that were sequence-verified and with a guaranteed final yield of 3 µg. Given that the in-house production of the ssDNA oligonucleotides was producing yields much lower than what was needed, and the products could not be sequenced, it was decided that it was more cost and time efficient to obtain the ssDNA templates from Integrated DNA Technologies. The ssDNA donor templates were produced according to the sequences designed in section 4.2.3.

4.2.6 Generation of eGFP-GR and eGFP-GP *C9orf72*-iPSC

Once the targeting of the gRNA had been verified and the ssDNA donor templates had been synthesised, these two components were used in combination with SpCas9 recombinant protein to insert eGFP downstream of the *C9orf72* repeat expansion in iPSCs. *C9orf72* patient-derived iPSCs were nucleofected with a pre-assembled ribonucleoprotein complex consisting of: 45 µM Guide 6 gRNA, 45 µM SpCas9 nuclease protein and 1 µg of either the eGFP-GR or eGFP-GP donor ssDNA. As a control, one population of cells were nucleofected without the SpCas9 nuclease. This population of cells was termed the 'nucleofection control population'.

Following nucleofection, the iPSCs were expanded and this population of cells was termed the 'Mixed Population', referring to that fact that some of the cells within the population will contain the eGFP sequence in their genome depending on whether they have undergone HDR following SpCas9 cleavage. From this population, individual cells were picked and expanded into clonal populations that were later screened for the eGFP insertion (Figure 4.8).
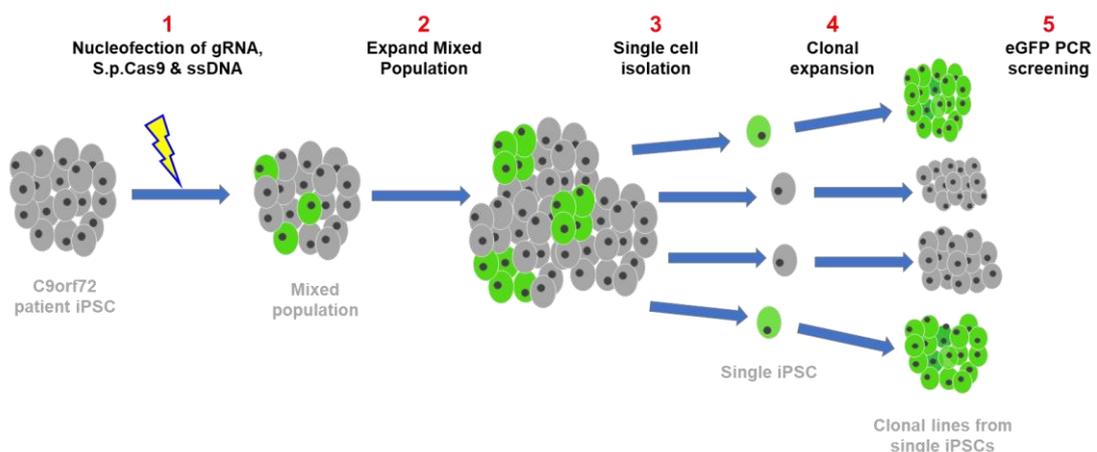
Figure 4.8 Generation of eGFP positive C9orf72-iPSC clones using CRISPR-Cas9 technology

1 – *C9orf72* patient-derived iPSC are nucleofected to insert the gRNA, SpCas9 and ssDNA template into the cells, generating a mixed population where some cells will have eGFP insertion. 2 – The mixed population is expanded to generate more eGFP positive cells. 3 – Single iPSC are selected from the mixed population. 4 – The single cells are expanded to form several clonal lines. 5 – The clonal iPSC lines are screened using PCR for the eGFP insertion.

4.2.7 Confirmation of eGFP insertion in mixed population

Before selecting individual colonies to screen for eGFP, the 'Mixed Population' of cells was screened for eGFP insertion by PCR. DNA was extracted and purified from a sub population of cells in the 'nucleofection control population', the 'eGFP-GP Mixed Population' and the 'eGFP-GR Mixed Population'. This DNA was used as a template for a PCR which was designed such that the forward primer was located in the eGFP sequence and the reverse primer was located in the genomic region downstream of the 3' homology arm. This ensured that a PCR product would only be produced if a) the eGFP sequence had been inserted and, b) it had inserted in the correct location in the genome (Figure 4.9 A).

Gel electrophoresis analysis of the PCR products showed that two separate samples of the eGFP-GR mixed population had the correct predicted product size of 542 bp (Figure 4.9 B). In contrast, the eGFP-GP population did not have a PCR product. These results indicated that HDR and insertion of the eGFP sequence had occurred in some cells in the 'eGFP-GR Mixed Population' (ssDNA designed to insert eGFP in frame with GR), but not in the 'eGFP-GP Mixed Population' (ssDNA designed to insert eGFP in frame with GP). Due to the negative result for eGFP-GP, all future experiments were carried out on the 'eGFP-GR Mixed Population' only.
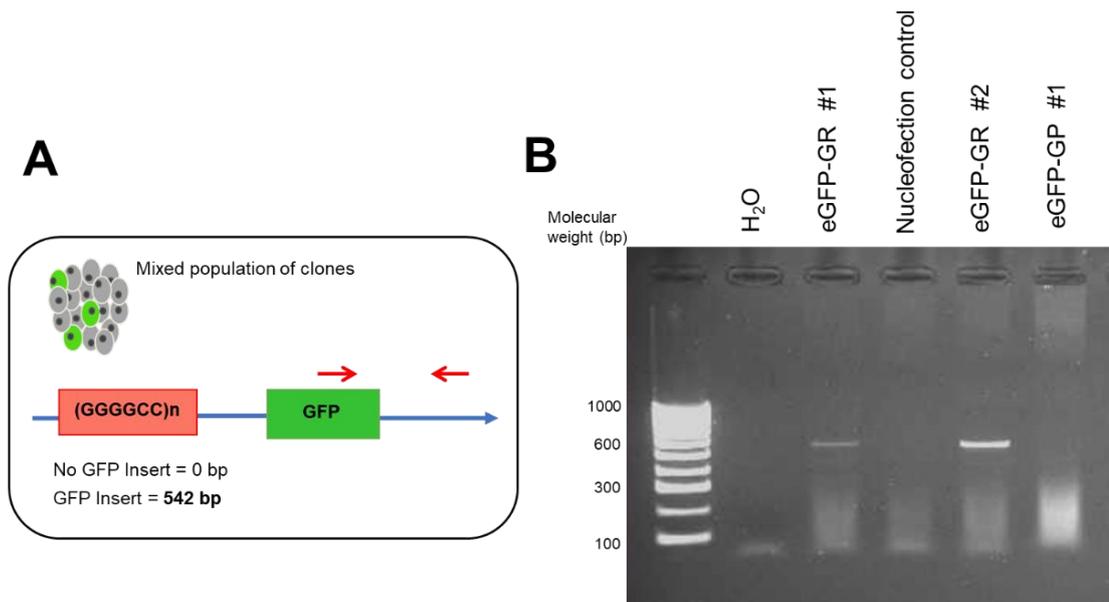
Figure 4.9 Confirmation of eGFP insertion with a mixed population of iPSCs

A) Schematic depicting the PCR strategy to detect eGFP in a mixed population of SpCas9 targeted iPSCs. The forward primer is in the eGFP sequence and the reverse primer is located downstream of the 3' homology arm. A 542 bp PCR product should only exist if eGFP is inserted in the correct position in the genome. **B**) PCR amplification products from iPSCs targeted by SpCas9 and template donor DNA. The predicted 542 bp product is present in the cells provided with the eGFP-GR donor DNA.

4.2.8 Single cell isolation and clonal expansion of CRISPR-Cas9 edited clones

After confirming the presence of eGFP in some cells in the eGFP-GR Mixed Population, the next step involved isolating the cells that contained the eGFP insertion by separating the cells into single cells followed by clonal expansion and screening.

One of the reasons eGFP was chosen as a tag was because it can be detected and used to sort cells into single cells by FACS. Unfortunately, no eGFP signal could be detected by FACS in the eGFP-GR mixed population, therefore an alternative method of generating single-cell derived colonies was used.

iPSCs from the eGFP-GR mixed population were separated into single cells by manual trituration and plated at a very low density. The plating of the iPSCs at a low

density, rather than alone as a single cell, was necessary because it is known that iPSCs have poor survival when plated alone. By plating the iPSCs as single cells but in a dish with other iPSCs, they can provide trophic support to each other, which improves survival.

Approximately 1 week after the low-density plating, the single cells had formed colonies that were big enough to be manually picked and plated into individual wells. Each colony was then expanded individually to form a number of clonal lines derived from the eGFP-GR mixed population. A total of 315 *C9orf72*-iPSC CRISPR-Cas9 edited clonal lines were generated.

4.2.9 eGFP colony PCR of *C9orf72*-iPSC clones

To determine which of the 315 clonal lines had an eGFP insertion, colony PCR were performed. Each clonal line was grown in separate wells of 96-well plates and lysed when confluent. 1 µl of the cell lysate was used as a DNA template in a PCR where the forward primer was located in the eGFP sequence and the reverse primer was located in the genomic region downstream of the 3' homology arm. Clones that had insertion of the eGFP in the correct place in the genome, generated a product of a 542 bp. In contrast, clones that did not have eGFP insertion, either formed no product, or multiple faint products of varying sizes (Figure 4.10). These aberrant products are thought to be a mixture of primer-dimers and smaller products caused by the forward primer having no template to bind to (because they do not contain eGFP), and therefore, randomly annealing to incorrect genomic sequences, producing random products.

4.2.10 PCR to confirm eGFP insertion in selected *C9orf72*-iPSC clones

The initial colony PCR screening identified 9 clones that showed the presence of a PCR product of approximately 542 bp. These were the clones numbered 17, 58, 68, 69, 70, 87, 114, 138 and 159. All other clones were discarded, and the 9 positive clones were expanded in culture for further analysis.

Due to the fact the colony PCR was performed on DNA from cells that had been lysed crudely and therefore was contaminated with cell debris and cellular proteins, the eGFP PCR was repeated on fresh DNA extracted and purified from the 9 identified clones. This showed that only 4 of the 9 clones had the 542 bp product and were therefore positive for eGFP (Figure 4.11).
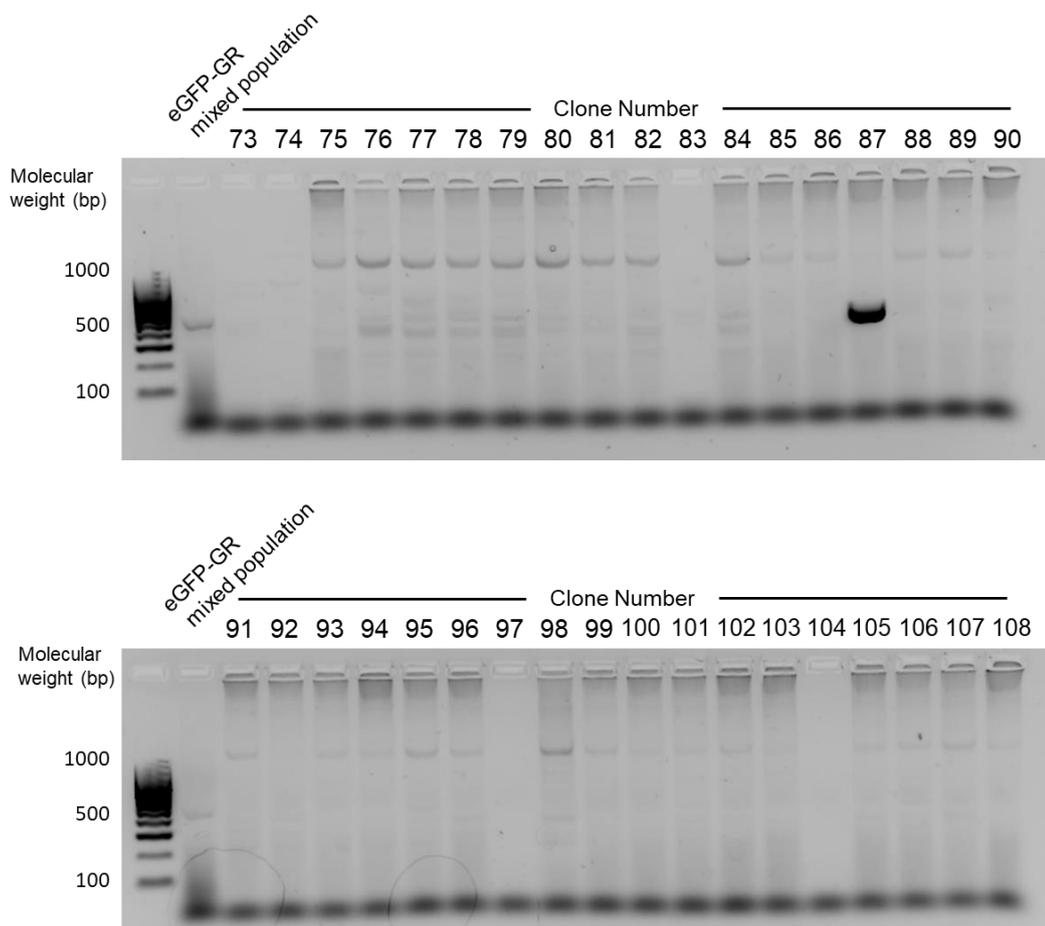
Figure 4.10 eGFP colony PCR screening of C9orf72-iPSC clones

Representative images of agarose gels showing the PCR amplification products of the eGFP PCR on the *C9orf72*-iPSC CRISPR-Cas9 edited clones. On these gels, only clone 87 and the eGFP-GR mixed population samples shows the presence of the 542 bp PCR product.
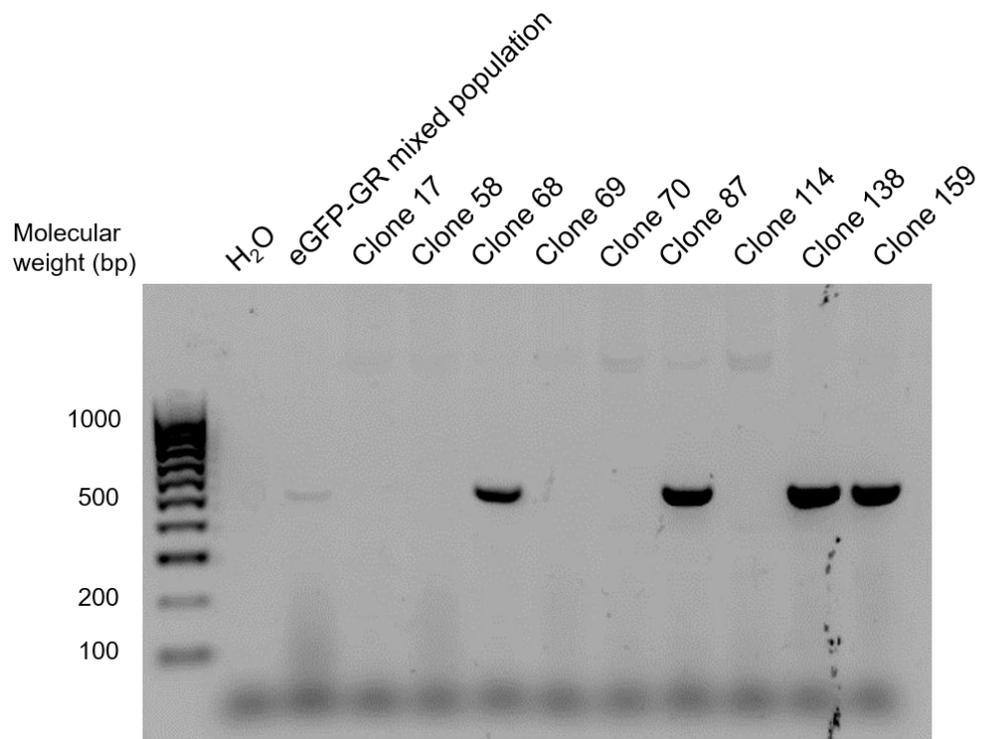
Figure 4.11 eGFP PCR screening of selected C9orf72-iPSC clones

Agarose gels showing the PCR amplification products of the eGFP PCR on the *C9orf72*-iPSC CRISPR-Cas9 edited clones that were selected for further screening. The eGFP-GR mixed population and clones 68, 87, 138 and 159 samples have the 542 bp PCR product.

4.3 Discussion

The results in this chapter describe the process of generating a CRISPR-Cas9 edited *C9orf72*-iPSC line, where eGFP has been inserted downstream of the repeat expansion to result in the production of C-terminally eGFP tagged poly-GR. This process involved designing and testing gRNAs, designing and synthesising three different single stranded donor DNA templates, inserting these components into *C9orf72*-iPSCs with SpCas9 to induced HDR, and isolating eGFP positive iPSC clones. These tasks were successfully achieved in this project, however further work is on-going in the Isaacs lab to confirm eGFP insertion and validate this model. Due to time-limitation, it was not possible to include the validation of the model in this thesis.

4.3.1 Confirmation of correct eGFP insertion

The next step in the process of generating this model will be to confirm that the eGFP sequence has been inserted in the allele carrying the *C9orf72* repeat expansion. The *C9orf72*-iPSC used in this study are heterozygous for the repeat expansion, and because the CRISPR-Cas9 targeting strategy in this study was not allele specific, it is possible that the eGFP sequence may have been inserted into the non-expanded allele. If this is the case, no eGFP will be detected because short GGGGCC repeats are not thought to be RAN translated, as is evident by the lack of DPR protein pathology in people with non-expanded *C9orf72* alleles (Ash et al., 2013; Mori et al., 2013c; Zu et al., 2013). To determine which allele the eGFP has been inserted into, a PCR strategy will be employed where the forward primer will be located upstream of the repeat expansion and the reverse primer located within the eGFP sequence. Due to the repetitive and GC rich nature of the *C9orf72* hexanucleotide repeat, the region is unable to be amplified by regular PCR. Therefore, if the eGFP sequence has inserted into the expanded allele, no PCR product will be expected because the

polymerase will be unable to read through the repeat expansion. In contrast, the non-expanded allele should have only a few copies of the GGGGCC repeat and should be short enough to allow for the PCR to be successful and produce a product. These PCR results in conjunction with the PCR that confirmed insertion of eGFP (Figure 4.11), would provide evidence to deduce which of the clonal iPSC lines have eGFP inserted into the expanded allele.

4.3.2 Screening for off-target effects

After confirmation of on-target eGFP insertion, it will also be necessary to confirm that eGFP has not been inserted into other sites in the genome. The most likely candidate regions for off-target effects were identified when designing and choosing gRNAs (Section 4.2). Although gRNAs with three or more mismatches with other regions of the genome were chosen, it is possible that these were able to bind to regions with similar sequence homology and guide the SpCas9 nuclease there for DNA cleavage. This could cause the insertion or deletion of sections of DNA in these regions during the DNA repair process, which may alter protein coding sequences or gene expression regulation sites. Although the frequency of off-target mutations is much lower than on-target mutations, when the cell has an intact DNA repair machinery (Hruscha et al., 2013; Veres et al., 2014; Yang et al., 2013), it is important to check the CRISPR-edited *C9orf72*-iPSCs for off-target effects prior to validation and phenotyping as any phenotypes identified may be a result of an off-target effect rather than a genuine effect of the cells expressing DPR proteins. Several methods have been developed to detect off-target effects, most of which rely on sequencing to detect mutations either in pre-selected regions or on a genome-wide scale (Zischewski et al., 2017). The method that will be used to check for off-target mutations for this project is PCR amplification and sequencing of pre-selected potential off-target sites previously identified using the Off-Spotter and CCTop CRISPR-Cas9 target online predictor algorithms (Section 4.2). However, this method

is biased towards one genomic locus and thus risks overlooking off-target effects at other genomic loci. Alternatively, whole-genome sequencing is an unbiased method of off-target mutation identification on a genome-wide scale. Although this method is expensive, the number of eGFP positive clones generated in this study is small and thus whole-genome sequencing may be worthwhile as it also enables screening for structural mutations, such as inversions, rearrangements, duplications and major deletions, which are unable to be detected by PCR based methods.

4.3.3 Validating poly-GR-eGFP protein expression

Once insertion of eGFP in the expanded allele has been verified and off-target effects ruled out, it will be necessary to demonstrate that eGFP-GR protein is being produced. The benefit of using a fluorescent tag is that this protein should be visible using a fluorescence microscope. However, given that no eGFP signal was detected during FAC sorting, it is possible that eGFP-GR will be unable to be detected by fluorescence microscopy in iPSCs. This could be because iPSCs may not produce, or have very little production of, DPR proteins. All previous studies that show detection of DPR proteins in *C9orf72*-patient derived cells using iPSC-derived neurons or iNeurons and do not demonstrate the presence of DPR proteins in the iPSCs themselves (Almeida et al., 2013; Dafinca et al., 2016; Donnelly et al., 2013; Lopez-Gonzalez et al., 2016; Simone et al., 2018; Su et al., 2014a; Westergard et al., 2016). It is possible that DPR proteins are only produced in mature neuronal cell types and it may, therefore, be necessary to differentiate the CRISPR-edited iPSCs into motor or cortical neurons before the eGFP-GR protein can be observed using fluorescence microscopy. Additional methods that will be employed to verify the production of eGFP-GR in either the edited *C9orf72*-iPSCs or *C9orf72*-iPSC-derived neurons will include western blotting and MSD immunoassays for both eGFP and poly-GR, which are established in the Isaacs lab (Simone et al., 2018).

Despite genetically confirming insertion of eGFP in the correct location downstream of the *C9orf72* repeat expansion, it may be possible that the C-terminally tagged poly-GR protein is not produced and, therefore, will not be detected by any of these methods. This could be because RAN translation is simply not occurring in the cells, or that RAN translation stops before the eGFP sequence is reached. To avoid this, the ssDNA templates were designed with stop codons removed between the repeat and eGFP sequence in the desired frame, however, given that relatively little is known about the mechanisms of RAN translation, and that the initiation of RAN translation does not utilise the canonical ATG start codon, it is possible that RAN translation also does not end at recognised canonical stop codons and may cease before reaching the eGFP sequence. If this occurs, poly-GR will be produced but will not be tagged with eGFP.

Finally, an additional reason why it may not be possible to detect eGFP-GR in these cells is that the eGFP tag may stabilise the poly-GR protein. Given that poly-GR is known to be one of the most toxic DPR proteins in several models (Kwon et al., 2014; Lopez-Gonzalez et al., 2016; Mizielinska et al., 2014), stabilisation of the poly-GR protein by eGFP may enhance its toxicity, causing cells that are expressing it to quickly die, leaving behind only the cells that are not expressing the eGFP-GR protein.

4.3.4 Generation of other eGFP tagged DPR protein *C9orf72*-iPSC lines

Following the generation and validation of the eGFP-poly-GR C9-iPSC line, the developed and optimised CRISPR-Cas9 protocol will be used to generate both eGFP-poly-GA and eGFP-poly-GP *C9orf72*-iPSC lines. The gRNAs and ssDNA donor templates for these have already been designed within this project and, therefore, generation of these cell lines should be relatively quick using the CRISPR-Cas9 protocol optimised in this project.

4.3.5 Future experiments using the DPR eGFP tagged C9orf72-iPSC lines

Once generated, the CRISPR-Ca9 modified *C9orf72*-iPSC lines expressing eGFP tagged endogenous DPR proteins will become an incredibly useful tool for investigating DPR protein toxicity in cell types affected in *C9orf72* FTD/ALS, such as cortical or motor neurons.

Several possibilities for future experiments with these cells exist. For example, live fluorescence longitudinal imaging could be performed to monitor the production, localisation and behaviour of the eGFP-tagged DPR proteins over time. This will provide spatial and temporal information of dynamic cellular events, rather than giving one time-point 'snapshots' that are generally obtained from immunofluorescence or immunoblotting methods. This could be particularly useful when studying poly-PR and poly-GR, as these proteins are known to be difficult to detect in neurons derived from *C9orf72* patient iPSCs and are rarely detected in post-mortem tissue, suggesting that perhaps neurons containing these proteins have died before they have been able to be studied. eGFP tagging of endogenous poly-PR and poly-GR will therefore allow for real-time studying of these proteins from production to cell death.

The eGFP tag could additionally be used to co-immunoprecipitate the tagged DPR proteins with their interacting proteins, which can then be identified by mass spectrometry. Deciphering the DPR protein interactome may provide more insight into the mechanisms by which the proteins induce toxicity. Several groups have previously published the interactomes of various DPR proteins (Kanekura et al., 2016; Lee et al., 2016; Lin et al., 2016; Lopez-Gonzalez et al., 2016; Yin et al., 2017), however these studies are based on the overexpression of the DPR proteins in human stable cell lines thus, the protein-protein interactions identified may not fully recapitulate the interactome of the endogenous levels of the DPR proteins in motor or cortical

neurons. Therefore, the generation of this CRISPR-Cas9 edited *C9orf72*-iPSC model will allow for interactome analysis of DPR proteins at endogenous levels in cell types affected by *C9orf72* FTD/ALS, such as iPSC-derived motor or cortical neurons.

Finally, the eGFP tag could be used to selectively knockdown the DPR proteins, in a process termed deGradFP (Caussinus and Affolter, 2016). This method harnesses the ubiquitin-proteasome pathway to achieve fast, selective depletion of eGFP-tagged proteins, and has been successfully used in mammalian cells and *Drosophila* models to knockdown selected eGFP-tagged proteins (Caussinus et al., 2011, 2013; Nagarkar-Jaiswal et al., 2015). In this study, the deGradFP could be employed to degrade an eGFP-tagged DPR protein and the effects of this could be studied. If a particular phenotype is observed in the *C9orf72* iPSC derived neurons compared to controls, it could be investigated whether degradation of certain DPR proteins rescues or enhances this phenotype. Possible phenotypes that have been observed in *C9orf72* iPSC derived neurons published by other groups include formation of RNA foci and production of RAN translated DPR proteins, increased sensitivity to cellular stress and DNA damage, and changes in electrophysiological properties of the neurons (Almeida et al., 2013; Devlin et al., 2015; Lopez-Gonzalez et al., 2016; Sareen et al., 2013).

4.3.6 Conclusion

In summary, this chapter describes the processes involved in the design and generation of the first endogenous eGFP-tagged DPR protein *C9orf72*-iPSC line. CRISPR-Cas9 genome engineering has been successfully used to edit the genome of a *C9orf72* ALS patient iPSC line to insert an eGFP tag in frame with the poly-GR DPR protein downstream of the hexanucleotide repeat expansion. Although further validation and characterisation of these cells is required, it is hoped that many research opportunities will result from the generation of this CRISPR-Cas9 modified

iPSC line, and that it will become a valuable tool for future investigations exploring the mechanisms of toxicity linked to DPR proteins in the pathogenesis of *C9orf72* FTD/ALS.

Chapter 5. Pathological assessment of arginine methylation of *C9orf72* dipeptide repeat proteins

5.1 Introduction

The methylation of arginine residues within a protein is a common post-translational modification that can have a significant impact on the localisation and behaviour of proteins. Given that the poly-GR and poly-PR proteins contain a significant amount of arginine residues and these DPR proteins have been consistently shown to be the most toxic DPR protein species in a number of model systems (Jovicic et al., 2015; Kwon et al., 2014; Lee et al., 2016; Mizielinska et al., 2014; Rudich et al., 2017; Tao et al., 2015; Wen et al., 2014; Xu and Xu, 2018), it is possible that the methylation of arginine residues within these proteins could play a significant role in their behaviour and influence toxicity. This chapter explores whether arginine methylation is a post-translational modification of the poly-GR protein in inclusions in *C9orf72* patient post-mortem brain tissue.

5.1.1 Protein arginine methylation

Arginine methylation is a post-translational modification that involves the addition of methyl groups to arginine residue within a protein. It is a highly prevalent post-translational modification found on both nuclear and cytoplasmic proteins and is thought to be as common as protein phosphorylation and ubiquitination in mammalian cells, with 0.7 – 1 % of the total arginine residues reported to be methylated (Bulau et al., 2006; Larsen et al., 2016; Uhlmann et al., 2012).

Arginine methylation is catalysed by the protein arginine methyltransferases (PRMTs) family of enzymes. PRMT enzymes catalyse the transfer of a methyl group (CH_3-) from S-adenosylmethionine (SAM) to the guanidino nitrogen atoms of arginine. Three distinct forms of methylated arginine residues can exist, depending on where the methyl groups are positioned on the nitrogen atoms of an arginine side chain (Blanc and Richard, 2017). The addition of a single methyl group to the terminal nitrogen atom produces monomethylarginine (MMA). This molecule can then be further modified by the addition of a second methyl group resulting in the formation of dimethylarginine. The methyl groups can be added in either an asymmetric or symmetric composition, depending on whether the methyl groups have been added to the same or different nitrogen atoms, respectively. Asymmetric dimethylarginine (ADMA) is the most prevalent of the three distinct types of arginine residue methylation, whilst MMA and symmetric dimethylarginine (SDMA) typically occur at levels of about 20 % to 50 % that of ADMA (Bedford and Clarke, 2009).
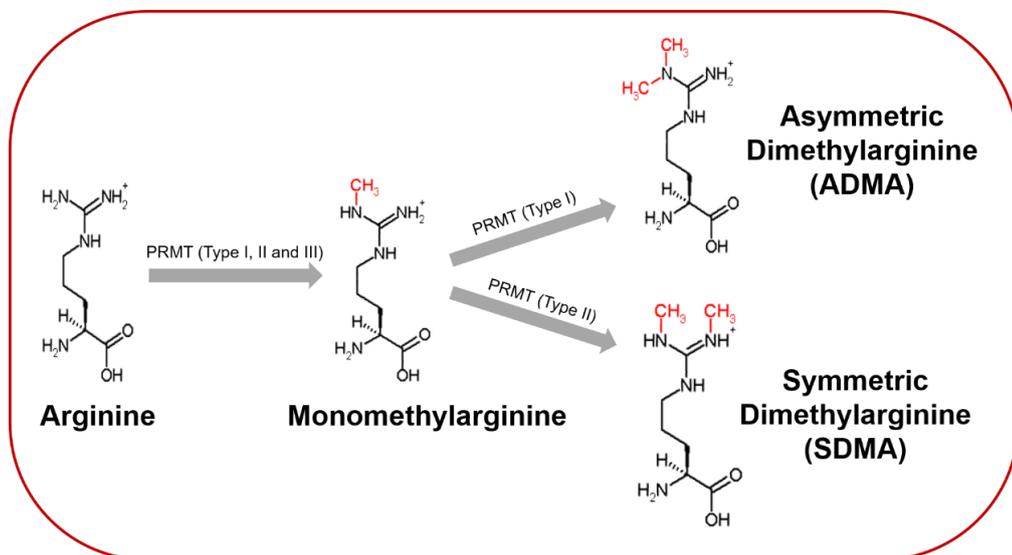
Figure 5.1 Process of protein arginine methylation

Type I, II and III protein arginine methyltransferases (PRMTs) catalyse the formation of a monomethylarginine intermediate by adding a methyl (-CH_3) group to a nitrogen atom within the positively charged guanidinium side chain of an arginine residue. This is followed by the addition of a second methyl group in a symmetric or asymmetric conformation, by Type II and Type I PRMTs respectively, to generate asymmetric dimethylarginine (ADMA) or symmetric dimethylarginine (SDMA).

The positive charge of an arginine residue is unaffected by the addition of methyl groups; however hydrogen bonding is reduced, which can affect intramolecular and intermolecular interactions (Lee and Stallcup, 2009). Arginine residues have five potential hydrogen-bond donors; thus, the addition of methyl groups alters interactions with hydrogen-bond acceptors of interacting partners, increases steric hindrance and increases overall hydrophobicity by decreasing hydrogen bonding (Fuhrmann et al., 2015; McBride and Silver, 2001). Arginine- and glycine- rich (RGG/RG) motifs, which function in both nucleic acid binding and in regulation of protein-protein interactions, are one of the most common amino acid sequences targeted by PRMTs (Thandapani et al., 2013). Consequently, the main biological effect of arginine methylation is alteration to a protein's interaction with its binding partners, such as other proteins, RNA and DNA.

Arginine methylation plays a significant role in a range of cellular processes, including transcriptional regulation, nucleocytoplasmic shuttling, mRNA splicing, DNA repair, and signal transduction (Bedford and Clarke, 2009; Blanc and Richard, 2017; Peng and Wong, 2017). Its importance in the regulation of essential cellular functions is reinforced by the observation that aberrant arginine methylation has been implicated in several diseases, including cancer, cardiovascular disease, viral pathogenesis and neurodegenerative diseases (Bedford and Clarke, 2009; Boulanger et al., 2005; Greenblatt et al., 2016; Kim et al., 2003; Peng and Wong, 2017; Stuhlinger et al., 2001).

The importance of arginine methylation in the pathogenesis of FTD and ALS has recently come to light as it has been demonstrated that the methylation of arginine residues within the FTD/ALS-linked protein, FUS, is an important regulator of its liquid-liquid phase transition and subsequent aggregation (Hofweber et al., 2018; Qamar et al., 2018).

5.1.2 Regulation of liquid-liquid phase transition by arginine methylation

Several groups have demonstrated that the FUS protein undergoes extensive arginine methylation within its RGG/RG motifs (Du et al., 2011; Hung et al., 2009; Ong et al., 2004; Rappsilber et al., 2003). This has previously been demonstrated to be important in the nuclear import and localisation of the FUS protein (Dormann et al., 2012), however recent studies have additionally shown that arginine methylation of the RGG3 domain within FUS is crucial for reducing the phase separation of the protein by disrupting the cation-π covalent bonds between C-terminal arginines and N-terminal tyrosines (Hofweber et al., 2018; Qamar et al., 2018). This arginine methylation is thought to be a physiological mechanism by which cells regulate the formation of dynamic membrane-less organelles, such as stress granules, by phase transition. However, in FTLD-FUS, pathological FUS inclusions are known to be hypomethylated (Dormann et al., 2012), giving rise to the hypothesis that an impairment in arginine methylation may have promoted phase separation and subsequent gelation of the FUS protein in this disease (Hofweber et al., 2018; Qamar et al., 2018). Several groups have proposed that the pathological aggregation of FUS is initiated from stress granules (Bentmann et al., 2013; Dormann et al., 2010; Li et al., 2013). The aberrant methylation of the arginines regulating the phase transition process may be one mechanism by which this switch from stress granule to pathological aggregate occurs. Interestingly, FUS aggregates in ALS-FUS show normal arginine methylation, implying that separate pathogenic mechanisms are involved in FTLD-FUS and ALS-FUS (Dormann et al., 2012).

The ability of arginine methylation to alter phase transition has only recently been described, thus its importance in the aggregation of other proteins linked to FTD and ALS is yet to be explored. However, several proteins associated with these diseases, which are capable of liquid-liquid phase transition, contain RGG/RG-rich domains and

are known targets of arginine methylation (Altmeyer et al., 2015; Hung et al., 2009; Molliex et al., 2015; Ong et al., 2004), suggesting that perhaps the regulation of phase transition by arginine methylation may be a mechanism common to these proteins. Potential candidate proteins include some members of the hnRNP A family, which aggregate in rare familial forms of ALS (Kim et al., 2013b; Mori et al., 2013b), and EWS and TAF15, both of which accumulate in pathological inclusions in FTLD-FUS (Neumann et al., 2011).

Although they do not contain RGG/RG motifs, the arginine-containing DPR proteins are also a potential candidate for arginine methylation given that every other amino acid in the protein is an arginine residue. Previous studies have demonstrated that the arginine-containing DPR proteins are able to undergo the process of liquid-liquid phase transition (Boeynaems et al., 2017; Lee et al., 2016; Lin et al., 2016), however little work has been carried out to determine if these proteins are methylated and whether their aggregation could be influenced by arginine methylation.

5.1.3 Arginine methylation of *C9orf72* arginine-containing DPR proteins

The post-translational modifications of DPR proteins in *C9orf72* patient brain tissue has not been extensively studied. To date, only three studies have investigated arginine methylation in relation to the arginine-containing DPR proteins (Boeynaems et al., 2016; Mori et al., 2013a; Sakae et al., 2018).

In one of the first published studies to describe DPR proteins in *C9orf72* post-mortem tissue, Mori *et al.* (2013) hypothesised that poly-GR aggregates may contain a mixture of unmethylated, SDMA and ADMA residues following immunohistochemical detection of DPR protein inclusions using antibodies designed to detect different arginine methylation states. It is not clear, however, whether these antibodies were

specific for the poly-GR protein or whether they were able to detect the methylation of other proteins which could be co-accumulating in the pathological inclusions.

Boeynaems *et al.* (2016) later identified a potential link between the toxicity of arginine-containing DPR proteins and arginine methylation when knockdown of four out of ten arginine methyltransferases enhanced the toxicity of poly-PR expression in *Drosophila*. One of these enzymes, PRMT1, was subsequently found to co-localise with both poly-PR and poly-GR in HeLa cells, suggesting that this enzyme may be involved in the methylation of these proteins. Furthermore, a commerically avaliable antibody, ASYM24, designed to detect asymmetric arginine dimethylation partially co-localised with accumulations of poly-GR, but not poly-PR, in transfected neuroblastoma cells. The detection of only one methylated arginine-containing DPR protein was explained by the fact that the ASYM24 antibody was originally raised against a peptide showing strong sequence similarity with poly-GR, but not poly-PR. To validate arginine methylation of poly-GR pathologically, immunohistochemistry was performed on brain tissue sections from *C9orf72* - FTLD patients using the ASYM24. This staining identified frequent inclusions with a similar morphology to DPR protein inclusions in the *C9orf72* patients, but not in control tissue, suggesting that arginine methylation of DPR protein could also occur *in vivo*.

Following this observation, Sakae *et al.* (2018) used the same ASYM24 antibody to extensively characterise the distribution of asymmetric arginine dimethylation in post-mortem brain tissue of 40 *C9orf72* patients classified into three different clinicopathological phenotypes; FTLD, FTLD-ALS, or ALS. The ASYM24 antibody labelled neuronal inclusions and many nuclei, the latter presumed to be detection of methylated nuclear proteins, such as histones. The inclusions detected had a similar cellular distribution to poly-GR inclusions with the majority being neuronal cytoplasmic inclusions, but sparse detection in glia and dystrophic neurites. Comparison of the

different clinicopathological subgroups revealed that the density of poly-GR and ASYM24 immunoreactive neuronal inclusions was greater in patients with FTLD-ALS compared to FTLD or ALS alone. Double immunofluorescence labelling found frequent co-localisation of poly-GR and ASYM24 in areas associated with *C9orf72* neurodegeneration, such as the frontal and motor cortices and the dentate gyrus of the hippocampus, suggesting that a significant portion of poly-GR inclusions may be methylated. The authors then went on to suggest that poly-GR aggregation may be modulated by this arginine methylation as they demonstrated that treating cells with adenosine dialdehyde (AdOx), a global methyltransferase inhibitor, decreased the formation of cytoplasmic aggregates of GFP tagged poly-GR consisting of 100 repeats in HEK 293 cells, while having no effect on the number of cells expressing GFP. If the toxicity of poly-GR arises from its aggregation into inclusions, then dimethylarginine modification of poly-GR could contribute to its toxicity.

Although the authors of both studies suggest that co-localisation of ASYM24 and poly-GR indicates that poly-GR undergoes asymmetric arginine methylation, it is important to note that the ASYM24 antibody is not specific for methylation of poly-GR and will recognise a broad range of proteins that are asymmetrically dimethylated. Therefore, it is impossible to determine whether the antibody is detecting the methylation of poly-GR, or the methylation of other proteins it associates with in pathological inclusions. Furthermore, these studies only provide insight into the asymmetric dimethylation modification of arginine and do not investigate symmetric arginine dimethylation. Nevertheless, these studies demonstrate that arginine methylation could be a significant post-translation modification of the arginine-containing DPR proteins and highlight a need for further study.

5.1.4 Chapter 5 aims

A number of studies indicate that arginine methylation can significantly alter the biophysical and biochemical properties of a protein. Given that the highly toxic arginine-containing DPR contain a high content of arginine residues, the presence of arginine methylation on these residues could significantly impact the toxicity of these proteins. Previous studies have suggested that poly-GR is likely to undergo asymmetric arginine methylation, but the specificity of this observation has been questioned and indicates a need for the generation of specific methylated-poly-GR antibodies. The current study utilises novel antibodies generated to specifically detect asymmetric and symmetric dimethylated forms of poly-GR in *C9orf72* patient post-mortem tissue. The frequency of inclusions containing these methylated forms of poly-GR was then correlated with various clinical information to determine whether arginine methylation of poly-GR has any influence on patient outcomes. The principle aims of Chapter 5 were to;

- Validate novel unmethylated-GR, ADMA-GR and SDMA-GR specific antibodies on *C9orf72* patient post-mortem tissue using immunohistochemistry.

- Quantify the number of ADMA-GR and SDMA-GR containing inclusions in the frontal and temporal cortices of *C9orf72* patient post-mortem brain.

- Determine whether ADMA-GR or SDMA-GR in the frontal and temporal cortices of *C9orf72* patient post-mortem brain significantly correlate with patient clinical information.

5.2 Results

5.2.1 Novel unmethylated and methylated poly-GR antibodies

To investigate the methylation specifically of poly-GR proteins, several new rabbit monoclonal antibodies were generated in collaboration with UCB. These included an antibody designed to target unmethylated-GR and two antibodies designed to target the two forms of dimethylated-GR. Each antibody was raised to a peptide antigen consisting of 8 repeats of glycine-arginine, with all the arginine residues being unmethylated, asymmetrically dimethylated or symmetrically dimethylated (Figure 5.2 A). These antibodies are henceforth referred to as unmethylated-GR, ADMA-GR and SDMA-GR, respectively.

To determine the specificity of each of these antibodies, quantitative ELISA analysis was performed by collaborators at UCB. This showed that each of the antibodies had high specificity for their respective peptide antigens. The unmethylated-GR antibody was highly specific for unmethylated poly-GR antigens and showed no affinity for methylated forms of poly-GR or poly-PR (Figure 5.2 B). Both ADMA-GR (Figure 5.2 C) and SDMA-GR (Figure 5.2 D) showed the highest affinity for their respective methylated poly-GR antigens, however they also showed some detection of asymmetrically dimethylated poly-PR (ADMA-PR) and symmetrically dimethylated poly-PR (SDMA-PR), respectively. Their affinity for methyl-PR, however, was much lower than for methyl-GR and, given that poly-PR is rarely detected in human post-mortem tissue by immunohistochemical methods, it was determined that these antibodies were unlikely to identify ADMA-PR and SDMA-PR in post-mortem tissue.
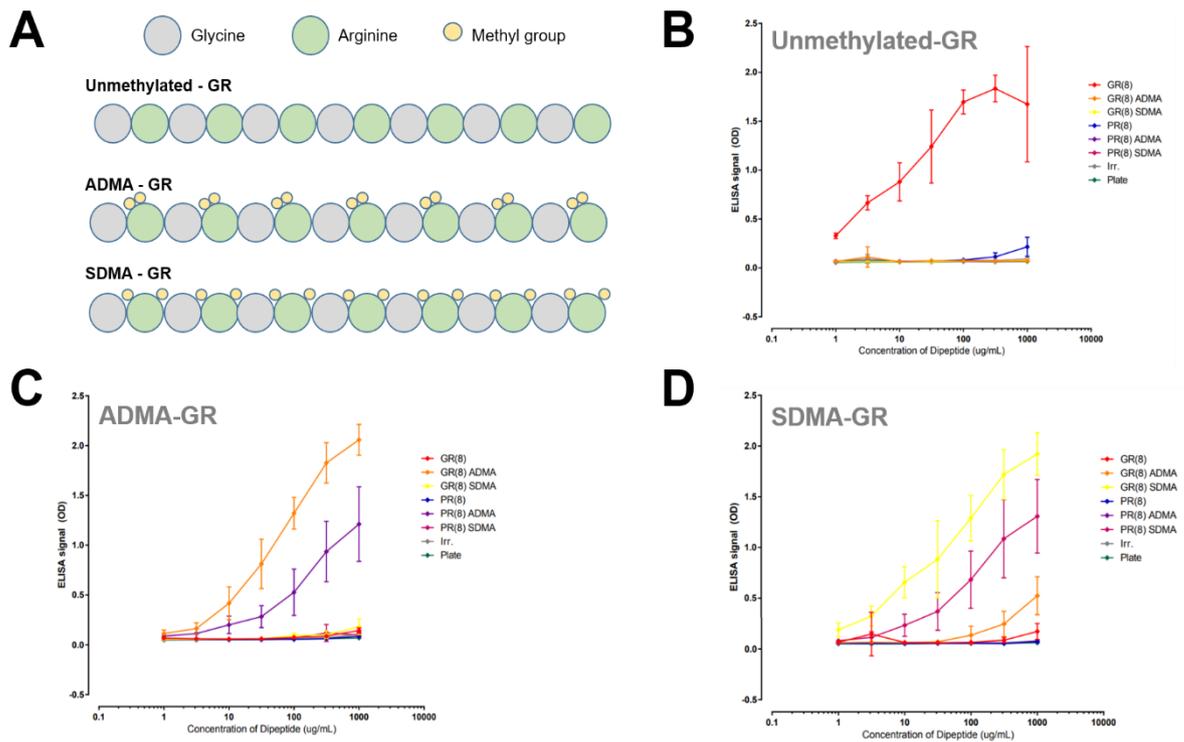
Figure 5.2 Specificity of novel unmethylated and dimethylated poly-GR antibodies

(**A**) Schematic of the peptide antigens used to generate unmethylated and dimethylated forms of poly-GR. Quantitative ELISA analysis for validation of (**B**) unmethylated-GR (**C**) ADMA-GR and (**D**) SDMA-GR antibody specificity. All three antibodies preferentially bind with highest affinity to their respective poly-GR antigens. Quantitative ELISA analysis was performed by collaborators at UCB.

5.2.2 Characterisation of unmethylated-GR antibody

Each antibody was characterised by immunohistochemistry in human post-mortem brain tissue of neurologically normal controls and *C9orf72* cases. The unmethylated-GR stained neuronal nuclei and was frequently observed as speckled puncta within the nucleus but did not detect DPR protein inclusions (Figure 5.3 A). The unmethylated-GR antibody did not detect neuronal cytoplasmic inclusions in the *C9orf72* cases, potentially indicating lack of sensitivity or that the targeted unmethylated poly-GR epitope was unable to be accessed by the antibody.

Visual observation of the staining suggested that there may be more of the speckled nuclei in the *C9orf72* cases. To determine if this was a *C9orf72* disease specific effect, the percentage of neurons with nuclear speckles was quantified in 5 *C9orf72* cases, 5 FTLD-TDP type A cases and 5 control cases, with ten fields of view analysed in each case (Figure 5.3 B). FTLD-TDP type A cases were included as a disease control. No significant difference in neurons containing nuclear unmethylated-GR speckles was found between the three groups, indicating the signal was not specific to poly-GR derived from RAN translation of the C9of72 repeat expansion.
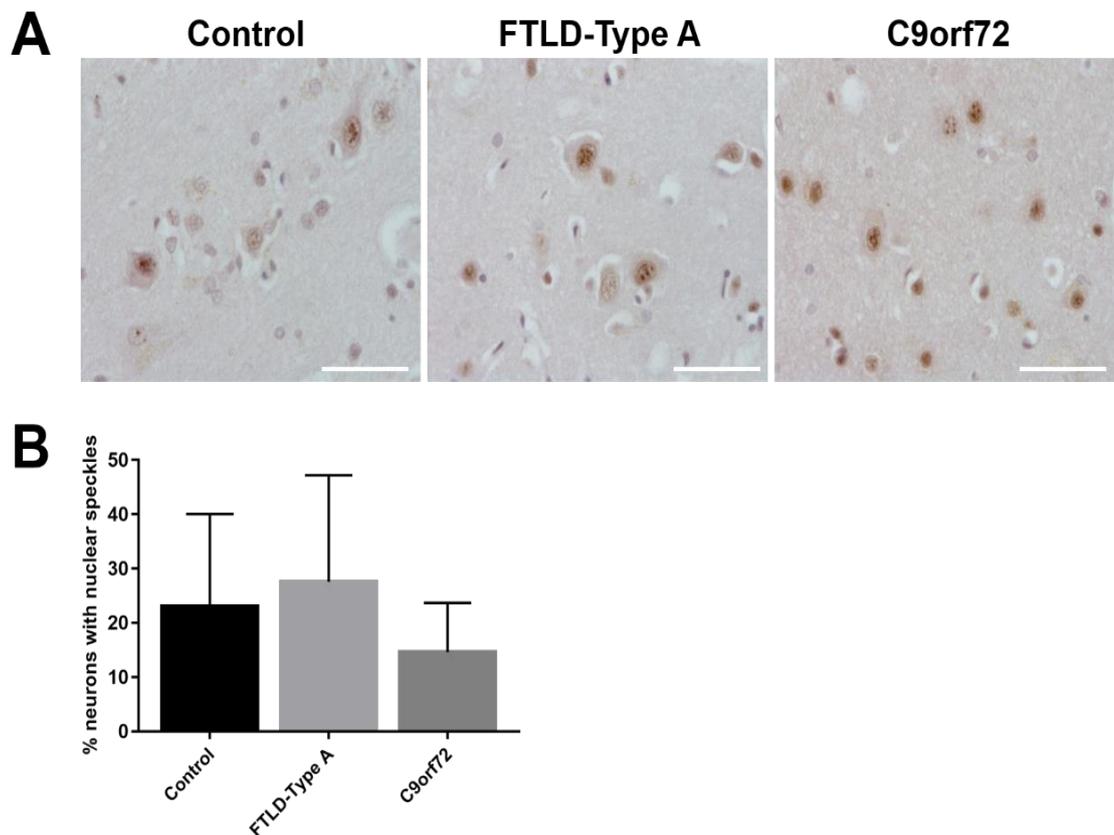
Figure 5.3 Unmethylated-GR is detected in neuronal nuclei and frequently forms nuclear speckles.

(**A**) Representative images of frontal cortex sections from a neurological normal control, a FTLD-Type A case and a heterozygous *C9orf72* case stained with unmethylated-GR. In both the control and disease cases, unmethylated-GR is detected in the nucleus of neurons and frequently forms a speckled pattern. Scale bars represent 50 μm. (**B**) Quantification of the percentage of neurons containing nuclear unmethylated-GR speckles in the frontal cortex. One-way ANOVA, p = 0.447, n = 5. Error bars represent standard error of the mean.

5.2.3 Characterisation of asymmetric dimethylated-GR antibody

Immunohistochemical staining of the frontal and temporal cortices of neurologically normal controls and C9orf72 cases revealed that the ADMA-GR antibody strongly stained neuronal nuclei and occasionally stained the neuronal cytoplasm. In both C9orf72 – FTLD and C9orf72 – ALS cases, frequent ADMA-GR immunoreactive neuronal cytoplasmic inclusions were detected (Figure 5.4). These inclusions had the characteristic star-like shape of DPR protein inclusions (Figure 5.6), and were not detected in controls, suggesting they were *C9orf72* specific DPR protein inclusions. Immunohistochemical signal was significantly reduced and DPR protein inclusions could no longer be detected when the ADMA-GR antibody was pre-incubated with poly-GR blocking peptide in an antibody pre-absorption assay (data not shown).
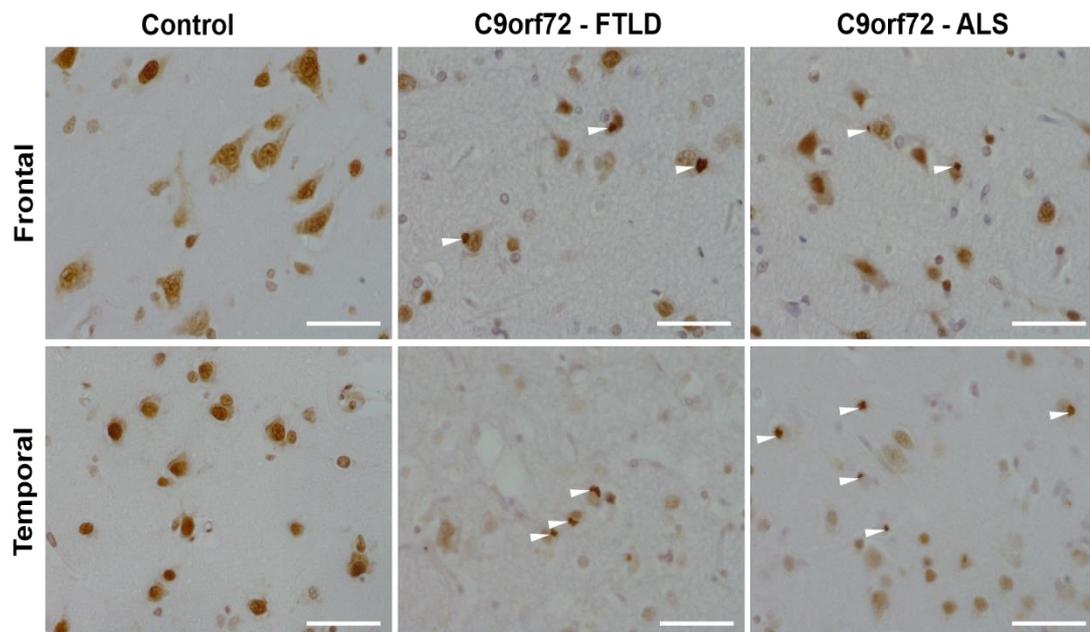
Figure 5.4 Detection of asymmetric dimethylated-GR (ADMA-GR) in C9orf72 post-mortem brain tissue.

Representative images of ADMA-GR immunohistochemical staining in the frontal and temporal cortices of a neurological normal control, *C9orf72* – FTLD case and *C9orf72* – ALS case. White arrow heads indicate neuronal cytoplasmic inclusions containing ADMA-GR. Scale bars represent 50 µm.

5.2.4 Characterisation of symmetric dimethylated-GR antibody

Immunohistochemical staining of the frontal and temporal cortices of neurologically normal controls and *C9orf72* cases showed that the SDMA-GR antibody weakly stained neuronal nuclei but frequently detected neuronal cytoplasmic inclusions in the *C9orf72* cases (Figure 5.5). These were less frequently observed than inclusions detected by ADMA-GR. Similar to the ADMA-GR staining, these inclusions had the characteristic star-like shape of DPR protein inclusions in both *C9orf72* – FTLD and *C9orf72* – ALS cases (Figure 5.6), and were not detected in controls, suggesting they were *C9orf72* specific DPR protein inclusions. An antibody pre-absorption assay resulted in reduced immunohistochemical signal and DPR protein inclusions could no longer be detected when the SDMA-GR antibody was pre-incubated with poly-GR blocking peptide (data not shown).
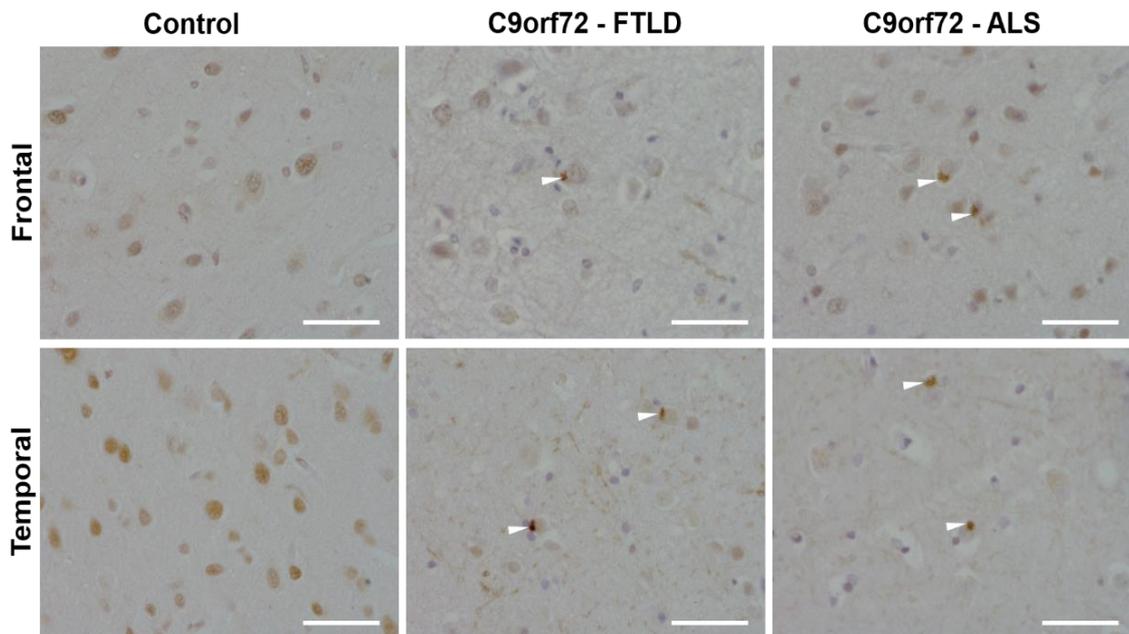
Figure 5.5 Detection of symmetric dimethylated-GR (SDMA-GR) in C9orf72 post-mortem brain tissue.

Representative images of SDMA-GR immunohistochemical staining in the frontal and temporal cortices of a neurological normal control, *C9orf72* – FTLD case and *C9orf72* – ALS case. White arrow heads indicate neuronal cytoplasmic inclusions containing SDMA-GR. Scale bars represent 50 μm.
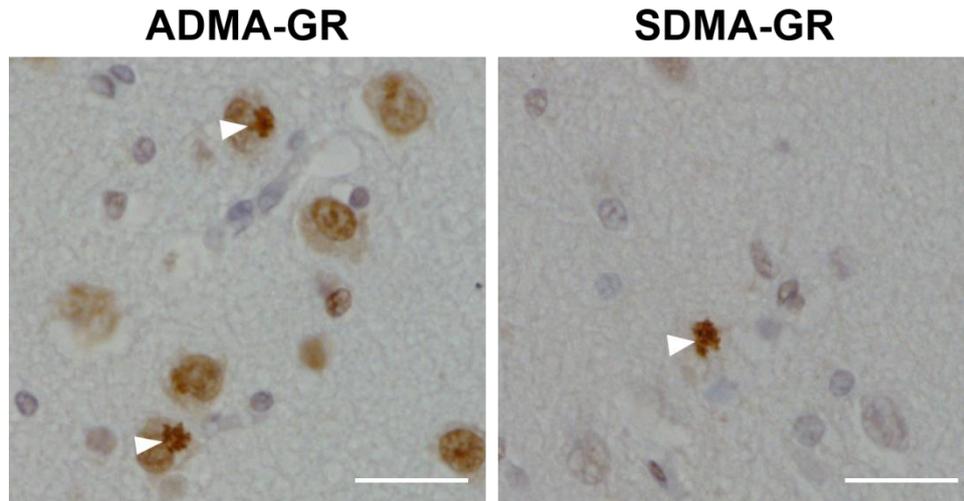
ADMA-GR   **SDMA-GR**

Figure 5.6 ADMA-GR and SDMA-GR antibodies detect characteristic star-shape inclusions in C9orf72 cases

Representative images of ADMA-GR and SDMA-GR immunohistochemical staining in the frontal cortex of *C9orf72* cases. White arrow heads indicate neuronal cytoplasmic inclusions that have a star-like shape, which are characteristic of *C9orf72* DPR protein inclusions. Scale bars represent 20 µm.

5.2.5 Co-localisation of poly-GR and methylated-GR

To determine whether the ADMA-GR and SDMA-GR antibodies were co-localising with poly-GR neuronal cytoplasmic inclusions, double immunofluorescence was performed on the frontal cortex of *C9orf72* cases using a previously validated pan poly-GR antibody with either ADMA-GR or SDMA-GR (Figure 5.7). Qualitative assessment of the fluorescence images showed that poly-GR co-localised with both ADMA-GR and SDMA-GR, however not all of the poly-GR inclusions were positive for a methylated form of poly-GR. Similar to the immunohistochemical staining, SDMA-GR immunoreactive inclusions were observed less frequently than ADMA-GR immunoreactive inclusions.
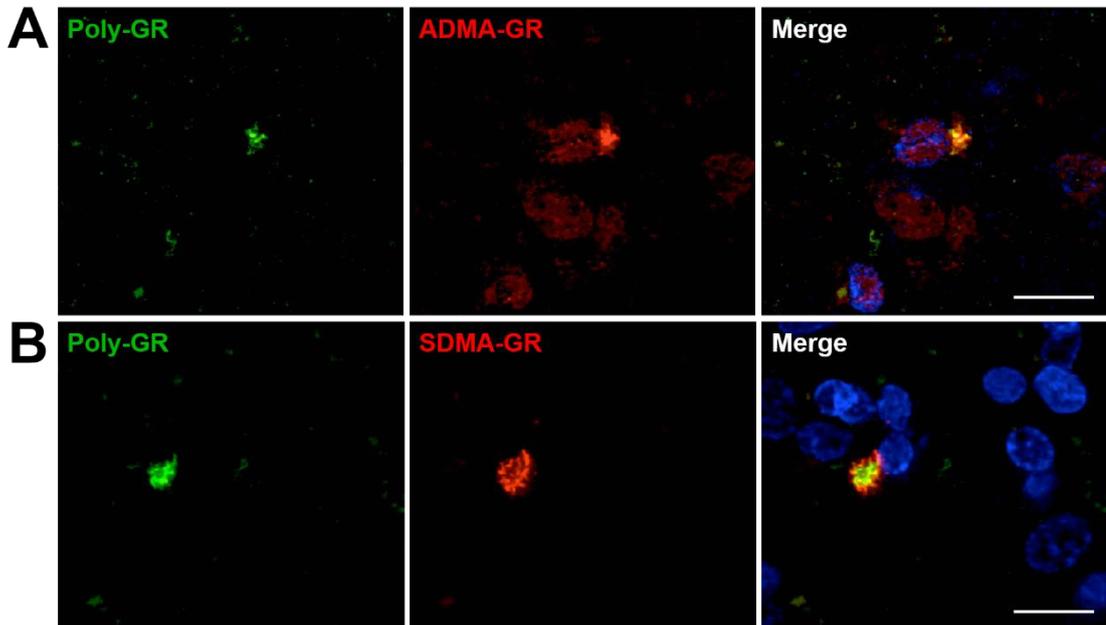
Figure 5.7 ADMA-GR and SDMA-GR colocalise with poly-GR inclusions in C9orf72 post-mortem brain tissue.

Representative images of double immunofluorescent staining of poly-GR (green) with either (**A**) ADMA-GR (red) or (**B)** SDMA-GR (red) in the frontal and temporal cortices of a *C9orf72* FTLD case. Cell nuclei are counterstained with DAPI (blue). Scale bar represents 10 µm.

5.2.6 Qualitative comparison between methylated-GR inclusions in *C9orf72* homozygous and heterozygous cases

Whilst qualitatively assessing the burden of methyl-GR pathology in the *C9orf72* cases, it was noticed that a rare *C9orf72* homozygous case had few SDMA-GR inclusions compared to heterozygous *C9orf72* cases (Figure 5.8). This homozygous case has previously been reported and was noted to have abundant DPR protein pathology and severe clinical features, although not outside the normal disease range (Fratta et al., 2013). The abundance of DPR proteins in the homozygous cases is demonstrated by p62 staining, which detects DPR protein inclusions in almost all neurons (Figure 5.8). It was surprising, therefore, to observe less SDMA-GR pathology in the homozygous compared to heterozygous cases. No obvious differences were visibly observed in the number of ADMA-GR inclusions between the homozygous case and heterozygous cases. This observation led to quantitative and correlative assessment of methyl-GR pathology in the *C9orf72* cases.
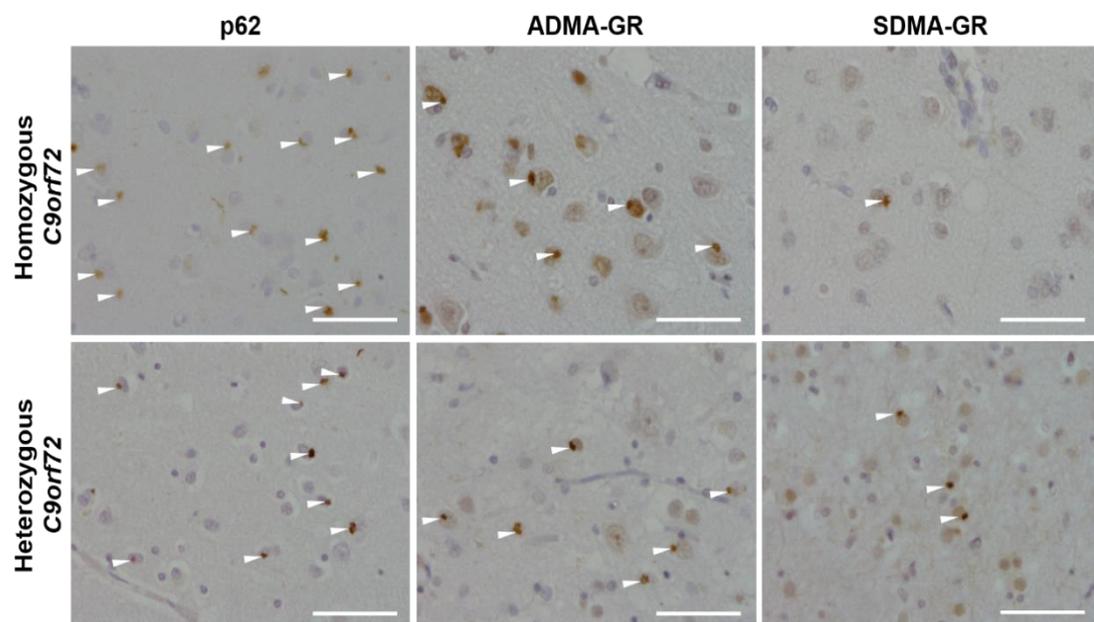
Figure 5.8 Comparison of methyl-GR staining in a homozygous and a heterozygous C9orf72 case.

Representative images of p62, ADMA-GR and SDMA-GR immunohistochemical staining in the frontal cortex of a *C9orf72* homozygous and *C9orf72* heterozygous case. White arrow heads indicate neuronal cytoplasmic inclusions. p62 staining is shown to indicate extent of DPR protein pathology in these cases. Scale bars represent 50 µm.

5.2.7 Quantification of ADMA-GR and SDMA-GR containing inclusions in QSBB cohort

To quantitatively assess methyl-GR pathology, the number of inclusions containing ADMA-GR and SDMA-GR were quantified in ten regions of interest in the frontal and temporal cortices of each of the *C9orf72* cases available for analysis at the Queen Square Brain Bank (QSBB) (*C9orf72* - FTLD n = 13, *C9orf72* - ALS n = 2) (Figure 5.9). As the number of *C9orf72* – ALS cases available to analyse was low, the quantification of inclusions for both *C9orf72* - ALS and *C9orf72* - FTLD have been included in the same analysis. The number of inclusions containing ADMA-GR was consistently higher than the number of inclusions containing SDMA-GR in both the frontal and temporal cortices of all cases. No significant differences were observed between the two brain regions for the number of ADMA-GR inclusions (p = 0.149) (Figure 5.9 A) or the number of SDMA-GR inclusions (p = 0.146) (Figure 5.9 B).
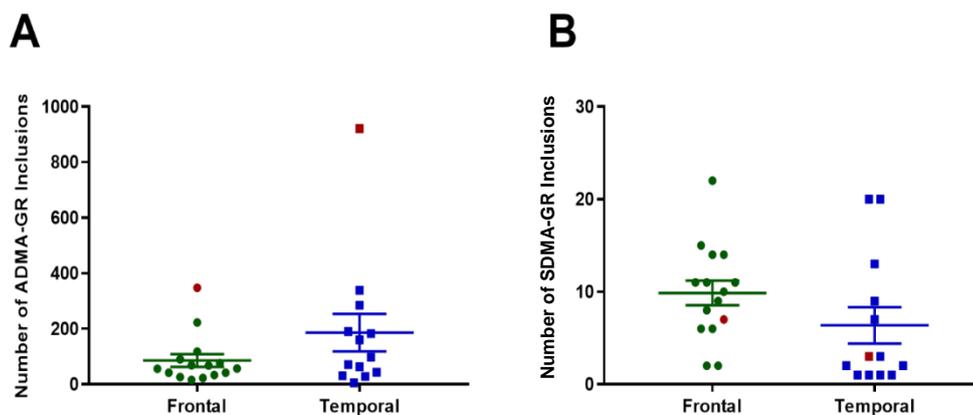
Figure 5.9 Quantification of ADMA-GR and SDMA-GR containing inclusions in QSBB cohort

(**A**) Total number of ADMA-GR inclusions across ten selected regions of interest in the frontal and temporal cortices of *C9orf72* cases in the Queen Square Brain Bank (QSBB) cohort. Unpaired T-test, n = 13, p = 0.149. (**B**) Total number of SDMA-GR inclusions across ten selected regions of interest in the frontal and temporal cortices of *C9orf72* cases in the QSBB cohort. Each point represents one *C9orf72* case. Unpaired T-test, n = 13, p = 0.146. In both A and B, the red point represents the *C9orf72* homozygous case. Error bars represent standard error of the mean.

5.2.8 Correlations between methyl-GR and case information in QSBB cohort

Given that the homozygous *C9orf72* case was observed to have few SDMA-GR inclusions and has severe clinical features, such as an early age of disease onset and short disease duration, it was questioned whether the methylated forms of poly-GR may correlate with these clinical features. To analyse this, the total number of ADMA-GR and SDMA-GR inclusions counted in the frontal and temporal cortices of each *C9orf72* case at the QSBB was correlated with their corresponding clinical information on age at disease onset, age at death, disease duration and post-mortem delay using Spearman's rank correlation coefficient. The results of this analysis are displayed in Table 5.1.

Two significant correlations were observed following Spearman's rank correlation analysis. The number of ADMA-GR inclusions in the frontal cortex was found to have a weak negative correlation with age at disease onset ($r = -0.5903$, $p = 0.0286$), suggesting that higher numbers of ADMA-GR containing inclusions correlate with earlier disease onset. Although, this correlation appears to be largely driven by one case with an early disease onset and a high number of ADMA-GR inclusions – the homozygous case. The significant correlation is abolished when this outlier is removed from the analysis ($r = -0.4868$, $p = 0.0930$). In contrast, the number of SDMA-GR inclusion in the frontal cortex was found to have a positive correlation with disease duration ($r = 0.7201$, $p = 0.0048$), suggesting that high numbers of SDMA-GR containing inclusions correlate with a longer disease duration. Neither of these correlations were replicated in the temporal cortex.
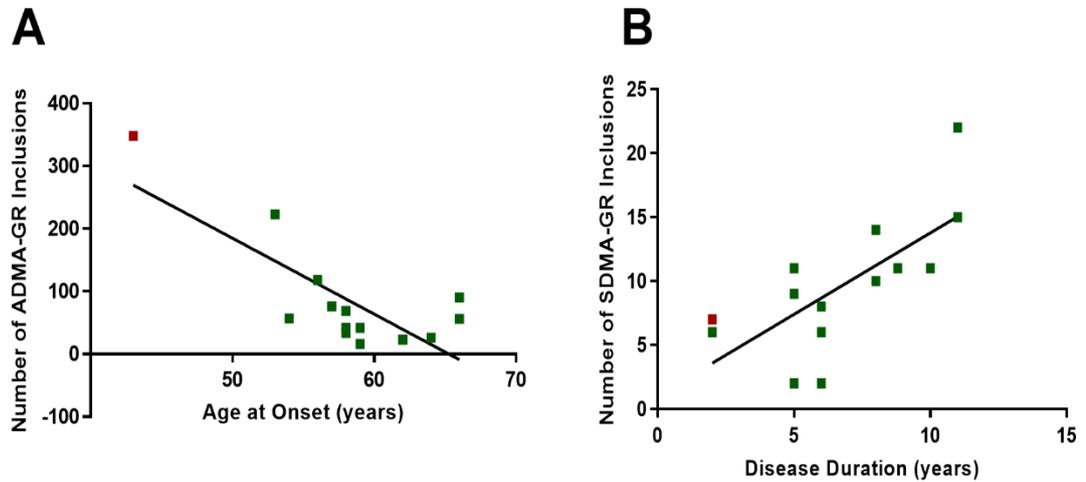
Figure 5.10 Statistically significant correlations in C9orf72 cases between methyl-GR and clinical information in the Queen Square Brain Bank cohort

(**A**) Association between number of ADMA-GR inclusions in the frontal cortex and age of disease onset for the *C9orf72* cases in the QSBB cohort. Spearman's rank correlation coefficient, r = -0.5903, p = 0.0286, n = 14. (**B**) Association between number of SDMA-GR inclusions in the frontal cortex and disease duration for the *C9orf72* cases in the QSBB cohort. Spearman's rank correlation coefficient, r = -0.7201, p = 0.0048, n = 14. In both A and B, each individual case is represented by a green square. The homozygous *C9orf72* case is shown as a red square.

Table 5.1 Associations of ADMA-GR and SDMA-GR inclusions in the frontal and temporal cortices of C9orf72 cases with clinical information in the Queen Square Brain Bank cohort.

Region	Association	ADMA-GR			SDMA-GR		
		n	Spearman's r (95% CI)	p value	n	Spearman's r (95% CI)	p value
Frontal Cortex							
	Age at onset	14	-0.59 (-0.86 to -0.070)	**0.029**	14	-0.25 (-0.70 to 0.34)	0.38
	Age at death	15	-0.32 (-0.72 to 0.25)	0.25	15	0.34 (-0.22 to 0.73)	0.21
	Disease duration	14	-0.09 (-0.60 to 0.48)	0.76	14	0.72 (0.29 to 0.91)	**0.0048**
	Post-mortem delay	15	-0.29 (-0.71 to 0.27)	0.28	15	0.033 (-0.50 to 0.55)	0.91
Temporal Cortex							
	Age at onset	12	-0.57 (-0.86 to 0.027)	0.057	12	-0.44 (0.81 to 0.19)	0.15
	Age at death	13	-0.53 (-0.84 to 0.054)	0.068	13	-0.26 (-0.71 to 0.36)	0.39
	Disease duration	12	-0.25 (-0.73 to 0.40)	0.43	12	-0.26 (-0.73 to 0.39)	0.41
	Post-mortem delay	13	-0.31 (-0.74 to 0.31)	0.30	13	-0.21 (-0.69 to 0.40)	0.49

Data are Spearman's correlation coefficient *r* (95 % confidence interval (CI)) and *p* value. Significance level was set at $p < 0.05$ (two-sided). Significant *p* values are indicated in bold. ADMA-GR = asymmetric dimethylated poly-GR. SDMA-GR = symmetric dimethylated poly-GR.

5.2.9 Correlations between methyl-GR and case information in a combined brain bank cohort

As the number of *C9orf72* cases available at QSBB was relatively low (n = 14), to determine whether the significant correlations previously identified were maintained in a larger cohort, the analysis was expanded to include *C9orf72* cases from the Manchester Brain Bank (MBB) (University of Manchester) and the MRC London Neurodegenerative Diseases Brain Bank (Institute of Psychiatry, Psychology & Neuroscience (IOP), King's College London). This increased the sample size to 37 *C9orf72* cases (*C9orf72*-FTLD n = 22, *C9orf72*-ALS n = 15). It should be noted, however, that some cases had missing clinical information, most commonly age of disease onset and disease duration, thus these cases were unable to be included in correlation analysis where relevant information was missing. Details of all cases used in this study can be found in Section 2.5.1 of Chapter 2.

Tissue sections from the frontal and temporal cortices of *C9orf72* cases from all three brain banks were stained with ADMA-GR and SDMA-GR and the number of inclusions containing these methylated forms of poly-GR were quantified. The inclusion counts were then correlated with age at disease onset, age at death, disease duration and post-mortem delay using Spearman's rank correlation coefficient. The results of this analysis are displayed in Table 5.2.

The negative correlation between ADMA-GR and age at disease onset in the frontal cortex no longer reached statistical significance when including data from a larger cohort. In contrast, the positive correlation between SDMA-GR and disease duration in the frontal cortex remained statistically significant in the combined brain bank cohort, although the strength of this correlation was reduced (r =0.5139, p = 0.0026). This analysis additionally found the number of SDMA-GR inclusions to positively correlate with age at death in the frontal cortex (r = 0.4568, p = 0.0045). Once again,

neither of the significant correlations observed in the frontal cortex were replicated in the temporal cortex.
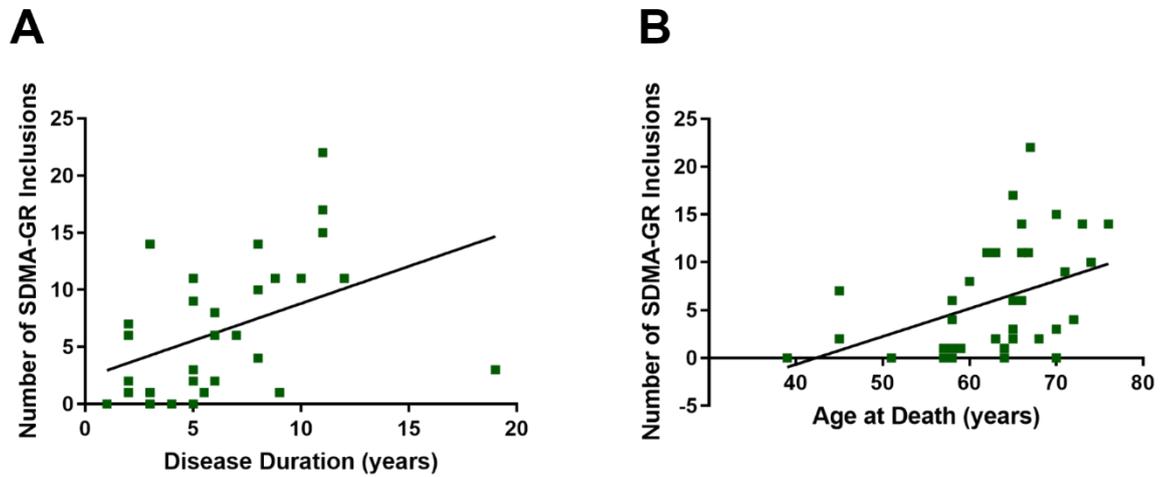
Figure 5.11 Statistically significant correlations in C9orf72 cases between SDMA-GR and clinical information in the combined brain bank cohort

(**A**) Association between number of SDMA-GR inclusions in the frontal cortex and disease duration in the combined cohort of *C9orf72* cases. Spearman's rank correlation coefficient, r =0.5139, p = 0.0026, n = 32. (**B**) Association between number of SDMA-GR inclusions in the frontal cortex and age at death in the combined cohort of *C9orf72* cases. Spearman's rank correlation coefficient, r = 0.4568, p = 0.0045, n = 37. In both A and B, each individual case is represented by a green square.

Table 5.2 Associations of ADMA-GR and SDMA-GR inclusions in the frontal and temporal cortices of C9orf72 cases with clinical information in the combined brain bank cohort.

Region	Association	ADMA-GR			SDMA-GR		
		n	Spearman's r (95% CI)	*p* value	n	Spearman's r (95% CI)	*p* value
Frontal Cortex							
	Age at onset	32	-0.098 (-0.41 to 0.27)	0.59	32	0.0051 (-0.35 to 0.36)	0.98
	Age at death	37	0.24 (-0.099 to 0.53)	0.15	37	0.46 (0.15 to 0.69)	**0.0045**
	Disease duration	32	0.095 (-0.27 to 0.43)	0.60	32	0.51 (0.19 to 0.74)	**0.0026**
	Post-mortem delay	31	0.20 (-0.18 to 0.53)	0.28	31	0.19 (-0.19 to 0.52)	0.32
Temporal Cortex							
	Age at onset	28	-0.12 (-0.48 to 0.28)	0.55	28	-0.087 (-0.45 to 0.31)	0.66
	Age at death	33	0.20 (-0.16 to 0.52)	0.26	33	0.18 (-0.18 to 0.50)	0.31
	Disease duration	28	0.21 (-0.18 to 0.55)	0.27	28	0.29 (-0.10 to 0.61)	0.13
	Post-mortem delay	29	0.19 (-0.20 to 0.53)	0.33	29	0.31 (-0.078 to 0.61)	0.11

Data are Spearman's correlation coefficient *r* (95 % confidence interval (CI)) and *p* value. Significance level was set at $p < 0.05$ (two-sided). Significant *p* values are indicated in bold. ADMA-GR = asymmetric dimethylated poly-GR. SDMA-GR = symmetric dimethylated poly-GR.

5.2.10 Correlation analysis when *C9orf72* ALS and FTLD cases are analysed independently

By combining *C9orf72* cases from the QSBB, MBB and IOP, this provided a large enough sample size of both *C9orf72* – ALS cases and *C9orf72* – FTLD cases to perform separate correlation analysis on the two different phenotypic groups. Such analysis could reveal important differences between the two disease groups, for example whether the association between disease duration and SDMA-GR is driven by one particular clinical phenotype. A total of 15 cases were analysed in the *C9orf72* – ALS groups, and 22 cases were analysed in the *C9orf72* – FTLD group. Again, however, some cases had missing clinical information, thus were unable to be included in correlation analysis where relevant information was missing. The number of cases used in each correlation analysis is shown in Table 5.3 and Table 5.4

Using the total number of ADMA-GR and SDMA-GR inclusions previously counted in the frontal and temporal cortices, these were correlated with clinical information of each *C9orf72* case on age at disease onset, age at death, disease duration and post-mortem delay using Spearman's rank correlation coefficient. The *C9orf72* – ALS cases were analysed independently of the *C9orf72* – FTLD cases. The results of these analyses are displayed in Table 5.3 for the FTLD cases and Table 5.4 for the ALS cases.

In the *C9orf72* – FTLD group, a significant negative correlation was found between ADMA-GR and age at disease onset, in both the frontal ($r = -0.4509$, $p = 0.0352$) and temporal ($r = -0.4877$, $p = 0.0291$) cortices. Additionally, a significant negative correlation was found between SDMA-GR and both age at disease onset ($r = -0.4463$, $p = 0.0486$) and age at death ($r = -0.4793$, $p = 0.0325$) in the temporal cortex. No other correlations reached statistical significance.

In the *C9orf72* – ALS group, no clinical information was found to significantly correlate with the number of SDMA-GR inclusions, however ADMA-GR positively correlated with age at disease onset in the frontal cortex (r = 0.7091, p = 0.0217), and age at death in both the frontal (r = 0.7828, p = 0.0006) and temporal (r = 0.6311, p = 0.0207) cortices.

Table 5.3 Associations of ADMA-GR and SDMA-GR inclusions in the frontal and temporal cortices of C9orf72-FTLD cases only with clinical information.

Region	Association	ADMA-GR			SDMA-GR		
		n	Spearman's r (95% CI)	*p* value	n	Spearman's r (95% CI)	*p* value
Frontal Cortex							
	Age at onset	22	-0.45 (-0.74 to -0.023)	**0.035**	22	-0.072 (-0.49 to 0.37)	0.75
	Age at death	22	-0.35 (-0.68 to 0.097)	0.11	22	0.22 (-0.23 to 0.60)	0.32
	Disease duration	22	0.055 (-0.39 to 0.48)	0.81	22	0.38 (-0.067 to 0.70)	0.084
	Post-mortem delay	18	-0.080 (-0.58 to 0.41)	0.75	18	0.058 (-0.43 to 0.52)	0.82
Temporal Cortex							
	Age at onset	20	-0.49 (-0.77 to -0.044)	**0.029**	20	-0.45 (-0.75 to 0.00094)	**0.049**
	Age at death	20	-0.39 (-0.72 to 0.073)	0.086	20	-0.48 (-0.77 to -0.033)	**0.033**
	Disease duration	20	0.040 (-0.43 to 0.48)	0.90	20	0.068 (-0.40 to 0.51)	0.78
	Post-mortem delay	17	-0.032 (-0.52 to 0.47)	0.90	17	0.24 (-0.29 to 0.65)	0.36

Data are Spearman's correlation coefficient *r* (95 % confidence interval (CI)) and *p* value. Significance level was set at $p < 0.05$ (two-sided). Significant *p* values are indicated in bold. ADMA-GR = asymmetric dimethylated poly-GR. SDMA-GR = symmetric dimethylated poly-GR.

Table 5.4 – Associations of ADMA-GR and SDMA-GR inclusions with clinical information in the frontal and temporal cortices of C9orf72-ALS cases only with clinical information.

Region	Association	ADMA-GR			SDMA-GR		
		n	Spearman's r (95% CI)	p value	n	Spearman's r (95% CI)	p value
Frontal Cortex							
	Age at onset	10	0.709	**0.022**	10	-0.28	0.43
	Age at death	15	0.78 (0.44 to 0.93)	**0.001**	15	0.26 (-0.31 to 0.69)	0.36
	Disease duration	10	-0.57	0.10	10	0.58	0.09
	Post-mortem delay	13	0.32 (-0.29 to 0.75)	0.28	13	-0.32 (-0.75 to 0.30)	0.29
Temporal Cortex							
	Age at onset	8	0.41	0.33	8	0.41	0.30
	Age at death	13	0.63 (0.10 to 0.88)	**0.021**	13	0.31 (-0.31 to 0.74)	0.31
	Disease duration	8	-0.18	0.66	8	-0.17	0.70
	Post-mortem delay	12	-0.13 (-0.67 to 0.49)	0.69	12	-0.11 (-0.65 to 0.51)	0.74

Data are Spearman's correlation coefficient r (95 % confidence interval (CI)) and p value. In instances where CI is not provided, this is due to a small sample size of less than 10 cases being analysed. Significance level was set at $p < 0.05$ (two-sided). Significant p values are indicated in bold. ADMA-GR = asymmetric dimethylated poly-GR. SDMA-GR = symmetric dimethylated poly-GR.

5.3 Discussion

This study describes the characterisation of two novel antibodies designed to detect unmethylated and dimethylated forms of poly-GR in *C9orf72* patient post-mortem tissue. In collaboration with UCB, three novel antibodies were generated that are the first to specifically detect unmethylated and two dimethylated forms of the poly-GR protein. These antibodies were tested on *C9orf72* patient post-mortem tissue and found to label a proportion of DPR protein neuronal cytoplasmic inclusions. An observation of few symmetrically dimethylated inclusions in a rare homozygous *C9orf72* case led to investigating whether the number of methyl-GR inclusions correlated with various pieces of clinical information. When *C9orf72* FTLD and ALS cases were analysed together, SDMA-GR inclusions in the frontal cortex were found to positively correlate with disease duration and age at death, suggesting that the symmetric demethylation of poly-GR could have protective effects. However, this correlation was not replicated in the temporal cortex, or when the two disease groups were analysed independently. The reason(s) for this discrepancy is currently unknown.

5.3.1 Methyl-GR specific antibodies

This is the first study to utilise two novel antibodies designed to target unmethylated and dimethylated forms of the poly-GR protein in *C9orf72* post-mortem tissue. Previous studies have suggested that this DPR protein may be post-translationally modified by the addition of methyl groups to arginine residues due to co-localisation of poly-GR with PRMT enzymes and immunoreactivity to an antibody which recognises proteins that are asymmetrically dimethylated. However, the antibody used in the previously published studies was not specific for the poly-GR protein, making it difficult to determine whether the poly-GR protein itself was methylated, or whether it associated with other proteins that can undergo arginine methylation (Boeynaems et al., 2016; Sakae et al., 2018).

ELISA analysis demonstrated that the antibodies developed by UCB to detect either symmetrically or asymmetrically modified poly-GR were specific for their respective peptide antigens and did not show reactivity to unmethylated forms of poly-GR or to the oppositely dimethylated poly-GR antigen. The ADMA-GR antibody showed some reactivity to ADMA-PR, and the SDMA-GR antibody showed some reactivity to SDMA-PR. However, given that poly-PR is rarely detected by immunohistochemistry in human post-mortem brain, and the antibodies had a higher affinity for the methylated poly-GR proteins, it was determined that these antibodies were unlikely to identify ADMA-PR and SDMA-PR in post-mortem tissue.

5.3.2 Characterisation of methyl-GR antibodies

Immunohistochemical assessment of the unmethylated-GR antibody showed detection of unmethylated-GR mostly in neuronal nuclei. The nuclear staining was non-uniform however, and speckled puncta or dot-like structures were frequently detected within the nuclei. This was an interesting observation given that several studies have reported short forms of over-expressed poly-GR to localise to the nucleolus in cellular models (Kwon et al., 2014; Lee et al., 2016; May et al., 2014; Tao et al., 2015; Wen et al., 2014), but this localisation has not been observed in *C9orf72* post-mortem tissue. However, immunohistochemical assessment and quantification of the neurons containing unmethylated-GR immunoreactive nuclear speckles in *C9orf72* cases, FTLD-TDP type A cases and control cases showed that there was no significant difference between these three groups. The presence of unmethylated-GR immunoreactive nuclear speckles in control and FTLD-TDP type A cases, suggests that this antibody is not specific to unmethylated forms of poly-GR produced by RAN translation of *C9orf72* hexanucleotide repeat expansion but rather it is likely detecting unmethylated epitopes of poly-GR present in other proteins. A protein BLAST search reveals an abundance of nuclear proteins with an epitope consisting of at least 8 glycine-arginine repeats. Potential candidate proteins that may

be being detected by the unmethylated-GR antibody include; small nuclear ribonucleoprotein Sm D (SNRPD1), chromatin target of PRMT1 protein (CHTOP) and 40S ribosomal protein S2.

The unmethylated-GR antibody did not detect cytoplasmic DPR protein inclusions. A potential reason for this could be that the structure of the inclusion makes the targeted epitope inaccessible. Alternatively, the antibody may lack sensitivity for unmethylated-GR by immunohistochemical methods. Given the lack of detected of DPR protein inclusions by the unmethylated-GR antibody in *C9orf72* cases, no further work was performed using the unmethylated-GR antibody.

Unlike the unmethylated-GR antibody, the ADMA-GR and SMDA-GR antibodies were specific to the *C9orf72* cases as they were able to detect neuronal cytoplasmic inclusions with the size and morphology of DPR protein inclusions that were not present in the control cases. Some nuclear staining and occasional diffuse cytoplasmic staining was observed in both control and *C9orf72* cases. This is most likely the detection of dimethylated epitopes of GR-containing proteins that often undergo methylation at sites with a consensus glycine/arginine rich motif. The detection of inclusions, however, was highly specific to the *C9orf72* cases.

Double immunofluorescence labelling of these inclusions demonstrated that both ADMA-GR and SDMA-GR co-localise with poly-GR, indicating that these antibodies are indeed detecting the DPR protein inclusions in the *C9orf72* cases. In agreement with the previous published studies on methylation of poly-GR inclusions, co-localisation was not observed for all poly-GR inclusions, suggesting that some inclusions may contain unmethylated or monomethlyated forms of poly-GR (Boeynaems et al., 2016; Sakae et al., 2018). The lack of methylation in some poly-GR inclusions may reflect the dynamic reversibility of arginine methylation. Although

arginine methylation was initially thought to be a permanent post-translational modification, a number of studies have emerged supporting the reversible nature of methylarginine and have identified putative arginine demethylases (Chang et al., 2007; Tsai et al., 2016; Walport et al., 2016; Wesche et al., 2017). The poly-GR inclusions detected that were unmethylated may therefore reflect poly-GR proteins yet to undergo arginine methylation or that have been demethylated. The ability to be reversibly methylated could have important physical or functional consequences for the poly-GR proteins. This has been demonstrated for FUS where arginine methylation alters the proteins ability to phase separate and changes its interaction with its nuclear importer (Hofweber et al., 2018; Qamar et al., 2018). Further biochemical work investigating the effect of arginine methylation and demethylation on poly-GR is required to assess the dynamics and importance of arginine methylation reversibility for this protein.

Quantification of ADMA-GR and SDMA-GR inclusions showed the number containing ADMA-GR to be consistently higher than number of inclusions containing SDMA-GR in both the frontal and temporal cortices. This is unsurprising given that asymmetric dimethylation of arginine residues is known to be the most prevalent form of methylation in physiological systems, with the monomethylated and symmetrically dimethylated forms thought to occur at levels of about 20 % to 50 % that of the asymmetrically dimethylated form (Bedford and Clarke, 2009).

In the present study, ADMA-GR and SDMA-GR containing inclusions were detected in both the frontal and temporal cortices of *C9orf72* – ALS and *C9orf72* – FTLD cases. This indicates that arginine methylation of poly-GR is not specific to a clinical phenotype and cannot be used to pathologically differentiate between *C9orf72* – ALS and *C9orf72* – FTLD cases, as is the case with FTLD-FUS and ALS-FUS, where

methylated FUS inclusions are only found in the ALS-FUS cases (Dormann et al., 2012).

5.3.3 Methylated-GR in the homozygous *C9orf72* case

One *C9orf72* case that was particularly interesting in this study was a patient that was homozygous for the *C9orf72* hexanucleotide repeat expansion. Previous clinical and pathological analysis of this case had shown that this patient had extensive DPR protein pathology and severe clinical characteristics, although these were not outside the normal range for *C9orf72*-linked disease (Fratta et al., 2013). When qualitatively assessing methylated-GR pathology in the homozygous case, it was noticed that this case had very few inclusions containing SDMA-GR, compared to heterozygous cases. This was intriguing, given the severe clinical phenotype of this case (Fratta et al., 2013), and the recent publications suggesting that arginine methylation is able to reduce the phase separation, and potentially the aggregation, of the FUS protein (Fratta et al., 2013; Hofweber et al., 2018; Qamar et al., 2018). Based on the assumption that arginine methylation may also reduce the phase separation and aggregation of poly-GR, it was hypothesised that the lack of methylation of poly-GR inclusions in the homozygous case may be associated with its severe clinical phenotype, for example an early age of onset and short disease duration. This hypothesis led to the subsequent correlational analysis to assess associations between the number of ADMA-GR and SDMA-GR inclusions and various pieces of clinical information.

5.3.4 Correlational analysis between methyl-GR and case information when *C9orf72* ALS and FTLD cases are analysed together

Due to resources available when conducting the study, the initial correlation analysis only included *C9orf72* cases available at the QSBB. This revealed two statistically significant correlations. To determine whether these correlations remained significant

in a larger cohort, *C9orf72* cases from the MBB and IOP were subsequently included in the analysis. In both of these analyses, all *C9orf72* cases were analysed together, regardless of clinical phenotype.

The first significant correlation identified in the QSBB cohort was ADMA-GR inclusions in the frontal cortex correlating negatively with age at disease onset, suggesting that higher numbers of ADMA-GR inclusions correlate with earlier disease onset. However, this correlation appeared to be largely driven by the homozygous case, which had a very early age of disease onset and a high number of ADMA-GR inclusions compared to the other cases. When this case was removed from the analysis, the correlation was no longer statistically significant. Furthermore, when a larger number of cases were included in the combined brain bank analysis, this correlation also did not reach statistical significance. This indicates that ADMA-GR inclusions are unlikely to be associated with age of disease onset in *C9orf72* cases.

The second statistically significant correlation identified in the QSBB cohort was a positive correlation between disease duration and the number of SDMA-GR inclusions in the frontal cortex. This correlation suggests that SDMA-GR may be protective with patients that have higher numbers of SDMA-GR inclusions in the frontal cortex having a longer disease duration. This fits with the initial observation of low SDMA-GR numbers in the homozygous case which had a disease duration of only two years. The correlation remained significant when the larger cohort of *C9orf72* cases was analysed, suggesting that SDMA-GR inclusions in the frontal cortex robustly correlates with disease duration in *C9orf72* cases. Furthermore, the correlation analysis in the combined brain bank cohort also revealed a significant positive correlation between SDMA-GR inclusions in the frontal cortex and age at death. Once again indicating the potentially protective nature of SDMA-GR inclusions in this brain region.

Given the recent publications that suggest arginine methylation may alter a proteins ability to phase separate and thus affect its propensity to aggregate (Hofweber et al., 2018; Qamar et al., 2018), the positive correlation of SDMA-GR in the frontal cortex with disease duration and age at death could be hypothesised to be due to the symmetric dimethylation of poly-GR reducing the toxic aggregation of poly-GR within neurons of this brain region. This could result in less neurodegeneration of this region, which may lead to longer disease duration and later age at death. However, it is important to stress that this hypothesis relies on several assumptions which would need further investigation before mechanistic conclusions surrounding this correlation can be drawn.

Firstly, it relies on the assumption that arginine methylation reduces the phase separation and aggregation of poly-GR, as has been demonstrated for FUS (Hofweber et al., 2018; Qamar et al., 2018). For the FUS protein, arginine methylation regulates phase separation by interfering with cation-π bonds formed between the guanidino moiety on the arginine side-chain and the benzene ring of tyrosine residues (Qamar et al., 2018). This mechanism is unlikely to be the same for poly-GR as the amino acid composition of the protein would not result in the formation of cation-π bonds. It is possible, however, that phase separation occurs by an alternative method for this protein, which arginine methylation may similarly interfere with. Although it is known that the arginine-containing DPR proteins can undergo phase separation (Boeynaems et al., 2017; Lee et al., 2016; Lin et al., 2016), the biophysical mechanism by which this occurs has not yet been identified. Further work is required to establish the driving force behind poly-GR phase separation and to determine whether this is affected by arginine methylation.

The hypothesised mechanism to link between SDMA-GR and increased disease duration also relies on the assumption that aggregation of poly-GR in neuronal

cytoplasmic inclusions is toxic and leads to neurodegeneration. Although poly-GR has been demonstrated to induce neuronal toxicity in several model systems (Kwon et al., 2014; Lee et al., 2016; Mizielinska et al., 2014; Tao et al., 2015), there is some debate as to whether poly-GR pathology in *C9orf72* post-mortem cases correlates with neurodegeneration in various brain regions (Mackenzie et al., 2013; Mackenzie et al., 2015; Saberi et al., 2018; Sakae et al., 2018). If symmetric dimethylation of poly-GR is able to reduce the toxic aggregation of poly-GR, it would be expected that less neurodegeneration would occur in cases with high abundance of SDMA-GR. In this study, the extent of neurodegeneration in each case was not assessed so it was not possible to determine whether SDMA-GR is also correlated with less neurodegeneration. This should be addressed in future studies.

Future studies are also required to investigate why it is only the less abundant symmetrically dimethylated form of poly-GR that is associated with longer disease duration and older age at death, and not the asymmetrically dimethylated form. There is little published research into the differences in biological roles and functional consequences of ADMA and SDMA modifications, which make it difficult to determine why one form of modification may be more protective than the other. It could be speculated, that the different conformations of the dimethylation may alter the biophysical properties of poly-GR in different ways, with only the SDMA-GR form reducing phase separation and aggregation. As previously mentioned, future investigations are required to establish the driving forces behind poly-GR phase separation and to determine how this is affected by the two forms of dimethylation in order to understand why one form may be more protective than the other.

It will also be necessary to establish why SDMA-GR correlates with disease duration and age at death only in the frontal cortex. If SDMA-GR is linked to reduced poly-GR aggregation and neurodegeneration, as previously hypothesised, it would be

expected that this association would also be replicated in the temporal cortex. However, no significant correlation was found for SDMA-GR in the temporal cortex for the analyses performed on either the QSBB only cohort or the combined brain bank cohort, despite the frontal and temporal cortices having a similar abundance of methylated-GR inclusions. The reason for this difference is unknown but could relate to potential differential levels of neurodegeneration between the two regions.

5.3.5 Correlational analysis between methyl-GR and case information when *C9orf72* ALS and FTLD cases are analysed independently

Following the correlation analysis on the combined brain bank *C9orf72* cohort, the cases were then split into two groups based on their pathological diagnosis of ALS or FTLD. The same correlation analyses were then performed on both groups independently to determine whether the association between SDMA-GR and disease duration or age at death was driven by one particular disease group.

Interestingly, when analysed separately, neither the ALS or FTLD cases showed a significant association between SDMA-GR and disease duration or age at death, indicating that neither disease group was driving this correlation. The reason for this is unknown but it is possible that the correlation was lost due to a reduction in sample size caused by analysing the two disease separately.

There were several correlations that were found to be significant when the ALS and FTLD groups were analysed separately. For example, ADMA-GR correlated significantly with age of disease onset in both the frontal and temporal cortices of the FTLD and ALS groups. However, this correlation was in the opposite direction for the two groups, with the FTLD group showing a negative correlation and the ALS group showing a positive correlation. The opposing directions of the correlation would explain why this association was not detected when the groups were analysed together, however it does not make biological sense that the same post-translation

modification would have opposing effects in a brain region where the DPR protein pathology of the two diseases is very similar. It is possible that the correlations identified in the separate ALS and FTLD analysis hold true for the small number of cases analysed in this study, however these associations should be tested in a larger *C9orf72* ALS and FTLD cohort to confirm their biological significance.

5.3.6 Limitations of the study and future work

Although this study is the first to identify specific methylation of poly-GR proteins in *C9orf72* post-mortem tissue, there are a number of limitations that should be addressed in future investigations.

ADMA-GR and SDMA-GR inclusions were identified in the frontal and temporal cortices, however lots of other brain regions are affected by DPR protein pathology and these have not yet been assessed for the presence of methylated-GR inclusions. More extensive characterisation of ADMA-GR and SDMA-GR pathology in other brain regions would be useful for assessing whether the methylation of poly-GR inclusions is associated with brain regions that are considered particularly vulnerable or resistant to neurodegeneration in *C9orf72*. For example, the cerebellum is known to have extensive DPR protein pathology but experiences relatively little neuronal loss in comparison to the frontal cortex. In this study, the extent of neurodegeneration in each case was not assessed or correlated with methyl-GR, however assessment of this could help to determine whether methylation of poly-GR is linked to pathological features of the disease.

Assessing the extent of neurodegeneration in each case is also important when considering the method of assessing ADMA-GR and SDMA-GR pathology. In this study, the number of inclusions containing ADMA-GR and SDMA-GR were quantified by manual counting, however when using post-mortem tissue, a general limitation is

that the tissue is a representation of the end-stage of disease and usually has extensive neuronal loss. This could, therefore, impact the number of inclusions quantified with more severe cases exhibiting less inclusions due to the presence of less neurons. The impact of neuronal loss was not controlled for in this study but could have important consequences on the quantification of pathological features in post-mortem tissue. One possible method of controlling for this would be to grade the extent of neurodegeneration in each case and then correlate each group with the methyl-GR counts. Similarly, the burden of DPR protein or TDP-43 pathology was not controlled for in this study but could also be assessed using a similar method or the number of methyl-GR inclusions could be normalised to the total number of poly-GR inclusions. Normalisation of ADMA-GR and SDMA-GR inclusions counts would help to account for the differences in extent of DPR protein pathology or neurodegeneration in each case.

The correlation analysis performed in this study revealed that SDMA-GR positively correlates with disease duration and age at death when *C9orf72* FTLD and ALS cases are analysed together, but this correlation is lost when the two disease groups are analysed separately. Importantly, the analyses performed in this study only asses the linear relationship between two variables and does not account for variables that may be confounding the relationship. For example, in this study the sex of the patient or the brain bank where the case was processed could have a significant effect on the correlations that were identified. To account for this, it would be better to perform multiple linear regression analysis on the relationships that were found to be significant, as this would enable the effect of cofounding factors to be considered. However, this was not possible in the current study because the data did not meet the assumptions or requirements for this analysis. This was mostly due to the small sample size, the fact that several cases had missing clinical data, and the fact that

several of the cofounding variables were correlated with each other, for example age at onset and age at death.

Finally, although this study has identified a relationship between SDMA-GR and disease duration, suggesting that this post-translational modification may have protective effects, the biological mechanisms underlying this relationship remain unknown. Although, based on the literature, it can be hypothesised that the arginine methylation may be altering the phase transition of the poly-GR protein, the physiological consequences of adding methyl groups to poly-GR is currently unknown and should be explored in future cellular and biochemical studies. This will be necessary before conclusions can be drawn about how symmetric dimethylation of poly-GR may lead to a longer disease duration

5.3.7 Conclusion

In summary, this study has identified that poly-GR proteins in *C9orf72* FTLD and ALS patients can be post-translationally modified by arginine methylation. Using novel antibodies designed to specifically target unmethylated and dimethylated forms of poly-GR, this study confirms that asymmetrically and symmetrically dimethylated forms of the arginine-rich DPR protein can be detected in the frontal and temporal cortices of *C9orf72* patient post-mortem brain tissue. Quantification of these inclusions showed ADMA-GR to be more abundant than SDMA-GR in both brain regions. Correlating the number of inclusions containing each of the methyl-GR isoforms with various pieces of clinical information revealed a positive correlation between SDMA-GR and disease duration in a large cohort of *C9orf72* cases, indicating that this form of methylation may be protective in disease. Further exploration of this relationship showed that this correlation was not present when *C9orf72* – FTLD or *C9orf72* – ALS cases were analysed separately; however, this may be attributed to the small size of the sample analysed and requires further

investigation in larger cohorts. The biological consequences of arginine methylation on poly-GR is currently unknown however, recent work has demonstrated that the methylation of arginine residues can have a significant impact on a protein's biophysical properties, such as its propensity to phase separate and aggregate. Future work is required to investigation whether specific forms of arginine methylation has a similar effect on the aggregation poly-GR as this may help determine the mechanism by which symmetrically dimethylated forms of poly-GR correlates with longer disease duration in *C9orf72*-linked disease.

Chapter 6.   Heterogeneous nuclear ribonucleoproteins R and Q label pathological inclusions in FTLD-FUS

6.1 Introduction

6.1.1 FTLD-FUS

FTLD-FUS is the term used to describe approximately 5 – 10 % of FTLD cases that contain the FUS protein within pathological inclusions (Mackenzie et al., 2011). Following the identification of several mutations in the *FUS* gene as causative of 26 cases of familial ALS (Kwiatkowski et al., 2009; Vance et al., 2009), immunohistochemical studies performed on a subset of tau- and TDP-43 negative FTLD cases identified the presence of the FUS protein in ubiquitin-positive neuronal cytoplasmic and intranuclear inclusions in aFTLD-U cases (Neumann et al., 2009a; Urwin et al., 2010). Subsequent screening of other FTLD cases with tau- and TDP-43 negative pathology also identified the FUS protein as a component in inclusions in neuronal filament inclusion disease (NIFID), and in basophilic inclusion body disease (BIBD) (Lashley et al., 2011; Munoz et al., 2009; Neumann et al., 2007; Neumann et al., 2009b). The identification of pathological FUS inclusions in these diseases led to the formation of the FTLD-FUS classification which encompasses the three pathological diagnoses; NIFID, BIBD, and aFTLD-U. Although there is some overlap, these diagnoses are considered clinically and neuropathologically distinct diseases, however most patients present with early-onset sporadic bvFTD and, pathologically, all cases contain FUS immunoreactive inclusions (Lashley et al., 2011; Mackenzie et al., 2011). Unlike ALS-FUS, to date no mutations in the *FUS* gene have been identified as causative of FTLD-FUS, however it is interesting to note that two related aFTLD-U cases have been reported (mother and son) and are included in the present

study. In addition to lacking *FUS* mutations, these cases were also negative for mutations in the *MAPT, GRN, VCP, CHMP2B* or *TARDBP* genes, suggesting that there may be other genetic causes of FTLD-FUS (Lashley et al., 2011; Rohrer et al., 2011).

6.1.2 FUS

The 53 kDa FUS protein (also known as translocated in liposarcoma (TLS)) is composed of 526 amino acids and is encoded by the *FUS* gene located on chromosome 16 (Aman et al., 1996; Crozat et al., 1993; Yang et al., 2010). It is a ubiquitously expressed, multi-functional DNA/RNA-binding protein that primarily localises to the nucleus but shuttles back and forth into the cytoplasm via interaction with nuclear importers and exporters (Aman et al., 1996; Andersson et al., 2008).

FUS is made up of multiple protein domains which regulate its biochemical behaviour and RNA and protein interactions (Deng et al., 2014; Iko et al., 2004; Sama et al., 2014). The N-terminus of the protein contains a serine-tyrosine-glycine-glutamine (SYGD) rich domain and an arginine-glycine (RGG) rich domain, which together are known as the low complexity domain (LCD), enabling the protein to undergo reversible liquid-liquid phase transition to form transient membrane-less organelles (Han et al., 2012; Kato et al., 2012; Murakami et al., 2015; Patel et al., 2015; Sun et al., 2011). The centre of the protein contains the single RNA recognition motif (RRM) and a nuclear export signal, followed by two more RRG domains, one zinc finger domain, and a non-classical proline-tyrosine nuclear localisation signal (PY-NLS) at the C-terminus of the protein. The arginine residues within the RGG domains of FUS are known to be extensively post-translationally methylated and are also thought to regulate the phase transition of FUS (Hofweber et al., 2018; Qamar et al., 2018). The multiple domains are required to mediate numerous protein–RNA and protein–protein

interactions that enable FUS to carry out a diverse range functions at both the transcriptional and post-transcriptional level (Ratti and Buratti, 2016).

FUS
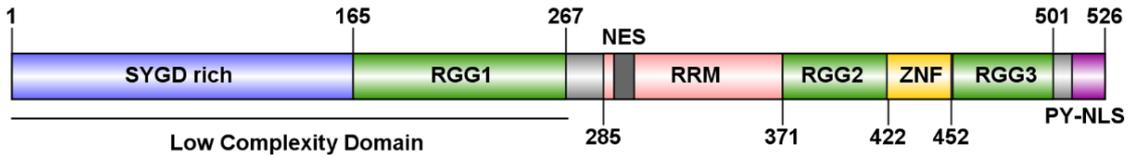
Figure 6.1 Protein domains of FUS

Diagram of the protein domains in FUS. Numbers correspond to amino acid number within the protein. SYGD = serine-tyrosine-glycine-glutamine rich domain, RRM = RNA recognition motif, RGG = arginine and glycine rich domain, NES = nuclear export signal, ZNF = zinc finger domain, PY-NLS = non-classical proline-tyrosine nuclear localisation signal. Diagrams produced using Illustrator for Biological Sequences software (Liu et al., 2015).

The physiological functions of FUS are wide ranging, owing to its ability to bind DNA, RNA and other proteins (Efimova et al., 2017; Ratti and Buratti, 2016; Sama et al., 2014). Through its interactions with both single- and double-stranded DNA, FUS is thought to play a crucial role in DNA repair, particularly in homologous recombination, and maintaining genomic stability (Baechtold et al., 1999; Hicks et al., 2000; Mastrocola et al., 2013). Additionally, FUS can bind to the promoters of over 1000 genes, indicating the importance of its role in transcriptional regulation. FUS also has a key role in many aspects of RNA metabolism and is thought to bind, or be involved in the processing and splicing of thousands of mRNAs (Colombrita et al., 2012; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012), including its own pre-mRNA to autoregulate its expression (Zhou et al., 2013).

The involvement of FUS in the splicing of pre-mRNAs has been widely reported and is supported by its binding to key components of the spliceosome, such as U1 small nuclear ribonucleoprotein and the Sm-snRNP complex (Gerbino et al., 2013; Yamazaki et al., 2012). FUS also contributes to the processing of several miRNAs,

and its binding to the 3'UTR sequence of selected target mRNAs indicates a role for FUS in regulating additional aspects in the life cycle of mRNA in the cytoplasm such as mRNA stability, transport and local translation. In neuronal cells, FUS is known to participate in the transport of mRNA out of the nucleus into the cytoplasm, and the movement of RNA into dendrites (Belly et al., 2005; Fujii and Takumi, 2005; Yasuda et al., 2013; Zinszner et al., 1997). The movement of FUS and associated RNA is believed to have functional importance for local protein translation, particularly in cells such as neurons, where protein translation occurs at a long distance from mRNA biogenesis.

The role of FUS in the formation and maintenance of cytoplasmic ribonucleoprotein (RNP) granules has been subject to intense research in recent years due to the emerging importance of RNA binding proteins and RNP granules in pathological aggregate formation in neurodegenerative diseases, particularly in ALS and FTLD. RNP granules are dynamic complexes formed between RNA and RNA-binding proteins (RBPs), such as stress granules, which are usually assembled in response to environmental or metabolic stress stimuli, allowing the cell to limit translation of cellular mRNAs to conserve resources and express the appropriate repertoire of proteins required to re-establish proteostasis (Alberti et al., 2017; Protter and Parker, 2016; Standart and Weil, 2018). As previously mentioned, the N-terminal LCD of FUS enables the protein to undergo reversible liquid-liquid phase transition to form membrane-less RNP organelles, such as stress granules, particularly in response to hyperosmolar stress (Sama et al., 2013). Whilst the accumulation of FUS into RNP granules is clearly an important cellular function of FUS, it has been proposed that aberrant and irreversible liquid-solid phase transition of FUS may be a mechanism initiating or contributing to the accumulation of FUS into pathological aggregates in FTLD-FUS and ALS-FUS (Bowden and Dormann, 2016; Murakami et al., 2015; Patel et al., 2015). Importantly, despite the presence of FUS aggregates in both of these

diseases, there is believed to be significant differences in the pathogenic mechanism contributing to FUS aggregation, largely due to differences in protein composition and methylation status of the FUS inclusions.

6.1.3 Methylation of FUS in FTLD-FUS and ALS-FUS

One important distinction between FUS inclusions in FTLD-FUS and ALS-FUS is their methylation status. As previously mentioned, the RGG domains of FUS can be post-translationally modified by the addition of methyl groups to arginine residues. By using methylated-FUS specific antibodies, it has been demonstrated that FUS inclusions in ALS-FUS patient tissue contains FUS that is asymmetrically dimethylated, whilst FUS inclusions in FTLD-FUS contains FUS that is unmethylated or hypomethylated (Dormann et al., 2012; Suarez-Calvet et al., 2016). Recent work has demonstrated that the methylation status of FUS strongly influences its propensity to phase separate. Loss of arginine methylation in the C-terminal domain of FUS was found to promote its phase separation and gelation *in vitro* and increases aggregation of FUS *in vivo* (Hofweber et al., 2018; Qamar et al., 2018). Additionally, methylation of FUS alters how it interacts with its nuclear importer, transportin (TRN1), with unmethylated and monomethylated FUS exhibiting much higher binding affinities to the importer (Suarez-Calvet et al., 2016). The distinction in methylation status has important implications for pathogenesis of FTLD-FUS and ALS-FUS and has led to the hypothesis that these two diseases are caused by different mechanisms that result in the same end point of FUS accumulation. This is also reflected in the composition of the pathological inclusions found in these diseases as several other proteins have been identified as a component of FUS inclusions specifically in FTLD-FUS cases but not ALS-FUS cases.

6.1.4 Other proteins in FTLD-FUS inclusions

In 2011 TRN1 was identified as a marker of FTLD-FUS as immunohistochemical and co-localisation studies revealed strong TRN1 immunoreactivity of FUS positive inclusions, indicating that it is likely that both FUS and TRN1 are major components of the inclusions in FTLD-FUS (Brelstaff et al., 2011). Transportin, also known as M9-interacting protein or karyopherin β2 (Kar β2), is the protein responsible for the transport of FUS, and several other proteins, from the cytoplasm to the nucleus. It binds to its cargoes via their PY-NLS and translocates into the nucleus through its association with the nuclear pore complex (Weis, 2003). Recently, TRN1 has also been shown to act as a molecular chaperone for FUS. Via its interaction with arginine residues within the RGG domains of FUS in the cytoplasm, TRN1 can alter the phase separation behaviour of FUS, thereby reducing the formation of FUS aggregates (Guo et al., 2018; Hofweber et al., 2018; Qamar et al., 2018; Yoshizawa et al., 2018).

Interestingly, TRN1 does not co-localise with FUS inclusions in ALS-FUS, which implicates two different diseases mechanisms (Neumann et al., 2012). The accumulation of FUS in ALS-FUS is hypothesised to be caused by mutations in the FUS protein that impair its ability to bind to TRN1 resulting in an accumulation of FUS in the cytosol (Dormann et al., 2010; Kwiatkowski et al., 2009; Vance et al., 2009). In contrast, the accumulation of FUS in FTLD-FUS is hypothesised to be driven by its hypomethylated status and increased propensity to phase separate. These physical characteristics, along with the high affinity of hypomethylated FUS for TRN1, are thought to result in the co-deposition of these proteins, and others, in pathological aggregates.

In addition to FUS, TRN1 is the nuclear import receptor for several other proteins and some of these have also been demonstrated to be present in FTLD-FUS inclusions (Gami-Patel et al., 2016; Neumann et al., 2011; Neumann et al., 2012). A screen of 13 different TRN1 cargoes in 2012 revealed that Ewing's sarcoma (EWS) protein and

TATA-binding protein-associated factor 15 (TAF15) protein are also present in FUS inclusions selectively in FTLD-FUS cases, although to a lesser extent than TRN1 (Neumann et al., 2012). EWS, TAF15 and FUS make up the FET family of proteins. These are highly conserved, nuclear proteins that are ubiquitously expressed and involved in various aspects of DNA and RNA metabolism, including RNA processing, transcription, splicing, transport and DNA repair (Andersson et al., 2008; Kovar, 2011; Law et al., 2006; Tan and Manley, 2009). FET proteins shuttle continuously between the nucleus and cytoplasm via the interaction between their PY-NLS, and their nuclear import protein, TRN1 (Jobert et al., 2009; Zakaryan and Gehring, 2006; Zinszner et al., 1997). FTLD-FUS inclusions were negative for all other TRN1 cargoes tested, implying that the accumulation of proteins in pathological aggregates in FTLD-FUS cases was specific to FET proteins (Neumann et al., 2012). This finding lead to the hypothesis that a generalised dysregulation of TRN1-mediated transport of FET proteins may underlie FTLD-FUS pathology (Neumann et al., 2012).

However, a more recent study has demonstrated that it is not only FET proteins that are components of pathological inclusions in FTLD-FUS cases. In 2016, Gami-Patel *et al.* identified an additional TRN1 cargo, hnRNP A1, to label a proportion of FUS-positive neuronal cytoplasmic and intranuclear inclusions (Gami-Patel et al., 2016). This finding extended the spectrum of proteins in FTLD-FUS inclusions to include non-FET proteins and led the authors to investigate the presence of other RNA binding proteins in FUS inclusions. The study focused on the heterogeneous nuclear ribonucleoprotein (hnRNP) family, also prompted by the fact that, in addition to being a FET protein, FUS can also be classified as an hnRNP (hnRNP P2) (Calvio et al., 1995). Pathological, biochemical and expression analysis of 11 hnRNPs indicated the presence of hnRNP D, G, I and L in pathological deposits in neurons of FTLD-FUS cases, in addition to hnRNP A1, (Gami-Patel et al., 2016). Like FUS, hnRNPs perform a diverse range of functions linked to RNA metabolism, therefore, the presence of

various hnRNPs in FUS inclusions implies that the pathogenesis of FTLD-FUS extends beyond FET proteins, TRN1 cargoes and dysfunctional nuclear import, and implicates a wider dysregulation of RBP metabolism.

6.1.5 Heterogeneous ribonuclear proteins

hnRNPs are a large family of primarily nuclear proteins, named hnRNPs A1 - U, that associate with RNA, but do not form a stable component of ribonucleoprotein complexes (Dreyfuss et al., 1988). They are a multi-functional and structurally diverse group of proteins, with the majority shuttling between the nucleus and cytoplasm to carry out a variety of functions linked to nucleic acid metabolism. Although hnRNPs do not all share a common structural domain, each hnRNP contains one, or more, of the following RNA binding domains; an RNA recognition motif (RRM), a quasi-RRM, a K-homology domain or an arginine and glycine-rich domain. Additionally, most hnRNPs contain at least one auxiliary domain with distinctive amino acid compositions, which are thought to mediate protein-protein interactions and localise the proteins within a cell (Dreyfuss et al., 1993).

hnRNPs are involved in nearly every step of mRNA processing and their functions can broadly be split into those that primarily occur in the nucleus – transcription, splicing, 5' capping and polyadenylation – and those that occur in the cytoplasm - mRNA transport, stability, translation and degradation. The different hnRNPs frequently have overlapping functions and often perform their functions as part of a larger co-operative protein complex, however they also have individual specialised roles that are dependent on specific RNA-protein or protein-protein interactions (Dreyfuss et al., 1993; Han et al., 2010). The ability to move between the nucleus and cytoplasm is crucial for hnRNPs to perform their variety of functions and, as such, the ability to undergo nucleocytoplasmic shuttling is thought to be a key characteristic of hnRNPs.

It is currently unknown whether the hnRNPs, or indeed any of the other proteins, identified in FUS inclusions contribute to the pathogenesis of disease or are innocent bystanders. Whilst the accumulation of TRN1 and FET proteins in FUS inclusions supports the notion of dysfunctional nuclear import in FTLD-FUS, the presence of hnRNPs in FUS inclusions suggests a wider dysregulation of RBPs which could be linked to many of their functions. This finding has highlighted the need for further investigation into the potentially pathogenic role of hnRNPs in FTLD.

Recently, FUS has been linked to an additional hnRNP, which has currently not been characterised in FTLD-FUS. In a series of individual nucleotide-resolution cross-linking and immunoprecipitation (iCLIP) experiments in mouse primary motor neurons and NSC34 cells, FUS was identified as an RNA interactor of hnRNP R (Briese et al., 2018). Although these experiments were carried out on healthy cells and not post-mortem tissue, an interaction between FUS RNA and hnRNP R *in vivo* implicates a further relationship between FUS and hnRNP R that should be explored.

6.1.6 hnRNP R

HnRNP R is an 80 kDa RNA-binding protein consisting of 633 amino acids that are arranged into three RRMs, one acidic rich domain and RGG domain (Figure 6.2) (Huang et al., 2005). It is widely expressed in neuronal tissue, with some evidence suggesting its expression is regulated in response to circadian rhythm (Lee et al., 2015; Peng et al., 2009). Little is known about this hnRNP in the context of neurodegenerative diseases and in neuronal cell death. hnRNP R is known to be involved in pre-mRNA processing as a necessary nuclear protein in the splicing machinery, and directly interacts with survival motor neuron (SMN) protein, with hnRNP R considered to be a linker between SMN and the spliceosome (Dombert et al., 2014; Mourelatos et al., 2001). HnRNP R also exists as a cytosolic protein in the axons of motor neurons, functioning in axonal RNA transport and processing

(Dombert et al., 2014; Rossoll et al., 2002). Additionally, hnRNP R is known to facilitate and regulate the transcription of genes including the proto-oncogene, c-fos (Fukuda et al., 2009; Fukuda et al., 2013; Huang et al., 2008)

hnRNP R has high sequence homology with another hnRNP protein, hnRNP Q. Sequence alignment of the canonical isoforms of these proteins reveals that they are approximately 81.2 % identical at the amino acid level (Mizutani et al., 2000). These two structurally alike proteins are expressed in parallel and recognise the same RNA binding sequence (UCUAUC). Consequently, they are known to have similar, but distinct, functional roles within the cell (Geuens et al., 2016).

6.1.7 hnRNP Q

hnRNP Q, also known as SYNCRIP, has three isoforms created by alternative splicing called, hnRNP Q1-Q3. The longest isoform, hnRNP Q3, is composed of 623 amino acids, and, like hnRNP R, also contains three RRMs, one acidic rich domain and an RGG domain (Figure 6.2). hnRNP Q also shares functional similarity to hnRNP R as it is also involved in the maintenance of circadian rhythms (Kim et al., 2011; Kim et al., 2010; Lee et al., 2012; Lim et al., 2016), interacts with the SMN protein, and is involved in pre-mRNA splicing as a component of the spliceosome (Chen et al., 2008; Mourelatos et al., 2001). Additionally, hnRNP Q has been shown to be involved in the regulation of mRNAs responsible for neuronal morphogenesis (Chen et al., 2012), exosomal sorting of specific miRNAs (Santangelo et al., 2016) and in the regulation of translation of p53 (Kim et al., 2013a).

hnRNP R
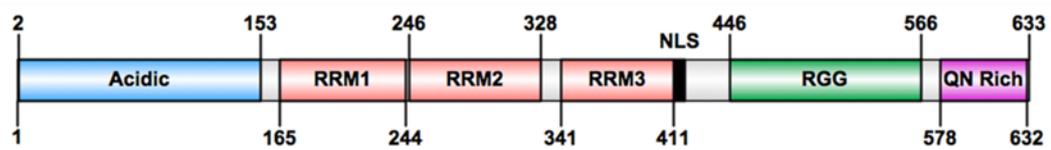
hnRNP Q
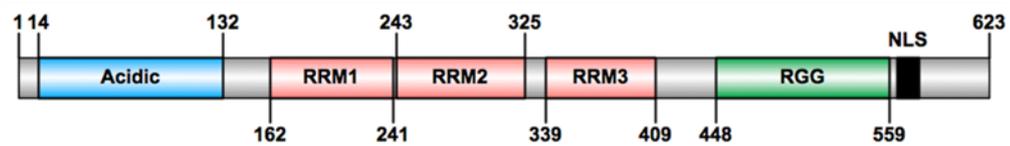
Figure 6.2 Protein domains of hnRNP R and hnRNP Q

Diagram of the protein domains present in the hnRNP R and hnRNP Q proteins. Numbers correspond to amino acid number within the proteins. Acidic = acidic rich domain, RRM = RNA recognition motif, RGG = arginine and glycine rich domain, NLS = nuclear localisation signal, QN = glutamine and asparagine rich domain. Diagrams produced using Illustrator for Biological Sequences software (Liu et al., 2015).

6.1.8 Chapter 6 aims

Given the identification of various hnRNP proteins in pathological inclusions in FTLD-FUS, it is clear that it is not only a dysfunction in FET protein transport that contributes to pathological protein aggregation in this disease. Whether they are a cause or consequence of disease pathogenesis remains to be explored, in addition to the characterisation of other potential hnRNPs and RBPs that may co-accumulate with FUS in pathological aggregates. The current study characterises the expression and accumulation of two additional hnRNP proteins, hnRNP R and hnRNP Q, in FTLD-FUS, using post-mortem tissue from pathologically confirmed FTLD-FUS cases. The principle aims of Chapter 6 were to;

- Compare the expression of hnRNP R mRNA in several FTLD subtypes.

- Utilise immunohistological techniques to compare the distribution of hnRNP R and hnRNP Q in several FTLD subtypes

- Characterise the accumulation of hnRNP R and hnRNP Q in FTLD-FUS using immunohistological and biochemical techniques.

6.2 Results

6.2.1 hnRNP R mRNA expression is increased in some FTLD subtypes

The mRNA expression of hnRNP R was analysed in the frontal and temporal cortices of FTLD-FUS and FTLD-TDP (type A, type B and type C) cases and compared to expression in neurologically normal control cases with no pathological abnormalities. mRNA expression was analysed using NanoString technology for high-sensitive capture of mRNA transcripts. Normalised expression indicated that hnRNP R mRNA expression was significantly increased in FTLD-TDP type A (p = 0.0005), FTLD-TDP type C (p = 0.0036) and FTLD-FUS (p = 00048) subtypes relative to controls (Figure 6.3). No significant difference in hnRNP R expression was found between controls and FTLD-TDP type B subtype (p = 0.5739).
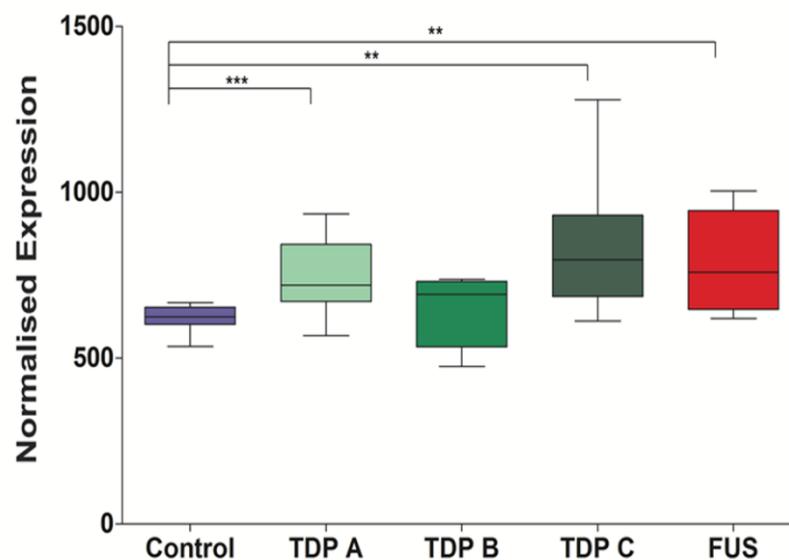
Figure 6.3 hnRNP R expression is increased in some FTLD subtypes

Nanostring expression analysis of hnRNP R mRNA levels in the frontal and temporal cortices of FTLD-TDP and FTLD-FUS cases compared to neurologically normal controls. A significant increase in expression of hnRNP R was identified in FTLD-TDP A (p = 0.0005, n = 19), FTLD-TDP C (p = 0.0036, n = 7) and FTLD-FUS (p = 0.0048, n = 5) compared to controls (n = 6). No significant difference was found between controls and FTLD-TDP B (p = 0.5739, n = 3). One-way analysis of variance.

6.2.2 Localisation of hnRNP R in FTLD subtypes

Given that hnRNP R expression was increased in several FTLD subtypes, the cellular distribution of hnRNP R was investigated using immunohistochemical staining in the frontal cortex and hippocampus of FTLD (TDP-A, TDP-B, TDP-C, and FUS) cases. Strong neuronal nuclear staining of hnRNP R was observed in controls and all FTLD subtypes, with occasional neurons additionally showing a weaker cytoplasm stain (Figure 6.4). The intensity of the neuronal staining varied among cases and was thought to be due to variation in fixation time; the shorter the fixation time, the higher the intensity of the staining. In addition to this normal pattern of staining, all FTLD-FUS cases additionally showed the presence of the hnRNP R protein in pathological inclusions. No inclusions containing hnRNP R were detected in neurologically normal control cases or the other FTLD subtypes.

6.2.3 Localisation of hnRNP Q in FTLD subtypes

Given the extensive homology between hnRNP R and hnRNP Q, immunohistological staining was also used to assess the cellular distribution of hnRNP Q in the frontal cortex and hippocampus of control and FTLD (TDP-A, TDP-B and FUS) cases. The staining replicated what was observed for hnRNP R with both controls and all FTLD subtypes showing strong neuronal nuclear staining and occasional weaker cytoplasm staining (Figure 6.5). hnRNP Q labelled pathological inclusions in the FTLD-FUS cases but not in any other FTLD subtypes or in neurological normal controls.
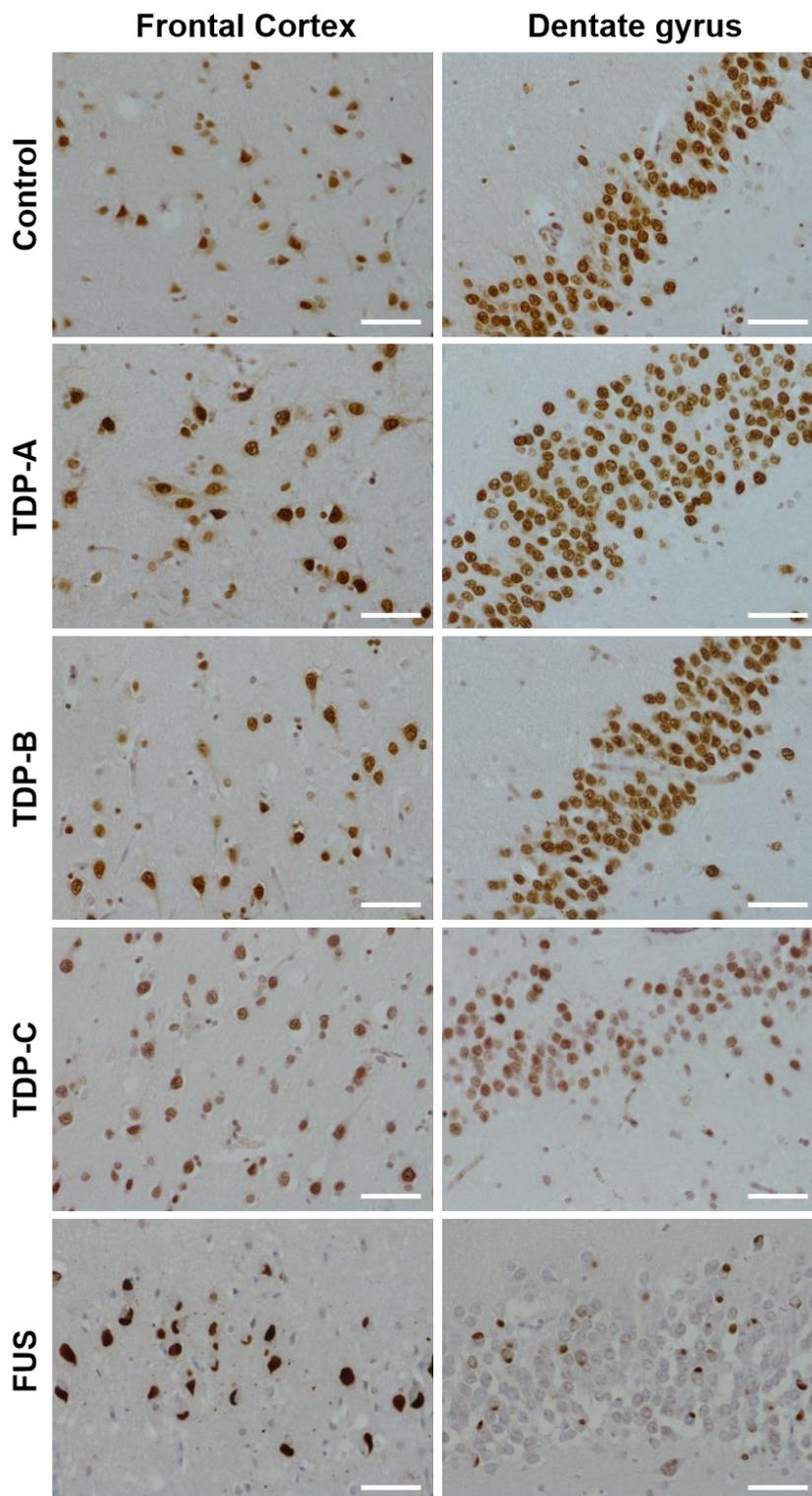
Figure 6.4 hnRNP R labels inclusions specifically in FTLD-FUS

Representative images of hnRNP R immunohistochemistry in the frontal cortex and granular cell layer of the dentate gyrus in a neurologically normal control and several FTLD subtypes (TDP-A, TDP-B, TDP-C and FUS). Strong neuronal nuclear hnRNP R staining is observed in controls and the FTLD-TDP subtypes, with occasional weak cytoplasmic staining in some neurons. FTLD-FUS cases show reduced nuclear hnRNP R staining and the presence of hnRNP R immunoreactive neuronal cytoplasmic inclusions. Scale bars represent 50 μm.
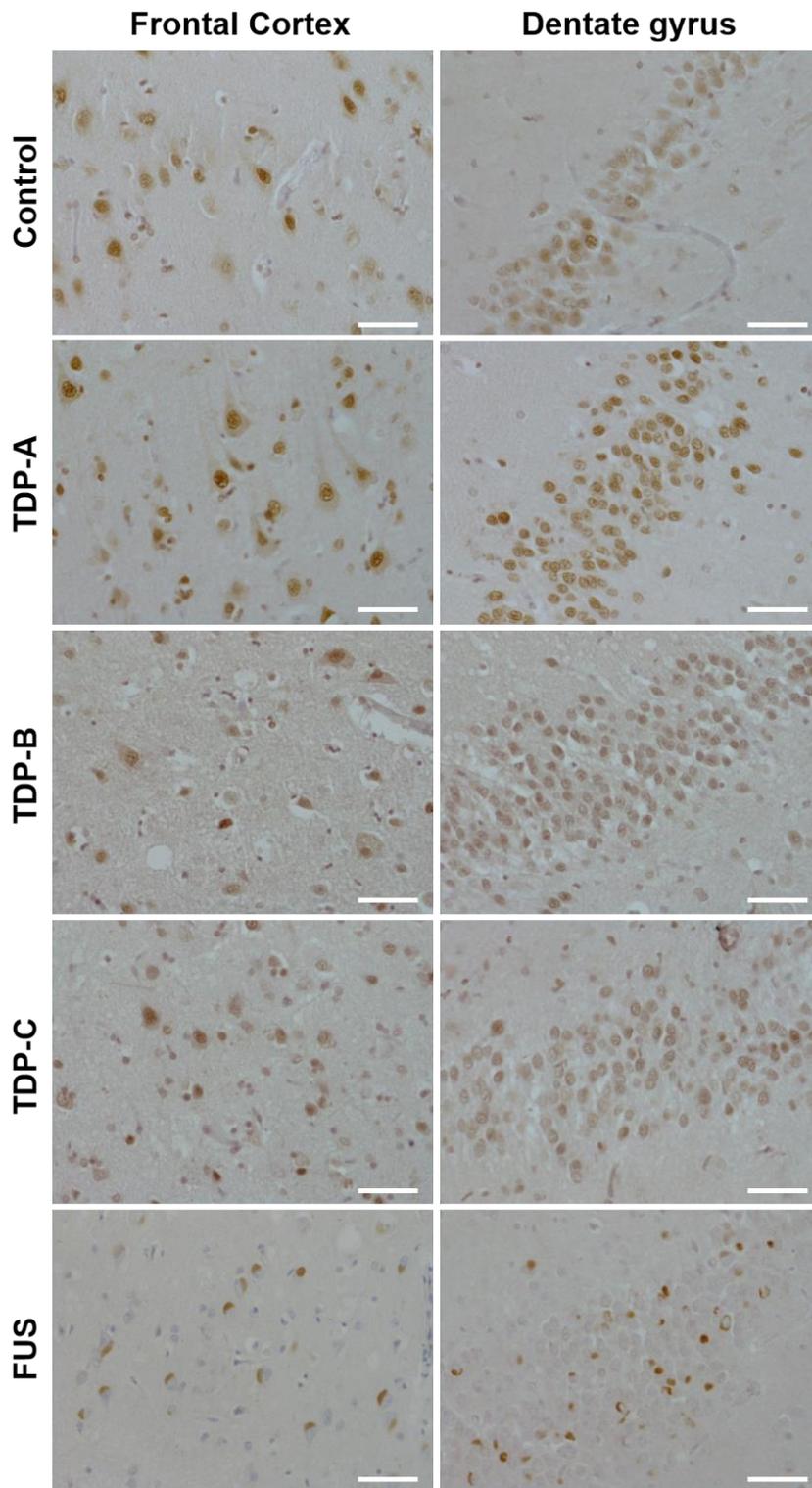
Figure 6.5 hnRNP Q labels inclusions specifically in FTLD-FUS

Representative images of hnRNP Q immunohistochemistry in the frontal cortex and granular cell layer of the dentate gyrus in a neurologically normal control and several FTLD subtypes (TDP-A, TDP-B, TDP-C and FUS). Strong neuronal nuclear hnRNP Q staining is observed in controls and the FTLD-TDP subtypes, with occasional weak cytoplasmic staining in some neurons. FTLD-FUS cases show reduced nuclear hnRNP Q staining and the presence of hnRNP R immunoreactive neuronal cytoplasmic inclusions. Scale bars represent 50 μm.

6.2.4 Characterisation of hnRNP R and hnRNP Q localisation in FTLD-FUS

Post-mortem brain tissue was available from six NIFID and seven aFTLD-U cases available at QSBB, and these were immunohistochemically assessed for hnRNP R and hnRNP Q pathology. No BIBD cases were available for assessment. All FTLD-FUS cases investigated, both of the NIFID and aFTLD-U subtype, had abundant hnRNP R and hnRNP Q immunoreactive neuronal cytoplasmic inclusions, and occasional neuronal intranuclear inclusions, in the frontal cortex and hippocampal granule cell layer (Figure 6.6 and Figure 6.8). In some, but not all, neurons in the FTLD-FUS cases, the presence of hnRNP R or hnRNP Q immunoreactive inclusions depleted the normal nuclear staining of these proteins. However, this was variable between cases, and is thought to be due to variation in tissue fixation time. In general, cases that had a longer fixation time tended to show weaker nuclear staining.

Between each case, the abundance and distribution of inclusions stained with hnRNP R or hnRNP Q varied but within each case the staining was comparable to FUS and TRN1 immunohistochemistry, which has previously been reported for these cases (Brelstaff et al., 2011; Lashley et al., 2011). As with FUS and TRN1 staining, the NIFID cases consistently showed more hnRNP R and hnRNP Q positive inclusions than aFTLD-U cases (Figure 6.12 and Figure 6.13)

The morphology of the hnRNP R and hnRNP Q inclusions was variable within and between each case, however, the inclusions formed structures resembling those previously seen for FUS and TRN1 staining in these cases. For example, Pick-like or crescent-shaped neuronal cytoplasmic inclusions, or rod-like structures in neuronal intranuclear inclusions (Figure 6.8 and Figure 6.9).

In the frontal cortex, crescent-shaped neuronal cytoplasmic inclusions surrounding the nucleus containing hnRNP R and Q were frequently observed in NIFID cases, as

well as dense, bean-shaped or larger Pick-like structures adjacent to the nucleus. Neuronal intranuclear inclusions were also observed in the NIFID cases but at a lower frequency than neuronal cytoplasmic inclusions. These typically formed a rod-like structure through the nucleus.

Similar to FUS and TRN1 staining, hnRNP R and hnRNP Q immunoreactive neuronal cytoplasmic and intranuclear inclusions were much less frequent in the cerebral cortex of aFTLD-U cases than NIFID cases. However, some aFTLD-U cases also showed hnRNP R and hnRNP Q staining of dystrophic neurites in the cerebral cortex.

In the hippocampus, hnRNP R and hnRNP Q immunoreactive neuronal cytoplasmic inclusions were observed in dentate gyrus granular cell layer in both NIFID and aFTLD-U subtypes, although the frequency of inclusions was much higher in the NIFID subtype. These inclusions were typically bean-shaped or Pick-like structures adjacent to the nucleus, however occasional crescent-shaped inclusions surrounding nuclei were also observed. In some cases, hnRNP R and hnRNP Q vermiform neuronal inclusions were also observed in the granular cell layer in both NIFID and aFTLD-U subtypes.
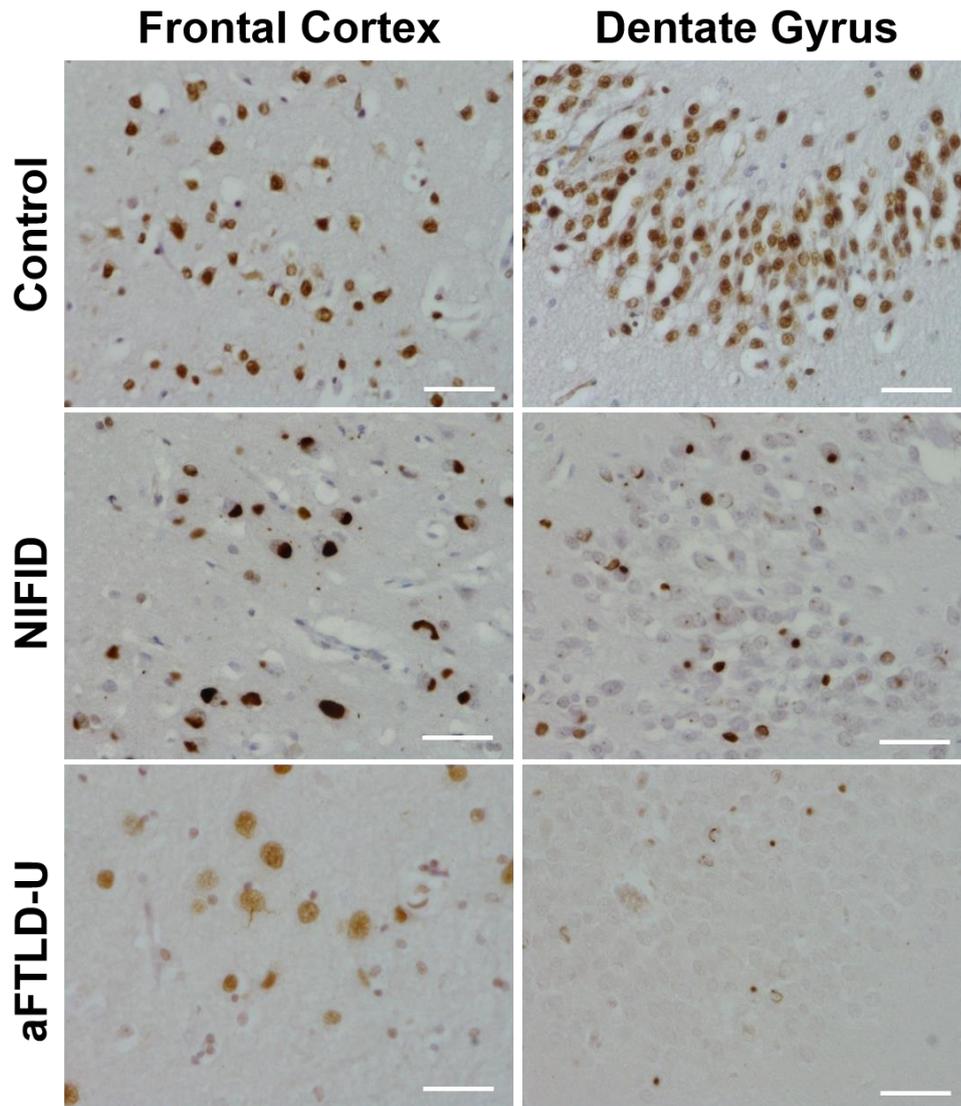
Figure 6.6 hnRNP R labels pathological inclusions in both FTLD-FUS subtypes

Representative images of hnRNP R immunohistochemistry in the frontal cortex and the granular cell layer of the dentate gyrus in the hippocampus of a neurologically normal control, NIFID and aFTLD-U case. hnRNP R stains the neuronal nuclei in neurologically normal controls but stains pathological cytoplasmic and intranuclear inclusions in both subtypes of FTLD-FUS investigated. Scale bars represent 50 µm.
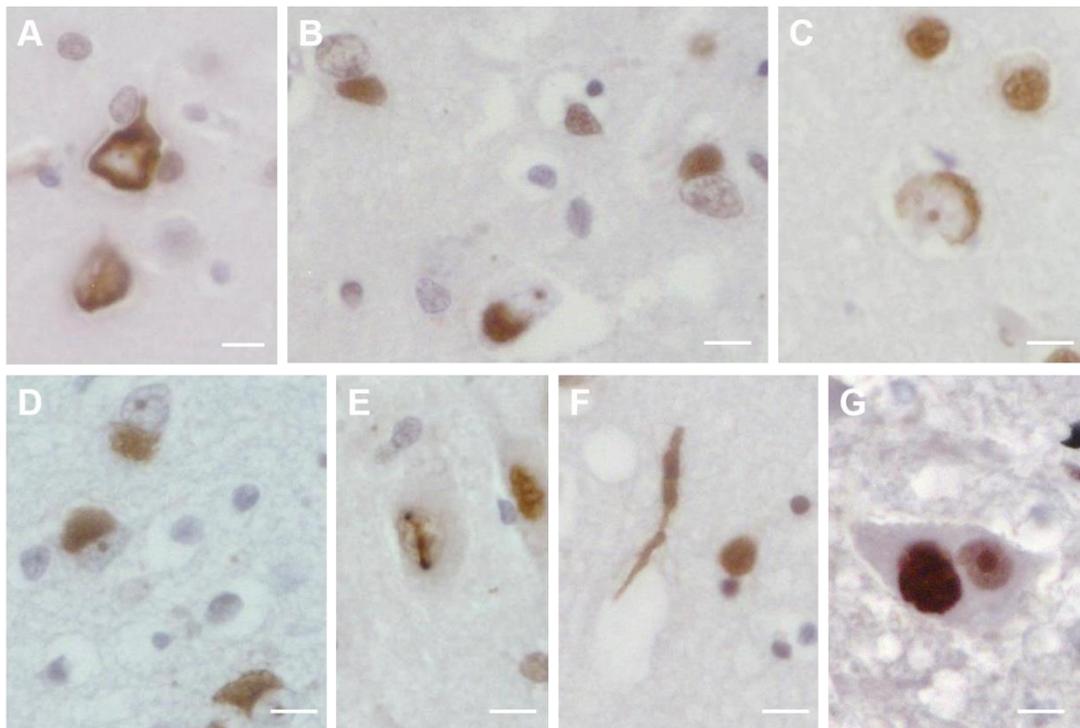
Figure 6.7 Different types of hnRNP R pathological inclusions in FTLD-FUS

Different classifications of neuronal cytoplasmic inclusions are shown (**A** - **D**), including bean-shaped (**B**), crescent shaped (**C**) and pick-like (**D**). Neuronal vermiform intranuclear inclusions (**E**) are also present along with dystrophic neurites (**F**). Pathological inclusions are also present in the upper and lower motor neurons (**G**). Scale bars represent 10 μm.

Frontal Cortex   **Dentate Gyrus**

Control

NIFID

aFTLD-U
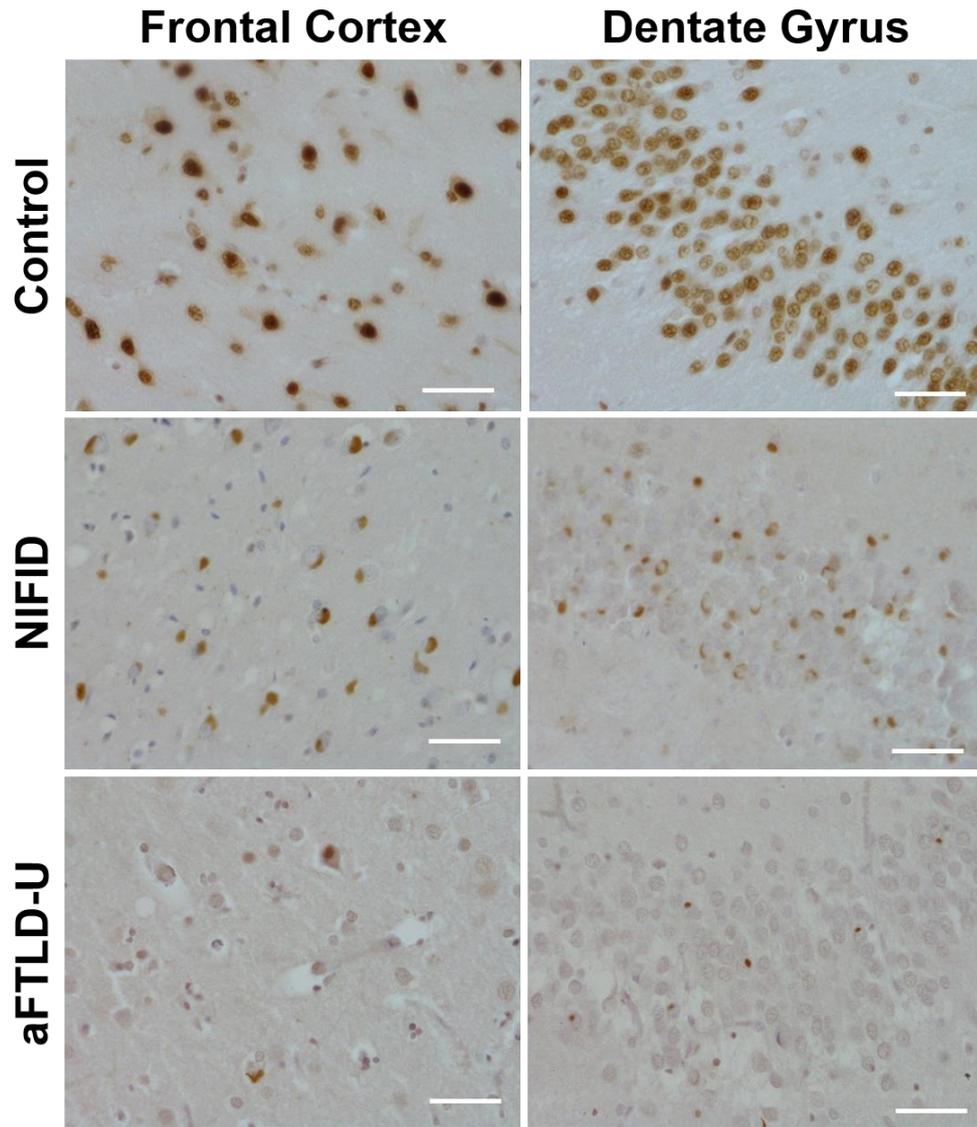
Figure 6.8 hnRNP Q labels pathological inclusions in both FTLD-FUS subtypes

Representative images of hnRNP Q immunohistochemistry in the frontal cortex and the granular cell layer of the dentate gyrus in the hippocampus of a neurological healthy control, NIFID and aFTLD-U case. hnRNP Q stains the neuronal nuclei in neurologically normal controls but stains pathological cytoplasmic and intranuclear inclusions in both subtypes of FTLD-FUS investigated. Scale bars represent 50 µm.
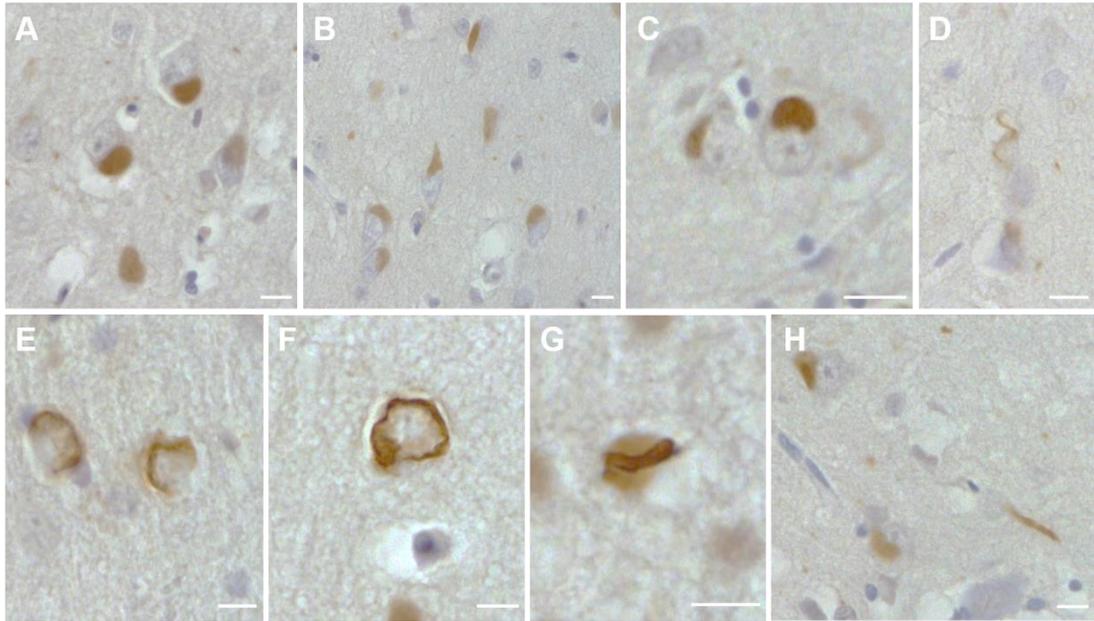
Figure 6.9 Different types of hnRNP Q pathological inclusions in FTLD-FUS

Different classifications of neuronal cytoplasmic inclusions are shown (**A** - **D** and **E** - **F**) including pick-like (**A** and **B**), bean-shaped (**C**) and crescent shaped (**E** and **F**). Neuronal vermiform intranuclear inclusions (**G**) are also present, along with dystrophic neurites (**D** and **H**). Scale bars represent 10 μm.

6.2.5 Co-localisation of hnRNP R with FUS inclusions in FTLD-FUS

To investigate whether the pathological inclusions labelled with hnRNP R also contained FUS, double immunofluorescence staining was performed on both NIFID (Figure 6.10) and aFTLD-U (Figure 6.11) cases. Due to the FUS antibody and hnRNP Q antibody being raised in the same species, double immunofluorescence staining with these two antibodies was unable to be performed. Qualitative assessment of the fluorescence images indicated good co-localisation of hnRNP R with FUS in neuronal cytoplasmic and intranuclear inclusions in the cerebral cortex and dentate gyrus, in both NIFID and aFTLD-U cases.
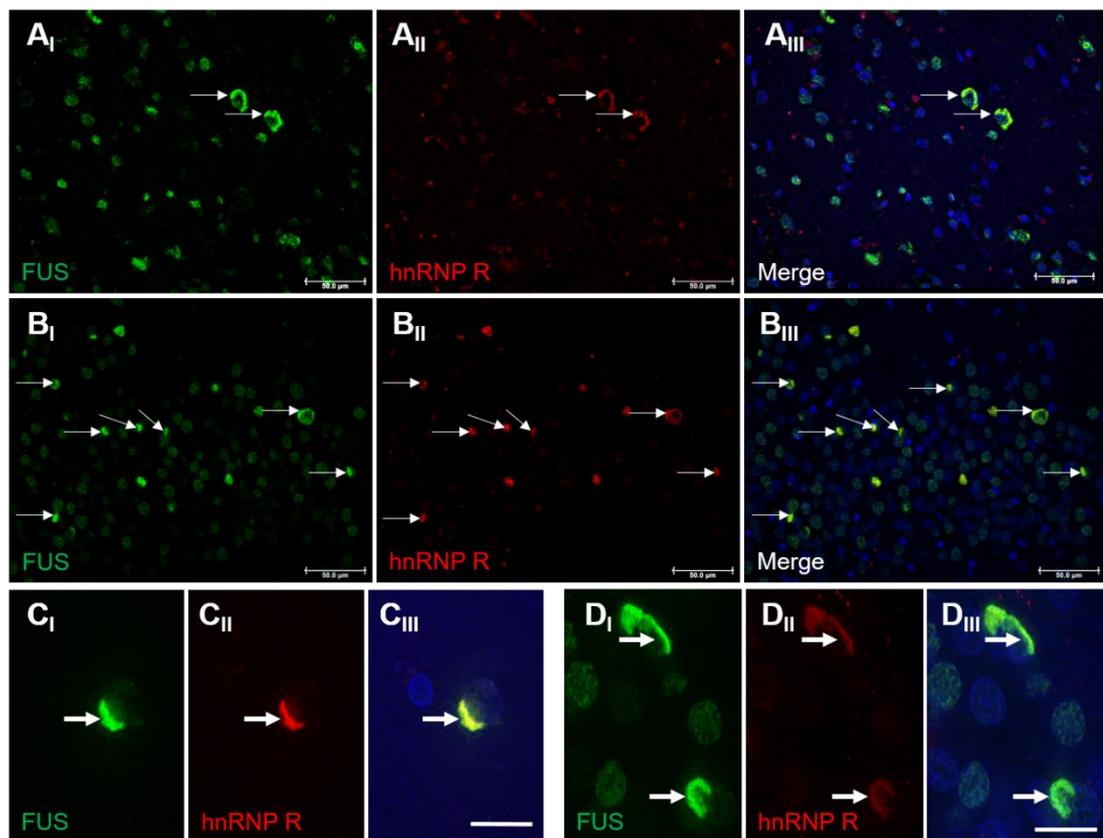
Figure 6.10 hnRNP R co-localises with FUS inclusions in NIFID cases.

Representative images of double-label immunofluorescence in the temporal cortex (**A** and **C**) and granular cell layer of the dentate gyrus (**B** and **D**) of a NIFID case demonstrating co-localisation of FUS (green) and hnRNP R (red) in neuronal cytoplasmic inclusions (white arrows). Neuronal nuclei are counterstained with DAPI. Scale bars in A and B represent 50 μm. Scale bars in C and D represent 20 μm.
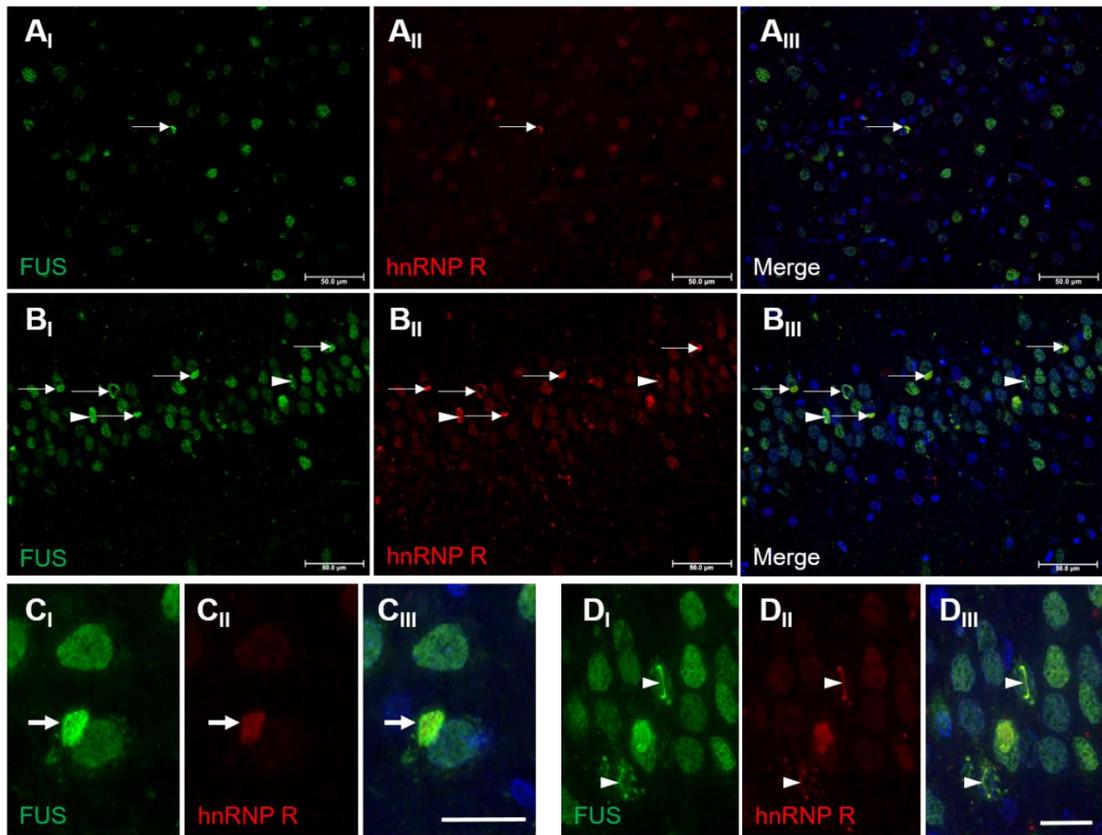
Figure 6.11 hnRNP R co-localises with FUS inclusions in aFTLD-U cases.

Representative images of double-label immunofluorescence in the temporal cortex (**A** and **C**) and granular cell layer of the dentate gyrus (**B** and **D**) of an aFTLD-U case demonstrating co-localisation of FUS (green) and hnRNP R (red) in neuronal cytoplasmic inclusions (white arrows). Neuronal nuclei are counterstained with DAPI. Scale bars in A and B represent 50 μm. Scale bars in C and D represent 20 μm.

6.2.6 Frequency of hnRNP R and hnRNP Q inclusions in FTLD-FUS

To quantitatively assess the abundance of pathological inclusions containing hnRNP R and hnRNP Q relative to FUS or TRN1 in FTLD-FUS cases, the number of FUS, TRN1, hnRNP R and hnRNP Q positive inclusions on immunohistochemically stained sections were counted in a defined area of the grey matter of the frontal cortex (Figure 6.12) and in the granule cell layer of the dentate gyrus in the hippocampus (Figure 6.13). Inclusion quantification indicated variation between cases in the number of all inclusions counted, however no statistically significant differences were found between the average number of FUS, TRN1, hnRNP R and hnRNP Q positive inclusions in both NIFID (frontal cortex $p = 0.7978$, dentate gyrus $p = 0.9723$) and aFTLD-U (frontal cortex $p = 0.2856$, dentate gyrus $p = 0.8934$) subtypes. As expected, the number of FUS, TRN1, hnRNP R and hnRNP Q positive inclusions were consistently higher in the NIFID cases than aFTLD-U. No inclusions containing FUS, TRN1, hnRNP R or hnRNP Q were observed in non-disease controls and have therefore not been quantified.
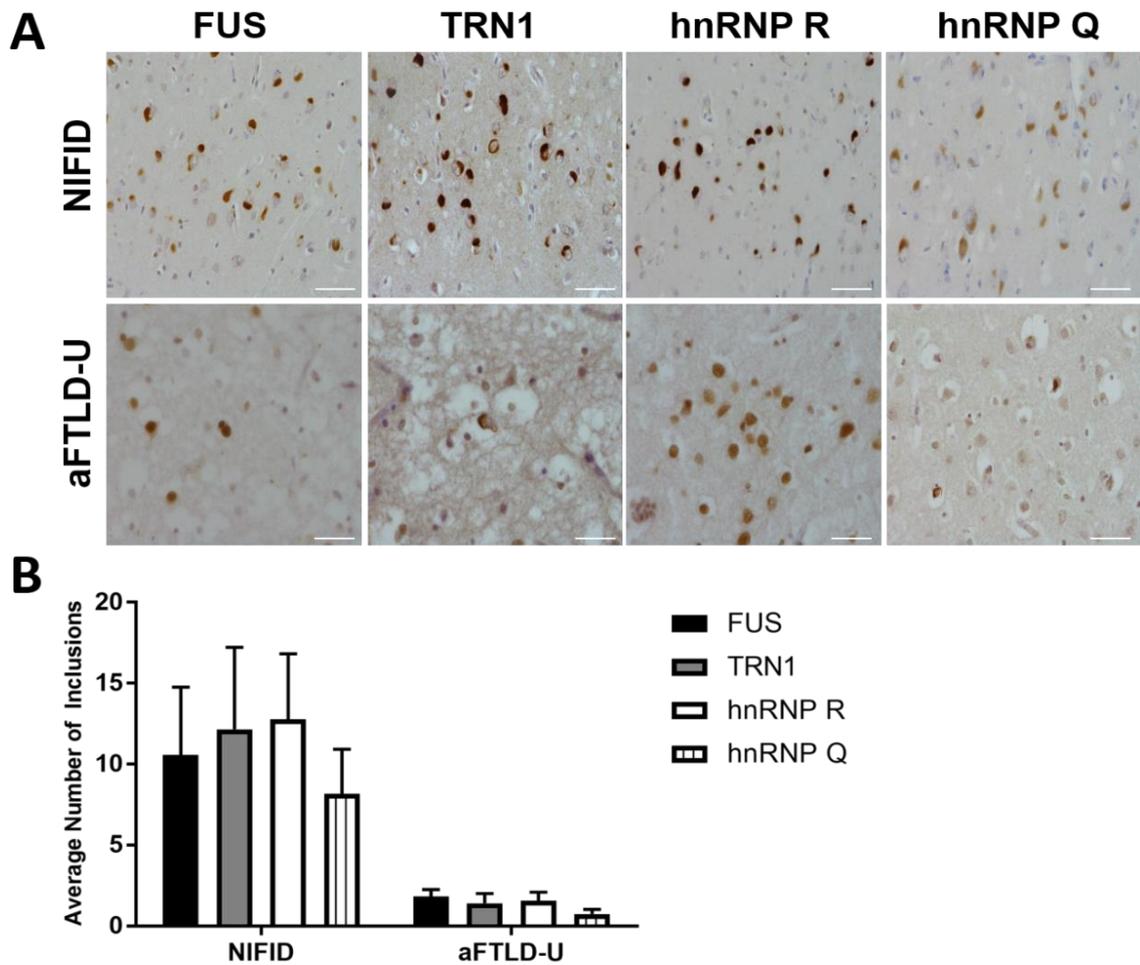
Figure 6.12 hnRNP R and hnRNP Q inclusions occur as frequently as FUS and TRN in the frontal cortex of FTLD-FUS cases.

A) Representative images of FUS, TRN, hnRNP R and hnRNP Q immunohistochemical staining in the frontal cortex of NIFID and aFTLD-U FTLD-FUS cases. Scale bars represent 50 µm. **B**) Quantification of FUS, TRN, hnRNP R and hnRNP Q in the frontal cortex of NIFID and atypical FTLD-FUS cases. No significant differences were found between the number of each inclusion type in either FTLD-FUS subtype. Kruscall-Wallis one-way analysis of variance. NIFID cases, $p = 0.7978$, $n = 5$. aFTLD-U cases, $p = 0.2856$, $n = 5$.
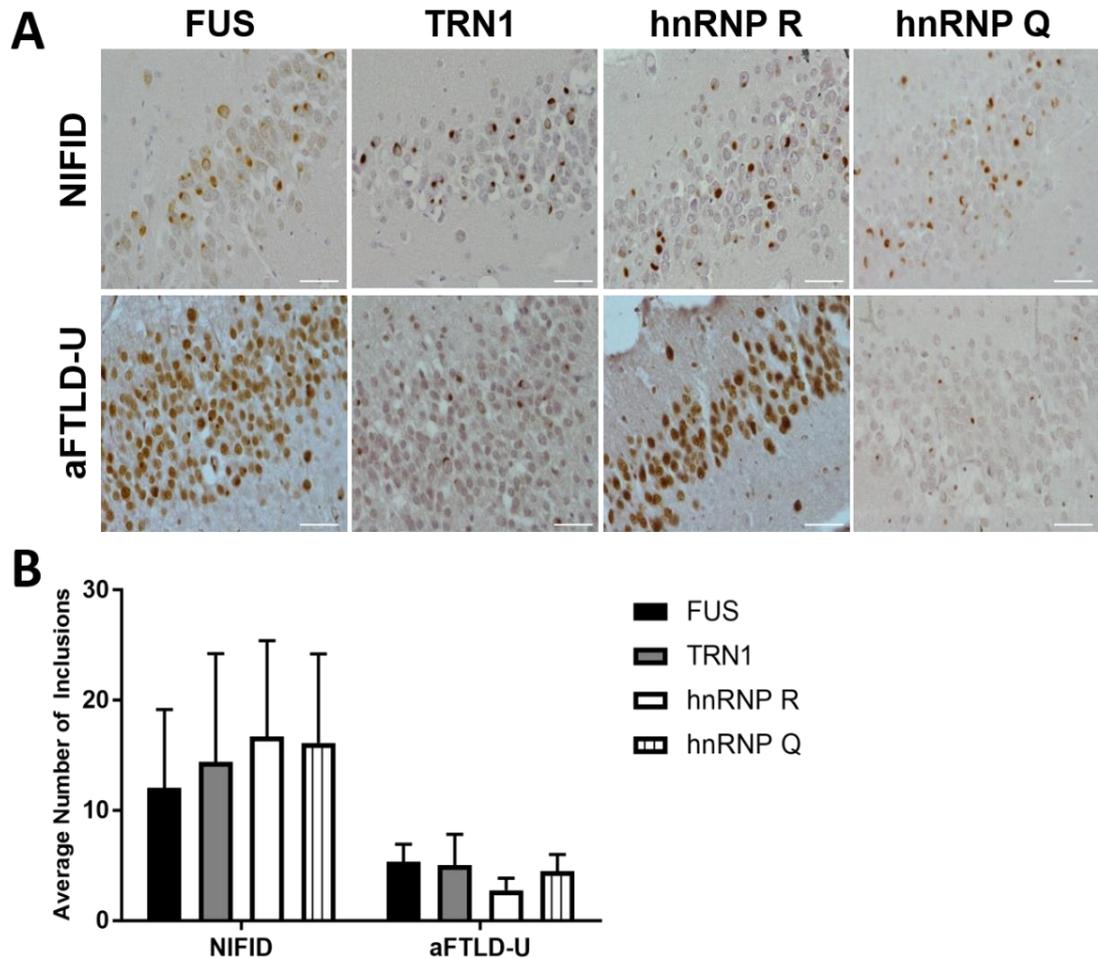
Figure 6.13 hnRNP R and hnRNP Q inclusions occur as frequently as FUS and TRN in the granular cell layer of the hippocampus of FTLD-FUS cases.

A) Representative images of FUS, TRN, hnRNP R and hnRNP Q immunohistochemical staining in the granular cell layer of the hippocampus of NIFID and atypical FTLD-FUS cases. Scale bars represent 50 µm. **B**) Quantification of FUS, TRN, hnRNP R and hnRNP Q in the granular cell layer of the hippocampus of NIFID and atypical FTLD-FUS cases. No significant differences were found between the number of each inclusion type in either FTLD-FUS subtype. Kruscall-Wallis one-way analysis of variance. NIFID cases, p = 0.9723, n = 6. aFTLD-U cases, p = 0.8934, n = 5.

6.3 Discussion

This study has shown for the first time that hnRNP R and hnRNP Q are mislocalised into pathological inclusions in two subtypes of FTLD-FUS. mRNA expression analysis revealed that hnRNP R expression is increased in several of the FTLD subtypes but immunohistochemical examination demonstrated that this protein, and the closely related hnRNP Q protein, only show pathological mislocalisation in the FTLD-FUS subtype. The pathological inclusions containing hnRNP R and hnRNP Q in the frontal cortex and hippocampus of these cases had a similar localisation pattern and morphological features as the previously described FUS and TRN1 inclusions. Quantification of the hnRNP R and hnRNP Q inclusions indicated they were as frequent as inclusions containing FUS or TRN1, and double-immunofluorescence confirmed the co-localisation of hnRNP R with FUS in both neuronal cytoplasmic and intranuclear inclusions. This study expands the spectrum of DNA/RNA binding proteins linked to the pathology of FTLD and reinforces the hypothesis that the pathogenesis of FTLD-FUS extends beyond FET proteins, TRN1 cargoes and dysfunctional nuclear import, but rather implicates a broader dysregulation of DNA/RNA binding proteins.

6.3.1 Expanding the spectrum of proteins in FTLD-FUS

A prominent hypothesis in the FTLD-FUS field is that pathological aggregation of FUS and the other FET proteins results from an impairment in arginine methylation of their RGG3 domains, causing overly tight binding of the FET proteins to their nuclear importer, TRN1, which prevents FET-TRN1 complex dissociation in the nucleus and subsequently results in the re-export and accumulation of FET proteins and TRN1 in the cytosol (Dormann and Haass, 2013; Dormann et al., 2012). Recent work has also suggested that impairment in arginine methylation enhances the phase transition of FUS, and potentially the other FET proteins, into liquid-like protein droplets which can

form solid, fibrous aggregates over time, promoting their pathological aggregation (Hofweber et al., 2018; Qamar et al., 2018).

Whilst this hypothesis explains the presence of TRN1 and the three FET proteins in pathological inclusions in FTLD-FUS, it cannot explain the pathological accumulation of non-FET proteins, such as hnRNP R, hnRNP Q and several other hnRNP proteins previously identified in these inclusions (Gami-Patel et al., 2016). With the exception of hnRNP A1 and hnRNP D, the majority of these proteins are not predicted to be imported by TRN1 (Lee et al., 2006; Pollard et al., 1996; Suzuki et al., 2005), and it is unclear to what extent these proteins are capable of liquid-liquid phase separation or subjected to the effects of arginine methylation. This suggests that it is not only FET proteins responsible for FTLD-FUS pathology but rather implicates dysfunction in a broader spectrum of RNA binding proteins. Unlike EWS, TAF15 and other hnRNPs, which are only found in a proportion of FUS inclusions, quantification of hnRNP R and hnRNP Q inclusions revealed that these proteins are found as frequently as FUS and TRN1 in inclusions. This suggests these proteins have a central role in the pathogenesis of FTLD-FUS, however it is currently unclear, whether the accumulation of these proteins is a trigger or consequence of FUS aggregation.

6.3.2 Hypothesised mechanisms of hnRNP R and hnRNP Q aggregation

FUS, hnRNP R and hnRNP Q are all RNA binding proteins that have a wide range of functions linked to various aspects of mRNA metabolism. They have all been reported to bind to the SMN protein spliceosome complex and are all known to shuttle between the nucleus and cytoplasm associated with mRNAs (Mourelatos et al., 2001; Rossoll et al., 2002; Yamazaki et al., 2012). This indicates these proteins have similar functions and could suggest that they interact with each other under physiological conditions within the cell. One hypothesis to explain the co-aggregation of FUS,

hnRNP R and hnRNP Q could be that these proteins associate with each other in a protein-RNA complex, either directly via a protein-protein interaction, or indirectly by binding to the same mRNA transcripts. If the proteins form part of the same complex, then aberrant aggregation of one of the proteins could trigger the co-deposition of the associated proteins. Future work should explore whether FUS and hnRNP R and hnRNP Q interact with each other, either directly or indirectly, under physiological conditions and investigate how this may change during disease. This could be performed using co-immunoprecipitation studies on brain homogenates from healthy controls and FTLD-FUS patients. Understanding how these proteins interact under normal and diseased conditions could provide useful information that may elucidate a cause for the co-aggregation of these proteins.

These studies could also help to explain the nanostring mRNA expression analysis results which showed that hnRNP R expression was increased in several FTLD subtypes compared to healthy controls but was only observed in pathological inclusions in FTLD-FUS. If co-immunoprecipitation studies reveal an interaction between hnRNP R and FUS but not between hnRNP R and TDP-43, this may explain why hnRNP R co-aggregates with FUS but not with TDP-43.

It may also be possible that FUS and hnRNP R and hnRNP Q do not interact with each other physiologically and localise to pathological aggregates independently. Endogenous FUS and hnRNP Q have both been shown to localise to cytoplasmic stress granules under specific cellular stress conditions (Quaresma et al., 2009; Sama et al., 2013). It may therefore be possible that these proteins co-localise in stress granules only during cellular stress and it is the aberrant disassembly of these granules that results in co-aggregation of these proteins in pathological inclusions. Several groups have proposed that the pathological accumulation of FUS, and other ALS/FTD linked proteins, is initiated by their assembly in stress granules or other

RNA granules (Bentmann et al., 2013; Dormann et al., 2010; Li et al., 2013). The mechanism by which FUS condenses into stress granules by liquid-liquid phase separation is well-characterised and is known to be driven by cation-π interactions between tyrosine residues in its N-terminal LCD domain and arginine residues in the C-terminal RGG domain (Hofweber et al., 2018; Qamar et al., 2018). Whether hnRNP R and hnRNP Q possess similar phase transitioning properties that enable them to condense into stress granules is currently unknown and requires further investigation. Both hnRNP proteins contain C-terminal RGG domains but they are not predicted to have N-terminal LCD domains. Their localisation into stress granules is therefore unlikely to be driven by the same mechanism as FUS but may be a consequence of alternative bio-physical interactions or association with other proteins capable of phase transitioning. Further experimental work is required to determine whether hnRNP R and hnRNP Q are capable of independent phase transition and subsequent aggregation, or whether it is their putative interaction with other proteins that enable this.

6.3.3 Addressing hnRNP R and hnRNP Q homology

One limitation of this study was that the immunohistochemical techniques used were unable to distinguish whether it is both hnRNP R and hnRNP Q accumulating in pathological inclusions, or whether it is only one of these proteins. hnRNP R and hnRNP Q are highly homologous, with 81.2 % similarity at the amino acid level (Mizutani et al., 2000). The antibodies targeting these proteins used in immunohistochemical staining in this study both target the C-terminus of these proteins, which, although different, consist of a similar amino acid sequence. Depending on the specificity of these antibodies, it is possible that they may be binding to the same protein and it is unclear from the current studies whether this is hnRNP R or hnRNP Q. These antibodies were chosen because they have been used and validated in previous published studies on hnRNP R and hnRNP Q (Dombert et

al., 2014; Lee et al., 2015; Reches et al., 2016). However, it is interesting to note that nearly every published study on hnRNP R or hnRNP Q fail to acknowledge the homology between these two proteins and do not address whether they are truly detecting the protein that the antibody is assumed to be detecting. This raises questions as to whether the functions ascribed to these proteins from these studies truly reflect their function or whether this is attributable to the homologous protein. Similar roles have been identified for hnRNP R and hnRNP Q, for example they are both thought to interact with SMN protein and be a component of the spliceosome (Mourelatos et al., 2001; Rossoll et al., 2002; Yamazaki et al., 2012). Although the homology between the two proteins indicates they are likely to have similar or overlapping functions, it would be useful for these studies to validate they are studying the intended protein. This will also need to be done for this study and future experiments should be performed to confirm the identity of the protein(s) in the pathological inclusions. One method of doing this could be to perform immunoprecipitation on brain tissue homogenates using either the hnRNP R or hnRNP Q antibody and then perform mass spectrometry to verify that the antibody is binding to its intended target. This will help to distinguish whether both hnRNP R and hnRNP Q are being detected in pathological inclusions in FTLD-FUS cases, or whether it is only one of them.

6.3.4 Additional future work

Although immunohistochemical methods clearly demonstrate the presence of hnRNP R and hnRNP Q in pathological inclusions in FTLD-FUS cases, the biochemical characteristics of these proteins in these cases has not been investigated here due to time limitations. Previous studies of FTLD-FUS cases have reported a shift in the solubility of the proteins that accumulate in FUS inclusions when protein levels from FTLD-FUS cases are analysed by immunoblotting following sequential solubility extraction. FUS, transportin, hnRNP A1, EWS and TAF15 all show a reduction in

solubility, with higher levels of the proteins found in the insoluble fraction in FTLD-FUS cases compared to controls (Brelstaff et al., 2011; Gami-Patel et al., 2016; Neumann et al., 2011). To determine whether hnRNP R and hnRNP Q also show the same shift towards insolubility following accumulation in pathological inclusions, biochemical assessment of the solubility of hnRNP R and hnRNP Q should be performed on FTLD-FUS and control cases. This can be assessed by immunoblotting for hnRNP R and hnRNP Q following the sequential extractions of proteins in increasingly stringent buffers to allow for the separation of proteins with different solubility characteristics.

The identification of hnRNP R and/or hnRNP Q in pathological inclusions in FTLD-FUS provide two candidate genes for genetic screening in FTLD. All FTLD-FUS cases used in this study have previously been screened for mutations in a variety of genes linked to FTD and ALS (Lashley et al., 2011; Rohrer et al., 2011), but to date, no genetic mutations have been identified as causative of FTLD-FUS. Screening the *HNRNPR* and *HNRNPQ* genes for mutations in these cases could identify mutations linked to disease. It would also be interesting to screen for mutations and assess hnRNP R and hnRNP Q pathology in ALS-FUS cases. To date, none of the additional proteins identified in FTLD-FUS inclusions have been found in ALS-FUS inclusions (Neumann et al., 2012). This is hypothesised to reflect the differing pathogenic mechanisms of the diseases, however, the end-point in both diseases is the pathological aggregation of FUS and it is possible that other proteins associated with FUS will also be common to both diseases. ALS-FUS cases should therefore be assessed for hnRNP R and hnRNP Q pathology to determine whether the dysregulation of these proteins is specific to FTLD-FUS or is a common feature shared by FUS pathologies.

6.3.5 Conclusion

In summary, the identification of hnRNP R and hnRNP Q in pathological inclusions in FTLD-FUS cases in this study adds two new proteins to the growing list of RNA binding proteins implicated in the pathogenesis of FTLD. The accumulation of these proteins in cytoplasmic and intranuclear neuronal inclusions was found to be specific to FTLD-FUS cases, although increased hnRNP R mRNA expression was seen in several FTLD subtypes. These inclusions were found to co-localise with and occur as frequently as inclusions containing FUS, suggesting these proteins may have a key role in the pathogenesis of FTLD-FUS. The relationship between FUS and these hnRNP proteins has not been previously explored and future experiments should be performed to establish whether these proteins directly or indirectly associate with FUS as this may help to establish the mechanism by which these proteins co-aggregate. Future biochemical experiments are also required to address whether it is both hnRNP R and hnRNP Q accumulating in these inclusions because the high level of homology between these proteins has made this difficult to decipher by immunohistochemical methods. Further functional understanding of these two new RNA binding proteins in FTLD-FUS aggregates may help to elucidate the mechanism by which these inclusions form and reveal novel functions for these hnRNP proteins.

Chapter 7. General Discussion

The identification of a hexanucleotide repeat expansion in the *C9orf72* gene as the most common genetic cause of FTD and ALS has led to a surge of research in recent years that has focused on identifying the molecular mechanism responsible for causing disease. Various *in vitro* and *in vivo* models have demonstrated that dissection of the disease mechanism is not straightforward, and it is likely that neurodegeneration results from a combination of loss-of-function and gain-of-function factors. Nevertheless, to fully understand the global disease mechanisms, it is necessary to appreciate the individual mechanisms by which each independent factor may contribute towards neurodegeneration. With this in mind, the work in this thesis focuses primarily on adding to the understanding of unique DPR proteins that are produced as a consequence of the repeat expansion by utilising cellular models and pathological tissue from *C9orf72* patients. Additionally, this thesis describes pathological investigation of the FTLD-FUS subtype of FTD and the identification of new RNA binding protein dysfunction in this disease. Key themes emerging throughout this work highlight the dysregulation of phase transition and aberrant RNA binding protein behaviour as a common mechanism linking many aspects of FTD and ALS biology.

7.1 Summary of main findings

7.1.1 Chapter 3

The use of a molecular chaperone as a potential method of mitigating the aggregation of DPR proteins in a cell-based model was explored in Chapter 3 because preliminary work had indicated that the molecular chaperone, HSJ1a, may be able to reduce the number of inclusions formed by overexpressed poly-GA in SK-N-SH cells (Christina Zarouchlioti, unpublished data). Building on this work, molecular sub-cloning was

used to add fluorescent tags to poly-GA, poly-PR and poly-GR expression construct to enable easier detection of both diffuse and aggregated forms of these DPR proteins within cells. Assessment of the effect of co-expression of HSJ1a with the fluorescently tagged DPR protein constructs, using a variety of quantification methods, found that the Hsp40 molecular co-chaperone was unable to significantly reduce the inclusion incidence of GFP-tagged poly-GA, although there was a trend towards inclusion reduction. It was hypothesised that the discrepancy between the ability of HSJ1a to reduce inclusions incidence of untagged but not GFP-tagged poly-GA was due to the addition of the GFP tag, which may be stabilising the protein and hindering access of the molecular chaperone. Further work on both tagged and untagged proteins, plus complimentary biochemical methods of aggregation assessment, are required to fully assess the effect of the addition of the GFP tag. Furthermore, it was noted that a larger screen of different molecular chaperones may be able to identify a protein that is more effective at reducing poly-GA aggregation than HSJ1a. For example, the beneficial effect of HSPB8 at reducing insoluble forms of the DPR proteins was reported in parallel with this research (Cristofani et al., 2017). The application of both of these studies, however, will depend on demonstrating that reduction in poly-GA aggregation leads to a reduction in cellular toxicity and death. There is currently conflicting evidence from cellular and animal models to suggest that poly-GA is toxic (Freibaum et al., 2015; Lee et al., 2017; Mizielinska et al., 2014; Wen et al., 2014; Zhang et al., 2014). In contrast, the arginine-containing DPR proteins induce toxicity in nearly all model systems. It may therefore be more beneficial to focus research efforts on reducing the toxicity of these proteins rather than reducing the aggregation of poly-GA.

7.1.2 Chapter 4

The work presented in Chapter 4 describes the development of methodology to add an eGFP tag to the C-terminal of endogenously produced poly-GR DPR proteins in

C9orf72 patient-derived iPSCs using CRISPR-Cas9 technology. Due to time limitations, it has not yet been confirmed whether the eGFP sequence has inserted into the expanded or non-expanded allele, and it is not yet known whether eGFP tagged poly-GR proteins are visible by fluorescence microscopy. Future work in the Isaacs laboratory will focus on extensively characterising these CRISPR-Cas9 edited iPSC lines. They will be utilised to study the production, localisation, toxicity and various other elements of poly-GR biology over time by fluorescence longitudinal imaging in disease relevant cell types, such as cortical or motor neurons. Furthermore, the CRISPR-Cas9 methodology optimised in this study to tag endogenous DPR proteins will be applied to the other DPR proteins with the aim of generating independent iPSC lines with different fluorescently tagged DPR proteins. This will be the first *C9orf72* model to enable each of the proteins to be studied at endogenous levels in disease relevant human cell types, which could provide novel insights into the molecular mechanisms of toxicity induced by these proteins.

7.1.3 Chapter 5

Post-translational modification of poly-GR was explored in Chapter 5 using novel monoclonal antibodies and *C9orf72* patient post-mortem brain tissue. In this study, antibodies raised against unmethylated, symmetrically dimethylated and asymmetrically dimethylated forms of poly-GR were validated in post-mortem tissue from the frontal and temporal cortices of both *C9orf72* – FTLD and *C9orf72* - ALS patients. The unmethylated-GR antibody did not label any neuronal cytoplasmic inclusions, however ADMA-GR and SDMA-GR positive inclusions were identified in *C9orf72* cases. Quantification of these inclusions revealed ADMA-GR to be the most abundant isoform in both *C9orf72* – FTLD and *C9orf72* – ALS patient tissue. Correlation analysis between the quantified methylated-GR inclusions and clinical data available indicated that SDMA-GR is significantly associated with disease duration and age at death in both the cohort of *C9orf72* cases at QSBB and in a larger

cohort. Further analysis in a larger cohort is therefore required to fully assess the association of SDMA-GR with disease duration and age at death. Furthermore, it will be necessary to perform experiments to determine the functional and biophysical consequences of the methylation of poly-GR. Recent work on FUS has indicated that arginine methylation is an important regulator of phase transition (Hofweber et al., 2018; Qamar et al., 2018). Given that poly-GR is known to also undergo phase separation (Boeynaems et al., 2017; Lee et al., 2016; Lin et al., 2016), it could be speculated that the methylation of arginine residues within this DPR protein may also play a role in regulating its phase transition. However, further *in vitro* and *in vivo* characterisation of this post-translational modification of poly-GR is required to better understand the significance arginine methylation may play in *C9orf72* disease pathogenesis.

7.1.4 Chapter 6

Chapter 6 described the identification of two RNA binding proteins, hnRNP R and hnRNP Q, in pathological inclusions specifically in FTLD-FUS patient post-mortem tissue. The proteins were found in both cytoplasmic and intranuclear neuronal inclusions in the cortex and granular cell layer of the hippocampus in both the NIFID and aFTLD-U pathological subtypes of FTLD-FUS. The inclusions co-localised with aggregates of the FUS protein and they were found to label inclusions as frequently as transportin, a common marker of FTLD-FUS pathology. This study implicates a wider dysregulation of RNA binding proteins in FTLD-FUS than just the members of the FET family, and it adds two new proteins to the growing list of RNA binding proteins implicated in FTD/ALS pathogenesis. Future work involving biochemical assessment of the solubility and phase transition ability of hnRNP R and hnRNP Q, plus identification of their binding partners in FTLD-FUS cases compared to other FTLD subtypes and healthy controls may help to reveal the mechanism by which these hnRNPs co-aggregate with the FUS protein. The fact that hnRNP R and hnRNP

Q are only identified in inclusions in FTLD-FUS cases suggests that the dysregulation of these proteins is specifically linked to FUS, however further work is required to determine the dynamics of this relationship, and whether the aberrant aggregation of these RNA binding proteins could have wider implications for ALS-FUS or other FTLD subtypes.

7.2 Emerging importance of RNA metabolism and phase separation in FTD/ALS biology

An emerging theme throughout the work in this thesis, and in the current field of ALS and FTD research, is the dysregulation in the metabolism of multiple RNA binding proteins. In particular, disease pathogenesis is strongly linked to the genetic or functional disruption of RNA binding proteins that contain low complexity domains (LCDs), such as TDP-43 and FUS. These LCDs contain motifs that are recognised by other proteins or nucleic acids, resulting in multivalent interactions that are essential for the biophysical phenomenon of liquid-liquid phase separation to drive the dynamic assembly of membrane-less organelles, such as stress granules, RNA transport granules, processing (P) bodies and nucleoli. These membrane-less organelles create dynamic, micron-scale environments that are essential for many biochemical reactions, particularly those critical for RNA metabolism (Boeynaems et al., 2018; Courchaine et al., 2016; Sawyer et al., 2018).

Accumulating evidence suggests that many of the disease-causing mutations in proteins linked to ALS and FTD alter the biophysical properties of LCDs and membrane-less organelles, resulting in both impairments in RNA metabolism and promoting the formation of stable long-lasting protein aggregates. For example, ALS-linked mutations within the LCD of TDP-43 have been associated with larger stress granules with faster kinetics (Dewey et al., 2011), increasing the viscosity of RNA transport granules (Alami et al., 2014; Gopal et al., 2017), and promoting protein

277

misfolding and stabilisation of TDP-43 within cytoplasmic inclusions (Johnson et al., 2009; Kabashi et al., 2010). Similarly, disease-linked mutant FUS has been shown to disrupt cytoplasmic RNA transport granules and reduce new protein translation (Murakami et al., 2015), as well as resulting in the formation of irreversible solid-state structures (Patel et al., 2015). Signification disruption to these usually reversible RNA-containing structures can lead to perturbed RNA metabolism, including deficits in the trafficking, splicing, and translation of RNA, all of which have been implicated in the pathogenesis of FTD and ALS.

In cases where disease is not caused by a mutation in an FTD/ALS-linked RNA binding protein, identifying a cause of pathological mislocalisation and aggregation of the RNA binding proteins implicated in disease, such as TDP-43 or FUS, has been difficult. However, recent work on the FUS protein has indicated a link between phase separation and potential aggregation of the protein, and the post-translational modification of arginine methylation in FTLD-FUS (Hofweber et al., 2018; Qamar et al., 2018). This suggests that LLPS and the assembly of essential membrane-less organelles is not only influenced by the composition and interactions of a protein but can also be regulated by post-translational modifications. Although currently FUS is the only FTD/ALS-linked protein that has been demonstrated to have its phase separation dynamics altered by post-translational modifications, these studies demonstrate that it is not only genetic mutations that can disrupt the dynamics of membrane-less organelles and consequently alter RNA metabolism. Future work should investigate whether similar post-translation modifications also impact the phase separation and aggregation behaviour of other proteins implicated in the pathogenesis of FTD/ALS. As mentioned in Chapter 5, for *C9orf72*-linked FTD/ALS it will be particularly interesting to address whether the asymmetric and symmetric dimethylation of arginine residues detected within poly-GR influences the phase separation behaviour of this toxic DPR protein.

For *C9orf72*-linked FTD/ALS, disrupted RNA metabolism cannot be attributed to mutations in RNA binding proteins however, the arginine-containing DPR proteins provide a link between the repeat expansion mutation, perturbed phase separation and RNA metabolism. Both poly-GR and poly-PR are capable of LLPS and have been shown to alter the biophysical and functional properties of several membrane-less organelles, including stress granules and the nucleolus (Boeynaems et al., 2017; Kwon et al., 2014). Both arginine-containing DPR proteins can strengthen the multivalent interactions that support the formation of stress granules in living cells, and the addition of arginine-containing DPR proteins can reduce the required concentrations of other proteins to phase separate (Lee et al., 2016). Furthermore, both arginine-containing DPR proteins are known to interact with the LCDs of several RNA binding proteins that are capable of phase separation, such as FUS, TIA-1 and hnRNPA1. The interaction with these proteins can alter their phase transition dynamics, making them more likely to remain in liquid droplet form, or lead to the formation of aggregating fibrils (Boeynaems et al., 2017; Lee et al., 2016). The disruption to membrane-less organelles caused by the arginine-containing DPR proteins not only alters their biophysical dynamics, but also interferes with their functional properties. For example, one study has demonstrated that the disruption caused to the nucleoli by poly-GR and poly-PR in a cell model resulted in impaired pre-mRNA splicing and the biogenesis of ribosomal RNA, which led to disrupted RNA processing and ultimately cell death (Kwon et al., 2014). Additionally, poly-GR and poly-PR can impair protein translation by blocking access of translation machinery to mRNA (Kanekura et al., 2016), which may provide an explanation for the increase in stress granule formation (a marker of translational arrest) when these proteins are overexpressed in immortalised human cell lines (Boeynaems et al., 2017; Kanekura et al., 2016; Lee et al., 2016; Wen et al., 2014).

It should also be noted that it is not only the arginine-containing DPR proteins that link *C9orf72*-related disease to phase transitioning membrane-less organelles. The C9orf72 protein itself has been shown to be required for stress granule formation in a mammalian cell model (Maharjan et al., 2016). CRISPR-Cas9 knockdown of *C9orf72* expression reduced transcript and protein levels of several stress granule-associated proteins, potentially also linking the loss of C9orf72 protein function to disrupted stress granule dynamics in *C9orf72*-linked disease. Furthermore*,* GGGGCC repeat containing RNA has also been shown to undergo phase transition to liquid droplets and induce phase transition of RNA granule proteins, even in the absence of LCD containing proteins (Jain and Vale 2017; Fay et al., 2017). The functional effect and potentially pathogenic consequences of this ability of repeat containing RNA remains to be determined.

Thus far, all research relating to the biophysical and functional effect of the arginine-containing DPR proteins on membrane-less organelles or liquid-like droplet structures has occurred *in vitro* or via overexpression of the proteins to high levels that are not likely to be representative of protein concentration in disease. The development of models that allow monitoring of endogenous DPR protein levels, such as the eGFP tagged CRISPR-Cas9 edited *C9orf72*-iPSC line developed in Chapter 4, will be crucial for understanding the effects of physiologically relevant levels of poly-GR and poly-PR on phase separation and membrane-less organelle dynamics. The evidence to date, however, indicates that the cellular toxicity attributed to the arginine-containing DPR proteins is, at least in part, due to their ability to interfere with the assembly, dynamics and function of phase separating membrane-less organelles, disturbing the balance of biochemical reactions critical for RNA processing and cellular function. This disruption to various aspects of RNA metabolism is a reoccurring theme across the FTD/ALS disease spectrum.

A prominent hypothesis in the FTD/ALS field linking disrupted RNA metabolism and the pathological aggregation of proteins seen post-mortem is the idea that dysregulation of phase separating structures, in particular stress granules, may be an initiating event which eventually leads to irreversible pathological aggregation (Dewey et al., 2012; Dormann et al., 2010; Li et al., 2013). In addition to the formation of pathological proteinaceous inclusions, several non-mutually exclusive hypotheses have been proposed to explain how the proposed stress granule 'precursors' of pathological aggregation could contribute to neurodegeneration. These include sustain translational arrest, the entrapment of important regulatory proteins and the seeding of more pathological aggregation (Bowden and Dormann, 2016). Evidence to support the hypothesis that stress granules may be precursors to pathological aggregation in FTD/ALS come from the observation that many disease-associated genes encode RNA binding proteins that are recruited to stress granules or are involved in the regulation of their dynamics (Guerreiro et al., 2015), and the co-localisation of stress granule markers with protein aggregates in FTLD and ALS patient tissue (Bentmann et al., 2012; Dormann et al., 2010; Fujita et al., 2008; Liu-Yesucevitz et al., 2010; McGurk et al., 2014).

Whilst this hypothesis can account for the presence of some stress granule associated or LCD-containing proteins in pathological inclusions, such as FUS, TDP-43 or the arginine-containing DPR proteins, it cannot account for the most prevalent DPR protein found in pathological inclusions in *C9orf72* patients – poly-GA – as this protein has not been shown to associate with stress granules. Instead, the accumulation of this DPR protein is likely to be due to its biophysical characteristics giving it a strong propensity to self-aggregate. Given that relatively few models demonstrate poly-GA to be toxic compared to the arginine-containing DPR proteins, this observation suggests that *C9orf72*-linked cellular toxicity is not attributable to the poly-GA protein aggregates seen post-mortem, but rather is a product of the disrupted

RNA processing and aberrant stabilisation of stress granules induced by the arginine-containing DPR proteins. This hypothesis suggests that the idea explored in Chapter 3 of utilising molecular chaperones to reduce poly-GA aggregation, with the aim of reducing cellular toxicity, may not be an effective approach. Future work should instead focus on methods that mitigate the biophysical characteristics of the arginine containing DPR proteins.

Although the hypothesis of stress granules being precursors to protein inclusions is supported by many groups working on FTD and ALS, the evolution of individual RNA stress granules into pathological aggregates has yet to be demonstrated *in vivo* (Coyne et al., 2017). It is also important to note that stress granules are not the only cellular structures affected by the dysregulation of phase separation, and disturbed membrane-less organelle function is not the only cellular mechanism linked to the pathogenesis of FTD or ALS. Thus, although the stress granule pre-cursor theory may account for some features of FTD/ALS pathology, it is likely that disease results from a combination of synergistic mechanisms.

7.3 Synergistic mechanisms of toxicity in *C9orf72* FTD/ALS

In addition to altered RNA and phase separation dynamics, it is important to remember that a number of other downstream cellular mechanisms have also been associated with *C9orf72*-linked FTD/ALS. These can be split into three main categories; problems with protein clearance systems, disrupted nucleocytoplasmic transport and alterations to neuronal characteristics (Balendra and Isaacs, 2018; Starr and Sattler, 2018; van Blitterswijk and Rademakers, 2015). Although much work has been done to elucidate the primary cause of neurodegeneration in *C9orf72*-linked disease, identifying a single cause has been difficult due to the complexities involved in modelling each mechanism independently. Given that there is evidence to suggest

dysfunction in multiple cellular mechanisms in *C9orf72*-linked disease, it is likely that several mechanisms are acting in combination to cause disease.

Numerous studies have linked the *C9orf72* repeat expansion mutation to defects in a variety of protein clearance pathway pathways. Reduction in the levels of the C9orf72 protein have been linked to impairments in autophagy and lysosomal dysfunction (Amick et al., 2016; Aoki et al., 2017; Ciura et al., 2013; Farg et al., 2014; Sellier et al., 2016) and some DPR proteins, in particularly poly-GA, have been shown to induce endoplasmic reticulum stress (Zhang et al., 2014) and cause impairments to the ubiquitin proteasome system (Gupta et al., 2017; May et al., 2014; Yamakawa et al., 2015; Zhang et al., 2016). Although several disease models have found these defects in protein clearance insufficient to fully recapitulate the neurodegeneration seen in *C9orf72*-linked FTD/ALS alone (possibly because other protein clearance systems are able to compensate for these impairments), it is likely that disruptions to cellular protein homeostasis is contributing to or exacerbating disease pathogenesis. The moderate toxicity of poly-GA indicated by some models (Lee et al., 2016; Mizielinska et al., 2014; Yamakawa et al., 2015), and the effect of *C9orf72* knock-out in mouse models (Atanasio et al., 2016; Burberry et al., 2016; Koppers et al., 2015; O'Rourke et al., 2016), supports this hypothesis because, although these models do not induce the severe neurodegenerative phenotype characteristic of *C9orf72* FTD/ALS, they result in the manifestation of other problems that are likely to synergistically impact neuronal toxicity, such as pathological protein aggregation or immune system dysfunction, respectively. This argues that therapeutic efforts to restore proteostasis, such as increasing molecular chaperones as explored in Chapter 3, may have some benefit in restoring proteostasis, but is unlikely to fully mitigate toxicity in *C9orf72*-linked disease due to the toxic effect of other impaired cellular mechanisms.

Disruption to nucleocytoplasmic transport is another reported mechanism that is likely to contribute towards pathogenesis. Both GGGGCC repeat RNA and DPR proteins

have been shown to cause dysfunctional nucleocytoplasmic transport in a variety of independent *C9orf72 in vivo* model systems (Boeynaems et al., 2016; Freibaum et al., 2015; Jovicic et al., 2015; Zhang et al., 2015). To support this finding, abnormal nuclear pore protein staining has been identified in *C9orf72* ALS patient post-mortem tissue and nucleocytoplasmic transport defects have been confirmed in *C9orf72* patient-derived cells (Li et al., 2018; Zhang et al., 2015). Furthermore, the short isoform of endogenous C9orf72 protein is known to localise to the nuclear envelope (Xiao et al., 2015), thus it could be speculated that a loss-of-function mechanism may also contribute to nucleocytoplasmic transport defects (Haeusler et al., 2016). It is currently unclear whether repeat RNA, DPR proteins, loss of C9orf72 protein or a combination of the three are the primary cause of nucleocytoplasmic transport defects. However, the observation of disrupted nucleocytoplasmic transport in numerous independent models and patient post-mortem tissue suggests that this mechanism is likely to be a significant contributing factor to *C9orf72* disease pathogenesis. Given that there is evidence to suggest dysfunctional nucleocytoplasmic transport is a key feature of *C9orf72*-linked disease, it is likely that this dysfunction is contributing towards pathogenesis in parallel with the other pathogenic mechanisms previously discussed.

Several alterations to neuronal characteristics have also been reported to contribute towards *C9orf72*-linked disease (Starr and Sattler, 2018). These include both structural problems, such as neuronal branching defects (Burguete et al., 2015; Ciura et al., 2013), dysfunctional axonal transport (Baldwin et al., 2016) and defects in growth cone dynamics (Sivadasan et al., 2016b), and excitability problems, such as hyper- and hypo- excitability (Devlin et al., 2015; Geevasinga et al., 2015; Sareen et al., 2013; Schanz et al., 2016), and an increased sensitivity to glutamate toxicity (Donnelly et al., 2013; Shi et al., 2018). Once again, these effects have been linked to a mixture of C9orf72 protein depletion, repeat RNA and DPR proteins, making it

difficult to deduce the primary cause of the neuronal dysfunction. As with the other pathogenic mechanisms discussed, it is likely that all of these alterations to normal neuronal behaviour are exacerbating neuronal death and neurodegeneration.

It is clear from the diverse range of mechanisms that have been shown to contribute to *C9orf72* pathogenesis that this is a highly complex disease. It is likely to be a multi-factorial disease involving dysfunction in a number of cellular pathways that leads to neurodegeneration. Future work should focus on elucidating the relative contribution of each of the dysfunctional mechanisms discussed. It may be possible that all pathogenic mechanisms are acting in parallel and with equal toxicity, however it is more likely that a select few molecular events trigger a pathological cascade of dysfunction that then subsequently leads to neurodegeneration. Identifying these early molecular triggers will aid in the sequential targeting of pathological events and will be crucial to fully understanding the mechanism of *C9orf72*-linked disease. Furthermore, C9orf72 protein depletion, repeat RNA and DPR proteins may also have non-cell-autonomous effects. It will, therefore, be important to study the effect of the *C9orf72* repeat expansion in non-neuronal cell types, both independently and in combination, to gain a wider systemic perspective on how the repeat expansion mutation may be causing disease.

7.4 Concluding remarks

The discovery of the *C9orf72* repeat expansion mutation has revolutionised the ALS and FTD research field in a relatively short space of time. It has led to the identification of *C9orf72*-specific pathological features, the development of several model systems and has resulted in the formation of many new and exciting hypothesised mechanism of disease. The work in this thesis has primarily focused on toxicity induced by unique DPR proteins produced by RAN translation in *C9orf72*-linked disease. By utilising post-mortem patient brain tissue and cellular systems, this work has explored a

possible method of mitigating DPR protein aggregation, generated a novel patient-derived iPSC model that can be utilised to study endogenous DPR protein toxicity *in vivo* and characterised a significant post-translational modification of these proteins. The CRISPR-Cas9 methodology developed in this project and the cell model generated will provide the Isaacs lab, and many international collaborating groups, with the first patient-derived tool to study endogenously produced DPR proteins in real-time. This could directly lead to screens for new therapies for *C9orf72*-linked FTD and ALS. Additionally, work in this thesis described the identification of two new hnRNP proteins as markers of pathological inclusions in FTLD-FUS. An emerging theme throughout all aspects of the work in this thesis, and in the wider FTD/ALS research community, is the dysregulation of RNA metabolism and phase separation in the pathogenesis of FTD and ALS. It is becoming clear that dysfunction in a range of RNA binding proteins, LCD-containing proteins and membrane-less organelles are a common feature in early FTD and ALS pathogenesis and this is hypothesised to precipitate the pathological protein aggregation seen at the end-stage of disease. It is crucial to note, however, that in *C9orf72*-linked FTD/ALS, dysregulation of RNA metabolism is just one of several mechanisms identified as a potential cause of disease and it is more likely to be a synergistic or sequential combination of pathogenic events that leads to the neurodegeneration seen in *C9orf72* patients. Although many questions still remain to be answered, it is hoped that the significant advancements made in recent years in our understanding of the mechanisms of FTD and ALS will ultimately aid in the development of successful disease-modifying treatments for these devastating neurodegenerative diseases.

Chapter 8. References

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*, 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*, 536-543.

Alberti, S., Mateju, D., Mediani, L., and Carra, S. (2017). Granulostasis: Protein Quality Control of RNP Granules. Front Mol Neurosci *10*, 84.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell, 4th edn (New York: Garland Science).

Almeida, S., Gascon, E., Tran, H., Chou, H.J., Gendron, T.F., Degroot, S., Tapper, A.R., Sellier, C., Charlet-Berguerand, N., Karydas, A.*, et al.* (2013). Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol *126*, 385-399.

Altmeyer, M., Neelsen, K.J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grofte, M., Rask, M.B., Streicher, W., Jungmichel, S., Nielsen, M.L.*, et al.* (2015). Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). Nat Commun *6*, 8088.

Aman, P., Panagopoulos, I., Lassen, C., Fioretos, T., Mencinger, M., Toresson, H., Hoglund, M., Forster, A., Rabbitts, T.H., Ron, D.*, et al.* (1996). Expression patterns of the human sarcoma-associated genes FUS and EWS and the genomic structure of FUS. Genomics *37*, 1-8.

Amick, J., Roczniak-Ferguson, A., and Ferguson, S.M. (2016). C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. Molecular biology of the cell *27*, 3040-3051.

Anderson, E.M., Haupt, A., Schiel, J.A., Chou, E., Machado, H.B., Strezoska, Z., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A.*, et al.* (2015). Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. J Biotechnol *211*, 56-65.

Andersson, M.K., Stahlberg, A., Arvidsson, Y., Olofsson, A., Semb, H., Stenman, G., Nilsson, O., and Aman, P. (2008). The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response. BMC Cell Biol *9*, 37.

Ansar, S., Burlison, J.A., Hadden, M.K., Yu, X.M., Desino, K.E., Bean, J., Neckers, L., Audus, K.L., Michaelis, M.L., and Blagg, B.S. (2007). A non-toxic Hsp90 inhibitor protects neurons from Abeta-induced toxicity. Bioorg Med Chem Lett *17*, 1984-1990.

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*, 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. Biochem Biophys Res Commun *351*, 602-611.

Arighi, A., Fumagalli, G.G., Jacini, F., Fenoglio, C., Ghezzi, L., Pietroboni, A.M., De Riz, M., Serpente, M., Ridolfi, E., Bonsi, R*., et al.* (2012). Early onset behavioral variant frontotemporal dementia due to the C9ORF72 hexanucleotide repeat expansion: psychiatric clinical presentations. J Alzheimers Dis *31*, 447-452.

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, J.W., 3rd, Rademakers, R*., et al.* (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron *77*, 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*, 23204.

Auluck, P.K., Chan, H.Y., Trojanowski, J.Q., Lee, V.M., and Bonini, N.M. (2002). Chaperone suppression of alpha-synuclein toxicity in a Drosophila model for Parkinson's disease. Science *295*, 865-868.

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. Neuropathol Appl Neurobiol *41*, 601-612.

Baechtold, H., Kuroda, M., Sok, J., Ron, D., Lopez, B.S., and Akhmedov, A.T. (1999). Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation. J Biol Chem *274*, 34337-34342.

Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting proteostasis for disease intervention. Science *319*, 916-919.

Baldwin, K.R., Godena, V.K., Hewitt, V.L., and Whitworth, A.J. (2016). Axonal transport defects are a common phenotype in Drosophila models of ALS. Human molecular genetics *25*, 2378-2392.

Balendra, R., and Isaacs, A.M. (2018). C9orf72-mediated ALS and FTD: multiple pathways to disease. Nature reviews Neurology *14*, 544-558.

Banez-Coronel, M., Ayhan, F., Tarabochia, A.D., Zu, T., Perez, B.A., Tusi, S.K., Pletnikova, O., Borchelt, D.R., Ross, C.A., Margolis, R.L*., et al.* (2015). RAN Translation in Huntington Disease. Neuron *88*, 667-677.

Barker, H.V., Niblock, M., Lee, Y.B., Shaw, C.E., and Gallo, J.M. (2017). RNA Misprocessing in C9orf72-Linked Neurodegeneration. Front Cell Neurosci *11*, 195.

Barral, J.M., Broadley, S.A., Schaffar, G., and Hartl, F.U. (2004). Roles of molecular chaperones in protein misfolding diseases. Seminars in cell & developmental biology *15*, 17-29.

Beck, J., Poulter, M., Hensman, D., Rohrer, J.D., Mahoney, C.J., Adamson, G., Campbell, T., Uphill, J., Borg, A., Fratta, P.*, et al.* (2013). Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. American journal of human genetics *92*, 345-353.

Bedford, M.T., and Clarke, S.G. (2009). Protein arginine methylation in mammals: who, what, and why. Mol Cell *33*, 1-13.

Belly, A., Moreau-Gachelin, F., Sadoul, R., and Goldberg, Y. (2005). Delocalization of the multifunctional RNA splicing factor TLS/FUS in hippocampal neurones: exclusion from the nucleus and accumulation in dendritic granules and spine heads. Neurosci Lett *379*, 152-157.

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 Neuropathol *126*, 895-905.

Benajiba, L., Le Ber, I., Camuzat, A., Lacoste, M., Thomas-Anterion, C., Couratier, P., Legallic, S., Salachas, F., Hannequin, D., Decousus, M.*, et al.* (2009). TARDBP mutations in motoneuron disease with frontotemporal lobar degeneration. Ann Neurol *65*, 470-473.

Bentmann, E., Haass, C., and Dormann, D. (2013). Stress granules in neurodegeneration--lessons learnt from TAR DNA binding protein of 43 kDa and fused in sarcoma. FEBS J *280*, 4348-4370.

Bentmann, E., Neumann, M., Tahirovic, S., Rodde, R., Dormann, D., and Haass, C. (2012). Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). The Journal of biological chemistry *287*, 23079-23094.

Benussi, L., Rossi, G., Glionna, M., Tonoli, E., Piccoli, E., Fostinelli, S., Paterlini, A., Flocco, R., Albani, D., Pantieri, R.*, et al.* (2014). C9ORF72 hexanucleotide repeat number in frontotemporal lobar degeneration: a genotype-phenotype correlation study. J Alzheimers Dis *38*, 799-808.

Blanc, R.S., and Richard, S. (2017). Arginine Methylation: The Coming of Age. Molecular cell *65*, 8-24.

Blokhuis, A.M., Koppers, M., Groen, E.J., van den Heuvel, D.M., 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 Neuropathol *132*, 175-196.

Boeve, B.F., Boylan, K.B., Graff-Radford, N.R., DeJesus-Hernandez, M., Knopman, D.S., Pedraza, O., Vemuri, P., Jones, D., Lowe, V., Murray, M.E.*, et al.* (2012). Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72. Brain *135*, 765-783.

Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., *et al.* (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol *28*, 420-435.

Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., Guharoy, M., De Decker, M., Jaspers, T., Ryan, V.H., *et al.* (2017). Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. Mol Cell *65*, 1044-1055 e1045.

Boeynaems, S., Bogaert, E., Michiels, E., Gijselinck, I., Sieben, A., Jovicic, 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*, 20877.

Boisvert, F.M., van Koningsbruggen, S., Navascues, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat Rev Mol Cell Biol *8*, 574-585.

Borroni, B., Bonvicini, C., Alberici, A., Buratti, E., Agosti, C., Archetti, S., Papetti, A., Stuani, C., Di Luca, M., Gennarelli, M., *et al.* (2009). Mutation within TARDBP leads to frontotemporal dementia without motor neuron disease. Hum Mutat *30*, E974-983.

Boulanger, M.C., Liang, C., Russell, R.S., Lin, R., Bedford, M.T., Wainberg, M.A., and Richard, S. (2005). Methylation of Tat by PRMT6 regulates human immunodeficiency virus type 1 gene expression. J Virol *79*, 124-131.

Bowden, H.A., and Dormann, D. (2016). Altered mRNP granule dynamics in FTLD pathogenesis. Journal of neurochemistry *138 Suppl 1*, 112-133.

Boxer, A.L., Mackenzie, I.R., Boeve, B.F., Baker, M., Seeley, W.W., Crook, R., Feldman, H., Hsiung, G.Y., Rutherford, N., Laluz, V., *et al.* (2011). Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. J Neurol Neurosurg Psychiatry *82*, 196-203.

Brangwynne, C.P., Tompa, P., and Pappu, R.V. (2015). Polymer physics of intracellular phase transitions. Nat Phys *11*, 899-904.

Brelstaff, J., Lashley, T., Holton, J.L., Lees, A.J., Rossor, M.N., Bandopadhyay, R., and Revesz, T. (2011). Transportin1: a marker of FTLD-FUS. Acta Neuropathol *122*, 591-600.

Brettschneider, J., Van Deerlin, V.M., Robinson, J.L., Kwong, L., Lee, E.B., Ali, Y.O., Safren, N., Monteiro, M.J., Toledo, J.B., Elman, L., *et al.* (2012). Pattern of ubiquilin pathology in ALS and FTLD indicates presence of C9ORF72 hexanucleotide expansion. Acta Neuropathol *123*, 825-839.

Briese, M., Saal-Bauernschubert, L., Ji, C., Moradi, M., Ghanawi, H., Uhl, M., Appenzeller, S., Backofen, R., and Sendtner, M. (2018). hnRNP R and its main interactor, the noncoding RNA 7SK, coregulate the axonal transcriptome of motoneurons. Proc Natl Acad Sci U S A *115*, E2859-E2868.

Broadley, S.A., and Hartl, F.U. (2009). The role of molecular chaperones in human misfolding diseases. FEBS letters *583*, 2647-2653.

Broe, M., Hodges, J.R., Schofield, E., Shepherd, C.E., Kril, J.J., and Halliday, G.M. (2003). Staging disease severity in pathologically confirmed cases of frontotemporal dementia. Neurology *60*, 1005-1011.

Bulau, P., Zakrzewicz, D., Kitowska, K., Wardega, B., Kreuder, J., and Eickelberg, O. (2006). Quantitative assessment of arginine methylation in free versus protein-incorporated amino acids in vitro and in vivo using protein hydrolysis and high-performance liquid chromatography. Biotechniques *40*, 305-310.

Buratti, E., and Baralle, F.E. (2008). Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. Front Biosci *13*, 867-878.

Burberry, A., Suzuki, N., Wang, J.Y., Moccia, R., Mordes, D.A., Stewart, M.H., Suzuki-Uematsu, S., Ghosh, S., Singh, A., Merkle, F.T.*, et al.* (2016). Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. Science translational medicine *8*, 347ra393.

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*, e08881.

Cacace, R., Van Cauwenberghe, C., Bettens, K., Gijselinck, I., van der Zee, J., Engelborghs, S., Vandenbulcke, M., Van Dongen, J., Baumer, V., Dillen, L.*, et al.* (2013). C9orf72 G4C2 repeat expansions in Alzheimer's disease and mild cognitive impairment. Neurobiol Aging *34*, 1712 e1711-1717.

Calvio, C., Neubauer, G., Mann, M., and Lamond, A.I. (1995). Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. RNA *1*, 724-733.

Carra, S., Seguin, S.J., Lambert, H., and Landry, J. (2008). HspB8 chaperone activity toward poly(Q)-containing proteins depends on its association with Bag3, a stimulator of macroautophagy. J Biol Chem *283*, 1437-1444.

Caussinus, E., and Affolter, M. (2016). deGradFP: A System to Knockdown GFP-Tagged Proteins. Methods Mol Biol *1478*, 177-187.

Caussinus, E., Kanca, O., and Affolter, M. (2011). Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol *19*, 117-121.

Caussinus, E., Kanca, O., and Affolter, M. (2013). Protein knockouts in living eukaryotes using deGradFP and green fluorescent protein fusion targets. Curr Protoc Protein Sci *73*, Unit 30 32.

Chang, B., Chen, Y., Zhao, Y., and Bruick, R.K. (2007). JMJD6 is a histone arginine demethylase. Science *318*, 444-447.

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. J Biol Chem *291*, 4903-4911.

Chapple, J.P., van der Spuy, J., Poopalasundaram, S., and Cheetham, M.E. (2004). Neuronal DnaJ proteins HSJ1a and HSJ1b: a role in linking the Hsp70 chaperone machine to the ubiquitin-proteasome system? Biochemical Society transactions *32*, 640-642.

Cheetham, M.E., Brion, J.P., and Anderton, B.H. (1992). Human homologues of the bacterial heat-shock protein DnaJ are preferentially expressed in neurons. Biochem J *284 (Pt 2)*, 469-476.

Cheetham, M.E., and Caplan, A.J. (1998). Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. Cell Stress Chaperones *3*, 28-36.

Cheetham, M.E., Jackson, A.P., and Anderton, B.H. (1994). Regulation of 70-kDa heat-shock-protein ATPase activity and substrate binding by human DnaJ-like proteins, HSJ1a and HSJ1b. European journal of biochemistry / FEBS *226*, 99-107.

Chen, H.H., Chang, J.G., Lu, R.M., Peng, T.Y., and Tarn, W.Y. (2008). The RNA binding protein hnRNP Q modulates the utilization of exon 7 in the survival motor neuron 2 (SMN2) gene. Mol Cell Biol *28*, 6929-6938.

Chen, H.H., Yu, H.I., Chiang, W.C., Lin, Y.D., Shia, B.C., and Tarn, W.Y. (2012). hnRNP Q regulates Cdc42-mediated neuronal morphogenesis. Mol Cell Biol *32*, 2224-2238.

Chen, H.J., Mitchell, J.C., Novoselov, S., Miller, J., Nishimura, A.L., Scotter, E.L., Vance, C.A., Cheetham, M.E., and Shaw, C.E. (2016). The heat shock response plays an important role in TDP-43 clearance: evidence for dysfunction in amyotrophic lateral sclerosis. Brain *139*, 1417-1432.

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 eIF2alpha phosphorylation. Nat Commun *9*, 51.

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*, 1151-1154.

Chi, S., Jiang, T., Tan, L., and Yu, J.T. (2016). Distinct neurological disorders with C9orf72 mutations: genetics, pathogenesis, and therapy. Neuroscience and biobehavioral reviews *66*, 127-142.

Chio, A., Borghero, G., Restagno, G., Mora, G., Drepper, C., Traynor, B.J., Sendtner, M., Brunetti, M., Ossola, I., Calvo, A.*, et al.* (2012). Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72. Brain *135*, 784-793.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res *24*, 132-141.

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. Ann Neurol *74*, 180-187.

Cleary, J.D., and Ranum, L.P. (2014). Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. Curr Opin Genet Dev *26*, 6-15.

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. J Biol Chem *287*, 15635-15647.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Cooper-Knock, J., Hewitt, C., Highley, J.R., Brockington, A., Milano, A., Man, S., Martindale, J., Hartley, J., Walsh, T., Gelsthorpe, C.*, et al.* (2012). Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. Brain *135*, 751-764.

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. (2015). Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. Acta Neuropathol *130*, 63-75.

Cooper-Knock, J., Shaw, P.J., and Kirby, J. (2014a). The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype. Acta neuropathologica *127*, 333-345.

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.* (2014b). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain *137*, 2040-2051.

Courchaine, E.M., Lu, A., and Neugebauer, K.M. (2016). Droplet organelles? The EMBO journal *35*, 1603-1612.

Cox, D., Carver, J.A., and Ecroyd, H. (2014). Preventing alpha-synuclein aggregation: the role of the small heat-shock molecular chaperone proteins. Biochimica et biophysica acta *1842*, 1830-1843.

Crippa, V., Sau, D., Rusmini, P., Boncoraglio, A., Onesto, E., Bolzoni, E., Galbiati, M., Fontana, E., Marino, M., Carra, S.*, et al.* (2010). The small heat shock protein B8 (HspB8) promotes autophagic removal of misfolded proteins involved in amyotrophic lateral sclerosis (ALS). Human molecular genetics *19*, 3440-3456.

Cristofani, R., Crippa, V., Vezzoli, G., Rusmini, P., Galbiati, M., Cicardi, M.E., Meroni, M., Ferrari, V., Tedesco, B., Piccolella, M.*, et al.* (2017). The small heat shock protein B8 (HSPB8) efficiently removes aggregating species of dipeptides produced in C9ORF72-related neurodegenerative diseases. Cell stress & chaperones.

Crivat, G., and Taraska, J.W. (2012). Imaging proteins inside cells with fluorescent tags. Trends Biotechnol *30*, 8-16.

Crozat, A., Aman, P., Mandahl, N., and Ron, D. (1993). Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature *363*, 640-644.

Cruts, M., Gijselinck, I., van der Zee, J., Engelborghs, S., Wils, H., Pirici, D., Rademakers, R., Vandenberghe, R., Dermaut, B., Martin, J.J.*, et al.* (2006). Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. Nature *442*, 920-924.

Dafinca, R., Scaber, J., Ababneh, N., Lalic, T., Weir, G., Christian, H., Vowles, J., Douglas, A.G., Fletcher-Jones, A., Browne, C.*, et al.* (2016). C9orf72 Hexanucleotide Expansions Are Associated with Altered Endoplasmic Reticulum Calcium Homeostasis and Stress Granule Formation in Induced Pluripotent Stem Cell-Derived Neurons from Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. Stem cells *34*, 2063-2078.

Davidson, Y., Robinson, A.C., Liu, X., Wu, D., Troakes, C., Rollinson, S., Masuda-Suzukake, M., Suzuki, G., Nonaka, T., Shi, J.*, et al.* (2016). Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. Neuropathol Appl Neurobiol *42*, 242-254.

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 Neuropathol Commun *2*, 70.

Day, R.N., and Davidson, M.W. (2009). The fluorescent protein palette: tools for cellular imaging. Chem Soc Rev *38*, 2887-2921.

DeJesus-Hernandez, M., Finch, N.A., Wang, X., Gendron, T.F., Bieniek, K.F., Heckman, M.G., Vasilevich, A., Murray, M.E., Rousseau, L., Weesner, R.*, et al.* (2017a). In-depth clinico-pathological examination of RNA foci in a large cohort of C9ORF72 expansion carriers. Acta Neuropathol *134*, 255-269.

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*, 245-256.

Delatycki, M.B., Williamson, R., and Forrest, S.M. (2000). Friedreich ataxia: an overview. J Med Genet *37*, 1-8.

Deng, H., Gao, K., and Jankovic, J. (2014). The role of FUS gene variants in neurodegenerative diseases. Nature reviews Neurology *10*, 337-348.

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*, 211-215.

Devenney, E., Hornberger, M., Irish, M., Mioshi, E., Burrell, J., Tan, R., Kiernan, M.C., and Hodges, J.R. (2014). Frontotemporal dementia associated with the C9ORF72 mutation: a unique clinical profile. JAMA Neurol *71*, 331-339.

Devlin, A.C., Burr, K., Borooah, S., Foster, J.D., Cleary, E.M., Geti, I., Vallier, L., Shaw, C.E., Chandran, S., and Miles, G.B. (2015). Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. Nature communications *6*, 5999.

Dewey, C.M., Cenik, B., Sephton, C.F., Dries, D.R., Mayer, P., 3rd, Good, S.K., Johnson, B.A., Herz, J., and Yu, G. (2011). TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. Mol Cell Biol *31*, 1098-1108.

Dewey, C.M., Cenik, B., Sephton, C.F., Johnson, B.A., Herz, J., and Yu, G. (2012). TDP-43 aggregation in neurodegeneration: are stress granules the key? Brain research *1462*, 16-25.

Diaz-Villanueva, J.F., Diaz-Molina, R., and Garcia-Gonzalez, V. (2015). Protein Folding and Mechanisms of Proteostasis. International journal of molecular sciences *16*, 17193-17230.

Dombert, B., Sivadasan, R., Simon, C.M., Jablonka, S., and Sendtner, M. (2014). Presynaptic localization of Smn and hnRNP R in axon terminals of embryonic and postnatal mouse motoneurons. PLoS One *9*, e110846.

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*, 415-428.

Dormann, D., and Haass, C. (2013). Fused in sarcoma (FUS): an oncogene goes awry in neurodegeneration. Mol Cell Neurosci *56*, 475-486.

Dormann, D., Madl, T., Valori, C.F., Bentmann, E., Tahirovic, S., Abou-Ajram, C., Kremmer, E., Ansorge, O., Mackenzie, I.R., Neumann, M.*, et al.* (2012). Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. EMBO J *31*, 4258-4275.

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

Doty, K.R., Guillot-Sestier, M.V., and Town, T. (2015). The role of the immune system in neurodegenerative disorders: Adaptive or maladaptive? Brain Res *1617*, 155-173.

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem *62*, 289-321.

Dreyfuss, G., Swanson, M.S., and Pinol-Roma, S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends Biochem Sci *13*, 86-91.

Du, K., Arai, S., Kawamura, T., Matsushita, A., and Kurokawa, R. (2011). TLS and PRMT1 synergistically coactivate transcription at the survivin promoter through TLS arginine methylation. Biochem Biophys Res Commun *404*, 991-996.

Efimova, A.D., Ovchinnikov, R.K., Roman, A.Y., Maltsev, A.V., Grigoriev, V.V., Kovrazhkina, E.A., and Skvortsova, V.I. (2017). [The FUS protein: Physiological functions and a role in amyotrophic lateral sclerosis]. Mol Biol (Mosk) *51*, 387-399.

Ellis, J. (1987). Proteins as molecular chaperones. Nature *328*, 378-379.

Evans, C.G., Wisen, S., and Gestwicki, J.E. (2006). Heat shock proteins 70 and 90 inhibit early stages of amyloid beta-(1-42) aggregation in vitro. The Journal of biological chemistry *281*, 33182-33191.

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*, 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*, 3573-3584.

Flores, B.N., Dulchavsky, M.E., Krans, A., Sawaya, M.R., Paulson, H.L., Todd, P.K., Barmada, S.J., and Ivanova, M.I. (2016). Distinct C9orf72-Associated Dipeptide Repeat Structures Correlate with Neuronal Toxicity. PLoS One *11*, e0165084.

Floris, G., Borghero, G., Cannas, A., Di Stefano, F., Costantino, E., Murru, M.R., Brunetti, M., Restagno, G., Traynor, B.J., Marrosu, M.G.*, et al.* (2012). Frontotemporal dementia with psychosis, parkinsonism, visuo-spatial dysfunction, upper motor neuron involvement associated to expansion of C9ORF72: a peculiar phenotype? J Neurol *259*, 1749-1751.

Forman, M.S., Mackenzie, I.R., Cairns, N.J., Swanson, E., Boyer, P.J., Drachman, D.A., Jhaveri, B.S., Karlawish, J.H., Pestronk, A., Smith, T.W.*, et al.* (2006). Novel ubiquitin neuropathology in frontotemporal dementia with valosin-containing protein gene mutations. J Neuropathol Exp Neurol *65*, 571-581.

Forsberg, K., Jonsson, P.A., Andersen, P.M., Bergemalm, D., Graffmo, K.S., Hultdin, M., Jacobsson, J., Rosquist, R., Marklund, S.L., and Brannstrom, T. (2010). Novel antibodies reveal inclusions containing non-native SOD1 in sporadic ALS patients. PLoS One *5*, e11552.

Fratta, P., Mizielinska, S., Nicoll, A.J., Zloh, M., Fisher, E.M., 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*, 1016.

Fratta, P., Polke, J.M., Newcombe, J., Mizielinska, S., Lashley, T., Poulter, M., Beck, J., Preza, E., Devoy, A., Sidle, K.*, et al.* (2015). Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion. Neurobiology of aging *36*, 546 e541-547.

Fratta, P., Poulter, M., Lashley, T., Rohrer, J.D., Polke, J.M., Beck, J., Ryan, N., Hensman, D., Mizielinska, S., Waite, A.J.*, et al.* (2013). Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. Acta Neuropathol *126*, 401-409.

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*, 129-133.

Frick, P., Sellier, C., Mackenzie, I.R.A., Cheng, C.Y., Tahraoui-Bories, J., Martinat, C., Pasterkamp, R.J., Prudlo, J., Edbauer, D., Oulad-Abdelghani, M.*, et al.* (2018). Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. Acta neuropathologica communications *6*, 72.

Fuhrmann, J., Clancy, K.W., and Thompson, P.R. (2015). Chemical biology of protein arginine modifications in epigenetic regulation. Chem Rev *115*, 5413-5461.

Fujii, R., and Takumi, T. (2005). TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. J Cell Sci *118*, 5755-5765.

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*, 439-445.

Fukuda, A., Nakadai, T., Shimada, M., and Hisatake, K. (2009). Heterogeneous nuclear ribonucleoprotein R enhances transcription from the naturally configured c-fos promoter in vitro. J Biol Chem *284*, 23472-23480.

Fukuda, A., Shimada, M., Nakadai, T., Nishimura, K., and Hisatake, K. (2013). Heterogeneous nuclear ribonucleoprotein R cooperates with mediator to facilitate transcription reinitiation on the c-Fos gene. PLoS One *8*, e72496.

Galimberti, D., Fenoglio, C., Serpente, M., Villa, C., Bonsi, R., Arighi, A., Fumagalli, G.G., Del Bo, R., Bruni, A.C., Anfossi, M.*, et al.* (2013). Autosomal dominant frontotemporal lobar degeneration due to the C9ORF72 hexanucleotide repeat expansion: late-onset psychotic clinical presentation. Biol Psychiatry *74*, 384-391.

Gami-Patel, P., Bandopadhyay, R., Brelstaff, J., Revesz, T., and Lashley, T. (2016). The presence of heterogeneous nuclear ribonucleoproteins in frontotemporal lobar degeneration with FUS-positive inclusions. Neurobiol Aging *46*, 192-203.

Gao, X.C., Zhou, C.J., Zhou, Z.R., Zhang, Y.H., Zheng, X.M., Song, A.X., and Hu, H.Y. (2011). Co-chaperone HSJ1a dually regulates the proteasomal degradation of ataxin-3. PLoS One *6*, e19763.

Geevasinga, N., Menon, P., Nicholson, G.A., Ng, K., Howells, J., Kril, J.J., Yiannikas, C., Kiernan, M.C., and Vucic, S. (2015). Cortical Function in Asymptomatic Carriers and Patients With C9orf72 Amyotrophic Lateral Sclerosis. JAMA neurology *72*, 1268-1274.

Gendron, T.F., Bieniek, K.F., Zhang, Y.J., Jansen-West, K., Ash, P.E., 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*, 829-844.

Gendron, T.F., van Blitterswijk, M., Bieniek, K.F., Daughrity, L.M., Jiang, J., Rush, B.K., Pedraza, O., Lucas, J.A., Murray, M.E., Desaro, P.*, et al.* (2015). Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. Acta neuropathologica *130*, 559-573.

Gerbino, V., Carri, M.T., Cozzolino, M., and Achsel, T. (2013). Mislocalised FUS mutants stall spliceosomal snRNPs in the cytoplasm. Neurobiol Dis *55*, 120-128.

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. Hum Genet *135*, 851-867.

Gijselinck, I., Engelborghs, S., Maes, G., Cuijt, I., Peeters, K., Mattheijssens, M., Joris, G., Cras, P., Martin, J.J., De Deyn, P.P.*, et al.* (2010). Identification of 2 Loci at chromosomes 9 and 14 in a multiplex family with frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Arch Neurol *67*, 606-616.

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 Neurol *11*, 54-65.

Gijselinck, I., Van Mossevelde, S., van der Zee, J., Sieben, A., Engelborghs, S., De Bleecker, J., Ivanoiu, A., Deryck, O., Edbauer, D., Zhang, M.*, et al.* (2016). The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry *21*, 1112-1124.

Gitcho, M.A., Baloh, R.H., Chakraverty, S., Mayo, K., Norton, J.B., Levitch, D., Hatanpaa, K.J., White, C.L., 3rd, Bigio, E.H., Caselli, R.*, et al.* (2008). TDP-43 A315T mutation in familial motor neuron disease. Ann Neurol *63*, 535-538.

Gomez-Deza, J., Lee, Y.B., 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 Neuropathol Commun *3*, 38.

Gopal, P.P., Nirschl, J.J., Klinman, E., and Holzbaur, E.L. (2017). Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. Proceedings of the National Academy of Sciences of the United States of America *114*, E2466-E2475.

Gorno-Tempini, M.L., Dronkers, N.F., Rankin, K.P., Ogar, J.M., Phengrasamy, L., Rosen, H.J., Johnson, J.K., Weiner, M.W., and Miller, B.L. (2004). Cognition and anatomy in three variants of primary progressive aphasia. Ann Neurol *55*, 335-346.

Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., Ogar, J.M., Rohrer, J.D., Black, S., Boeve, B.F.*, et al.* (2011). Classification of primary progressive aphasia and its variants. Neurology *76*, 1006-1014.

Green, K.M., Glineburg, M.R., Kearse, M.G., Flores, B.N., Linsalata, A.E., Fedak, S.J., Goldstrohm, A.C., Barmada, S.J., and Todd, P.K. (2017). RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. Nat Commun *8*, 2005.

Green, M.R., and Sambrook, J. (2012). Molecular Cloning, 4th edn (USA: Cold Spring Harbor Laboratory Press).

Greenblatt, S.M., Liu, F., and Nimer, S.D. (2016). Arginine methyltransferases in normal and malignant hematopoiesis. Exp Hematol *44*, 435-441.

Guerreiro, R., Bras, J., and Hardy, J. (2015). SnapShot: Genetics of ALS and FTD. Cell *160*, 798 e791.

Guo, L., Kim, H.J., Wang, H., Monaghan, J., Freyermuth, F., Sung, J.C., O'Donovan, K., Fare, C.M., Diaz, Z., Singh, N.*, et al.* (2018). Nuclear-Import Receptors Reverse

Aberrant Phase Transitions of RNA-Binding Proteins with Prion-like Domains. Cell *173*, 677-692 e620.

Gupta, R., Lan, M., Mojsilovic-Petrovic, J., Choi, W.H., Safren, N., Barmada, S., Lee, M.J., and Kalb, R. (2017). The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome. eNeuro *4*.

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*, 195-200.

Haeusler, A.R., Donnelly, C.J., and Rothstein, J.D. (2016). The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. Nat Rev Neurosci *17*, 383-395.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. Biochem J *430*, 379-392.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G*., et al.* (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768-779.

Hardiman, O., van den Berg, L.H., and Kiernan, M.C. (2011). Clinical diagnosis and management of amyotrophic lateral sclerosis. Nature reviews Neurology *7*, 639-649.

Harms, M.B., Cady, J., Zaidman, C., Cooper, P., Bali, T., Allred, P., Cruchaga, C., Baughn, M., Libby, R.T., Pestronk, A*., et al.* (2013). Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Neurobiol Aging *34*, 2234 e2213-2239.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324-332.

Hautbergue, T., Puel, O., Tadrist, S., Meneghetti, L., Pean, M., Delaforge, M., Debrauwer, L., Oswald, I.P., and Jamin, E.L. (2017). Evidencing 98 secondary metabolites of Penicillium verrucosum using substrate isotopic labeling and high-resolution mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci *1071*, 29-43.

He, F., Jones, J.M., Figueroa-Romero, C., Zhang, D., Feldman, E.L., Goutman, S.A., Meisler, M.H., Callaghan, B.C., and Todd, P.K. (2016). Screening for novel hexanucleotide repeat expansions at ALS- and FTD-associated loci. Neurology Genetics *2*, e71.

Henry, R.G., Shieh, M., Okuda, D.T., Evangelista, A., Gorno-Tempini, M.L., and Pelletier, D. (2008). Regional grey matter atrophy in clinically isolated syndromes at presentation. Journal of neurology, neurosurgery, and psychiatry *79*, 1236-1244.

Hensman Moss, D.J., Poulter, M., Beck, J., Hehir, J., Polke, J.M., Campbell, T., Adamson, G., Mudanohwo, E., McColgan, P., Haworth, A*., et al.* (2014). C9orf72 expansions are the most common genetic cause of Huntington disease phenocopies. Neurology *82*, 292-299.

Herranz-Martin, S., Chandran, J., Lewis, K., Mulcahy, P., Higginbottom, A., Walker, C., Valenzuela, I.M.Y., Jones, R.A., Coldicott, I., Iannitti, T.*, et al.* (2017). Viral delivery of C9ORF72 hexanucleotide repeat expansions in mice lead to repeat length dependent neuropathology and behavioral deficits. Dis Model Mech.

Hetz, C., Chevet, E., and Oakes, S.A. (2015). Proteostasis control by the unfolded protein response. Nature cell biology *17*, 829-838.

Hicks, G.G., Singh, N., Nashabi, A., Mai, S., Bozek, G., Klewes, L., Arapovic, D., White, E.K., Koury, M.J., Oltz, E.M.*, et al.* (2000). Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. Nat Genet *24*, 175-179.

Himuro, N., Kozuka, N., and Mori, M. (2013). Measurement of family-centred care: translation, adaptation and validation of the Measure of Processes of Care (MPOC-56 and -20) for use in Japan. Child Care Health Dev *39*, 358-365.

Hofweber, M., Hutten, S., Bourgeois, B., Spreitzer, E., Niedner-Boblenz, A., Schifferer, M., Ruepp, M.D., Simons, M., Niessing, D., Madl, T.*, et al.* (2018). Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. Cell *173*, 706-719 e713.

Hogan, D.B., Jette, N., Fiest, K.M., Roberts, J.I., Pearson, D., Smith, E.E., Roach, P., Kirk, A., Pringsheim, T., and Maxwell, C.J. (2016). The Prevalence and Incidence of Frontotemporal Dementia: a Systematic Review. Can J Neurol Sci *43 Suppl 1*, S96-S109.

Hoshino, T., Murao, N., Namba, T., Takehara, M., Adachi, H., Katsuno, M., Sobue, G., Matsushima, T., Suzuki, T., and Mizushima, T. (2011). Suppression of Alzheimer's disease-related phenotypes by expression of heat shock protein 70 in mice. The Journal of neuroscience : the official journal of the Society for Neuroscience *31*, 5225-5234.

Howarth, J.L., Kelly, S., Keasey, M.P., Glover, C.P., Lee, Y.B., Mitrophanous, K., Chapple, J.P., Gallo, J.M., Cheetham, M.E., and Uney, J.B. (2007). Hsp40 molecules that target to the ubiquitin-proteasome system decrease inclusion formation in models of polyglutamine disease. Molecular therapy : the journal of the American Society of Gene Therapy *15*, 1100-1105.

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., and Schmid, B. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. Development *140*, 4982-4987.

Hsiung, G.Y., DeJesus-Hernandez, M., Feldman, H.H., Sengdy, P., Bouchard-Kerr, P., Dwosh, E., Butler, R., Leung, B., Fok, A., Rutherford, N.J.*, et al.* (2012). Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p. Brain *135*, 709-722.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Huang, J., Chen, X.H., Wu, K., and Xu, P. (2005). Cloning and expression of a novel isoform of heterogeneous nuclear ribonucleoprotein-R. Neuroreport *16*, 727-730.

Huang, J., Li, S.J., Chen, X.H., Han, Y., and Xu, P. (2008). hnRNP-R regulates the PMA-induced c-fos expression in retinal cells. Cell Mol Biol Lett *13*, 303-311.

Hubers, A., Marroquin, N., Schmoll, B., Vielhaber, S., Just, M., Mayer, B., Hogel, J., Dorst, J., Mertens, T., Just, W.*, et al.* (2014). Polymerase chain reaction and Southern blot-based analysis of the C9orf72 hexanucleotide repeat in different motor neuron diseases. Neurobiol Aging *35*, 1214 e1211-1216.

Hung, C.J., Lee, Y.J., Chen, D.H., and Li, C. (2009). Proteomic analysis of methylarginine-containing proteins in HeLa cells by two-dimensional gel electrophoresis and immunoblotting with a methylarginine-specific antibody. Protein J *28*, 139-147.

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*, 702-705.

Iko, Y., Kodama, T.S., Kasai, N., Oyama, T., Morita, E.H., Muto, T., Okumura, M., Fujii, R., Takumi, T., Tate, S.*, et al.* (2004). Domain architectures and characterization of an RNA-binding protein, TLS. J Biol Chem *279*, 44834-44840.

Isaacs, A.M., Johannsen, P., Holm, I., Nielsen, J.E., and consortium, F.R. (2011). Frontotemporal dementia caused by CHMP2B mutations. Curr Alzheimer Res *8*, 246-251.

Ishigaki, S., Masuda, A., Fujioka, Y., Iguchi, Y., Katsuno, M., Shibata, A., Urano, F., Sobue, G., and Ohno, K. (2012). Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. Scientific reports *2*, 529.

Ishiura, H., Takahashi, Y., Mitsui, J., Yoshida, S., Kihira, T., Kokubo, Y., Kuzuhara, S., Ranum, L.P., Tamaoki, T., Ichikawa, Y.*, et al.* (2012). C9ORF72 repeat expansion in amyotrophic lateral sclerosis in the Kii peninsula of Japan. Arch Neurol *69*, 1154-1158.

Jana, N.R., Tanaka, M., Wang, G., and Nukina, N. (2000). Polyglutamine length-dependent interaction of Hsp40 and Hsp70 family chaperones with truncated N-terminal huntingtin: their role in suppression of aggregation and cellular toxicity. Human molecular genetics *9*, 2009-2018.

Jeng, W., Lee, S., Sung, N., Lee, J., and Tsai, F.T. (2015). Molecular chaperones: guardians of the proteome in normal and disease states. F1000Research *4*.

Jensen, E.C. (2012). Use of fluorescent probes: their effect on cell biology and limitations. Anat Rec (Hoboken) *295*, 2031-2036.

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*, 535-550.

Jiang, Y., Lv, H., Liao, M., Xu, X., Huang, S., Tan, H., Peng, T., Zhang, Y., and Li, H. (2012). GRP78 counteracts cell death and protein aggregation caused by mutant huntingtin proteins. Neurosci Lett *516*, 182-187.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jobert, L., Argentini, M., and Tora, L. (2009). PRMT1 mediated methylation of TAF15 is required for its positive gene regulatory function. Exp Cell Res *315*, 1273-1286.

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. The Journal of biological chemistry *284*, 20329-20339.

Johnson, L., Miller, J.W., Gkazi, A.S., Vance, C., Topp, S.D., Newhouse, S.J., Al-Chalabi, A., Smith, B.N., and Shaw, C.E. (2012). Screening for OPTN mutations in a cohort of British amyotrophic lateral sclerosis patients. Neurobiology of aging *33*, 2948 e2915-2947.

Josephs, K.A., Whitwell, J.L., Murray, M.E., Parisi, J.E., Graff-Radford, N.R., Knopman, D.S., Boeve, B.F., Senjem, M.L., Rademakers, R., Jack, C.R., Jr*., et al.* (2013). Corticospinal tract degeneration associated with TDP-43 type C pathology and semantic dementia. Brain *136*, 455-470.

Jovicic, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S.B., Paul, J.W., 3rd, Sun, S., Herdy, J.R., Bieri, G*., et al.* (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat Neurosci *18*, 1226-1229.

Kabashi, E., Lin, L., Tradewell, M.L., Dion, P.A., Bercier, V., Bourgouin, P., Rochefort, D., Bel Hadj, S., Durham, H.D., Vande Velde, C*., 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*, 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. Nat Genet *40*, 572-574.

Kampinga, H.H., and Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol *11*, 579-592.

Kanekura, K., Yagi, T., Cammack, A.J., Mahadevan, J., Kuroda, M., Harms, M.B., Miller, T.M., and Urano, F. (2016). Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. Human molecular genetics *25*, 1803-1813.

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*, 753-767.

Kato, S., Horiuchi, S., Liu, J., Cleveland, D.W., Shibata, N., Nakashima, K., Nagai, R., Hirano, A., Takikawa, M., Kato, M*., et al.* (2000). Advanced glycation endproduct-modified superoxide dismutase-1 (SOD1)-positive inclusions are common to familial amyotrophic lateral sclerosis patients with SOD1 gene mutations and transgenic mice expressing human SOD1 with a G85R mutation. Acta neuropathologica *100*, 490-505.

Kearse, M.G., Green, K.M., Krans, A., Rodriguez, C.M., Linsalata, A.E., Goldstrohm, A.C., and Todd, P.K. (2016). CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. Mol Cell *62*, 314-322.

Kertesz, A., Ang, L.C., Jesso, S., MacKinley, J., Baker, M., Brown, P., Shoesmith, C., Rademakers, R., and Finger, E.C. (2013). Psychosis and hallucinations in frontotemporal dementia with the C9ORF72 mutation: a detailed clinical cohort. Cogn Behav Neurol *26*, 146-154.

Khosravi, B., Hartmann, H., May, S., Mohl, C., Ederle, H., Michaelsen, M., Schludi, M.H., Dormann, D., and Edbauer, D. (2016). Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD. Human molecular genetics.

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*, 942-955.

Kim, D.Y., Kim, W., Lee, K.H., Kim, S.H., Lee, H.R., Kim, H.J., Jung, Y., Choi, J.H., and Kim, K.T. (2013a). hnRNP Q regulates translation of p53 in normal and stress conditions. Cell Death Differ *20*, 226-234.

Kim, D.Y., Kwak, E., Kim, S.H., Lee, K.H., Woo, K.C., and Kim, K.T. (2011). hnRNP Q mediates a phase-dependent translation-coupled mRNA decay of mouse Period3. Nucleic Acids Res *39*, 8901-8914.

Kim, D.Y., Woo, K.C., Lee, K.H., Kim, T.D., and Kim, K.T. (2010). hnRNP Q and PTB modulate the circadian oscillation of mouse Rev-erb alpha via IRES-mediated translation. Nucleic Acids Res *38*, 7068-7078.

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.* (2013b). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*, 467-473.

Kim, J.K., Mastronardi, F.G., Wood, D.D., Lubman, D.M., Zand, R., and Moscarello, M.A. (2003). Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis. Mol Cell Proteomics *2*, 453-462.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res *24*, 1012-1019.

Kohli, M.A., John-Williams, K., Rajbhandary, R., Naj, A., Whitehead, P., Hamilton, K., Carney, R.M., Wright, C., Crocco, E., Gwirtzman, H.E.*, et al.* (2013). Repeat expansions in the C9ORF72 gene contribute to Alzheimer's disease in Caucasians. Neurobiology of aging *34*, 1519 e1515-1512.

Konno, T., Shiga, A., Tsujino, A., Sugai, A., Kato, T., Kanai, K., Yokoseki, A., Eguchi, H., Kuwabara, S., Nishizawa, M.*, et al.* (2013). Japanese amyotrophic lateral sclerosis patients with GGGGCC hexanucleotide repeat expansion in C9ORF72. Journal of neurology, neurosurgery, and psychiatry *84*, 398-401.

Koppers, M., Blokhuis, A.M., Westeneng, H.J., Terpstra, M.L., Zundel, C.A., Vieira de Sa, 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*, 426-438.

Kovar, H. (2011). Dr. Jekyll and Mr. Hyde: The Two Faces of the FUS/EWS/TAF15 Protein Family. Sarcoma *2011*, 837474.

Koyama, S., Arawaka, S., Chang-Hong, R., Wada, M., Kawanami, T., Kurita, K., Kato, M., Nagai, M., Aoki, M., Itoyama, Y*., et al.* (2006). Alteration of familial ALS-linked mutant SOD1 solubility with disease progression: its modulation by the proteasome and Hsp70. Biochemical and biophysical research communications *343*, 719-730.

Krasowska, J., Olasek, M., Bzowska, A., Clark, P.L., and Wielgus-Kutrowska, B. (2010). The comparison of aggregation and folding of enhanced green fluorescent protein (EGFP) by spectroscopic studies. Spectroscopy *24*.

Kurian, K.M., Forbes, R.B., Colville, S., and Swingler, R.J. (2009). Cause of death and clinical grading criteria in a cohort of amyotrophic lateral sclerosis cases undergoing autopsy from the Scottish Motor Neurone Disease Register. J Neurol Neurosurg Psychiatry *80*, 84-87.

Kwiatkowski, T.J., Jr., 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*, 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*, 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 Neurol *9*, 978-985.

Labbadia, J., Novoselov, S.S., Bett, J.S., Weiss, A., Paganetti, P., Bates, G.P., and Cheetham, M.E. (2012). Suppression of protein aggregation by chaperone modification of high molecular weight complexes. Brain : a journal of neurology *135*, 1180-1196.

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. Proc Natl Acad Sci U S A *110*, E4530-4539.

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

Larsen, S.C., Sylvestersen, K.B., Mund, A., Lyon, D., Mullari, M., Madsen, M.V., Daniel, J.A., Jensen, L.J., and Nielsen, M.L. (2016). Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. Sci Signal *9*, rs9.

Lashley, T., Rohrer, J.D., Bandopadhyay, R., Fry, C., Ahmed, Z., Isaacs, A.M., Brelstaff, J.H., Borroni, B., Warren, J.D., Troakes, C.*, et al.* (2011). A comparative clinical, pathological, biochemical and genetic study of fused in sarcoma proteinopathies. Brain *134*, 2548-2564.

Lashley, T., Rohrer, J.D., Mead, S., and Revesz, T. (2015). Review: an update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations. Neuropathol Appl Neurobiol *41*, 858-881.

Lattante, S., Rouleau, G.A., and Kabashi, E. (2013). TARDBP and FUS mutations associated with amyotrophic lateral sclerosis: summary and update. Hum Mutat *34*, 812-826.

Law, W.J., Cann, K.L., and Hicks, G.G. (2006). TLS, EWS and TAF15: a model for transcriptional integration of gene expression. Brief Funct Genomic Proteomic *5*, 8-14.

Le Ber, I., Camuzat, A., Berger, E., Hannequin, D., Laquerriere, A., Golfier, V., Seilhean, D., Viennet, G., Couratier, P., Verpillat, P.*, et al.* (2009). Chromosome 9p-linked families with frontotemporal dementia associated with motor neuron disease. Neurology *72*, 1669-1676.

Lee, B.J., Cansizoglu, A.E., Suel, K.E., Louis, T.H., Zhang, Z., and Chook, Y.M. (2006). Rules for nuclear localization sequence recognition by karyopherin beta 2. Cell *126*, 543-558.

Lee, H.R., Kim, T.D., Kim, H.J., Jung, Y., Lee, D., Lee, K.H., Kim, D.Y., Woo, K.C., and Kim, K.T. (2015). Heterogeneous ribonucleoprotein R regulates arylalkylamine N-acetyltransferase synthesis via internal ribosomal entry site-mediated translation in a circadian manner. J Pineal Res *59*, 518-529.

Lee, K.H., Woo, K.C., Kim, D.Y., Kim, T.D., Shin, J., Park, S.M., Jang, S.K., and Kim, K.T. (2012). Rhythmic interaction between Period1 mRNA and hnRNP Q leads to circadian time-dependent translation. Mol Cell Biol *32*, 717-728.

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*, 774-788 e717.

Lee, S.J., Lim, H.S., Masliah, E., and Lee, H.J. (2011). Protein aggregate spreading in neurodegenerative diseases: problems and perspectives. Neurosci Res *70*, 339-348.

Lee, V.M., and Trojanowski, J.Q. (2001). Transgenic mouse models of tauopathies: prospects for animal models of Pick's disease. Neurology *56*, S26-30.

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*, 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*, 1178-1186.

Lee, Y.H., and Stallcup, M.R. (2009). Minireview: protein arginine methylation of nonhistone proteins in transcriptional regulation. Mol Endocrinol *23*, 425-433.

Lesage, S., Le Ber, I., Condroyer, C., Broussolle, E., Gabelle, A., Thobois, S., Pasquier, F., Mondon, K., Dion, P.A., Rochefort, D.*, et al.* (2013). C9orf72 repeat expansions are a rare genetic cause of parkinsonism. Brain *136*, 385-391.

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*, 499-503.

Li, H.L., Gee, P., Ishida, K., and Hotta, A. (2016). Efficient genomic correction methods in human iPS cells using CRISPR-Cas9 system. Methods *101*, 27-35.

Li, H.R., Chen, T.C., Hsiao, C.L., Shi, L., Chou, C.Y., and Huang, J.R. (2018). The physical forces mediating self-association and phase-separation in the C-terminal domain of TDP-43. Biochimica et biophysica acta *1866*, 214-223.

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

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S.*, et al.* (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J Biotechnol *208*, 44-53.

Lillo, P., and Hodges, J.R. (2009). Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia *16*, 1131-1135.

Lim, I., Jung, Y., Kim, D.Y., and Kim, K.T. (2016). HnRNP Q Has a Suppressive Role in the Translation of Mouse Cryptochrome1. PLoS One *11*, e0159018.

Lin, Y., Mori, E., Kato, M., Xiang, S., Wu, L., Kwon, I., and McKnight, S.L. (2016). Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. Cell *167*, 789-802 e712.

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

Liu-Yesucevitz, L., Bilgutay, A., Zhang, Y.J., Vanderweyde, T., Citro, A., Mehta, T., Zaarur, N., McKee, A., Bowser, R., Sherman, M.*, et al.* (2010). Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. PloS one *5*, e13250.

Liu, F., Liu, Q., Lu, C.X., Cui, B., Guo, X.N., Wang, R.R., Liu, M.S., Li, X.G., Cui, L.Y., and Zhang, X. (2016a). Identification of a novel loss-of-function C9orf72 splice site mutation in a patient with amyotrophic lateral sclerosis. Neurobiol Aging.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y.*, et al.* (2015). IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics *31*, 3359-3361.

Liu, Y., Pattamatta, A., Zu, T., Reid, T., Bardhi, O., Borchelt, D.R., Yachnis, A.T., and Ranum, L.P. (2016b). C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. Neuron *90*, 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. J Neurol Neurosurg Psychiatry *81*, 385-390.

Lomen-Hoerth, C., Anderson, T., and Miller, B. (2002). The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. Neurology *59*, 1077-1079.

Lopez-Gonzalez, R., Lu, Y., Gendron, T.F., Karydas, A., Tran, H., Yang, D., Petrucelli, L., Miller, B.L., Almeida, S., and Gao, F.B. (2016). Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron.

Luo, Y., Jiao, B., Wang, J., Du, J., Yan, X., Xia, K., Tang, B., and Shen, L. (2014). C9orf72 hexanucleotide repeat expansion analysis in Chinese spastic paraplegia patients. J Neurol Sci *347*, 104-106.

Luty, A.A., Kwok, J.B., Thompson, E.M., Blumbergs, P., Brooks, W.S., Loy, C.T., Dobson-Stone, C., Panegyres, P.K., Hecker, J., Nicholson, G.A.*, et al.* (2008). Pedigree with frontotemporal lobar degeneration--motor neuron disease and Tar DNA binding protein-43 positive neuropathology: genetic linkage to chromosome 9. BMC Neurol *8*, 32.

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 Neuropathol *126*, 859-879.

Mackenzie, I.R., Frick, P., Grasser, 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*, 845-861.

Mackenzie, I.R., Munoz, D.G., Kusaka, H., Yokota, O., Ishihara, K., Roeber, S., Kretzschmar, H.A., Cairns, N.J., and Neumann, M. (2011). Distinct pathological subtypes of FTLD-FUS. Acta neuropathologica *121*, 207-218.

Mackenzie, I.R., and Neumann, M. (2016). Molecular neuropathology of frontotemporal dementia: insights into disease mechanisms from postmortem studies. Journal of neurochemistry *138 Suppl 1*, 54-70.

Maharjan, N., Kunzli, C., Buthey, K., and Saxena, S. (2016). C9ORF72 Regulates Stress Granule Formation and Its Deficiency Impairs Stress Granule Assembly, Hypersensitizing Cells to Stress. Molecular neurobiology.

Mahoney, C.J., Beck, J., Rohrer, J.D., Lashley, T., Mok, K., Shakespeare, T., Yeatman, T., Warrington, E.K., Schott, J.M., Fox, N.C.*, et al.* (2012a). Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. Brain *135*, 736-750.

Mahoney, C.J., Downey, L.E., Ridgway, G.R., Beck, J., Clegg, S., Blair, M., Finnegan, S., Leung, K.K., Yeatman, T., Golden, H.*, et al.* (2012b). Longitudinal neuroimaging

and neuropsychological profiles of frontotemporal dementia with C9ORF72 expansions. Alzheimer's research & therapy *4*, 41.

Majounie, E., Renton, A.E., Mok, K., Dopper, E.G., Waite, A., Rollinson, S., Chio, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J*., et al.* (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. The Lancet Neurology *11*, 323-330.

Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C., and Thornton, C.A. (2002). Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. Mol Cell *10*, 35-44.

Marangi, G., and Traynor, B.J. (2015). Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. Brain Res *1607*, 75-93.

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. J Biol Chem *288*, 24731-24741.

May, S., Hornburg, D., Schludi, M.H., Arzberger, T., Rentzsch, K., Schwenk, B.M., Grasser, 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 Neuropathol *128*, 485-503.

McBride, A.E., and Silver, P.A. (2001). State of the arg: protein methylation at arginine comes of age. Cell *106*, 5-8.

McFarland, N.R., Dimant, H., Kibuuka, L., Ebrahimi-Fakhari, D., Desjardins, C.A., Danzer, K.M., Danzer, M., Fan, Z., Schwarzschild, M.A., Hirst, W*., et al.* (2014). Chronic treatment with novel small molecule Hsp90 inhibitors rescues striatal dopamine levels but not alpha-synuclein-induced neuronal cell loss. PloS one *9*, e86048.

McGurk, L., Lee, V.M., Trojanowksi, J.Q., Van Deerlin, V.M., Lee, E.B., and Bonini, N.M. (2014). Poly-A binding protein-1 localization to a subset of TDP-43 inclusions in amyotrophic lateral sclerosis occurs more frequently in patients harboring an expansion in C9orf72. J Neuropathol Exp Neurol *73*, 837-845.

McLean, P.J., Kawamata, H., Shariff, S., Hewett, J., Sharma, N., Ueda, K., Breakefield, X.O., and Hyman, B.T. (2002). TorsinA and heat shock proteins act as molecular chaperones: suppression of alpha-synuclein aggregation. Journal of neurochemistry *83*, 846-854.

Meisler, M.H., Grant, A.E., Jones, J.M., Lenk, G.M., He, F., Todd, P.K., Kamali, M., Albin, R.L., Lieberman, A.P., Langenecker, S.A*., et al.* (2013). C9ORF72 expansion in a family with bipolar disorder. Bipolar Disord *15*, 326-332.

Millecamps, S., Boillee, S., Le Ber, I., Seilhean, D., Teyssou, E., Giraudeau, M., Moigneu, C., Vandenberghe, N., Danel-Brunaud, V., Corcia, P*., et al.* (2012). Phenotype difference between ALS patients with expanded repeats in C9ORF72 and patients with mutations in other ALS-related genes. J Med Genet *49*, 258-263.

Miller, J.W., Urbinati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., and Swanson, M.S. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. EMBO J *19*, 4439-4448.

Mitchell, J.D., and Borasio, G.D. (2007). Amyotrophic lateral sclerosis. Lancet *369*, 2031-2041.

Miura, H., Gurumurthy, C.B., Sato, T., Sato, M., and Ohtsuka, M. (2015). CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. Scientific reports *5*, 12799.

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*, 1192-1194.

Mizielinska, S., Lashley, T., Norona, F.E., Clayton, E.L., Ridler, C.E., Fratta, P., and Isaacs, A.M. (2013). C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol *126*, 845-857.

Mizielinska, S., Ridler, C.E., Balendra, R., Thoeng, A., Woodling, N.S., Grasser, F.A., Plagnol, V., Lashley, T., Partridge, L., and Isaacs, A.M. (2017). Bidirectional nucleolar dysfunction in C9orf72 frontotemporal lobar degeneration. Acta Neuropathol Commun *5*, 29.

Mizutani, A., Fukuda, M., Ibata, K., Shiraishi, Y., and Mikoshiba, K. (2000). SYNCRIP, a cytoplasmic counterpart of heterogeneous nuclear ribonucleoprotein R, interacts with ubiquitous synaptotagmin isoforms. J Biol Chem *275*, 9823-9831.

Mohajeri, M.H., Madani, R., Saini, K., Lipp, H.P., Nitsch, R.M., and Wolfer, D.P. (2004). The impact of genetic background on neurodegeneration and behavior in seizured mice. Genes Brain Behav *3*, 228-239.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell *163*, 123-133.

Montuschi, A., Iazzolino, B., Calvo, A., Moglia, C., Lopiano, L., Restagno, G., Brunetti, M., Ossola, I., Lo Presti, A., Cammarosano, S.*, et al.* (2015). Cognitive correlates in amyotrophic lateral sclerosis: a population-based study in Italy. J Neurol Neurosurg Psychiatry *86*, 168-173.

Mori, K., Arzberger, T., Grasser, 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*, 881-893.

Mori, K., Lammich, S., Mackenzie, I.R., Forne, I., Zilow, S., Kretzschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M.*, et al.* (2013b). hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol *125*, 413-423.

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.* (2013c). The C9orf72

GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science *339*, 1335-1338.

Morita, M., Al-Chalabi, A., Andersen, P.M., Hosler, B., Sapp, P., Englund, E., Mitchell, J.E., Habgood, J.J., de Belleroche, J., Xi, J.*, et al.* (2006). A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology *66*, 839-844.

Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., and Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. EMBO J *20*, 5443-5452.

Mukherjee, O., Das, G., Sen, S., Dutt, A., Alladi, S., and Ghosh, A. (2015). C9orf72 mutations may be rare in frontotemporal lobar degeneration patients in India. Amyotroph Lateral Scler Frontotemporal Degener *17*, 151-153.

Munoz, D.G., Neumann, M., Kusaka, H., Yokota, O., Ishihara, K., Terada, S., Kuroda, S., and Mackenzie, I.R. (2009). FUS pathology in basophilic inclusion body disease. Acta Neuropathol *118*, 617-627.

Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T., Michel, C.H.*, et al.* (2015). ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron *88*, 678-690.

Murray, M.E., DeJesus-Hernandez, M., Rutherford, N.J., Baker, M., Duara, R., Graff-Radford, N.R., Wszolek, Z.K., Ferman, T.J., Josephs, K.A., Boylan, K.B.*, et al.* (2011). Clinical and neuropathologic heterogeneity of c9FTD/ALS associated with hexanucleotide repeat expansion in C9ORF72. Acta Neuropathol *122*, 673-690.

Nagarkar-Jaiswal, S., Lee, P.T., Campbell, M.E., Chen, K., Anguiano-Zarate, S., Gutierrez, M.C., Busby, T., Lin, W.W., He, Y., Schulze, K.L.*, et al.* (2015). A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. eLife *4*.

Nandi, D., Tahiliani, P., Kumar, A., and Chandu, D. (2006). The ubiquitin-proteasome system. J Biosci *31*, 137-155.

Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P.H., Albert, M.*, et al.* (1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology *51*, 1546-1554.

Neary, D., Snowden, J.S., and Mann, D.M. (2000). Cognitive change in motor neurone disease/amyotrophic lateral sclerosis (MND/ALS). J Neurol Sci *180*, 15-20.

Neumann, M., Bentmann, E., Dormann, D., Jawaid, A., DeJesus-Hernandez, M., Ansorge, O., Roeber, S., Kretzschmar, H.A., Munoz, D.G., Kusaka, H.*, et al.* (2011). FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. Brain *134*, 2595-2609.

Neumann, M., Mackenzie, I.R., Cairns, N.J., Boyer, P.J., Markesbery, W.R., Smith, C.D., Taylor, J.P., Kretzschmar, H.A., Kimonis, V.E., and Forman, M.S. (2007). TDP-43 in the ubiquitin pathology of frontotemporal dementia with VCP gene mutations. J Neuropathol Exp Neurol *66*, 152-157.

Neumann, M., Rademakers, R., Roeber, S., Baker, M., Kretzschmar, H.A., and Mackenzie, I.R. (2009a). A new subtype of frontotemporal lobar degeneration with FUS pathology. Brain *132*, 2922-2931.

Neumann, M., Roeber, S., Kretzschmar, H.A., Rademakers, R., Baker, M., and Mackenzie, I.R. (2009b). Abundant FUS-immunoreactive pathology in neuronal intermediate filament inclusion disease. Acta Neuropathol *118*, 605-616.

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*, 130-133.

Neumann, M., Valori, C.F., Ansorge, O., Kretzschmar, H.A., Munoz, D.G., Kusaka, H., Yokota, O., Ishihara, K., Ang, L.C., Bilbao, J.M*., et al.* (2012). Transportin 1 accumulates specifically with FET proteins but no other transportin cargos in FTLD-FUS and is absent in FUS inclusions in ALS with FUS mutations. Acta Neuropathol *124*, 705-716.

Nichols, L.O., Martindale-Adams, J., Burns, R., Graney, M.J., and Zuber, J. (2011). Translation of a dementia caregiver support program in a health care system--REACH VA. Arch Intern Med *171*, 353-359.

Nielsen, T.T., Svenstrup, K., Duno, M., and Nielsen, J.E. (2014). Hereditary spastic paraplegia is not associated with C9ORF72 repeat expansions in a Danish cohort. Spinal Cord *52*, 77-79.

Novoselov, S.S., Mustill, W.J., Gray, A.L., Dick, J.R., Kanuga, N., Kalmar, B., Greensmith, L., and Cheetham, M.E. (2013). Molecular chaperone mediated late-stage neuroprotection in the SOD1(G93A) mouse model of amyotrophic lateral sclerosis. PloS one *8*, e73944.

O'Rourke, J.G., Bogdanik, L., Muhammad, A.K., 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*, 892-901.

O'Rourke, J.G., Bogdanik, L., Yanez, A., Lall, D., Wolf, A.J., Muhammad, A.K., 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*, 1324-1329.

Ohki, Y., Wenninger-Weinzierl, A., Hruscha, A., Asakawa, K., Kawakami, K., Haass, C., Edbauer, D., and Schmid, B. (2017). Glycine-alanine dipeptide repeat protein contributes to toxicity in a zebrafish model of C9orf72 associated neurodegeneration. Mol Neurodegener *12*, 6.

Ong, S.E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods *1*, 119-126.

Onyike, C.U., and Huey, E.D. (2013). Frontotemporal dementia and psychiatry. Int Rev Psychiatry *25*, 127-129.

Pasinelli, P., and Brown, R.H. (2006). Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat Rev Neurosci *7*, 710-723.

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*, 1066-1077.

Patel, Y.J., Payne Smith, M.D., de Belleroche, J., and Latchman, D.S. (2005). Hsp27 and Hsp70 administered in combination have a potent protective effect against FALS-associated SOD1-mutant-induced cell death in mammalian neuronal cells. Brain Res Mol Brain Res *134*, 256-274.

Pearson, J.P., Williams, N.M., Majounie, E., Waite, A., Stott, J., Newsway, V., Murray, A., Hernandez, D., Guerreiro, R., Singleton, A.B.*, et al.* (2011). Familial frontotemporal dementia with amyotrophic lateral sclerosis and a shared haplotype on chromosome 9p. J Neurol *258*, 647-655.

Penagarikano, O., Mulle, J.G., and Warren, S.T. (2007). The pathophysiology of fragile x syndrome. Annu Rev Genomics Hum Genet *8*, 109-129.

Peng, C., and Wong, C.C. (2017). The story of protein arginine methylation: characterization, regulation, and function. Expert Rev Proteomics, 1-14.

Peng, Z.Y., Huang, J., Lee, S.C., Shi, Y.L., Chen, X.H., and Xu, P. (2009). The expression pattern of heterogeneous nuclear ribonucleoprotein R in rat retina. Neurochem Res *34*, 1083-1088.

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*, 902-909.

Piggott, M.A., Ballard, C.G., Rowan, E., Holmes, C., McKeith, I.G., Jaros, E., Perry, R.H., and Perry, E.K. (2007). Selective loss of dopamine D2 receptors in temporal cortex in dementia with Lewy bodies, association with cognitive decline. Synapse *61*, 903-911.

Pikkarainen, M., Hartikainen, P., and Alafuzoff, I. (2010). Ubiquitinated p62-positive, TDP-43-negative inclusions in cerebellum in frontotemporal lobar degeneration with TAR DNA binding protein 43. Neuropathology *30*, 197-199.

Pillai, S.G., Chiano, M.N., White, N.J., Speer, M., Barnes, K.C., Carlsen, K., Gerritsen, J., Helms, P., Lenney, W., Silverman, M.*, et al.* (2006). A genome-wide search for linkage to asthma phenotypes in the genetics of asthma international network families: evidence for a major susceptibility locus on chromosome 2p. European journal of human genetics : EJHG *14*, 307-316.

Pollard, V.W., Michael, W.M., Nakielny, S., Siomi, M.C., Wang, F., and Dreyfuss, G. (1996). A novel receptor-mediated nuclear protein import pathway. Cell *86*, 985-994.

Portet, F., Cadilhac, C., Touchon, J., and Camu, W. (2001). Cognitive impairment in motor neuron disease with bulbar onset. Amyotroph Lateral Scler Other Motor Neuron Disord *2*, 23-29.

Protter, D.S.W., and Parker, R. (2016). Principles and Properties of Stress Granules. Trends Cell Biol *26*, 668-679.

Proudfoot, M., Gutowski, N.J., Edbauer, D., Hilton, D.A., Stephens, M., Rankin, J., and Mackenzie, I.R. (2014). Early dipeptide repeat pathology in a frontotemporal dementia kindred with C9ORF72 mutation and intellectual disability. Acta Neuropathol *127*, 451-458.

Putcha, P., Danzer, K.M., Kranich, L.R., Scott, A., Silinski, M., Mabbett, S., Hicks, C.D., Veal, J.M., Steed, P.M., Hyman, B.T.*, et al.* (2010). Brain-permeable small-molecule inhibitors of Hsp90 prevent alpha-synuclein oligomer formation and rescue alpha-synuclein-induced toxicity. J Pharmacol Exp Ther *332*, 849-857.

Qamar, S., Wang, G., Randle, S.J., Ruggeri, F.S., Varela, J.A., Lin, J.Q., Phillips, E.C., Miyashita, A., Williams, D., Strohl, F.*, et al.* (2018). FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-pi Interactions. Cell *173*, 720-734 e715.

Quaresma, A.J., Bressan, G.C., Gava, L.M., Lanza, D.C., Ramos, C.H., and Kobarg, J. (2009). Human hnRNP Q re-localizes to cytoplasmic granules upon PMA, thapsigargin, arsenite and heat-shock treatments. Exp Cell Res *315*, 968-980.

Rabinovici, G.D., and Miller, B.L. (2010). Frontotemporal lobar degeneration: epidemiology, pathophysiology, diagnosis and management. CNS Drugs *24*, 375-398.

Rademakers, R., Neumann, M., and Mackenzie, I.R. (2012). Advances in understanding the molecular basis of frontotemporal dementia. Nature reviews Neurology *8*, 423-434.

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

Rappsilber, J., Friesen, W.J., Paushkin, S., Dreyfuss, G., and Mann, M. (2003). Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. Anal Chem *75*, 3107-3114.

Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., van Swieten, J.C., Seelaar, H., Dopper, E.G., Onyike, C.U.*, et al.* (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain *134*, 2456-2477.

Ratti, A., and Buratti, E. (2016). Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. Journal of neurochemistry *138 Suppl 1*, 95-111.

Ravikumar, B., Sarkar, S., Davies, J.E., Futter, M., Garcia-Arencibia, M., Green-Thompson, Z.W., Jimenez-Sanchez, M., Korolchuk, V.I., Lichtenberg, M., Luo, S.*, et al.* (2010). Regulation of mammalian autophagy in physiology and pathophysiology. Physiol Rev *90*, 1383-1435.

Reches, A., Nachmani, D., Berhani, O., Duev-Cohen, A., Shreibman, D., Ophir, Y., Seliger, B., and Mandelboim, O. (2016). HNRNPR Regulates the Expression of Classical and Nonclassical MHC Class I Proteins. J Immunol *196*, 4967-4976.

Reddy, K., Zamiri, B., Stanley, S.Y., Macgregor, R.B., Jr., 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. J Biol Chem *288*, 9860-9866.

Renton, A.E., Chio, A., and Traynor, B.J. (2014). State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci *17*, 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*, 257-268.

Richter, K., Haslbeck, M., and Buchner, J. (2010). The heat shock response: life on the verge of death. Mol Cell *40*, 253-266.

Ridler, C.E. (2016). C9orf72 frontotemporal dementia and amyotrophic lateral sclerosis: investigating repeat pathology in cell culture models and human post-mortem brain. In Department of Neurodegenerative Disease (London: University College London).

Ringholz, G.M., Appel, S.H., Bradshaw, M., Cooke, N.A., Mosnik, D.M., and Schulz, P.E. (2005). Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology *65*, 586-590.

Robberecht, W., and Philips, T. (2013). The changing scene of amyotrophic lateral sclerosis. Nat Rev Neurosci *14*, 248-264.

Rohrer, J.D., Caso, F., Mahoney, C., Henry, M., Rosen, H.J., Rabinovici, G., Rossor, M.N., Miller, B., Warren, J.D., Fox, N.C.*, et al.* (2013). Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia. Brain Lang *127*, 121-126.

Rohrer, J.D., Guerreiro, R., Vandrovcova, J., Uphill, J., Reiman, D., Beck, J., Isaacs, A.M., Authier, A., Ferrari, R., Fox, N.C.*, et al.* (2009). The heritability and genetics of frontotemporal lobar degeneration. Neurology *73*, 1451-1456.

Rohrer, J.D., Isaacs, A.M., Mizielinska, S., Mead, S., Lashley, T., Wray, S., Sidle, K., Fratta, P., Orrell, R.W., Hardy, J.*, et al.* (2015). C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. The Lancet Neurology *14*, 291-301.

Rohrer, J.D., Lashley, T., Holton, J., Revesz, T., Urwin, H., Isaacs, A.M., Fox, N.C., Rossor, M.N., and Warren, J. (2011). The clinical and neuroanatomical phenotype of FUS associated frontotemporal lobar degeneration. J Neurol Neurosurg Psychiatry *82*, 1405-1407.

Rollinson, S., Halliwell, N., Young, K., Callister, J.B., Toulson, G., Gibbons, L., Davidson, Y.S., Robinson, A.C., Gerhard, A., Richardson, A.*, et al.* (2012). Analysis of the hexanucleotide repeat in C9ORF72 in Alzheimer's disease. Neurobiology of aging *33*, 1846 e1845-1846.

Rollinson, S., Rizzu, P., Sikkink, S., Baker, M., Halliwell, N., Snowden, J., Traynor, B.J., Ruano, D., Cairns, N., Rohrer, J.D.*, et al.* (2009). Ubiquitin associated protein 1 is a risk factor for frontotemporal lobar degeneration. Neurobiology of aging *30*, 656-665.

Rose, J.M., Novoselov, S.S., Robinson, P.A., and Cheetham, M.E. (2011). Molecular chaperone-mediated rescue of mitophagy by a Parkin RING1 domain mutant. Human molecular genetics *20*, 16-27.

Rosen, D.R. (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature *364*, 362.

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*, 59-62.

Ross, C.A., and Poirier, M.A. (2005). Opinion: What is the role of protein aggregation in neurodegeneration? Nat Rev Mol Cell Biol *6*, 891-898.

Rossi, S., Serrano, A., Gerbino, V., Giorgi, A., Di Francesco, L., Nencini, M., Bozzo, F., Schinina, 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. J Cell Sci *128*, 1787-1799.

Rossoll, W., Kroning, A.K., Ohndorf, U.M., Steegborn, C., Jablonka, S., and Sendtner, M. (2002). Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? Human molecular genetics *11*, 93-105.

Rubino, E., Rainero, I., Chio, A., Rogaeva, E., Galimberti, D., Fenoglio, P., Grinberg, Y., Isaia, G., Calvo, A., Gentile, S.*, et al.* (2012). SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Neurology *79*, 1556-1562.

Rudich, P., Snoznik, C., Watkins, S.C., Monaghan, J., Pandey, U.B., and Lamitina, S.T. (2017). Nuclear localized C9orf72-associated arginine-containing dipeptides exhibit age-dependent toxicity in C. elegans. Human molecular genetics *26*, 4916-4928.

Rutherford, N.J., Heckman, M.G., Dejesus-Hernandez, M., Baker, M.C., Soto-Ortolaza, A.I., Rayaprolu, S., Stewart, H., Finger, E., Volkening, K., Seeley, W.W.*, et al.* (2012). Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype. Neurobiology of aging *33*, 2950 e2955-2957.

Sabatelli, M., Conforti, F.L., Zollino, M., Mora, G., Monsurro, M.R., Volanti, P., Marinou, K., Salvi, F., Corbo, M., Giannini, F.*, et al.* (2012). C9ORF72 hexanucleotide repeat expansions in the Italian sporadic ALS population. Neurobiology of aging *33*, 1848 e1815-1820.

Saberi, S., Stauffer, J.E., Jiang, J., Garcia, S.D., Taylor, A.E., Schulte, D., Ohkubo, T., Schloffman, C.L., Maldonado, M., Baughn, M.*, et al.* (2018). Sense-encoded poly-GR dipeptide repeat proteins correlate to neurodegeneration and uniquely co-localize with TDP-43 in dendrites of repeat-expanded C9orf72 amyotrophic lateral sclerosis. Acta Neuropathol *135*, 459-474.

Saberi, S., Stauffer, J.E., Schulte, D.J., and Ravits, J. (2015). Neuropathology of Amyotrophic Lateral Sclerosis and Its Variants. Neurol Clin *33*, 855-876.

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol *14*, 630-642.

Sakae, N., Bieniek, K.F., Zhang, Y.J., Ross, K., Gendron, T.F., Murray, M.E., Rademakers, R., Petrucelli, L., and Dickson, D.W. (2018). Poly-GR dipeptide repeat polymers correlate with neurodegeneration and Clinicopathological subtypes in C9ORF72-related brain disease. Acta Neuropathol Commun *6*, 63.

Sama, R.R., Ward, C.L., and Bosco, D.A. (2014). Functions of FUS/TLS from DNA repair to stress response: implications for ALS. ASN Neuro *6*.

Sama, R.R., Ward, C.L., Kaushansky, L.J., Lemay, N., Ishigaki, S., Urano, F., and Bosco, D.A. (2013). FUS/TLS assembles into stress granules and is a prosurvival factor during hyperosmolar stress. J Cell Physiol *228*, 2222-2231.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355.

Santangelo, L., Giurato, G., Cicchini, C., Montaldo, C., Mancone, C., Tarallo, R., Battistelli, C., Alonzi, T., Weisz, A., and Tripodi, M. (2016). The RNA-Binding Protein SYNCRIP Is a Component of the Hepatocyte Exosomal Machinery Controlling MicroRNA Sorting. Cell reports *17*, 799-808.

Sareen, D., O'Rourke, J.G., Meera, P., Muhammad, A.K., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A.*, et al.* (2013). Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Science translational medicine *5*, 208ra149.

Sawyer, I.A., Bartek, J., and Dundr, M. (2018). Phase separated microenvironments inside the cell nucleus are linked to disease and regulate epigenetic state, transcription and RNA processing. Seminars in cell & developmental biology.

Schanz, O., Bageac, D., Braun, L., Traynor, B.J., Lehky, T.J., and Floeter, M.K. (2016). Cortical hyperexcitability in patients with C9ORF72 mutations: Relationship to phenotype. Muscle & nerve *54*, 264-269.

Schludi, M.H., Becker, L., Garrett, L., Gendron, T.F., Zhou, Q., Schreiber, F., Popper, B., Dimou, L., Strom, T.M., Winkelmann, J.*, et al.* (2017). Spinal poly-GA inclusions in a C9orf72 mouse model trigger motor deficits and inflammation without neuron loss. Acta Neuropathol *134*, 241-254.

Schludi, M.H., May, S., Grasser, F.A., Rentzsch, K., Kremmer, E., Kupper, C., Klopstock, T., German Consortium for Frontotemporal Lobar, D., Bavarian Brain Banking, A., Arzberger, T.*, et al.* (2015). Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol *130*, 537-555.

Sedelis, M., Hofele, K., Auburger, G.W., Morgan, S., Huston, J.P., and Schwarting, R.K. (2000). MPTP susceptibility in the mouse: behavioral, neurochemical, and histological analysis of gender and strain differences. Behav Genet *30*, 171-182.

Sellier, C., Buijsen, R.A.M., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziane, H., Vincent, A.*, et al.* (2017). Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron *93*, 331-347.

Sellier, C., Campanari, M.L., 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.

Selvaraj, B.T., Livesey, M.R., and Chandran, S. (2017). Modeling the C9ORF72 repeat expansion mutation using human induced pluripotent stem cells. Brain Pathol *27*, 518-524.

Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. Nat Methods *2*, 905-909.

Shatunovskii, M.I., and Ruban, G.I. (2010). [Intraspecific variability of life strategies in boreal fishes inhabiting wide ranges]. Izv Akad Nauk Ser Biol, 486-497.

Shi, X., Tian, B., Ma, C., Liu, L., Zhang, N., Na, Y., Li, J., Lu, J., and Qiao, Y. (2017). GSK3beta activity is essential for senescence-associated heterochromatin foci (SAHF) formation induced by HMGA2 in WI38 cells. Am J Transl Res *9*, 167-174.

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. Nat Med *24*, 313-325.

Sieben, A., Van Langenhove, T., Engelborghs, S., Martin, J.J., Boon, P., Cras, P., De Deyn, P.P., Santens, P., Van Broeckhoven, C., and Cruts, M. (2012). The genetics and neuropathology of frontotemporal lobar degeneration. Acta neuropathologica *124*, 353-372.

Simon-Sanchez, J., Dopper, E.G., Cohn-Hokke, P.E., Hukema, R.K., Nicolaou, N., Seelaar, H., de Graaf, J.R., de Koning, I., van Schoor, N.M., Deeg, D.J.*, et al.* (2012). The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. Brain *135*, 723-735.

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.* (2018). G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. EMBO Mol Med *10*, 22-31.

Sivadasan, R., Hornburg, D., Drepper, C., Frank, N., Jablonka, S., Hansel, A., Lojewski, X., Sterneckert, J., Hermann, A., Shaw, P.J.*, et al.* (2016a). C9ORF72 interaction with cofilin modulates actin dynamics in motor neurons. Nat Neurosci *19*, 1610-1618.

Skibinski, G., Parkinson, N.J., Brown, J.M., Chakrabarti, L., Lloyd, S.L., Hummerich, H., Nielsen, J.E., Hodges, J.R., Spillantini, M.G., Thusgaard, T.*, et al.* (2005). Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. Nat Genet *37*, 806-808.

Sleegers, K., Roks, G., Theuns, J., Aulchenko, Y.S., Rademakers, R., Cruts, M., van Gool, W.A., Van Broeckhoven, C., Heutink, P., Oostra, B.A.*, et al.* (2004). Familial clustering and genetic risk for dementia in a genetically isolated Dutch population. Brain : a journal of neurology *127*, 1641-1649.

Smith, H.L., Li, W., and Cheetham, M.E. (2015). Molecular chaperones and neuronal proteostasis. Seminars in cell & developmental biology *40*, 142-152.

Snapp, E. (2005). Design and use of fluorescent fusion proteins in cell biology. Curr Protoc Cell Biol *Chapter 21*, Unit 21 24.

Snowden, J.S., Rollinson, S., Thompson, J.C., Harris, J.M., Stopford, C.L., Richardson, A.M., Jones, M., Gerhard, A., Davidson, Y.S., Robinson, A.*, et al.* (2012). Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. Brain *135*, 693-708.

Soto, C. (2003). Unfolding the role of protein misfolding in neurodegenerative diseases. Nat Rev Neurosci *4*, 49-60.

Spillantini, M.G., and Goedert, M. (2013). Tau pathology and neurodegeneration. Lancet Neurol *12*, 609-622.

Standart, N., and Weil, D. (2018). P-Bodies: Cytosolic Droplets for Coordinated mRNA Storage. Trends Genet *34*, 612-626.

Starr, A., and Sattler, R. (2018). Synaptic dysfunction and altered excitability in C9ORF72 ALS/FTD. Brain research *1693*, 98-108.

Stewart, H., Rutherford, N.J., Briemberg, H., Krieger, C., Cashman, N., Fabros, M., Baker, M., Fok, A., DeJesus-Hernandez, M., Eisen, A.*, et al.* (2012). Clinical and pathological features of amyotrophic lateral sclerosis caused by mutation in the C9ORF72 gene on chromosome 9p. Acta Neuropathol *123*, 409-417.

Strong, M.K., Southwell, A.L., Yonan, J.M., Hayden, M.R., Macgregor, G.R., Thompson, L.M., and Steward, O. (2012). Age-Dependent Resistance to Excitotoxicity in Htt CAG140 Mice and the Effect of Strain Background. J Huntingtons Dis *1*, 221-241.

Stuhlinger, M.C., Tsao, P.S., Her, J.H., Kimoto, M., Balint, R.F., and Cooke, J.P. (2001). Homocysteine impairs the nitric oxide synthase pathway: role of asymmetric dimethylarginine. Circulation *104*, 2569-2575.

Su, Z., Zhang, Y., Gendron, T.F., Bauer, P.O., Chew, J., Yang, W.Y., Fostvedt, E., Jansen-West, K., Belzil, V.V., Desaro, P.*, et al.* (2014a). Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. Neuron *84*, 239.

Suarez-Calvet, M., Neumann, M., Arzberger, T., Abou-Ajram, C., Funk, E., Hartmann, H., Edbauer, D., Kremmer, E., Gobl, C., Resch, M.*, et al.* (2016). Monomethylated and unmethylated FUS exhibit increased binding to Transportin and distinguish FTLD-FUS from ALS-FUS. Acta Neuropathol *131*, 587-604.

Sudria-Lopez, E., Koppers, M., de Wit, M., van der Meer, C., Westeneng, H.J., Zundel, C.A., Youssef, S.A., Harkema, L., de Bruin, A., Veldink, J.H.*, et al.* (2016). Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. Acta Neuropathol *132*, 145-147.

Sullivan, P.M., Zhou, X., Robins, A.M., Paushter, D.H., Kim, D., Smolka, M.B., and Hu, F. (2016). The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. Acta Neuropathol Commun *4*, 51.

Sun, Z., Diaz, Z., Fang, X., Hart, M.P., Chesi, A., Shorter, J., and Gitler, A.D. (2011). Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. PLoS Biol *9*, e1000614.

Suzuki, M., Iijima, M., Nishimura, A., Tomozoe, Y., Kamei, D., and Yamada, M. (2005). Two separate regions essential for nuclear import of the hnRNP D nucleocytoplasmic shuttling sequence. FEBS J *272*, 3975-3987.

Swaminathan, A., Bouffard, M., Liao, M., Ryan, S., Callister, J.B., Pickering-Brown, S.M., Armstrong, G.A.B., and Drapeau, P. (2018). Expression of C9orf72-related dipeptides impairs motor function in a vertebrate model. Human molecular genetics *27*, 1754-1762.

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 Neuropathol *135*, 427-443.

Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C.Z., Lin, C.C., Jiang, J., Jansen-West, K., Abou-Hamdan, H.*, et al.* (2018). CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. Nat Commun *9*, 152.

Tan, A.Y., and Manley, J.L. (2009). The TET family of proteins: functions and roles in disease. J Mol Cell Biol *1*, 82-92.

Tanaka, K., and Matsuda, N. (2014). Proteostasis and neurodegeneration: the roles of proteasomal degradation and autophagy. Biochim Biophys Acta *1843*, 197-204.

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*, 2426-2441.

Tebbenkamp, A.T., and Borchelt, D.R. (2009). Protein aggregate characterization in models of neurodegenerative disease. Methods Mol Biol *566*, 85-91.

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

Theuns, J., Verstraeten, A., Sleegers, K., Wauters, E., Gijselinck, I., Smolders, S., Crosiers, D., Corsmit, E., Elinck, E., Sharma, M.*, et al.* (2014). Global investigation and meta-analysis of the C9orf72 (G4C2)n repeat in Parkinson disease. Neurology *83*, 1906-1913.

Thompson, J.C., Stopford, C.L., Snowden, J.S., and Neary, D. (2005). Qualitative neuropsychological performance characteristics in frontotemporal dementia and Alzheimer's disease. Journal of neurology, neurosurgery, and psychiatry *76*, 920-927.

Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K.C., Scaglione, K.M., Basrur, V.*, et al.* (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron *78*, 440-455.

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*, 1207-1214.

Troakes, C., Maekawa, S., Wijesekera, L., Rogelj, B., Siklos, L., Bell, C., Smith, B., Newhouse, S., Vance, C., Johnson, L*., et al.* (2012). An MND/ALS phenotype associated with C9orf72 repeat expansion: abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline. Neuropathology : official journal of the Japanese Society of Neuropathology *32*, 505-514.

Tsai, C.P., Soong, B.W., Tu, P.H., Lin, K.P., Fuh, J.L., Tsai, P.C., Lu, Y.C., Lee, I.H., and Lee, Y.C. (2012). A hexanucleotide repeat expansion in C9ORF72 causes familial and sporadic ALS in Taiwan. Neurobiol Aging *33*, 2232 e2211-2232 e2218.

Tsai, W.C., Gayatri, S., Reineke, L.C., Sbardella, G., Bedford, M.T., and Lloyd, R.E. (2016). Arginine Demethylation of G3BP1 Promotes Stress Granule Assembly. The Journal of biological chemistry *291*, 22671-22685.

Tsien, R.Y. (1998). The green fluorescent protein. Annu Rev Biochem *67*, 509-544.

Ugolino, J., Ji, Y.J., Conchina, K., Chu, J., Nirujogi, R.S., Pandey, A., Brady, N.R., Hamacher-Brady, A., and Wang, J. (2016). Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling. PLoS Genet *12*, e1006443.

Uhlmann, T., Geoghegan, V.L., Thomas, B., Ridlova, G., Trudgian, D.C., and Acuto, O. (2012). A method for large-scale identification of protein arginine methylation. Mol Cell Proteomics *11*, 1489-1499.

Urwin, H., Josephs, K.A., Rohrer, J.D., Mackenzie, I.R., Neumann, M., Authier, A., Seelaar, H., Van Swieten, J.C., Brown, J.M., Johannsen, P*., et al.* (2010). FUS pathology defines the majority of tau- and TDP-43-negative frontotemporal lobar degeneration. Acta Neuropathol *120*, 33-41.

Uryu, K., Richter-Landsberg, C., Welch, W., Sun, E., Goldbaum, O., Norris, E.H., Pham, C.T., Yazawa, I., Hilburger, K., Micsenyi, M*., et al.* (2006). Convergence of heat shock protein 90 with ubiquitin in filamentous alpha-synuclein inclusions of alpha-synucleinopathies. Am J Pathol *168*, 947-961.

Valdmanis, P.N., Dupre, N., Bouchard, J.P., Camu, W., Salachas, F., Meininger, V., Strong, M., and Rouleau, G.A. (2007). Three families with amyotrophic lateral sclerosis and frontotemporal dementia with evidence of linkage to chromosome 9p. Arch Neurol *64*, 240-245.

van Blitterswijk, M., DeJesus-Hernandez, M., Niemantsverdriet, E., Murray, M.E., Heckman, M.G., Diehl, N.N., Brown, P.H., Baker, M.C., Finch, N.A., Bauer, P.O*., et al.* (2013). Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol *12*, 978-988.

van Blitterswijk, M., Gendron, T.F., Baker, M.C., DeJesus-Hernandez, M., Finch, N.A., Brown, P.H., Daughrity, L.M., Murray, M.E., Heckman, M.G., Jiang, J*., et al.* (2015). Novel clinical associations with specific C9ORF72 transcripts in patients with repeat expansions in C9ORF72. Acta Neuropathol *130*, 863-876.

van Blitterswijk, M., and Rademakers, R. (2015). Neurodegenerative disease: C9orf72 repeats compromise nucleocytoplasmic transport. Nature reviews Neurology.

Van Deerlin, V.M., Sleiman, P.M., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M.*, et al.* (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat Genet *42*, 234-239.

van der Zee, J., Gijselinck, I., Dillen, L., Van Langenhove, T., Theuns, J., Engelborghs, S., Philtjens, S., Vandenbulcke, M., Sleegers, K., Sieben, A.*, et al.* (2013). A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats. Hum Mutat *34*, 363-373.

van Es, M.A., Veldink, J.H., Saris, C.G., Blauw, H.M., van Vught, P.W., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J., Huisman, M.H.*, et al.* (2009). Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet *41*, 1083-1087.

Van Langenhove, T., van der Zee, J., Gijselinck, I., Engelborghs, S., Vandenberghe, R., Vandenbulcke, M., De Bleecker, J., Sieben, A., Versijpt, J., Ivanoiu, A.*, et al.* (2013). Distinct clinical characteristics of C9orf72 expansion carriers compared with GRN, MAPT, and nonmutation carriers in a Flanders-Belgian FTLD cohort. JAMA Neurol *70*, 365-373.

Van Langenhove, T., van der Zee, J., Sleegers, K., Engelborghs, S., Vandenberghe, R., Gijselinck, I., Van den Broeck, M., Mattheijssens, M., Peeters, K., De Deyn, P.P.*, et al.* (2010). Genetic contribution of FUS to frontotemporal lobar degeneration. Neurology *74*, 366-371.

Van Langenhove, T., van der Zee, J., and Van Broeckhoven, C. (2012). The molecular basis of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum. Ann Med *44*, 817-828.

Van Mossevelde, S., van der Zee, J., Gijselinck, I., Sleegers, K., De Bleecker, J., Sieben, A., Vandenberghe, R., Van Langenhove, T., Baets, J., Deryck, O.*, et al.* (2017). Clinical Evidence of Disease Anticipation in Families Segregating a C9orf72 Repeat Expansion. JAMA Neurol *74*, 445-452.

van Rheenen, W., van Blitterswijk, M., Huisman, M.H., Vlam, L., van Doormaal, P.T., Seelen, M., Medic, J., Dooijes, D., de Visser, M., van der Kooi, A.J.*, et al.* (2012). Hexanucleotide repeat expansions in C9ORF72 in the spectrum of motor neuron diseases. Neurology *79*, 878-882.

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*, 1208-1211.

Vatsavayai, S.C., Yoon, S.J., Gardner, R.C., Gendron, T.F., Vargas, J.N., Trujillo, A., Pribadi, M., Phillips, J.J., Gaus, S.E., Hixson, J.D.*, et al.* (2016). Timing and significance of pathological features in C9orf72 expansion-associated frontotemporal dementia. Brain *139*, 3202-3216.

Venditti, A., Frezza, C., Celona, D., Bianco, A., Serafini, M., Cianfaglione, K., Fiorini, D., Ferraro, S., Maggi, F., Lizzi, A.R.*, et al.* (2017). Polar constituents, protection against reactive oxygen species, and nutritional value of Chinese artichoke (Stachys affinis Bunge). Food Chem *221*, 473-481.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell *15*, 27-30.

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. Neurobiol Aging *35*, 1779 e1775-1779 e1713.

Walport, L.J., Hopkinson, R.J., Chowdhury, R., Schiller, R., Ge, W., Kawamura, A., and Schofield, C.J. (2016). Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. Nature communications *7*, 11974.

Weber, S.C., and Brangwynne, C.P. (2015). Inverse size scaling of the nucleolus by a concentration-dependent phase transition. Current biology : CB *25*, 641-646.

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

Weis, K. (2003). Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. Cell *112*, 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*, 1213-1225.

Wen, X., Westergard, T., Pasinelli, P., and Trotti, D. (2017). Pathogenic determinants and mechanisms of ALS/FTD linked to hexanucleotide repeat expansions in the C9orf72 gene. Neurosci Lett *636*, 16-26.

Wesche, J., Kuhn, S., Kessler, B.M., Salton, M., and Wolf, A. (2017). Protein arginine methylation: a prominent modification and its demethylation. Cellular and molecular life sciences : CMLS *74*, 3305-3315.

Westergard, T., Jensen, B.K., Wen, X., Cai, J., Kropf, E., Iacovitti, L., Pasinelli, P., and Trotti, D. (2016). Cell-to-Cell Transmission of Dipeptide Repeat Proteins Linked to C9orf72-ALS/FTD. Cell reports *17*, 645-652.

Westhoff, B., Chapple, J.P., van der Spuy, J., Hohfeld, J., and Cheetham, M.E. (2005). HSJ1 is a neuronal shuttling factor for the sorting of chaperone clients to the proteasome. Current biology : CB *15*, 1058-1064.

Whitwell, J.L., Jack, C.R., Jr., Parisi, J.E., Knopman, D.S., Boeve, B.F., Petersen, R.C., Dickson, D.W., and Josephs, K.A. (2011). Imaging signatures of molecular pathology in behavioral variant frontotemporal dementia. J Mol Neurosci *45*, 372-378.

Wilhelmus, M.M., Boelens, W.C., Otte-Holler, I., Kamps, B., Kusters, B., Maat-Schieman, M.L., de Waal, R.M., and Verbeek, M.M. (2006). Small heat shock protein HspB8: its distribution in Alzheimer's disease brains and its inhibition of amyloid-beta protein aggregation and cerebrovascular amyloid-beta toxicity. Acta Neuropathol *111*, 139-149.

Wojciechowska, C., Wodniecki, J., Wojnicz, R., Romuk, E., Jachec, W., Tomasik, A., Skrzep-Poloczek, B., Spinczyk, B., and Nowalany-Kozielska, E. (2014a). Neopterin and beta-2 microglobulin relations to immunity and inflammatory status in nonischemic dilated cardiomyopathy patients. Mediators Inflamm *2014*, 585067.

Wojciechowska, M., Olejniczak, M., Galka-Marciniak, P., Jazurek, M., and Krzyzosiak, W.J. (2014b). RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. Nucleic Acids Res *42*, 11849-11864.

Woollacott, I.O., and Rohrer, J.D. (2016). The clinical spectrum of sporadic and familial forms of frontotemporal dementia. Journal of neurochemistry *138 Suppl 1*, 6-31.

Woolley, J.D., Khan, B.K., Murthy, N.K., Miller, B.L., and Rankin, K.P. (2011). The diagnostic challenge of psychiatric symptoms in neurodegenerative disease: rates of and risk factors for prior psychiatric diagnosis in patients with early neurodegenerative disease. J Clin Psychiatry *72*, 126-133.

Wright, A.V., Nunez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell *164*, 29-44.

Wyttenbach, A., Sauvageot, O., Carmichael, J., Diaz-Latoud, C., Arrigo, A.P., and Rubinsztein, D.C. (2002). Heat shock protein 27 prevents cellular polyglutamine toxicity and suppresses the increase of reactive oxygen species caused by huntingtin. Human molecular genetics *11*, 1137-1151.

Xi, Z., Zinman, L., Moreno, D., Schymick, J., Liang, Y., Sato, C., Zheng, Y., Ghani, M., Dib, S., Keith, J.*, et al.* (2013). Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. American journal of human genetics *92*, 981-989.

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. Ann Neurol *78*, 568-583.

Xiao, S., MacNair, L., McLean, J., McGoldrick, P., McKeever, P., Soleimani, S., Keith, J., Zinman, L., Rogaeva, E., and Robertson, J. (2016). C9orf72 isoforms in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration. Brain Res *1647*, 43-49.

Xu, W., and Xu, J. (2018). C9orf72 dipeptide repeats cause selective neurodegeneration and cell-autonomous excitotoxicity in Drosophila glutamatergic neurons. J Neurosci.

Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S.*, et al.* (2013). Expanded GGGGCC repeat RNA associated with

amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc Natl Acad Sci U S A *110*, 7778-7783.

Yamakawa, M., Ito, D., Honda, T., Kubo, K., Noda, M., Nakajima, K., and Suzuki, N. (2015). Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. Human molecular genetics *24*, 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*, 799-806.

Yang, D., Abdallah, A., Li, Z., Lu, Y., Almeida, S., and Gao, F.B. (2015). FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. Acta Neuropathol *130*, 525-535.

Yang, F., Moss, L.G., and Phillips, G.N., Jr. (1996). The molecular structure of green fluorescent protein. Nat Biotechnol *14*, 1246-1251.

Yang, H., and Hu, H.Y. (2016). Sequestration of cellular interacting partners by protein aggregates: implication in a loss-of-function pathology. FEBS J *283*, 3705-3717.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370-1379.

Yang, L., Yang, J.L., Byrne, S., Pan, J., and Church, G.M. (2014). CRISPR/Cas9-Directed Genome Editing of Cultured Cells. Curr Protoc Mol Biol *107*, 31 31 31-17.

Yang, M., Liang, C., Swaminathan, K., Herrlinger, S., Lai, F., Shiekhattar, R., and Chen, J.F. (2016). A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. Science advances *2*, e1601167.

Yang, S., Warraich, S.T., Nicholson, G.A., and Blair, I.P. (2010). Fused in sarcoma/translocated in liposarcoma: a multifunctional DNA/RNA binding protein. Int J Biochem Cell Biol *42*, 1408-1411.

Yasuda, K., Zhang, H., Loiselle, D., Haystead, T., Macara, I.G., and Mili, S. (2013). The RNA-binding protein Fus directs translation of localized mRNAs in APC-RNP granules. J Cell Biol *203*, 737-746.

Yerbury, J.J., Ooi, L., Dillin, A., Saunders, D.N., Hatters, D.M., Beart, P.M., Cashman, N.R., Wilson, M.R., and Ecroyd, H. (2016). Walking the tightrope: Proteostasis and neurodegenerative disease. Journal of neurochemistry.

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*, 2244-2256.

Yoshimi, K., Kunihiro, Y., Kaneko, T., Nagahora, H., Voigt, B., and Mashimo, T. (2016). ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. Nat Commun *7*, 10431.

Yoshizawa, T., Ali, R., Jiou, J., Fung, H.Y.J., Burke, K.A., Kim, S.J., Lin, Y., Peeples, W.B., Saltzberg, D., Soniat, M.*, et al.* (2018). Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. Cell *173*, 693-705 e622.

Yu, C.H., Teulade-Fichou, M.P., and Olsthoorn, R.C. (2014). Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. Nucleic Acids Res *42*, 1887-1892.

Zakaryan, R.P., and Gehring, H. (2006). Identification and characterization of the nuclear localization/retention signal in the EWS proto-oncoprotein. J Mol Biol *363*, 27-38.

Zarouchlioti, C., Parfitt, D.A., Li, W., Gittings, L.M., and Cheetham, M.E. (2018). DNAJ Proteins in neurodegeneration: essential and protective factors. Philos Trans R Soc Lond B Biol Sci *373*.

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. Front Genet *3*, 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*, 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. Nat Neurosci *19*, 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 Neuropathol *128*, 505-524.

Zhou, Y., Liu, S., Liu, G., Ozturk, A., and Hicks, G.G. (2013). ALS-associated FUS mutations result in compromised FUS alternative splicing and autoregulation. PLoS Genet *9*, e1003895.

Zinszner, H., Sok, J., Immanuel, D., Yin, Y., and Ron, D. (1997). TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling. J Cell Sci *110 (Pt 15)*, 1741-1750.

Zischewski, J., Fischer, R., and Bortesi, L. (2017). Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. Biotechnol Adv *35*, 95-104.

Zou, Z.Y., Li, X.G., Liu, M.S., and Cui, L.Y. (2013). Screening for C9orf72 repeat expansions in Chinese amyotrophic lateral sclerosis patients. Neurobiol Aging *34*, 1710 e1715-1716.

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. Proc Natl Acad Sci U S A *110*, E4968-4977.

Chapter 9.   Appendices

9.1 Appendix A

Table 9.1 FTLD patient post mortem case details.

Case Number	Brain Bank	Gender	AAO	AAD	Duration (years)	PM delay (hrs:mins:sec)	Clinical Diagnosis	Pathological Diagnosis	Mutations
1	QSBB	F	58	66.8	8.8	115:00:00	MND	FTLD-TDP A	GRN / C9orf72
2	QSBB	F	66	74	8	85:50:00	Pick's Disease	FTLD-TDP A	C9orf72
3	QSBB	M	54	60	6	32:20:00	Pick's Disease	FTLD-TDP A	C9orf72
4	QSBB	M	62	68	6	99:00:00	Pick's Disease	FTLD-TDP A	C9orf72
5	QSBB	M	43	45	2	25:53:00	Non-specific Dementia	FTLD-TDP A	C9orf72
6	QSBB	M	53	63	10	77:20:00	Pick's Disease	FTLD-TDP A	C9orf72
7	QSBB	F	64	66	2	94:05:00	FTD	FTLD-TDP B	C9orf72
8	QSBB	F	56	67	11	85:35:00	FTD	FTLD-TDP A	C9orf72
9	QSBB	F	57	62	5	63:05:00	PNFA FTD	FTLD-TDP A	C9orf72
10	QSBB	M	66	71	5	51:52:00	FTD/MND	FTDL-TDP A	C9orf72
11	QSBB	M	59	65	6	30:00:00	FTD	FTLD-TDP A	C9orf72
12	QSBB	F	58	66	8	107:05:00	bvFTD	FTLD-TDP A	C9orf72
13	QSBB	M	59	70	11	44:05:00	FTD	FTLD-TDP A	C9orf72
14	Man	M	49	58	9	Not provided	FTD	FTLD-TDP A	C9orf72
15	Man	M	60	62	2	Not provided	FTD/MND	FTLD-TDP B	C9orf72
16	Man	M	60	68	8	Not provided	FTD	FTLD-TDP A	C9orf72
17	Man	F	59	64	5	Not provided	FTD	FTLD-TDP A	C9orf72
18	Man	M	54	66	12	Not provided	FTD	FTLD-TDP A	C9orf72
19	Man	M	64	72	8	Not provided	FTD	FTLD-TDP A	C9orf72
20	Man	F	51	70	19	48:00:00	FTD	FTLD-TDP B	C9orf72
21	Man	F	63	65	2	50:00:00	MND/FTD	FTLD-TDP B	C9orf72

Case Number	Brain Bank	Gender	AAO	AAD	Duration (years)	PM delay (hrs:mins:sec)	Clinical Diagnosis	Pathological Diagnosis	Mutations
22	Man	F	70	73	3	50:00:00	MND/FTD	FTLD-TDP B	C9orf72
23	Man	M	63	65	2	94:30:00	FTD	FTLD-TDP A	C9orf72
24	Man	M	54	65	11	81:00:00	FTD	FTLD-TDPA	C9orf72
25	QSBB	M	47	53	6	33:40:00	MND	FTLD-TDP A	None known
26	QSBB	F	70	83	13	07:50:00	FTD	FTLD-TDP A	None known
27	QSBB	M	57	62	5	92:55:00	FTD	FTLD-TDP A	None known
28	QSBB	F	57	63	6	85:20:00	Familial Alzheimer's	FTLD-TDP A	None known
29	QSBB	M	49	55	6	29:20:00	Familial Alzheimer's	FTLD-TDP A	None known
30	QSBB	F	75	78	3	36:15:00	CBD	FTLD-TDP A	None known
31	QSBB	F	83	87	4	68:55:00	PSP	FTLD-TDP A	None known
32	QSBB	F	62	68	6	99:45:00	Pick's Disease	FTLD-TDP A	None known
33	QSBB	M	53	61	8	72:35:00	bvFTD	FTLD-TDP A	None known
34	QSBB	M	62	72	10	97:25:00	bvFTD	FTLD-TDP A	None known
35	QSBB	F	72	79	7	49:55:00	Non-specific Dementia	FTLD-TDP A	None known
36	QSBB	M	67	69	2	70:10:00	MND	FTLD-TDP B	None known
37	QSBB	F	63	67	4	45:30:00	MND/FTD	FTLD-TDP B	None known
38	QSBB	F	63	83	20	45:05:00	Pick's Disease	FTLD-TDP B	None known
39	QSBB	F	58	73	15	37:55:00	Pick's Disease	FTLD-TDP C	None known
40	QSBB	F	59	73	14	83:40:00	Pick's Disease	FTLD-TDP C	None known
41	QSBB	M	64	78	14	26:50:00	Pick's Disease	FTLD-TDP C	None known
42	QSBB	M	64	74	10	19:00:00	FTD / Semantic Dementia	FTLD-TDP C	None known
43	QSBB	M	50	65	15	51:50:00	Semantic Dementia	FTLD-TDP C	None known
44	QSBB	M	61	66	5	70:45:00	FTD	FTLD-TDP C	None known
45	QSBB	M	44	67	23	76:00:00	Semantic Dementia	FTLD-TDP C	None known

Case Number	Brain Bank	Gender	AAO	AAD	Duration (years)	PM delay (hrs:mins:sec)	Clinical Diagnosis	Pathological Diagnosis	Mutations
46	QSBB	F	41	43	2	55:00:00	Unknown	NIFID FTLD-FUS	None known
47	QSBB	M	44	46	2	96:00:00	MND	NIFID FTLD-FUS	None known
48	QSBB	F	63	59	6	02:00:00	MND	NIFID FTLD-FUS	None known
49	QSBB	F	43	46	3	30:00:00	Unknown	NIFID FTLD-FUS	None known
50	QSBB	F	69	72	3	90:15:00	MND	NIFID FTLD-FUS	None known
52	QSBB	F	49	55	6	03:30:00	MND/FTD	Atypical FTLD-FUS	None known
53	QSBB	F	43	53	10	96:00:00	FTD	Atypical FTLD-FUS	None known
54	QSBB	F	55	58	3	Not available	bvFTD	Atypical FTLD-FUS	None known
55	IOP	M	40	51	11	12:00:00	bvFTD	Atypical FTLD-FUS	None known
56	IOP	M	44	51	7	24:00:00	bvFTD	Atypical FTLD-FUS	None known
57	IOP	M	47	53	6	05:00:00	bvFTD	Atypical FTLD-FUS	None known
58	IOP	M	51	60	9	48:00:00	bvFTD	Atypical FTLD-FUS	None known

PM = post mortem, AAO = age at disease onset, AAD = age at death, QSBB = Queen Square Brain Bank, Man = Manchester Brain Bank, IOP = King's College Institute of Psychiatry Brain Bank, CBD = corticobasal degeneration, FTD = frontotemporal dementia, FTLD = frontotemporal lobar degeneration, bvFTD = behavioural variant frontotemporal dementia, MND = motor neuron disease, PNFA FTD = progressive non-fluent aphasia frontotemporal dementia, PSP = progressive supranuclear palsy.

9.2 Appendix B

Table 9.2 MND patient post mortem details

Case Number	Brain Bank	Gender	AAO	AAD	Duration (years)	PM delay (hrs:mins:sec)	Clinical Diagnosis	Pathological Diagnosis	Mutations
59	QSBB	F	58	63	5	66:30:00	MND	MND	*C9orf72*
60	QSBB	M	/	76	/	26:25:00	MND	MND	*C9orf72*
61	IOP	F	35	39	4	69:30:00	MND	MND	C9orf72
62	IOP	F	56	59	3	34:30:00	MND	MND	C9orf72
63	IOP	M	69	70	1	38:00:00	MND	MND	C9orf72
64	IOP	M	62	64	2	68:00:00	MND	MND	C9orf72
65	IOP	M	51	57	5.5	22:30:00	MND	MND	C9orf72
66	IOP	F	53	58	5	14:00:00	MND	MND	C9orf72
67	IOP	F	67	70	3	60:00:00	MND	MND	C9orf72
68	IOP	M	/	51	/	64:00:00	MND	MND	C9orf72
69	Man	M	60	65	5	Not provided	MND	MND	*C9orf72*
70	Man	M	/	45	/	36:00:00	MND	MND	*C9orf72*
71	Man	M	/	57	/	Not provided	MND	MND	*C9orf72*
72	Man	M	/	58	/	35:30:00	MND	MND	*C9orf72*
73	Man	M	51	58	7	31:30:00	MND	MND	*C9orf72*
74	QSBB	F	68	70	2	25:30:00	MND	MND	None known
75	QSBB	M	56	58	2	49:10:00	MND	MND	None known
76	QSBB	F	68	69	1	48:35:00	MND	MND	None known
77	QSBB	M	52	54	2	39:40:00	MND	MND	None known
78	QSBB	M	73	76	3	31:45:00	MND	MND	None known
79	QSBB	F	74	76	2	22:55:00	MND	MND	None known

PM = post mortem, AAO = age at disease onset, AAD = age at death, QSBB = Queen Square Brain Bank, IOP = King's College Institute of Psychiatry Brain Bank, Man = Manchester Brain Bank MND = motor neuron disease, FTD = frontotemporal dementia, FTLD = frontotemporal lobar degeneration

9.3 Appendix C

Table 9.3 Control patient post mortem case details.

Case Number	Brain Bank	Gender	AAD	PM Delay (hrs:mins:sec)	Clinical Diagnosis	Cause of Death	Pathological Diagnosis
80	QSBB	M	69	171:00:00	Normal	Myocardial infarction	Pathological ageing
81	QSBB	M	81	50:55:00	Normal	Acute Myeloid Leukaemia	Pathological ageing with incidental Lewy bodies
82	QSBB	F	73	24:00:00	Normal	Pulmonary embolism	Pathological ageing
83	QSBB	F	80	49:10:10	Normal	Pancreatic carcinoma	Pathological ageing
84	QSBB	F	86	120:00:00	Normal	Not provided	Pathological ageing
85	QSBB	F	83	99:00:00	Normal	Bronchopneumonia / Renal Failure	Pathological ageing with incidental Lewy bodies
86	QSBB	M	84	79:10:00	Normal	Not provided	Pathological ageing
87	QSBB	F	88	32:25:00	Normal	Colonic carcinoma	Pathological ageing
88	QSBB	F	72	72:00:00	Normal	Not provided	Pathological ageing
89	QSBB	M	77	40:10:00	Normal	Not provided	Pathological ageing

PM = post mortem, AAD = age at death, QSBB = Queen Square Brain Bank