ALS/FTD associated C9orf72 expansions induce heat shock response activation and induce SOD1 proteinopathy

By Matthew Shaw

Student number: 140260830

Submitted for the degree of Doctor of Philosophy (PhD)

May 2019

Sheffield Institute for Translational Neuroscience

Department of Neuroscience

University of Sheffield

Acknowledgements

Abstract

A hexanucleotide repeat expansion (HRE) within the C9orf72 gene is the most prevalent cause of amyotrophic lateral sclerosis/fronto-temporal dementia (ALS/FTD). Current evidence suggests that HREs induce neurodegeneration through the accumulation of RNA foci and/or dipeptide repeat proteins (DPRs), though there is also evidence for haploinsufficiency of the C9orf72 protein.

Here, we modelled toxic gain of function disease mechanisms by generating and characterising two distinct transgenic zebrafish models. Whilst both model systems express HREs that lead to RNA foci production, only one model expresses ATG-driven DPRs. The two model systems were termed "RNA-only" and "RNA+DPR" zebrafish. As a readout of cellular stress, we utilised a fluorescent-reporter construct that expresses DsRed under the control of a heat shock promoter, which we then used to screen for potential therapeutic compounds. We also used immunoblotting techniques to characterise DPR, TDP-43 and SOD1 proteinopathies in human ALS post-mortem tissue.

Our RNA-only zebrafish model did not display any evidence of reduced survival at 26 months post-fertilisation. In contrast, the RNA+DPR zebrafish model displayed muscle atrophy, motor neuron loss and reduced survival at 36 months post-fertilisation. In line with increased toxicity, we identified that RNA+DPR zebrafish activate the heat shock response (HSR), which we also found to be true in C9orf72-ALS patient samples. HSR activation correlated with disease progression in our RNA+DPR zebrafish model and, through drug screening, we found that this could be attenuated using the compounds ivermectin, selamectin and riluzole. Finally, we report that SOD1 proteinopathy is detectable in disease relevant areas of *C9orf72* patients CNS, and that this correlates with TDP-43 proteinopathy.

We conclude that our RNA+DPR zebrafish model system shows motor neuron pathology, and the HSR readout in this zebrafish model may prove useful as a tool for evaluating potential neuroprotective compounds prior to mammalian testing. Finally, our neuropathological findings suggest a potential link between *C9orf72*-HREs and SOD1 as well as TDP-43 proteinopathies.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
ALS	Amyotrophic lateral sclerosis
APP	Amyloid precursor protein
ATM	ataxia telangiectasia-mutated
ATXN2	Ataxin-2
BAC	Bacterial artificial chromosome
C9orf72	Chromosome 9 open reading frame 72
Cas9	CRISPR associated protein 9
CHMP2B	Charged multivesicular body protein 2B
CNS	Central nervous system
CRISPR	Clustered regularly interspaced short palindromic repeats
DENN	differential in normal and neoplastic cells
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DPF	days post-fertilisation
DPR	Dipeptide repeat protein
DSB	Double strand break
DsRed	Discosoma red fluorescent protein
FTD	Frontotemporal dementia
FUS	Fused in sarcoma
GFP	Green fluorescent protein
H&E	Haematoxylin and eosin
HMW-SOD1	High molecular weight SOD1
HPF	Hours post-fertilisation
HRE	Hexanucleotide repeat expansion
HSP70	Heat shock protein 70
HSR	Heat shock response
LRRK2	Leucine-rich repeat kinase 2
MN	Motor neuron
MND	Motor neuron disease
mRNA	Messenger RNA
mSOD1	Mutant SOD1
MW	Molecular weight
NEK1	NIMA related kinase 1
NGS	Normal goat serum
NPC	Nuclear pore complex
NUP	Nucleoporin
PAR	Poly (ADP-ribose)
PARP1	Poly (ADP-ribose) polymerase 1
pATM	Phosphorylated ataxia telangiectasia mutated
PBS	Phosphate buffered saline
PINK1	PTEN-induced kinase 1
Poly(GA)	Poly glycine-alanine

Poly(GP)	Poly glycine-proline
Poly(GR)	Poly glycine-arginine
Poly(PA)	Poly proline-alanine
Poly(PR)	Poly proline-arginine
PSEN1	Presenilin 1
PSEN2	Presenilin 2
RAN-translation	Repeat-associated non-ATG translation
RNA	Ribonucleic acid
sALS	Sporadic ALS
SCAN-1	Spinocerebellar ataxia with axonal neuropathy-1
SMA	Spinal muscular atrophy
SOD1	Superoxide dismutase 1
SQSTM1	Sequestosome 1
TARDBP	TAR DNA binding protein
TBST	Tris buffered saline with tween 20
TDP-1	Tyrosyl-DNA phosphodiesterase 1
TDP-35	35KDa fragmented form of TDP-43
TDP-43	TAR DNA-binding protein 43
TIGAR	TP53-inducible glycolysis and apoptosis regulator
TILLING	targeting induced local lesions in genomes
wtSOD1	wildtype SOD1
XRCC1	X-ray repair cross-complementing protein 1
yH2AX	Phosphorylated histone H2AX

1.   Chapter 1: General introduction

1.1.   Amyotrophic lateral sclerosis onset and symptoms

Amyotrophic lateral sclerosis (ALS) or motor neuron disease (MND), is a progressive neurodegenerative disorder primarily causing loss of the motor neurons (MNs) supplying voluntary muscles. ALS results in progressive muscle weakness which eventually leads to paralysis and death, usually due to respiratory failure. A 2013 meta-analysis of ALS epidemiological studies showed that the median incidence of ALS in European populations is 2.08 per 100,000 persons per year, and that the median prevalence of ALS in the same population is 5.4 per 100,000 persons (Chio et al., 2013). From onset of symptoms, the median survival time for ALS patients is 30 months, however this figure comes with the caveat that there is considerable variability in ALS patient survival rates (with some patients surviving for decades after diagnosis) (Logroscino et al., 2008). Currently in the United Kingdom, riluzole is the only approved disease modifying treatment for ALS, which is thought to extend the lifespan of ALS patients by an average of 3-4 months (Bensimon et al., 1994). In both the United States and Japan, Edaravone (Radicava) has also been approved for the treatment of ALS patients (Hardiman and van den Berg, 2017). At the time of writing, data on average lifespan extension for ALS patients being treated with Edaravone is not available. However, Edaravone treated patients showed a significantly smaller decline on the revised ALS Functional Rating Scale (a measure of physical disability in ALS patients), when compared to placebo control patients (Cedarbaum et al., 1999, Abe et al., 2017).

There are two types of MNs: the upper MNs (primary MNs or Betz cells) are located in the primary motor cortex and do not exit the central nervous system (CNS), whilst the lower MNs exit the CNS to innervate muscles. Bulbar lower motor neurons supply the muscles of the face and throat and reside in the brainstem. Spinal lower motor neurons supply the voluntary muscles of the rest of the body and reside in the ventral horn of the spinal cord. In ALS it is common for the degeneration of one MN population to precede the degeneration of another (Gordon, 2013). ALS onset in lower MNs typically presents as fasciculation, cramps, muscle atrophy and marked muscle weakness (Gordon, 2013). Upper MN onset ALS typically presents as spasticity, hyperreflexia and modest weakness (Gordon, 2013). Additionally, ALS patients are commonly classified as having spinal or bulbar onset referring to whether muscle weakness is first observed in the limbs or the bulbar (speech and swallowing) muscles (Hardiman et al., 2017). Despite the discovery of several causative genetic mutations, there is still no clear link between causative mutation and upper vs lower MN onset, nor spinal vs bulbar onset, in ALS patients.

ALS remains a purely motor disorder in some patients, however approximately 40% of ALS patients exhibit some cognitive impairment throughout the disease course (Phukan et al.,

2012). Almost 60% of ALS patients with cognitive decline show impaired executive function, with the remaining 40% showing decline in other cognitive areas such as memory or language (Phukan et al., 2012). Indeed, ALS and frontotemporal dementia (FTD) share many common causative genetic mutations and pathomechanisms, and for this reason the two disorders are now often considered to be at opposite ends of a continuous disease spectrum (Neumann et al., 2006, Kabashi et al., 2008, Sreedharan et al., 2008, Borroni et al., 2009, Ferrari et al., 2011).

1.2.   The genetic aetiology of ALS

ALS cases are most commonly sporadic in origin (80-95% of all cases) and less frequently are familial in origin (5-20% of all cases), these numbers vary depending on the definitions of sporadic and familial used, and depending on geographic location (Byrne et al., 2011, Byrne et al., 2013). Sporadic cases are thought to be caused by complex interactions between genetic and environmental risk factors, whereas familial cases have a direct heritable cause which typically affects multiple members of the same family (Byrne et al., 2013).

The earliest identified genetic cause of ALS was mutations in the *superoxide dismutase 1* (*SOD1*) gene (Rosen et al., 1993). SOD1 is the first in a family of three superoxide enzymes which are all involved in catalysing the conversion of superoxide anions into hydrogen peroxide, thus detoxifying free radicals generated from metabolism (Zelko et al., 2002). Despite sharing many functional similarities with SOD1, mutations in *SOD2* or *SOD3* genes are not linked with ALS. On the other hand, over 170 mutations spanning almost every region of the *SOD1* gene have been reported to cause ALS (Rotunno and Bosco, 2013, Wei et al., 2017). Due to being discovered first, *SOD1* is currently the most well studied ALS gene. As DNA sequencing technologies improved, the discovery rate of novel ALS genes greatly increased. Mutations in two genes coding for RNA-binding proteins were both identified as causing ALS. The first gene in which mutations were identified was TAR DNA binding protein 43 (*TARDBP)* which codes for the RNA-binding protein TDP-43 (Sreedharan et al., 2008). The second gene in which mutations were identified was fused in sarcoma (*FUS*), which encodes an RNA-binding protein of the same name (Kwiatkowski et al., 2009, Vance et al., 2009). Both TDP-43 and FUS are predominately nuclear, and both have an array of functions which are mostly related to RNA processing (Kapeli et al., 2017). Later, tandem studies identified the most commonly known cause of ALS to date, a hexanucleotide expansion in the chromosome 9 open reading frame 72 (*C9orf72*) gene (DeJesus-Hernandez et al., 2011, Renton et al., 2011). Since its discovery, *C9orf72* and its' hexanucleotide expansion have become the fastest growing area of ALS research. *C9orf72* is the most common ALS gene by a wide margin and in the coming years will likely overtake *SOD1* as the most well studied ALS gene (Chió et al., 2012). Shortly thereafter, mutations in

the *SQSTM1* gene, which encodes the p62 protein, were identified as causing ALS (Fecto et al., 2011). p62 functions in the protein degradation pathway by binding to proteins and targeting them for degradation by the autophagosome, p62 positive inclusions are also often observed in ALS post-mortem tissues (Bjorkoy et al., 2006, Mizuno et al., 2006a). Later, ALS causative mutations were identified in the genes encoding RNA binding proteins Heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1), Heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNPA2B1), TATA-Box Binding Protein Associated Factor 15 (TAF15) and Ewing sarcoma breakpoint region 1 (EWSR1) (Ticozzi et al., 2011, Couthouis et al., 2012, Kim et al., 2013a). Thus, further implicating RNA processing in the pathobiology of ALS. Axonal transport defects have been reported in both mutant SOD1 and mutant TARDBP expressing mouse models (Magrane et al., 2014). Later, ALS causative mutations were identified in the axonal transport associated genes TUBA4A and KIF5A (Smith et al., 2014, Nicolas et al., 2018).

The above section gives an outline of the ALS genetic discoveries which are relevant to data or discussion presented in later chapters of this thesis. For a comprehensive list of ALS genes and their associated pathomechanisms see **Table 1.1** below.

Gene name	Encoded protein	Potential pathomechanism	References
ALS2	Alsin	Endosomal processing	(Hadano et al., 2001, Yang et al., 2001, Sato et al., 2018)
ANG	Angiogenin	RNA processing	(Greenway et al., 2006, Thiyagarajan et al., 2012)
ATXN2	Ataxin-2	RNA processing	(Elden et al., 2010, van den Heuvel et al., 2014)
C9orf72	C9orf72	Autophagy, DNA damage, nucleocytoplasmic transport, RNA processing	(DeJesus-Hernandez et al., 2011, Renton et al., 2011, Zhang et al., 2015b, Webster et al., 2016a, Walker et al., 2017)
CCNF	Cyclin F	Ubiquitin/proteasome system	(Williams et al., 2016)

CHCHD10	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10	Mitochondrial dysfunction	(Bannwarth et al., 2014)
CHMP2B	Charged multivesicular body protein 2B	Autophagy	(Parkinson et al., 2006, Cox et al., 2010)
ERBB4	Receptor tyrosine-protein kinase erbB-4	Neuronal development	(Takahashi et al., 2013)
EWSR1	Ewing sarcoma breakpoint region 1	RNA processing	(Couthouis et al., 2012)
FIG4	Polyphosphoinositide phosphatase	Endosomal processing	(Chow et al., 2009)
FUS	Fused in sarcoma	RNA processing, DNA damage, mitochondrial dysfunction	(Kwiatkowski et al., 2009, Vance et al., 2009, Deng et al., 2015, Shang and Huang, 2016)
GLT8D1	Glycosyltransferase 8 domain containing 1	Unknown	(Cooper-Knock et al., 2019)
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	RNA processing	(Kim et al., 2013a)
HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2/B1	RNA processing	(Kim et al., 2013a)
NEK1	NIMA related kinase 1	DNA damage	(Kenna et al., 2016, Higelin et al., 2018)
KIF5A	Kinesin family member 5A	Cytoskeleton	(Nicolas et al., 2018)
MATR3	Matrin 3	RNA processing	(Johnson et al., 2014)
OPTN	Optineurin	Autophagy, neuroinflammation	(Maruyama et al., 2010, Markovinovic et al., 2017)
PFN1	Profilin 1	Cytoskeleton	(Wu et al., 2012)
SETX	Senataxin	RNA processing, nucleocytoplasmic transport	(Chen et al., 2004, Bennett et al., 2018)
SIGMAR1	Sigma non-opioid intracellular receptor 1	Ubiquitin/proteasome system, mitochondrial dysfunction	(Luty et al., 2010, Fukunaga et al., 2015)
SOD1	Superoxide dismutase 1	Oxidative stress, mitochondrial dysfunction	(Rosen et al., 1993, Kraft et al., 2007)

SPG11	Spastic paraplegia-11	Lysosomal dysfunction	(Orlacchio et al., 2010, Branchu et al., 2017)
SQSTM1	p62	Autophagy	(Fecto et al., 2011)
TAF15	TATA-Box Binding Protein Associated Factor 15	RNA-processing	(Ticozzi et al., 2011)
TARDBP	TAR DNA-binding protein 43	RNA-processing, mitochondrial dysfunction	(Kabashi et al., 2008, Sreedharan et al., 2008)
TBK1	TANK-binding kinase 1	Autophagy, neuroinflammation	(Cirulli et al., 2015, Freischmidt et al., 2015, Oakes et al., 2017)
TUBA4A	Tubulin alpha-4A	Cytoskeleton	(Smith et al., 2014)
UBQLN2	Ubiquilin 2	Autophagy, Ubiquitin/proteasome system	(Deng et al., 2011, Chen et al., 2018, Kim et al., 2018)
VAPB	Vesicle-associated membrane protein-associated protein B/C	Endoplasmic reticulum dysfunction	(Nishimura et al., 2004, Fasana et al., 2010)
VCP	Valosin-containing protein	Autophagy, DNA damage	(Johnson et al., 2010, Meerang et al., 2011, Kim et al., 2013b)

Table 1.1: Genetic discoveries and associated pathomechanisms in ALS

1.3. Risk factors associated with ALS

Risk factors may be broadly categorised as genetic or non-genetic (environmental) in nature. Genetic risk factors for ALS can be thought of as being on a continuous spectrum with the identified disease causing mutations, the location of the gene on the spectrum simply depends on the penetrance of the identified mutations. Precise limits have not been well defined, but generally ALS gene mutations with a high penetrance are considered causative ALS genes, whereas those with low penetrance are considered risk factors.

Well known general risk factors for ALS include increasing age, being male and having a family history of ALS (Armon, 2003). Currently, there are many genes in which mutations have been identified as increasing the risk of developing ALS (low penetrance ALS genes), these include charged multi vesicular protein 2B (*CHMP2B*), *ATXN2, SPG11*, *TAF15*, *OPTN*, *VAPB*, *UBQLN2* and *NEK1* (Kenna et al., 2013, Kenna et al., 2016, Wang et al., 2017). As with causative ALS genes, the pathological molecular mechanisms of ALS genetic risk factors are not well understood. As well as genetic risks factors, diverse non-

genetic risk factors have also been associated with ALS. ALS is not one of the most common diseases caused by smoking, however a significant link between smoking and ALS has been reported in more than one study (Armon, 2003, Factor-Litvak et al., 2013). Other lifestyle factors may also influence ALS, consumption of some antioxidants including vitamin E have been reported to associate with a lower risk of ALS (Ascherio et al., 2005, Veldink et al., 2007). Conversely, treatment of ALS patients with riluzole supplemented with vitamin E failed to prolong survival in comparison to riluzole treatment alone (Graf et al., 2005). Low BMI and frequent strenuous exercise have been reported to associate with an increased risk of ALS (Granieri et al., 1988, Turner, 2013). Additionally, the incidence of ALS has been reported to be higher in populations of professional athletes and armed forces members (Beard and Kamel, 2015, Lacorte et al., 2016). Exposure to pesticides and heavy metals have also been reported to increase the risk of developing ALS (Kamel et al., 2003, Kamel et al., 2012).

1.4. Histopathology of ALS

Histopathological characterisation of CNS tissues from ALS patients has revealed that large motor neurons are lost from the spinal cord, brainstem and motor cortex (Hammer et al., 1979, Nihei et al., 1993). In the ventral horn of the spinal cord, neuronal loss is observed across multiple types of morphologically distinct neurons, and is not limited to motor neurons (Stephens et al., 2006). Loss of myelinated axons is also commonly observed in the ventral and lateral spinal cord of ALS patients (Saberi et al., 2015). In patients with a pure ALS phenotype, degeneration is predominately confined to motor areas, but in patients with ALS/FTD neuronal and axonal loss in the frontal and temporal lobes also occurs (Brown and Al-Chalabi, 2017). In both the brain and spinal cord of sALS patients, degenerating motor neurons are often surrounded by reactive astrocytes expressing glial fibrillary acidic protein (Kawamata et al., 1992, Schiffer et al., 1996). Additionally, activated microglia are present in degenerating regions CNS regions of sALS patients (Troost et al., 1990, Kawamata et al., 1992). In sALS patients, mitochondria located in the ventral horn of the spinal cord often exhibit morphological abnormalities, such as appearing swollen with increased cristae or appearing stubby with short protrusions from the outer mitochondrial membrane (Sasaki and Iwata, 2007).

CNS tissues from ALS patients frequently contain abnormal cellular inclusions. Approximately 85% of sALS cases show small round or oval eosinophillic inclusions termed Bunina bodies (Bunina, 1962). Bunina bodies are located in the cell bodies and dendrites of motor neurons of the spinal cord and brain stem nuclei (Piao et al., 2003, Okamoto et al., 2008). Bunina bodies are p62 and TDP-43 negative, but occasionally contain ubiquitin

(Lowe et al., 1988, Leigh et al., 1991, Mizuno et al., 2006a). Furthermore, Bunina bodies are positive for proteins cytostatin c and transferrin (Okamoto et al., 1993, Mizuno et al., 2006b). It is currently unknown how Bunina bodies develop, or whether they actively contribute to ALS progression. Ubiquitin positive inclusions are also observed in neurons (and less frequently in glial cells) of the majority of fALS and sALS cases (Leigh et al., 1988, Lowe et al., 1988). These ubiquitin positive inclusions are frequently also positive for p62 and are primarily comprised of fragmented/aggregated TDP-43 protein (Arai et al., 2006, Mizuno et al., 2006a, Neumann et al., 2006). p62 positive inclusions are also detected in the ventral horn of the spinal cord in the majority of ALS cases, and are frequently (but not always) ubiquitin positive (Arai et al., 2003, Mizuno et al., 2006a).

Muscle tissue from ALS patients shows evidence of denervation, and attempted reinnervation by nerve sprouting of surviving motor neurons (Dadon-Nachum et al., 2011). Areas of type I or type II muscle fibre grouping due to reinnervation of muscle fibres is commonly observed in ALS patients (Soraru et al., 2008, Jensen et al., 2016). Muscle fibres in ALS patients are frequently atrophic, although compensatory muscle hypertrophy as an attempt at muscle regeneration may also be observed (Soraru et al., 2008, Jensen et al., 2016). Similarly to ALS patient CNS tissue, muscle tissue from ALS patients also exhibits evidence of mitochondrial abnormalities and infiltration of immune cells (Wiedemann et al., 1998, Al-Sarraj et al., 2014). Also in line with CNS histopathology, p62 positive intracellular inclusions are also frequently detected in muscle fibres of ALS patients (Al-Sarraj et al., 2014). Conversely, no aggregation, fragmentation or cytoplasmic mislocalisation of TDP-43 has been detected in muscle tissue from ALS patients, thus TDP-43 proteinopathy appears to be limited to CNS tissues of ALS patients (Soraru et al., 2010, Al-Sarraj et al., 2014).

1.5. Common proteinopathies linking ALS sub-types

Many causative ALS genes have now been identified. The protein products of these genes have diverse ranging functions which are not always well understood. An additional layer of complexity arises from the fact that mutations may not cause ALS by loss of endogenous function, but rather by an additional gained toxic function of the mutant protein product. The lack of understanding of both the endogenous functions and possible gained toxic functions of ALS genes has made it difficult to identify whether different ALS genes share common pathological species. Nevertheless, multiple previous studies have identified evidence for similar proteinopathies in different genetic variants of ALS which were previously thought to be unrelated, these studies are discussed in detail below.

The SOD1 protein is essential in regulating the oxidation levels of the cell (Rosen et al., 1993). For this reason the first suspected mechanism of mutant SOD1 (mSOD1) toxicity

was due to a loss of dismutase enzymatic function leading to increased intracellular reactive oxygen species (Rosen et al., 1993). However, evidence has since shown that many ALS causing variants of mSOD1 do not have reduced enzymatic activity (Borchelt et al., 1994, Ratovitski et al., 1999, Ramesh et al., 2010). Additionally, a *SOD1* knockout mouse model does not show a neurodegenerative phenotype (Reaume et al., 1996). This evidence has led to a gain of toxic function mechanism being favoured for *SOD1*-ALS. Many variants of mSOD1 are known to be less structurally stable than wildtype SOD1 (wtSOD1), and have been shown to form abnormal conformations which cause novel interactions between mSOD1 and proteins such as bcl2, among others (Pasinelli et al., 2004, Pedrini et al., 2010). Furthermore, the abnormal conformation, or misfolding, of mSOD1 makes the protein prone to oligomerisation, and formation of insoluble intracellular aggregates (Anzai et al., 2017). Although, recent evidence suggests that aggregation of SOD1 may be neuroprotective, and that soluble forms of misfolded mSOD1 are the more toxic species (Gill et al., 2019). Thus, gain of toxic function through novel protein-protein interactions or aggregation of mSOD1 are both proposed mechanisms of mSOD1 mediated MN toxicity. Other ALS genes cause a range of neurological phenotypes, for example *C9orf72*, *VCP*, *FUS* and *TARDBP* mutations result in a spectrum of ALS-fronto-temporal dementia (FTD) (Liscic, 2015). Conversely, in >170 reported *SOD1* mutations, a pure ALS phenotype predominates, this has led some researchers to hypothesise that SOD1 proteinopathy may be a crucial aspect of selective motor neuron toxicity and that SOD1 proteinopathy may also contribute to motor neuron pathology in non-*SOD1* forms of ALS (Rotunno and Bosco, 2013, Wei et al., 2017). To test the hypothesis that misfolded/aggregated SOD1 may underlie other forms of ALS, various antibodies which specifically recognise misfolded SOD1 have been generated. Multiple studies have used several different misfolded SOD1 specific antibodies to detect misfolded/aggregated SOD1 in spinal cord tissue from patients with sporadic ALS (sALS) (Bosco et al., 2010, Forsberg et al., 2010, Pokrishevsky et al., 2012, Grad et al., 2014, Maier et al., 2018). These data suggest that misfolded SOD1 may contribute to pathology in non-*SOD1* forms of ALS. Moreover, a study using patient derived oligodendrocytes demonstrated that co-culture of oligodendrocytes derived from sALS and *TARDBP* mutation ALS patients conferred toxicity to wildtype mouse motor neurons upon co-culture (Ferraiuolo et al., 2016). Interestingly, knockdown of the SOD1 protein in these oligodendrocytes prior to co-culture, ameliorated the toxicity usually conferred to the motor neurons. These data suggest that within this *in vitro* system, the SOD1 protein directly or indirectly contributes to the motor neuron toxicity conferred by oligodendrocytes derived from non-*SOD1* ALS patients.

In contrast however, there is also evidence that misfolded/aggregated wildtype SOD1 is not involved in non-*SOD1* ALS. Other studies using similar misfolded SOD1 specific antibodies to those discussed previously, have not been able to identify specific staining in the spinal cord from sALS patients (Kerman et al., 2010, Brotherton et al., 2012, Liu et al., 2012, Ayers et al., 2014). The antibodies, staining techniques and sample numbers used vary from one study to another, and this may explain some of the observed discrepancies in the datasets. To address this issue, one recent study used 7 different misfolded SOD1 specific antibodies and stained a large cohort of over 40 ALS patient samples in total (Da Cruz et al., 2017). This comprehensive study did not identify any increase in misfolded SOD1 staining in sALS patients in comparison to non-neurological controls, despite the specificity of every antibody being validated in mSOD1 patient tissue. Thus, whether SOD1 misregulation has a role in non-*SOD1* ALS remains a highly controversial topic.

The *TARDBP* gene codes for the ubiquitously expressed protein TDP-43 which has roles in mRNA splicing, RNA stability, gene transcription and many other functions (Lagier-Tourenne and Cleveland, 2009). Mutations in the *TARDBP* gene are known to cause ALS (Sreedharan et al., 2008). Additionally, nearly all sporadic and familial ALS cases (except *SOD1* and *FUS* ALS cases) show ubiquitinated cytoplasmic TDP-43 positive aggregates which co-localise with disease affected areas (Arai et al., 2006, Neumann et al., 2006, Mackenzie et al., 2007, King et al., 2015). TDP-43 is often mislocalised from the nucleus to the cytoplasm and aggregated, leading to the hypothesis that loss of nuclear function and aggregation mediated toxicity may be potential ALS causing mechanisms (Arai et al., 2006, Neumann et al., 2006). In addition to TDP-43 many other RNA-binding proteins have been implicated in ALS pathogenesis. Mutations in the genes which code for RNA-binding proteins FUS, hnRNPA1, hnRNPA2B1, TAF15 and EWSR1 are all known to cause ALS (Kwiatkowski et al., 2009, Vance et al., 2009, Ticozzi et al., 2011, Couthouis et al., 2012, Kim et al., 2013a). All of the above mentioned RNA-binding proteins contain low-complexity prion-like domains, which allow for protein-protein interactions and self-assembly. The majority of ALS causing mutations in RNA-binding protein genes are clustered in the regions coding for the low complexity prion-like domains of the protein (Kapeli et al., 2017). Similarly to TDP-43 proteinopathy, all of the above mentioned RNA binding proteins have been shown to aggregate or mislocalise to the cytoplasm in a diverse range of ALS cases and models, albeit this is observed much less frequently than TDP-43 proteinopathy (Dormann et al., 2010, Couthouis et al., 2011, Couthouis et al., 2012, Kato et al., 2012, Kim et al., 2013a). Notably, two studies from the same group have suggested that RNA-binding proteins (TDP-43 and FUS) which are mislocalised to the cytoplasm can induce misfolding of wtSOD1 (Pokrishevsky et al., 2012, Pokrishevsky et al., 2016). Thus, dysregulation of multiple RNA

binding proteins and the SOD1 protein have been reported across diverse genetic subtypes of ALS, although evidence for SOD1 proteinopathy in non-*SOD1* ALS remains a highly controversial topic at this point.

1.6. C9orf72 expansions and potential routes of toxicity

Tandem repeat disorders are caused by a microsatellite expansion located within the identified disease gene. A microsatellite expansion is a short sequence of DNA (usually 3 to 6 base pairs in length) which is repeated multiple times consecutively within a genome (Richard et al., 2008). Huntington's disease, fragile X tremor ataxia syndrome and myotonic dystrophy are all examples of neurodegenerative disorders caused by microsatellite expansions (Mirkin, 2007). Recently, an intronic hexanucleotide expansion within the *C9orf72* gene was identified as the most common known cause of ALS and frontotemporal dementia (FTD) (DeJesus-Hernandez et al., 2011, Renton et al., 2011). In the case of the *C9orf72* gene, the microsatellite expansion consists of repeats of the hexanucleotide sequence GGGGCC (G_4C_2), the number of these repeats can vary typically from 1-10 in healthy people and up to >1000 repeats in some ALS patients (DeJesus-Hernandez et al., 2011). The repeats are thought to become pathogenic from a length of approximately 30 (Renton et al., 2011). The microsatellite expansion is located in the first intron of the *C9orf72* gene, and due to alternate splicing there are 3 possible mRNA transcripts which may be produced from *C9orf72* (DeJesus-Hernandez et al., 2011, Renton et al., 2011). These mRNAs give rise to two distinct protein variants. Both C9orf72 proteins contain a DENN (differential in normal and neoplastic cells) domain which led to the early hypothesis that *C9orf72* may be a guanine exchange factor for Rab GTPases (Levine et al., 2013). Later, C9orf72 proteins were discovered to associate with Rab family proteins and regulate processes such as endocytosis and autophagy (Farg et al., 2014). Most recently, the C9orf72 protein was shown to act as an effector for Rab proteins, ultimately involved in the initiation of autophagy (Sellier et al., 2016, Webster et al., 2016a). In C9orf72 expansion bearing patients, expression of mRNA encoding endogenous C9orf72 gene products is reduced (DeJesus-Hernandez et al., 2011).

Microsatellite expansions are transcribed bi-directionally (He et al., 2008). Bidirectional transcription of the C9orf72 expansion results in the production of sense and antisense RNA transcripts very rich in bases G and C. These GC rich RNA transcripts have a high propensity to form secondary structures such as hairpins and G-quadruplexes, as well RNA-DNA hybrids termed r-loops (Fratta et al., 2012, Reddy et al., 2013, Zhou et al., 2018). In *C9orf72*-ALS patients, the GC rich RNA transcripts produced from C9orf72 expansions are known to form insoluble, predominately intranuclear inclusions termed RNA foci (DeJesus-

Hernandez et al., 2011) (**Figure 1.1a+b**). Both sense and antisense transcribed RNA foci can be detected in neurons of multiple CNS regions including motor cortex, spinal cord, frontal cortex, cerebellum and hippocampus (Mizielinska et al., 2013, Zu et al., 2013, DeJesus-Hernandez et al., 2017). Additionally, RNA foci frequently sequester RNA-binding proteins such as Pur-alpha, hnRNP-H and hnRNP-A1 (Lee et al., 2013, Sareen et al., 2013, Xu et al., 2013, Cooper-Knock et al., 2014).

RNA transcripts produced from microsatellite expansions also undergo repeat associated non-ATG (RAN) translation (Zu et al., 2011). RAN translation is the initiation of translation in the absence of the ordinarily necessary ATG start codon, and is triggered exclusively in microsatellite expansions. As RAN translation is not bound by the requirement of an ATG codon, it is not limited to translation of a single frame of an mRNA transcript as in normal physiological translation. Rather, RAN translation is capable of translating all frames of an mRNA transcript and can therefore produce several distinct proteins from a single transcript (Zu et al., 2011). As microsatellite expansions are transcribed bi-directionally this will give rise to a $(G_4C_2)_n$ mRNA strand and a $(C_4G_2)_n$ mRNA strand. From each of these mRNAs, 3 different frames can be translated into proteins. Thus, the C9orf72 hexanucleotide expansion will give rise to 2 mRNAs and 6 proteins. However, in actuality one frame of G_4C_2 and C_4G_2 RNA transcripts share a common base pair sequence, so only 5 distinct protein species are produced in total, these are poly glycine-alanine (GA), poly glycine-arginine (GR), poly glycine-proline (PA), poly proline-alanine (PA) and poly proline-arginine (PR), these are collectively termed dipeptide repeat proteins (Ash et al., 2013, Mori et al., 2013, Zu et al., 2013) (**Figure 1.1c+d**). Dipeptide repeat proteins form p62/ubiquitin positive, TDP-43 negative neuronal inclusions in multiple brain regions including the spinal cord, motor cortex, frontal cortex and temporal cortex, basal ganglia and cerebellum of ALS/FTD patients (Al-Sarraj et al., 2011, Bigio et al., 2013, Mackenzie et al., 2013, Mann et al., 2013, Mori et al., 2013, May et al., 2014, Baborie et al., 2015, Davidson et al., 2016).

In summary, the three major, non-mutually exclusive, proposed mechanisms of C9orf72 expansion mediated neurodegeneration are: 1) Toxicity derived from expansion mediated reduction of endogenous C9orf72 gene products, 2) Toxicity derived from RNA transcribed from the C9orf72 expansion and 3) Toxicity derived from the proteins translated from the C9orf72 expansion RNA.
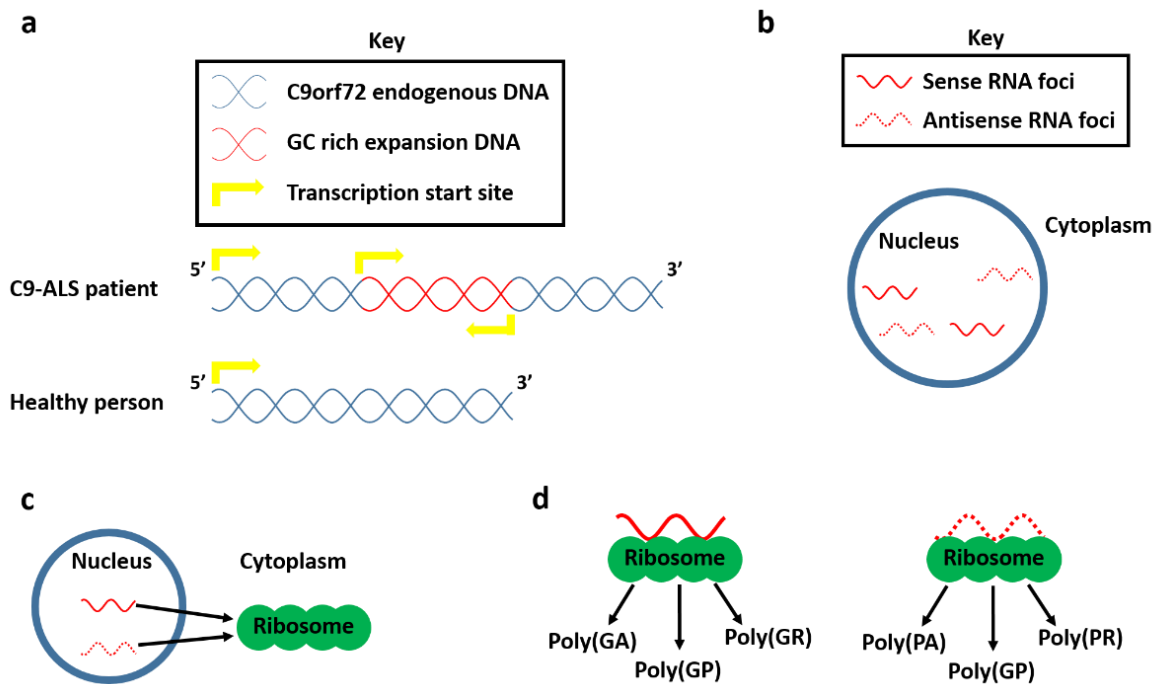
Figure 1.1: Schematic of transcription and translation of the C9orf72 expansion

(**a**) An expanded GC rich DNA region is located within the *C9orf72* gene of C9-ALS patients. In addition to the endogenous transcription start site present in the 5' promotor region of the *C9orf72* gene, the presence of C9orf72 GC rich expansions also allows the transcription machinery to bind and transcribe RNA in the local expansion region (the exact binding site of the transcription machinery in the local expansion region is not known). Additionally, expansion mediated initiation of transcription may occur in both sense and antisense directions. (**b**) Sense and antisense RNA transcripts produced from the GC rich expansion may form insoluble sense and antisense foci respectively. These RNA foci are predominately nuclear, and less frequently also occur in the cytoplasm. (**c**) In addition to forming intranuclear RNA foci, some RNA transcripts transcribed from the GC rich expansion are exported to the cytoplasm and then trafficked to the ribosome. (**d**) Sense and antisense RNA transcripts transcribed from the GC rich expansion bind to the ribosome where peptides are produced in all 3 possible reading frames (the reason why canonical rules dictating reading frame are not respected in the case of these GC rich RNA transcripts is not currently known).

1.7.   Evidence for loss or gain of function toxicity in *C9orf72* models

In addition to post-mortem tissue from C9orf72 expansion ALS patients, various *in vitro* and *in vivo* models have now been developed in order to help study different aspects of the role of *C9orf72* in ALS pathogenesis. In order to target therapeutic interventions and future research appropriately, it is important to ascertain which element(s) of the C9orf72 expansion drive ALS pathogenesis. Therefore the majority of ALS models to date have focussed on modelling one or a combination of RNA foci, DPRs and C9orf72 endogenous protein reduction.

1.7.1. Evidence for loss of C9orf72 endogenous function causing toxicity

Loss of function toxicity refers to a mutation which disrupts the endogenous functioning of the affected gene or its gene products. Whether loss of function of the endogenous C9orf72 protein products is neurotoxic has not been clearly defined. In zebrafish, morpholino-mediated knockdown of the zebrafish orthologue of *C9orf72* caused a motor phenotype which could be rescued upon expression of human *C9orf72* mRNA (Ciura et al., 2013). Furthermore, injection of *C9orf72* RNA containing deletions in the DENN domain into zebrafish blastomeres resulted in locomotor defects and neuronal apoptosis (Yeh et al., 2018). Injection of a C9orf72 morpholino also phenocopied overexpression of the DENN deleted RNAs (Yeh et al., 2018). Additionally, knockout of the C. elegans orthologue of *C9orf72* resulted in motor neuron loss (Therrien et al., 2013). These data suggest that haploinsufficiency may play a role in *C9orf72*-ALS pathogenesis. Conversely, early reports from stable *C9orf72* knockout zebrafish suggest that complete loss of C9orf72 protein expression does not induce any motor or neurodegenerative phenotype (Stepto et al., 2014). Additionally, 6 independently generated *C9orf72* knockout mice, displayed no motor or behavioural abnormalities and no neurodegenerative phenotype (Koppers et al., 2015, Atanasio et al., 2016, Jiang et al., 2016, O'Rourke et al., 2016, Sudria-Lopez et al., 2016, Ji et al., 2017). Further work is clearly needed to explain these directly contradictory datasets.

1.7.2. Evidence for gain of function toxicity in C9orf72 expansion models

Gain of function toxicity occurs when a genetic mutation results in the affected gene, or gene products, possessing new molecular functions. Toxicity related to RNA foci or DPR proteins are both examples of gain of function toxicity as RNA foci and DPR proteins are not expressed in people who possess a wildtype *C9orf72* gene. Modelling of gain of function mechanisms (RNA foci or DPRs) and loss of function mechanisms (C9orf72 protein levels) separately can be achieved with relative ease due to the independent nature of these pathways. However, RNA foci and DPR pathology are intimately linked, therefore

expressing RNA without DPR or vice versa has proven to be a challenge in identifying the exact species which are responsible for driving ALS progression. To overcome the difficulty of expressing RNA foci without DPRs being produced (or vice versa) researchers have developed various techniques. One such technique involves introducing stop codons immediately prior to the expansion, which should therefore allow expression of RNA foci, but prevent translation of the RNA into DPR protein (Mizielinska et al., 2014). Another technique involves expressing constructs which take advantage of codon degeneracy to avoid expressing 100% GC rich RNA transcripts, for example rather than being encoded by GGG-GCC repeats (100% GC content), poly(GA) may be encoded by GGA-GCA repeats (66% GC content), thus allowing large DPR proteins to be encoded by RNA which has a more moderate GC content (Mizielinska et al., 2014, Wen et al., 2014). Such constructs are henceforth referred to as DPR encoding constructs.

Multiple cell lines expressing hexanucleotide expansions have now been generated, these cell lines show RNA foci and/or produce DPRs, which results in toxicity as measured by induction of apoptosis and direct reduction in cell viability. (Lee et al., 2013, Kwon et al., 2014, Stopford et al., 2017). Certainly these studies demonstrate the potent toxicity of RNA foci and/or DPR species and therefore support gain of function mediated toxicity in ALS. However, due to RAN translation these models will likely not only express RNA foci but also some DPR proteins, indeed this was proven to be the case in one cell line but was not investigated in another (Lee et al., 2013, Stopford et al., 2017).

Cell lines are often not representative of the complex biology of an entire organism, for this reason many *in vivo* models of *C9orf72*-ALS have also been generated (see **Appendix A** for a table summarising the *in vivo C9orf72* models discussed in the section below). These *in vivo* models have also been utilised in an attempt to unravel the contributions to pathogenicity of RNA foci and DPR toxicity.

Transgenic drosophila expressing C9orf72 expansion containing constructs have helped to shed light on the relative toxicities of RNA and DPR species. Drosophila expressing over 200 G_4C_2 repeats showed RNA foci pathology, but RAN-translation of DPRs was prevented by the insertion of a bidirectional multi-frame stop codons throughout the C9-expansion (Mizielinska et al., 2014). The drosophila expressing over 200 repeats with a stop codon did not show any DPR production or neurodegenerative phenotype. However, in the same study drosophila expressing 103 repeats with no stop codons, displayed RAN translated poly(GR) peptides accompanied by a neurodegenerative phenotype and reduced survival (Mizielinska et al., 2014). Additionally, treatment of the 103 repeat (no stop codon) drosophila with an agent which partially blocks protein synthesis ameliorated the decrease in

their survival (Mizielinska et al., 2014). Similarly, another drosophila model demonstrated that expression of 160 G_4C_2 repeats (no stop codon) was sufficient to produce RNA foci, but again did not cause a neurodegenerative phenotype (Tran et al., 2015). In the same study however, expression of 34 repeats in the context of a poly(A) tail allowed RNA to be exported to the cytoplasm and translated into DPR, and was potently neurotoxic. Finally, Tran et al. demonstrated that growing the drosophila containing 160 repeats at a higher temperature (in order to increase the rate of protein synthesis) increased the amount of DPRs produced, and resulted in a corresponding increase in neurodegenerative phenotype (Tran et al., 2015).

In addition to drosophila models, zebrafish models have also been utilised in attempting to unravel RNA and DPR toxicity. Early data generated in zebrafish transiently expressing G_4C_2 RNA showed activation of apoptotic pathways (Lee et al., 2013). Later, zebrafish stably expressing G_4C_2 expansions showed a modest toxic phenotype, whereas zebrafish expressing ATG-driven DPR proteins displayed a much more severe toxic phenotype (Ohki et al., 2017). Additionally, injection of sense (35, 70 and 90 G_4C_2 repeats) and antisense (35 and 70 C_4G_2 repeats) RNA at the 1-2 cell stage resulted in motor axonopathy in zebrafish (Swinnen et al., 2018). In the same study, injection of an equimolar concentration (with respect to sense and antisense RNAs) of codon optimised (use of codon degeneracy to produce RNA transcripts with the lowest possible propensity to form secondary structures) DPR encoding RNA, resulted in motor axonopathy when 50 repeats of poly(PR) or poly(GR) encoding RNA was injected, but not when 50 repeats of poly(PA), poly(GA) or poly(GP) encoding RNA was injected. Intriguingly, Swinnen and other authors confirmed that sense or antisense RNA injection did not result in the production of detectable levels of the toxic poly(GR) DPR or poly(PR) DPR respectively (Swinnen et al., 2018). Furthermore, co-injection of Pur-alpha mRNA resulted in rescue of the motor axonopathy observed with injection of sense and antisense repeats, but had no effect on poly(GR) or poly(PR) DPR mediated motor axonopathy. Thus, Swinnen and other authors demonstrate that in this zebrafish model, C9orf72 expansion RNA and DPRs are both toxic to motor neurons, and may mediate their toxicity through independent mechanisms.

Mammalian models have also offered insights into the debate over RNA and DPR toxicity. Viral mediated expression of 66 G_4C_2 repeats in the mouse central nervous system accurately recapitulated major aspects of human ALS (Chew et al., 2015). The mice expressing 66 G_4C_2 repeats displayed RNA foci with the characteristic TDP-43 pathology frequently observed in ALS patients, the mice also showed intracellular inclusions of RAN translated proteins. These mice also displayed mild motor impairment and neuronal cell loss (Chew et al., 2015). Similarly, viral mediated expression of 102 G_4C_2 in the mouse CNS

resulted in expression of RNA foci, RAN-translated DPR species and caused gait and memory disturbances in these mice (Herranz-Martin et al., 2017). A further mouse model virally expressed a DPR encoding construct (50 poly(GA) repeats), which resulted in mice showing anxiety-like behaviour and reduced latency to fall in the rotarod test (Zhang et al., 2016). Likewise, a similar mouse model which virally expressed a DPR encoding construct (69 poly(GA) repeats) showed motor impairment and neurodegeneration (Walker et al., 2017). Thus, virally driven mouse models of C9orf72 expansion ALS show a motor phenotype whether G_4C_2 expansions or DPR coding constructs are used.

In addition to the mouse models described above, transgenic mouse models have also been generated from a bacterial artificial chromosome (BAC) containing a human ALS patient *C9orf72* gene complete with G_4C_2 expansion and flanking regions. Two BAC mouse models demonstrate the reduced survival, neuronal loss and motor deficits observed in human C9-ALS/FTD (Jiang et al., 2016, Liu et al., 2016). Liu and other authors, generated BAC mice expressing a maximum expansion size of 500 G_4C_2 repeats, within the full length human *C9orf72* gene and flanked by 52 Kb of upstream sequence and 19 Kb of downstream sequence (Liu et al., 2016). Whereas Jiang and other authors, generated BAC mice expressing a maximum of 450 G_4C_2 repeats, within exons 1-5 of the human *C9orf72* gene and flanked by 140 Kb of upstream sequence (Jiang et al., 2016). Unexpectedly, a further two independently generated BAC transgenic mouse models showed no signs of neuronal loss or reduced survival (O'Rourke et al., 2015, Peters et al., 2015). O'Rourke and other authors, generated BAC mice with a maximum expansion size of 1000 G_4C_2 repeats within the full length *C9orf72* gene (O'Rourke et al., 2015). Whereas Peters and colleagues, generated BAC mice with a maximum expansion size of 500 G_4C_2 repeats (these mice contained an additional 300 repeat expansion) within exons 1-6 of the human *C9orf72* gene, and flanked by 140.5 Kb of upstream sequence (Peters et al., 2015). All 4 BAC transgenic mice showed expression of RNA foci and RAN translated DPR species. While the two BAC mouse models which show an ALS/FTD like phenotype are consistent with the RNA foci and/or DPR mediated gain of function toxicity observed in cell, zebrafish, drosophila and previous viral expression mouse models. It is still not known why two seemingly similar BAC mouse models expressed RNA foci and DPR proteins but did not develop any ALS/FTD symptoms (see **Appendix A** for a table of information on all 4 BAC mouse, and other *in vivo* *C9orf72* models). Taken together, current *in vivo* data show that both RNA foci and DPR species are toxic, with DPRs likely being the more potently cytotoxic species. However, two of the 4 BAC mouse models generated, clearly demonstrate that expression of RNA foci and DPRs are not always sufficient to cause an ALS/FTD phenotype.

In summary, there is evidence suggesting that RNA foci, DPR species and haploinsufficiency of the C9orf72 protein products may all contribute to ALS pathogenesis. As these are none mutually exclusive pathways, it is possible that all three pathways may contribute to ALS pathology.

1.8. Differential toxicity of DPR species

Each DPR species is found to be differentially expressed in post-mortem CNS tissue from *C9orf72*-ALS patients (Mackenzie et al., 2013, Mackenzie et al., 2015, Schludi et al., 2015). For this reason it is important to determine what the relative toxicity of each DPR species is. Expression of DPR coding constructs in induced neuronal-like cells showed that poly(PA) was not cytotoxic but did cause electrophysiological abnormalities when expressed at over 1000 repeats (Bennion Callister et al., 2016). In the same study modest toxicity and electrophysiological changes were observed with similar repeat lengths of poly(GA). However, the most toxic DPRs were found to be the arginine rich poly(GR) and poly(PR) as these DPRs induced severe cytotoxicity at repeat lengths of ~1000 repeats, and caused electrophysiological changes from as few as 284 repeats (Bennion Callister et al., 2016). Similar results have also been obtained from drosophila expressing DPR encoding constructs. Expression of DPR encoding constructs containing 100 repeats coding for either poly(PA) or poly(GA) had no effect, whereas expression of 100 repeats coding for either poly(GR) or poly(PR) resulted in reduced viability of the drosophila and the induction of a neurodegenerative phenotype (Mizielinska et al., 2014). Additionally, independently generated drosophila expressing DPR encoding constructs also showed that poly(PR) DPRs induced a neurotoxic phenotype whereas poly(PA) and poly(GA) did not (Wen et al., 2014). Similarly, RNA injection of 50 repeats of codon optimised poly(PR) or poly(GR) resulted in motor axonopathy in zebrafish, however injection of equimolar concentrations of 50 repeats of poly(GA), poly(PA) or poly(GP) had no effect (Swinnen et al., 2018). Independently, generated zebrafish models injected with DPR encoding DNA constructs identified that 200 poly(GR) repeats more frequently induced death and swimming defects than constructs encoding 200 repeats of poly(GA, PA or PR) (Swaminathan et al., 2018). In the same study, transgenic zebrafish expressing 100 poly(GR) encoding repeats showed reduced swimming distance at 7 dpf and reduced axonal length at 2 dpf, in comparison to non-poly(GR) expressing zebrafish (Swaminathan et al., 2018). Importantly, induced motor neurons from *C9orf72*-ALS patients produce poly(PR) DPRs, and poly(PR) positive staining has also been identified in hippocampal neurons from *C9orf72* patients (Mann et al., 2013, Wen et al., 2014). However, poly(PR) positive inclusions have not been identified in motor neurons in post-mortem tissue from *C9orf72*-ALS patients. These data suggest that differential toxicity

exists between the various DPR species. Although, there is evidence for direct cytotoxicity or a contribution to cell pathology in all DPR species.

1.9. Pathways implicated in C9orf72 expansion pathology

While it is important to ascertain which species resulting from the C9orf72 expansion is most toxic to cell and animal models, it is also important to identify the pathways through which these species mediate their toxicity. Dysregulation of RNA processing, stress granule formation, nucleocytoplasmic transport, autophagy and DNA repair are among the most common pathways recurrently implicated in C9orf72 expansion pathology. Data from post-mortem *C9orf72* patient tissue and cell/animal models of *C9orf72* have all helped contribute to our understanding of how C9orf72 expansions affect these pathways.

1.9.1. RNA processing

Dysregulation of RNA processing in C9orf72 expansion pathology is thought to be caused by RNA foci mediated sequestration of RNA-binding proteins such as Pur-alpha, hnRNP-H and hnRNP-A1 (Lee et al., 2013, Sareen et al., 2013, Xu et al., 2013, Cooper-Knock et al., 2014) (**Figure 1.2a**). Indeed, cell lines derived from C9orf72 expansion harbouring patients were shown to have an increased rate of splicing errors in comparison to non-C9orf72 expansion cells (Cooper-Knock et al., 2015a) (**Figure 1.2b-c**). Overexpression of the RNA-binding protein Pur-alpha suppressed the toxic phenotype otherwise observed when expressing G_4C_2 repeat expansions in neuro-2a cells, drosophila or zebrafish (Xu et al., 2013, Swinnen et al., 2018). Splicing errors in *C9orf72*-ALS patients may not be solely due to RNA foci sequestration of RNA-binding proteins, as nuclear extracts incubated with arginine rich DPR proteins showed defective spliceosome assembly and therefore impairment of normal splicing (Yin et al., 2017). Moreover, expression of arginine-rich DPRs has been shown to impair splicing (Kwon et al., 2014). As mentioned previously, cytoplasmic mislocalisation of RNA-binding factors such as TDP-43 and FUS among others, is a common theme occurring in the vast majority of ALS cases (Lagier-Tourenne and Cleveland, 2009, Kapeli et al., 2017). Additionally, loss of nuclear TDP-43 has been shown to cause widespread splicing dysregulation in motor neurons (Highley et al., 2014). Thus suggesting that loss of TDP-43 may cause splicing dysregulation across multiple ALS sub-types. Whether RNA foci or DPRs are the main initiators of splicing dysregulation in *C9orf72*-ALS remains to be determined.
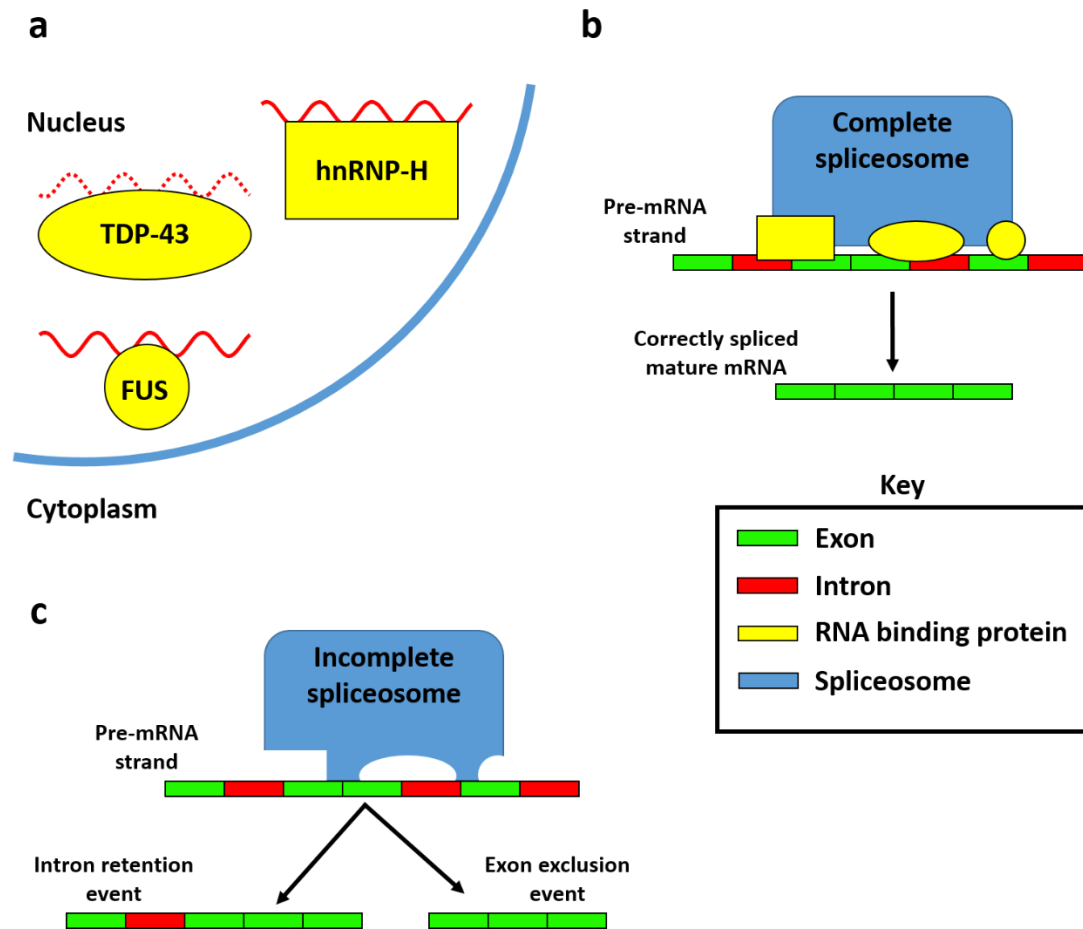
Figure 1.2: Schematic of sequestration of RNA binding proteins by RNA foci and subsequent increase in splicing errors

(**a**) In *C9orf72*-ALS patients, nuclear sense and antisense RNA foci bind and sequester RNA binding proteins such as TDP-43, FUS and hnRNP-H among many others. (**b**) During normal functioning, RNA binding proteins assemble with other splicing factors to form the spliceosome. The spliceosome binds to pre-mRNA strands and removes the introns before then ligating the exons back together, thus forming a final mature mRNA strand which is then trafficked to the ribosome to begin translation. (**c**) In *C9orf72*-ALS patients, RNA binding proteins are sequestered by sense and antisense RNA foci and therefore are not available during spliceosome assembly. An incomplete spliceosome is then formed in the absence of the RNA binding proteins, and splicing continues with the incomplete spliceosome, thus leading to an increase in the rate of splicing errors such as intron retention and exon exclusion events.

1.9.2. Formation and function of membraneless organelles

Many RNA binding proteins, including TDP-43, FUS and hnRNP-A1, contain a low complexity prion-like domain (Kapeli et al., 2017). Proteins containing a low complexity prion-like domain can undergo liquid-liquid phase separation to form membraneless organelles (Molliex et al., 2015). Liquid-liquid phase separation refers to the process by which a single liquid undergoes demixing and transitions into two or more distinct liquid structures (Hyman et al., 2014). Well known examples of liquid-liquid phase separation in biology include the formation of liquid-like nucleoli from the liquid which comprises the nucleoplasm, and the formation of liquid-like stress granules from the cytoplasm/nucleoplasm (Hyman et al., 2014). Stress granules are formed from a mixture of RNA-binding proteins, RNA and various protein/RNA modifying enzymes (Jain et al., 2016). The precise role of stress granules in the cell is not fully understood. However, stress granules are thought to contribute to the activation of stress response related proteins, sequestration of specific signalling molecules in order to modulate signalling pathways and the metabolism of RNA (Protter and Parker, 2016). Both poly(GR) and poly(PR) DPR proteins are known to interact with proteins containing low complexity prion like domains, and have been shown to interfere with the dynamics and function of stress granules both *in vitro* and *in vivo* (Lee et al., 2016, Lin et al., 2016, Boeynaems et al., 2017). Additionally, stress granule assembly has been associated with defects in nucleocytoplasmic transport, due to stress granule mediated sequestration of key nucleocytoplasmic transport factors such as the Ran protein (**Figure 1.3a**) (Zhang et al., 2018).

The nucleolus has functions including ribosome biogenesis and response to cell stress (Pederson, 2011). Both RNA foci and poly(GR)/poly(PR) proteins co-localise with nucleolin protein, a principle component of the nucleolus (Haeusler et al., 2014, Lee et al., 2016). In line with this, induced motor neurons from C9orf72-ALS patients, exhibit dispersed nucleolin immunostaining patterns, in comparison to the dense and well-defined immunostaining pattern observed in induced motor neurons from healthy controls (Haeusler et al., 2014). Furthermore, in C9orf72-ALS patient post-mortem tissue, neurons containing RNA foci or poly(GR) proteins exhibit enlarged nucleoli (Mizielinska et al., 2017). Nucleolar localised poly(GR)/poly(PR) disrupt the normal functioning of the nucleolar protein nucleophosmin 1, which then disrupts the organisation of the nucleolus and impedes the normal maturation of ribosomal RNA transcripts, (Kwon et al., 2014, Tao et al., 2015, White et al., 2019). Thus, C9orf72 expansions disrupt both the structure and functioning of the nucleolus.

1.9.3. Nucleocytoplasmic transport

Disruptions in nucleocytoplasmic transports pathways are another recurrent theme in *C9orf72* related ALS pathology. Nucleocytoplasmic transport is the shuttling of protein and

RNA cargoes between the cytoplasm and the nucleoplasm. Nucleocytoplasmic transport primarily occurs through large protein complexes known as nuclear pore complexes (NPCs). These complexes are formed from approximately 30 subunits which are individually known as nucleoporins (NUPs) (Cronshaw et al., 2002). Molecules smaller than ~30kDa can freely diffuse through the NPC, molecules larger than this require specialised transporter proteins which are also associated with, and traverse through, the NPC (Cautain et al., 2015). Induced-neuronal cells derived from patients with C9orf72 expansions showed retention of nuclear RNA, suggesting dysregulation of the normal routes of RNA export through the NPC (Freibaum et al., 2015). The same study also replicated this finding in drosophila models carrying C9orf72 expansions. Using the same C9orf72 expansion drosophila model, a genetic modifier screen identified 18 genes which enhanced or suppressed the neurodegenerative phenotype, all of these genes coded for components of the NPC (Freibaum et al., 2015). An additional and independently generated C9orf72 expansion drosophila model, also identified a key nucleocytoplasmic transport regulator as a potent suppressor of the neurodegenerative phenotype (Zhang et al., 2015b). Furthermore, depletion of the nuclear export adapter protein SRSF1 suppressed the motor and rough eye phenotype observed in C9orf72 expansion drosophila (Hautbergue et al., 2017). Also, depletion of SRSF1 in induced *C9orf72* patient motor neurons resulted in reduced DPR production by the cells and increased motor neuron survival (Hautbergue et al., 2017). SRSF1 co-localised with sense RNA foci in spinal cord post-mortem tissue from *C9orf72*-ALS patients, thus suggesting that SRSF1 mediates its effect on DPR production through processing of G_4C_2 RNA (Hautbergue et al., 2017). RanGAP1 a regulator of nucleocytoplasmic transport has also been shown to co-localise with sense RNA foci in induced neurons derived from *C9orf72*-ALS patients (Zhang et al., 2015b). Additionally, RanGAP1 and the nucleoporin Nup205 were shown to mislocalise into perinuclear puncta in induced neurons and brain tissue from *C9orf72*-ALS patients (Zhang et al., 2015b). However, DPR species may also interact with the nucleocytoplasmic transport system. Genetic modifier screens carried out in drosophila expressing DPR constructs encoding poly(PR) also identified genes coding for both components of the NPC and nucleocytoplasmic transport regulators as modifiers of the cytotoxic phenotype (Jovicic et al., 2015). Moreover, both poly(PR) and poly(GR) DPR proteins have been reported to bind with multiple nucleoporin proteins in two independent protein interaction assays (Lee et al., 2016, Lin et al., 2016). Collectively, these data show that nucleocytoplasmic transport pathways are important disease modifiers of C9orf72 expansion pathology, and may interact with both RNA foci and DPR proteins (**Figure 1.3b**). Deletion of genes relating to nucleocytoplasmic transport could either enhance or suppress the toxic phenotype, the reason why deleting seemingly similar genes can often times generate opposing effects is not currently known.
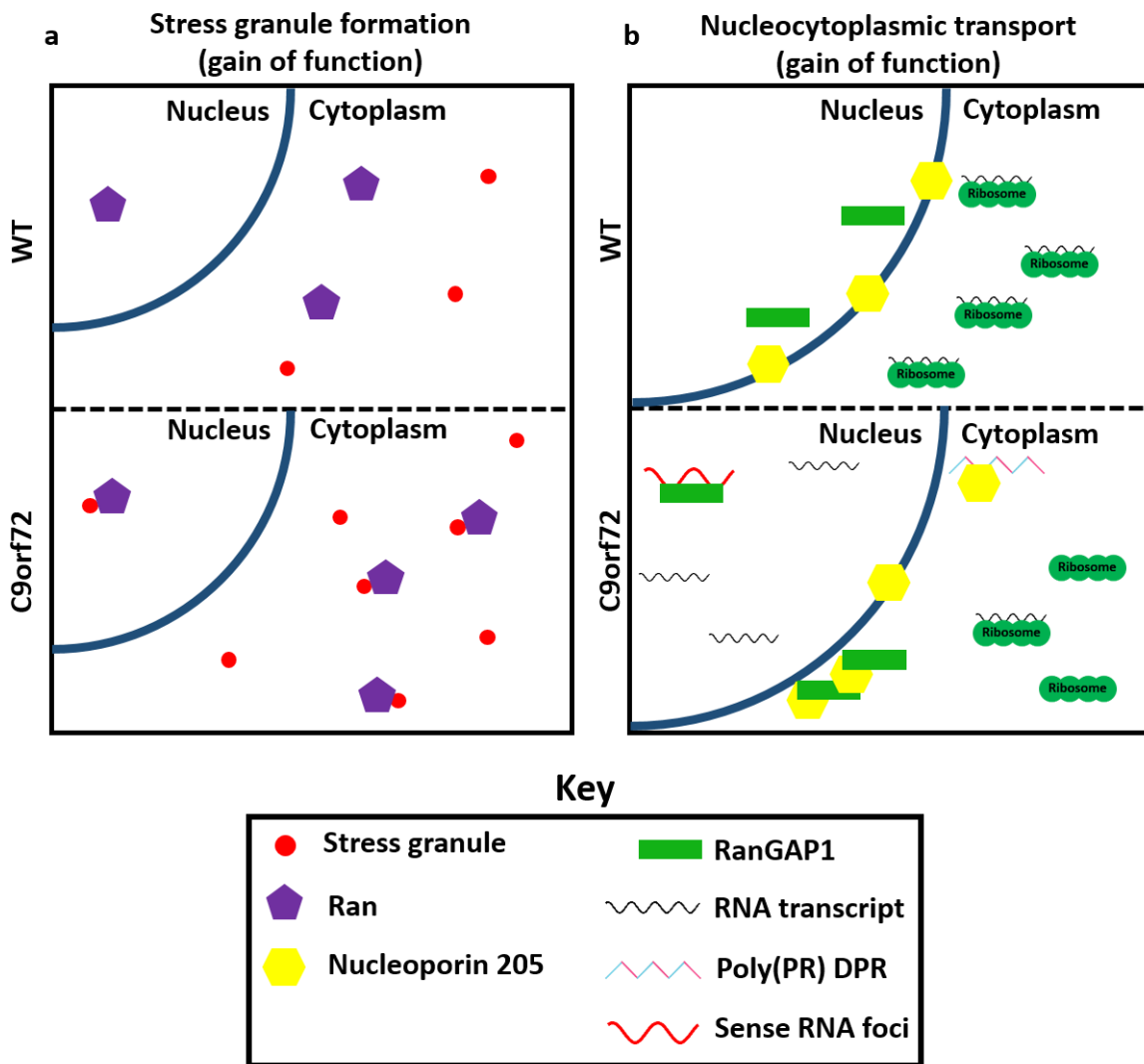
Figure 1.3: Schematic of the C9orf72 expansion mediated stress granule formation and nucleocytoplasmic transport defects

a) C9orf72 expansions drive increased stress granule formation. Increased stress granule formation results in increased sequestration of nucleocytoplasmic transport factors such as the Ran protein. **b**) Expression of C9orf72 expansions results in the sequestration of nucleocytoplasmic transport associated proteins nucleoporin 205 and RanGAP1 into perinuclear puncta. Poly(PR) DPRs translated from C9orf72 expansions bind with nucleoporin 205, and sense RNA foci sequester RanGAP1, it is not known whether the binding of these proteins by DPR/RNA foci is related to their formation of perinuclear puncta. Expression of C9orf72 transcripts also results in nuclear retention of RNA transcripts, suggesting defects in RNA export. Gain of function denotes that the pathway disruption is caused by the gained function of the C9orf72 expansion or its gene products.

A final notable point about nucleocytoplasmic transport in ALS, is that it has been hypothesised that disrupted nucleocytoplasmic transport may be related to the pathological primarily cytoplasmic mislocalisation of many ALS associated RNA-binding proteins (Kim and Taylor, 2017).

1.9.4. Autophagy

Data obtained from human cell lines suggests that an important function of the C9orf72 protein is in the initiation of autophagy. Knockdown of the C9orf72 protein results in reduced autophagosome formation (Sellier et al., 2016, Webster et al., 2016a). In a similar fashion, overexpression of the C9orf72 protein results in increased autophagosome formation (Webster et al., 2016a). Intriguingly, *C9orf72* knockout mice are reported to have defective autophagy functioning (Ji et al., 2017). Furthermore, induced neurons derived from *C9orf72* patients show defective autophagy and increased sensitivity to autophagy inhibition (Almeida et al., 2013, Webster et al., 2016a). Additionally, reduced autophagy mediated clearance of DPR proteins has been identified as contributing to neurodegeneration in induced motor neurons derived from *C9orf72* patients (**Figure 1.4a**) (Shi et al., 2018). Defective autophagy has also been closely linked with Parkinson's disease, and rare *C9orf72* expansion harbouring patients have been reported to show parkinsonian symptoms (Zhang et al., 2015a, Wilke et al., 2016). Finally, accumulation of p62 positive intracellular inclusions have been hypothesised to be caused by defective autophagy mediated clearance of p62 tagged aggregates, thus suggesting a role for defective autophagy signalling across multiple ALS sub-types (Arai et al., 2003, Mizuno et al., 2006a).

1.9.5. DNA damage and repair

Damage to the DNA of a cell can arise from a myriad of chemical and physical sources which can induce single strand breaks or double strands breaks (DSBs) depending on the nature of the interaction (Ciccia and Elledge, 2010). Because of this, cells have evolved complex DNA repair mechanisms which are capable repairing these DNA breaks through various mechanisms which faithfully restore the DNA to its original base pair sequence (Ciccia and Elledge, 2010). Chronic DNA damage or impairment of the machinery of the DNA repair response, can ultimately result in DNA damage induced cell death (Roos and Kaina, 2006). Induced neurons derived from *C9orf72* patients have increased expression of the DNA DSB marker yH2AX (phosphorylated histone 2AX) (Lopez-Gonzalez et al., 2016). Additionally, the same authors found that expression of poly(GR) repeats in control cell lines was also sufficient to induce DNA damage. The number of DSBs detected in these cells could be reduced by treatment with an antioxidant, suggesting that poly(GR) induced DSBs
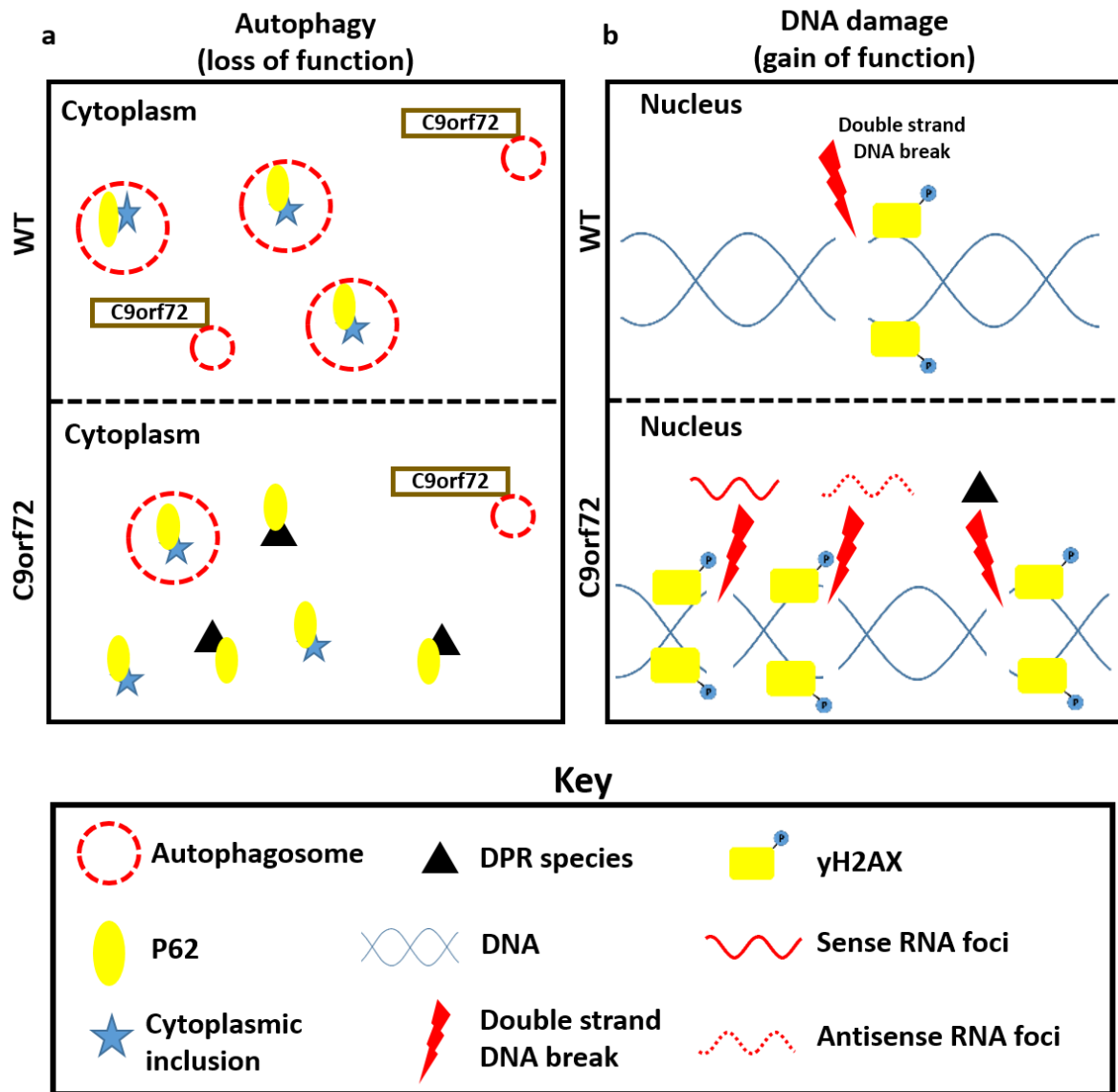
Figure 1.4: Schematic of the C9orf72 expansion mediated defects in autophagy and DNA damage

a) Haploinsufficiency of the C9orf72 protein caused by the presence of C9orf72 expansions results in decreased initiation of autophagy and therefore decreased clearance of p62 tagged cellular inclusions. Additionally, autophagy dysfunction may be exacerbated by an increase in the number of p62 positive inclusions requiring degradation due to cytoplasmic inclusions of DPR proteins produced from the C9orf72 expansion. **b**) Sense RNA foci, antisense RNA foci and DPR proteins (or a combination of these species) increase the frequency of double strand DNA breaks, resulting in increased expression of yH2AX (phosphorylated histone 2AX). Loss of function denotes that the pathway disruption is caused by the loss of function of C9orf72 endogenous gene products. Gain of function denotes that the pathway disruption is caused by the gained function of the C9orf72 expansion or its gene products.

may be mediated in part by reactive oxygen species. Furthermore, increased expression of key DNA repair proteins such as yH2AX, 53BP1 and phosphorylated ATM (pATM), have been reported in the spinal cords of *C9orf72*-ALS patients (Farg et al., 2017). A more recent report however, supported the finding that increased yH2AX levels are induced by C9orf72 expansions, but also reported that recruitment of 53BP1 and pATM to DNA repair foci was impaired by the C9orf72 expansion (Walker et al., 2017). The same study showed that even upon treatment with camptothecin (a potent inducer of DNA breaks), 53BP1 and pATM would not be recruited to repair foci in the presence of C9orf72 expansions. These findings suggest that, not only do *C9orf72* expansions induce DNA damage, but that they may simultaneously impair critical aspects of the DNA damage response. The precise reason why 53BP1 recruitment was increased in the presence of C9orf72 expansions in one study (Farg et al., 2017), and reported to be decreased in a later study (Walker et al., 2017), is not currently known. Interestingly, mutation of the ALS/FTD causative genes *VCP* and *FUS* have also been shown to increase markers of DNA damage whilst simultaneously impairing the rate of DNA repair (Meerang et al., 2011, Wang et al., 2013, Naumann et al., 2018, Wang et al., 2018). These data indicate that induction of DNA damage with simultaneous impairment of the DNA damage response may be involved in multiple ALS sub-types (**Figure 1.4b**).

The precise molecular mechanism by which DNA damage is detected and repaired is dependent on the nature of the insult (Ciccia and Elledge, 2010). Various forms of DNA damage can be detected by PARP1 and PARP2, these enzymes respond to DNA damage by attaching poly-ADP-ribose (PAR) polymers to the local DNA region (Dantzer et al., 2006). Subsequently, these PAR chains recruit proteins which are essential in the DNA repair process (Dantzer et al., 2000, Li and Yu, 2013). PARP1 catalyses the formation of PAR chains, a process which consumes a single molecule of NAD^+ for each PAR molecule added to the chain (Martire et al., 2015). Hundreds of PAR molecules can be joined together in a single chain, therefore it is possible for over-activation of PARP1 to deplete cellular stores of NAD^+ and induce cell death (Hoch et al., 2017). Neurodegeneration induced by PARP1 over-activation has been reported in patients suffering from cerebellar ataxia (Hoch et al., 2017). Additionally, mutations in the DNA repair factor tyrosyl phosphodiesterase 1 (*TDP1*) inhibit the ability of this protein to repair DNA single strand breaks and cause spinocerebellar ataxia with axonal neuropathy-1 (SCAN-1). Despite their known roles in other forms of neurodegeneration, whether over activation of PARP1 or loss of function of TDP1 may contribute to *C9orf72*-ALS progression has yet to be investigated.

1.9.6. Neuroinflammation

Activated astrocytes and microglia have long been implicated in the pathology of sALS (Troost et al., 1990, Kawamata et al., 1992, Schiffer et al., 1996). Recent evidence has also linked C9-ALS with neuroinflammation. Rather than resulting in a neurodegenerative phenotype, knockout of *C9orf72* in mice caused dysregulation of the immune system, including causing splenomegaly, lymphadenopathy, increased circulating levels of proinflammatory cytokines and abnormal immune cell functioning (Atanasio et al., 2016, Burberry et al., 2016, O'Rourke et al., 2016, Sudria-Lopez et al., 2016). RNA-seq analysis revealed an upregulation of inflammatory related pathways in *C9orf72* knockout mice in comparison to wildtype mice (O'Rourke et al., 2016). Similarly, RNA-seq analysis carried out in ALS patient CNS tissue, identified a significant upregulation of inflammatory pathways in C9-ALS patients when compared to sALS patients (O'Rourke et al., 2016). Furthermore, increased levels of activated microglia are observed in the white matter motor cortex of C9-ALS patient's brains when compared to the same regions in sALS patient's brains (Brettschneider et al., 2012). Thus, neuroinflammation may become dysregulated in C9orf72 expansion pathology.

1.9.7. Axonal transport and mitochondrial dysfunction

Mutations in the axonal transport related genes KIF5A and TUBA4A are causative of ALS (Smith et al., 2014, Nicolas et al., 2018). Additionally, dysfunction in the axonal transport of cargoes such as mitochondria, have been reported in transgenic mouse models of ALS expressing mutant forms of either SOD1 or TDP-43 proteins (Magrane et al., 2014). These data suggest a role of axonal transport in ALS pathobiology. In line with this, impaired mitochondrial transport has also been reported in C9orf72 drosophila models expressing either 39 poly(PR) repeats or 39 G_4C_2 repeats (Baldwin et al., 2016). However, drosophila expressing 39 RNA only G_4C_2 repeats did not exhibit significantly altered mitochondrial transport, suggesting that impaired axonal transport may be caused by DPR species in this model (Baldwin et al., 2016).

In addition to defects in mitochondrial transport, defects in mitochondrial morphology have been observed in sALS post-mortem tissue, and defects in both mitochondrial structure and function have been observed in mouse models of SOD1-ALS (Higgins et al., 2003, Pasinelli et al., 2004, Sasaki and Iwata, 2007). Increased production of reactive oxygen species by mitochondria is observed in induced motor neurons derived from C9orf72-ALS patients, when compared to induced neurons from healthy controls (Lopez-Gonzalez et al., 2016). This increase in reactive oxygen species production eventually lead to increased DNA damage in these motor neurons (Lopez-Gonzalez et al., 2016). Furthermore, in a C9orf72 mouse model expressing poly(GR) DPRs, morphological abnormalities in mitochondria were

observed at 3 months old (Choi et al., 2019). Using the same mouse model, poly(GR) was discovered to interact with the mitochondrial protein Atp5a1, this interaction resulted in the ubiquitination and proteasomal degradation of Atp5a1 (Choi et al., 2019). Moreover, Atp5a1 protein expression was significantly reduced in the frontal cortex of both poly(GR) expressing mice and C9orf72-ALS patients, and overexpression of Atp5a1 rescued mitochondrial morphological abnormalities in cortical neurons derived from poly(GR) expressing mice (Choi et al., 2019). Taken together these data highlight an emerging role for DPR mediated mitochondrial dysfunction in C9orf72 pathobiology. **Table 1.2** summarises pathways which have been recurrently implicated in C9orf72 pathobiology.

Dysregulated pathway	Reported cause of defect	Links to other pathways	References
RNA processing	RNA foci sequester RNA binding proteins resulting in increased splicing errors	Sequestered RNA binding proteins also function in stress granule formation	(Lee et al., 2013, Cooper-Knock et al., 2014)
Membraneless organelle formation/function	DPR proteins/RNA foci expression alters stress granule dynamics	Stress granules formation can cause nucleocytoplasmic transport defects by sequestering key proteins	(Lee et al., 2016, Boeynaems et al., 2017, Zhang et al., 2018)
Nucleocytoplasmic transport	DPR proteins/RNA foci cause mislocalisation of nucleocytoplasmic transport proteins. DPR/RNA foci induced toxicity can be enhance/suppressed by the deletion of various nucleocytoplasmic transport genes	Both RNA processing and DNA repair factors are frequently shuttled between the nucleus and the cytoplasm	(Freibaum et al., 2015, Zhang et al., 2015b)
Autophagy	Loss of *C9orf72* gene products results in defective autophagy initiation	The autophagy associated protein p62 co-localises with DPR aggregates	(Baborie et al., 2015, Davidson et al., 2016, Webster et al., 2016a)
DNA repair	DPR proteins/RNA foci cause DNA damage and impair the normal recruitment and functioning of DNA repair factors	Aggregation of the autophagy associated protein p62 has been reported to cause DNA repair defects	(Farg et al., 2017, Walker et al., 2017)
Neuroinflammation	Chronic activation of inflammatory pathways and microglia in the CNS	Haploinsufficiency of C9orf72 protein may contribute to immune dysregulation	(Brettschneider et al., 2012, Burberry et al., 2016, O'Rourke et al., 2016)

Mitochondrial dysfunction	Defects in mitochondrial transport, morphology and regulation of reactive oxygen species production	Defects in axonal transport pathways result in impaired mitochondrial transport	(Baldwin et al., 2016, Lopez-Gonzalez et al., 2016, Choi et al., 2019)

Table 1.2 Summary of pathways implicated in C9orf72 pathobiology

DPR: Dipeptide repeat protein

1.10. Generation of genetic animal models

There are many methods for generating transgenic or mutant animals, each method having advantages and disadvantages. Due to the low efficiency of generating most transgenic animals, often one method will be used in conjunction with another in an attempt to increase efficiency. Some of the most commonly used methods are DNA microinjection (with or without the use of transposons), transfection with lentiviral vectors and addition of transgenic pluripotent stem cells. DNA microinjection involves injecting highly concentrated DNA in nl volumes, preferably at the one cell stage immediately following fertilisation (Ramesh et al., 2010). Following microinjection the exogenous DNA incorporates into the animal's genome at an unknown location (Ramesh et al., 2010). An advantage of microinjection is that its low efficiency means that the transgene construct incorporating multiple times in different genetic loci is less likely. Likewise, the low efficiency of microinjection means that generating a transgenic animal by this method can be more time consuming.

Homologous recombination is an endogenous DNA repair process during which a complementary DNA strand is synthesised and then used to replace a damaged genomic DNA sequence (Szostak et al., 1983). By matching short sequences of transgenic DNA to a specific sequence present in the genomic DNA, the endogenous homologous recombination machinery can be exploited to accurately deliver exogenous DNA into a specific genetic locus (Thomason et al., 2007). Homologous recombination of exogenous DNA into a specific locus has an extremely low efficiency (Garate et al., 2013).

In order to increase efficiency of microinjections it is possible to add transposons to the DNA sequence being injected. Transposons are short DNA sequences which are replicated and repeatedly incorporated into the same genome (McClintock, 1950). By attaching the desired transgene to a transposon prior to microinjection, the efficiency by which the transgene incorporates into the host genome is increased (Largaespada, 2003). Unfortunately due to the high efficiency of this process, the transgene will often incorporate into multiple genomic loci of a single organism (Ivics et al., 2014). In the case of multiple transgene incorporations an additional outcrossing step is necessitated. The transgenic animals are bred with

wildtype animals and screened for those offspring who possess the transgene at only one locus, this can be a very time consuming process (Ivics et al., 2014).

Transfection with lentiviral vectors is another commonly used method of generating transgenic animals (Hofmann et al., 2003). Lentiviruses have the ability to incorporate their own genes into the host's genome (Goff and Berg, 1976). This feature of lentiviruses is exploited in molecular biology, where genes of interest can be inserted into the lentivirus and then incorporated into the genome of the prospective animal model. This method has previously been used to insert genes into early zebrafish embryos (Kawasaki et al., 2009).

Clustered regularly interspaced short palindromic repeats (CRISPR) technology, is a recently developed tool which is used in the generation of model animals (Jinek et al., 2012). A synthetic guide RNA molecule is used to direct the endonuclease CRISPR associated protein 9 (cas9) to a specific genetic locus which is then cleaved by cas9 (Jinek et al., 2012). CRISPR therefore allows highly specific removal of short lengths of DNA and can be used to efficiently generate knockout mutations of specific genes (Auer and Del Bene, 2014, Lebedeva et al., 2017). After cas9 cleavage, but before DNA ligation, additional genes can also be incorporated at the cleavage locus, as so CRISPR can be used to knockout or insert genes, allowing generation of both knockout and transgenic animal models with this technique (Kimura et al., 2014). A disadvantage of the CRISPR-cas9 system is that off-target cleavage events may occur, the frequency and loci of off-target cleavage events is unpredictable, thus making off target cleavages difficult to detect (Sander and Joung, 2014). CRISPR-cas9 has previously been used to modify the genome of zebrafish (Auer and Del Bene, 2014, Lebedeva et al., 2017).

A further commonly used method of generating transgenic animals is by the use of pluripotent stem cells. Genes of interest can first be inserted into the pluripotent cells using one of the aforementioned methods (e.g. homologous recombination). The pluripotent cells can then be inserted into a developing wildtype embryo at the blastocyst stage (Doetschman et al., 1987, Thompson et al., 1989). Pluripotent stem cells have a high capacity for self-renewal and are capable of contributing to a large proportion of the developing embryo including contributing to cells in all three germ layers, and incorporating successfully into the germline (Bradley et al., 1984, Beddington and Robertson, 1989). The chimeric animal is then bred, and if the pluripotent cells gave rise to a portion of the germ line this will result in a proportion of the offspring being full transgenic animals.

Zebrafish are frequently used as model animals of multiple diseases (Bradford et al., 2017). The predominant method of generating transgenic zebrafish is by microinjection, which is likely due to the fact that the zebrafish offers several specific advantages with this technique

but fewer advantages with others. The constantly improving efficiency and specificity of CRISPR-cas9 techniques indicates that in the near future a large proportion of zebrafish and other animal models may well be generated using CRISPR.

1.11. Zebrafish as model organisms

Zebrafish have been used as a research tool since the 1960s, and following their genetic characterisation have become a widely used in the modelling of human diseases (Streisinger et al., 1981, Howe et al., 2013, Bradford et al., 2017). The following section describes the development of zebrafish post-fertilisation and identifies the advantages and disadvantages of their use as model organisms.

Following fertilisation, zebrafish development begins rapidly and the yolk of the egg cell provides nourishment for the first five days (Fraher et al., 2016). However, the yolk is an extraembryonic structure and will not form any of the developing zebrafish tissues (Kimmel et al., 1995). All tissues of the zebrafish arise from the yolk free cytoplasm which begins to segregate immediately after fertilisation. At 24 hours post-fertilisation (hpf), early developmental versions of many organs are now visible and diverse cellular morphogenesis is occurring throughout the embryo (Kimmel et al., 1995). Also, at this time, spinal motor axons have grown into the myotomes and formed functional connections (Kimmel et al., 1988). As functional motor neuron connections have been formed, it is possible to commence electrophysiological studies in the zebrafish motor system from 24hpf onwards. Thus early motor unit formation makes zebrafish amenable to studying motor pathologies.

In comparison to zebrafish, functional neuromuscular junctions in rodents are not observed until approximately 12 days post-fertilisation (dpf) (Lin et al., 2000). In addition to longer developmental periods, study of mammalian embryos is further complicated by the nature of *in utero* development, as embryos must be extracted from the uterus of the parent prior to any investigations being possible. In the case of DNA micro injection when attempting to generate transgenic mice, the chimeric embryos must also be re-implanted into the uterus of the parent mouse. As the zebrafish embryo develops externally, the simplicity of carrying out microinjections, and therefore generating transgenic animals, is greatly increased. The minimum generation time in zebrafish is 2 months, this is similar to the generation times of both rats and mice, therefore transgenic zebrafish lines can be produced reasonably efficiently (Humphreys et al., 1976, Lambert, 2009, Lawrence et al., 2012). Another advantage of using zebrafish as model animals is that the zebrafish genome has been fully sequenced and approximately 70% of human genes have at least one zebrafish orthologue (Howe et al., 2013). Generation of transgenic animals often involves the addition of DNA (often a fluorescent reporter gene) which incorporates into the zebrafish genome. In

41

zebrafish most tissues are optically clear which greatly aids in florescence, and other types of detailed microscopy. A further benefit of using zebrafish as a model system stems from their relatively small size. During the first few days of life zebrafish are only a few millimetres in length and can therefore be conveniently placed into multi-well plates. This makes zebrafish very efficient when used in assays where a high throughput is required, such as during *in vivo* drug screening (McGown et al., 2013). Owing to their rapid development, short generation time, externally developing embryos, well characterised genetics, optically clear tissues, small size and comparatively low maintenance costs, zebrafish offer unique advantages when used as animal models of disease.

As expected, there are also disadvantages to the use of zebrafish in modelling human diseases. Evolutionarily, zebrafish are further removed from humans than other typical mammalian lab animals such as mice and rats. Therefore there is an expectation that a greater proportion of human disease genes may be lacking or have a modified function in the zebrafish genome. Indeed, during evolutionary history zebrafish underwent a gene duplication event which did not occur in mammals (Meyer and Schartl, 1999, Taylor et al., 2003). Duplicate genes in the zebrafish may possess redundant functions which could potentially complicate the use of zebrafish in research, in particular the use of zebrafish in gene knockout studies (Force et al., 1999). One example of the zebrafish genome duplication causing unexpected complications was in the study of the *TARDBP* gene. *TARDBP* is a human gene encoding the TDP-43 protein, mutations in *TARDBP* are linked with the eventual development of human motor neuron disease (Kabashi et al., 2008, Sreedharan et al., 2008). Upon knockout of the zebrafish orthologue of *TARDBP*, it was discovered that no motor phenotype developed. Later it was discovered this was due to a second gene in the zebrafish genome which coded for a truncated version of TDP-43, which upon knockout of *TARDBP* was able to produce a novel splice variant which functionally replaced TDP-43 (Hewamadduma et al., 2013). This is one example of how variations between zebrafish and human genomes can cause unforeseen problems when attempting to generate models of disease. Additional complications when using zebrafish as models of human disease may stem from fundamental anatomical and physiological differences between the two species. For example, zebrafish muscle is polyneuronally innervated, whereas in humans each myotome receives innervation from a single motor neuron (Westerfield et al., 1986). Additionally, zebrafish motor neurons exhibit a markedly higher capacity for regeneration in comparison with human motor neurons (Reimer et al., 2009). Clearly these two factors may cause unforeseen complications when using zebrafish to model human motor system disorders. It is therefore vital to ensure that data generated in zebrafish are confirmed in mammalian models before considering its relevance to humans.

1.12. Zebrafish as models of neurodegenerative disorders

Zebrafish are utilised in labs worldwide as model organisms of wide-ranging diseases. The exact reasons for using zebrafish as a model animal vary depending on the disease being studied. However, a common reason zebrafish are used is due to their remarkable ability to regenerate damaged tissues. Zebrafish have a very high regenerative capacity, and unlike humans, retain their ability to activate regenerative pathways in all tissue types through to adulthood (Reimer et al., 2009). To date, zebrafish have been used in the study of multiple diseases including heart disease, liver disease, vascular disease, cancer and many neurodegenerative disorders (Liu and Leach, 2011, Wilkins and Pack, 2013, Wilkinson et al., 2014, Wilkinson and van Eeden, 2014). Zebrafish are commonly used to model disorders of the central nervous system, among these are zebrafish models of Alzheimer's disease, Parkinson's disease and spinal muscular atrophy (Bretaud et al., 2007, Boon et al., 2009, Pu et al., 2017).

Alzheimer's disease results in progressive cognitive impairment and memory dysfunction, eventually leading to severe disturbances in speech, perception and normal social functioning (Forstl and Kurz, 1999). In its most advanced stages, Alzheimer's disease results in patients becoming bedridden and incapable of caring for themselves until eventually dying from an external cause, frequently due to infection (Forstl and Kurz, 1999). Alzheimer's disease is caused by neurodegeneration in regions of the parietal, temporal and frontal lobes, as well as in sub-cortical regions such as the locus coeruleus (Wenk, 2003, Braak and Del Tredici, 2012). Alzheimer's disease molecular pathology is characterised by the presence of hallmark abnormal proteins, these are intracellular neurofibrillary tangles composed of hyper-phosphorylated tau protein (tau proteins are encoded by the *MAPT* gene) and extracellular aggregates of misfolded β-amyloid protein (β-amyloid proteins result from proteolytic cleavage of the APP protein) (Goedert et al., 1988, De Strooper et al., 1998, Blennow et al., 2006). A small proportion of Alzheimer's cases (<1%) are known to have a monogenic aetiology due to mutations in *amyloid precursor protein* (*APP*), *presenilin 1* (*PSEN1*) or *presenilin 2* (*PSEN2*) genes, collectively these genes are responsible for the majority of familial Alzheimer's disease (Scheuner et al., 1996, Robinson et al., 2017). However, the vast majority of Alzheimer's cases are referred to as sporadic and are caused by complex interactions between disease genes and environmental factors (Robinson et al., 2017). In zebrafish, knockdown of the *PSEN1* or *PSEN2* genes causes abnormal brain development and deficits in neuronal functioning (Nornes et al., 2003, Nornes et al., 2009). Additionally, knockdown of *PSEN1* in zebrafish larvae results in cognitive defects, increased β-amyloid deposition and reduced expression of synaptic marker protein PSD-95 (Nery et

al., 2017). Transgenic zebrafish expressing a mutant human *APP* gene, displayed amyloid deposition and angiopathy in the brain, which eventually lead to neurodegeneration (Pu et al., 2017). Similarly, transgenic zebrafish expressing a pathogenic human *MAPT* gene (MAPT-P301L) displayed the hyper-phosphorylated and aggregated tau protein characteristic of human Alzheimer's disease (Paquet et al., 2009). In the same study, the mutant human tau expressing zebrafish model was used to screen thousands of potentially therapeutic compounds, one of which one was found to powerfully inhibit a tau kinase which contributes to hyper-phosphorylation (Paquet et al., 2009). Thus, zebrafish models of three major familial Alzheimer's disease genes (*APP*, *PSEN1* and *PSEN2*), recapitulate some aspects of human Alzheimer's disease pathology including the hallmark deposition of β-amyloid protein in the brain.

Parkinson's disease is primarily characterised by motor symptoms including bradykinesia, resting tremor, postural instability and muscle rigidity (Sveinbjornsdottir, 2016). However, Parkinson's disease can eventually result in non-motor symptoms such as cognitive decline, psychiatric symptoms, neuropathic pain and sleep disorders (Moreno et al., 2012, Kalia and Lang, 2015). The primary molecular hallmark of Parkinson's disease is the presence of intracellular inclusions of the α-synuclein protein, often termed Lewy bodies, or Lewy neurites, for cell body and neural process inclusions respectively (Spillantini et al., 1997, Goedert et al., 2013). Notably, Parkinson's disease patients possessing mutations in *LRKK2* or homozygous/compound heterozygous *PRKN* mutations frequently do not possess Lewy body pathology (Poulopoulos et al., 2012). This has led some to hypothesise that soluble α-synuclein oligomers, or other misfolded protein species may also be important in Parkinson's disease pathogenesis (Kalia and Lang, 2015). Parkinson's disease can be caused by autosomal dominant mutations in the genes *SNCA* (encodes α-synuclein) and *LRKK2*, or by autosomal recessive mutations in genes such as *PRKN*, *PINK1* and *DJ-1* (Kalia and Lang, 2015). Zebrafish models of Parkinson's disease most commonly involve silencing of one of the zebrafish orthologues of the above mentioned genes. The ventral diencephalon of the zebrafish brain contains dopaminergic neurons which are analogous to the dopaminergic neurons of the mammalian nigrostriatal pathway (Rink and Wullimann, 2002). *LRKK2* knockdown in zebrafish resulted in neurodegeneration which included dopaminergic neuron loss (Prabhudesai et al., 2016). Similarly, knockdown of the *PRKN* orthologue in embryonic zebrafish caused substantial and selective loss of dopaminergic neurons, and reduced activity of complex I of the mitochondrial respiratory chain (Flinn et al., 2009). Loss of dopaminergic neurons and reduced complex I activity, are both features typical of human Parkinson's disease (Schapira et al., 1990, Haas et al., 1995, Kalia and Lang, 2015). Two independently generated models in which the zebrafish orthologue of *PINK1* was knocked-

down, resulted in neurodevelopmental impairment (Anichtchik et al., 2008, Xi et al., 2010). Additionally, knockout of the *PINK1* gene in zebrafish caused dopaminergic neuron loss and deficiencies in mitochondrial complex I and III activities (Flinn et al., 2013). Furthermore, knockdown of *TigarB* (the zebrafish orthologue of human *TIGAR*) led to rescue of both mitochondrial dysfunction and dopaminergic neuron loss in *PINK1* null zebrafish (Flinn et al., 2013). Intriguingly, the TIGAR protein has previously been identified as a component of Lewy bodies and Lewy neurites in human Parkinson's disease patients (Lopez et al., 2019). Loss of PINK1 signalling has also been pharmacologically rescued in zebrafish models by utilising drugs which help restore mitochondrial calcium homeostasis or autophagy functioning (Soman et al., 2017, Zhang et al., 2017). Finally, knockdown of *DJ-1* in zebrafish did not result in dopaminergic neuron loss (Bretaud et al., 2007). However, when *DJ-1* knockdown zebrafish were exposed to a second stressor, such as oxidative stress or proteasomal inhibition, significant loss of dopaminergic neurons was observed (Bretaud et al., 2007). Taken together, these studies indicate that zebrafish models of multiple genetic sub-types of Parkinson's disease recapitulate the dopaminergic neuron loss associated with human disease, and have aided in identifying potential therapeutic targets and compounds.

The clinical presentation of spinal muscular atrophy (SMA) varies widely. Type I SMA typically results in paralysis and death at under 2 years old, whilst type II SMA results in the inability to walk during childhood and typically causes death after 2 years old, whereas types III and IV SMA result in milder motor symptoms and death during adulthood (Lunn and Wang, 2008, Ibrahim et al., 2012). Greater than 95% of SMA cases are caused by homozygous mutations of the *SMN1* gene, while the remaining cases are caused by mutations of 32 other causative SMA genes (Brzustowicz et al., 1990, Farrar and Kiernan, 2015). In the case of *SMN1* mutation related SMA, reduced expression of the SMN protein causes selective degeneration of spinal and cranial motor neurons (Lefebvre et al., 1997, Lunn and Wang, 2008). Human SMN protein functions as an RNA binding protein (Fallini et al., 2012). The zebrafish SMN protein shares 52% homology with the human SMN protein and also possesses RNA binding activity (Bertrandy et al., 1999). Three independently generated *SMN1* knockdown zebrafish models showed defects in axonal outgrowth and branching (McWhorter et al., 2003, Chitramuthu et al., 2010, Powis et al., 2016). Furthermore, the expression of the proteasome associated protein UBA1 was decreased in SMA-patient induced motor neurons, this reduced UBA1 expression was also mirrored in *SMN1* knockdown zebrafish, and co-injection of UBA1 mRNA rescued the abnormal axonal outgrowth and axonal branching phenotype of these zebrafish (Powis et al., 2016). Zebrafish that are homozygous for knockout of the *SMN1* gene display abnormal neuromuscular junction structure and die during the larval stage of development (Boon et al.,

2009). Using the same *SMN1* knockout zebrafish, it was identified that loss of SMN protein expression caused a severe reduction in the expression of the plastin 3 protein, and that partial rescue of plastin 3 expression was sufficient to rescue the neuromuscular junction defects of *SMN1* knockout zebrafish (Hao le et al., 2012). Humans possess an additional SMN protein encoding gene named *SMN2*, this gene is capable of producing small amounts of functional SMN protein (Mailman et al., 2002). Conversely, zebrafish do not possess an *SMN2* gene, and consequently, homozygous knockout of the *SMN1* gene in zebrafish ablates all SMN encoding genes from the genome, a genotype which is not representative of human SMA (Rochette et al., 2001). In an attempt to rectify this inconsistency between human and zebrafish biology, the human *SMN2* gene was expressed in the background of an *SMN1* null zebrafish (Hao le et al., 2011). Transgenic expression of human *SMN2* in *SMN1* null zebrafish delayed onset of neuromuscular junction defects and extended survival (Hao le et al., 2011). Hence, expression of the *SMN2* gene reduced disease severity caused by *SMN1* gene loss in zebrafish, as is also the case in human SMA (Mailman et al., 2002, Jedrzejowska et al., 2009). Thus, the SMN protein plays an important role in motor neuron development and survival in both humans and zebrafish.

In summary, these studies indicate that there is a disease relevant overlap between zebrafish and human neurobiology, which led to zebrafish models providing mechanistic insights, therapeutic targets and potential treatments for human diseases. The disease relevant overlap applied to neurodegenerative disorders affecting multiple brain regions, motor specific brain regions and motor neuron specific disorders, as evidenced by the successful generation of zebrafish models of Alzheimer's disease, Parkinson's disease and SMA respectively.

1.13. Zebrafish as models of amyotrophic lateral sclerosis

Multiple models of known ALS causative disease genes have been generated in zebrafish using a variety of techniques. ALS has been modelled in zebrafish by overexpression of ALS genes, by expression of physiological levels of mutant ALS genes, and by knockdown/knockout of ALS genes. Current zebrafish models of ALS are outlined below. There are multiple mutations in Cu-Zn superoxide dismutase 1 (*SOD1*) that have been identified in ALS patients (Rosen et al., 1993, Elshafey et al., 1994). Mutant *SOD1* RNA injected into zebrafish at the 2-4 cell stage results in dose dependent defects in motor neuron outgrowth (Lemmens et al., 2007). Motor axon outgrowth defects were consistently observed upon injection of RNA from G93A, G37A or A4V mutant *SOD1* variants (Lemmens et al., 2007). Intriguingly, co-expression of wildtype SOD1 and A4V mutant SOD1 caused more severe axonal outgrowth defects than expression of A4V mutant SOD1 alone

(Lemmens et al., 2007). A further two independently generated mutant *SOD1* RNA injection zebrafish models, also reported axonal outgrowth defects at 2 dpf (Sakowski et al., 2012, Robinson et al., 2018). In addition to transient mutant SOD1 zebrafish, germline mutant *SOD1* zebrafish have also been generated. Overexpression of the G93A SOD1 mutation variant in transgenic zebrafish caused neuromuscular junction defects by 30 weeks post-fertilisation and motor neuron loss at 40 weeks post-fertilisation (Sakowski et al., 2012). Similarly, transgenic overexpression of the human G93R mutant *SOD1* gene in zebrafish caused defects in the structure of neuromuscular junctions, reduced swimming endurance, evidence of motor neuron loss and reduced life expectancy (Ramesh et al., 2010). Using the same G93R model zebrafish, it was identified that a sub-populations of spinal interneurons begins to show stress response activation before any detectable motor neuron abnormalities were observed (McGown et al., 2013). The discovery that interneuron dysfunction precedes motor neuron dysfunction led the authors to the hypothesis that both mutant SOD1 protein pathology and loss of interneuron innervation may contribute to motor neuron injury (McGown et al., 2013). There are, however, potential drawbacks when overexpressing mutant SOD1 in zebrafish, firstly because the expression levels of mutant SOD1 may be significantly higher than would occur endogenously, and secondly because expression of both copies of wildtype zebrafish *SOD1* may mask any potential loss of function caused by *SOD1* mutations. In order to address these potential drawbacks, an additional SOD1 zebrafish model was generated using the targeting induced local lesions in genomes (TILLING) genome editing technique to generate zebrafish expressing a pathological T70I mutation in the endogenous *SOD1* zebrafish gene (Da Costa et al., 2014). Neither heterozygous nor homozygous T70I mutation of endogenous zebrafish *SOD1* affected overall SOD1 protein expression levels in comparison to wildtype zebrafish (Da Costa et al., 2014). Physiological expression levels of T70I mutant SOD1 protein caused abnormal neuromuscular junction formation, spinal motor neuron loss and motor impairment in homozygous zebrafish (Da Costa et al., 2014).

Mutations in the *TARDBP* gene (encoding TDP-43 protein) are a known cause of ALS (Kabashi et al., 2008). Overexpression of mutant *TARDBP* transcripts, or TARDBP morpholino mediated knockdown in zebrafish, both resulted in motor axonopathies and defects in touch evoked escape response at 2 dpf (Kabashi et al., 2010). Moreover, *TARDBP* knockdown induced motor axon and touched evoked escape response defects could be rescued by co-injection of wildtype but not mutant *TARDBP* transcripts (Kabashi et al., 2010). A further independently generated zebrafish model, identified that overexpression of wildtype or mutant *TARDBP* caused defects in motor axon outgrowth and increased axonal branching, axonal defects caused by mutant *TARDBP* overexpression could be

rescued by co-expression of human progranulin (Laird et al., 2010). In another independently generated zebrafish model, mutant *TARDBP* or mutant *FUS* expression induced motor axonopathies, which were again rescued by co-expression of human progranulin (Chitramuthu et al., 2017). Mutations in the GRN gene which encodes progranulin are a known cause of frontotemporal dementia (Baker et al., 2006, Cruts et al., 2006). Progranulin has known functions in cell survival and immune signalling among various others functions (Ryan et al., 2009, Zhu et al., 2013). Intriguingly, overexpression of human progranulin also rescued motor axon truncation and branching defects in SMN1 knockdown zebrafish models of spinal muscular atrophy (Chitramuthu et al., 2010). Zebrafish homozygous for a *TARDBP* null mutation were viable to adulthood and did not show significant motor axonopathies when examined at 36 hpf (Hewamadduma et al., 2013). Zebrafish possess an additional gene named *TARDBPL* which encodes a truncated variant of the TDP-43 protein, *TARDBP* knockout precipitated alternative splicing of the *TARDBPL* pre-mRNA which led to the production of a novel mRNA transcript and novel TDP-43-like protein, this novel protein was termed TARDBPL-FL (Hewamadduma et al., 2013). Morpholino mediated knockdown of TARDBP-FL expression in *TARDBP* null zebrafish, caused axonal outgrowth defects at 36 hpf and death in the early larval stage of development, suggesting that TARDBP-FL was at least partially replacing TDP-43 functionality (Hewamadduma et al., 2013). Additionally, an independently generated *TARDBP* null zebrafish model confirmed that *TARDBP* knockout did not result in significant motor axonopathy at 30 hpf, and this was also confirmed to be due to at least partial functional replacement of TDP-43 by TARDBP-FL generated through a novel splicing event (Schmid et al., 2013). Furthermore, knockout of both *TARDBP* and *TARDBPL* genes caused motor axon outgrowth defects at 28 hpf, muscle degeneration at 2 dpf and death in the early larval stage of development (Schmid et al., 2013).

Mutations in the RNA-binding protein FUS also cause ALS (Kwiatkowski et al., 2009, Vance et al., 2009). Both knockdown of the ALS associated gene *FUS*, and over expression of mutant *FUS* in zebrafish, resulted in deficits in the touch evoked escape response and impaired motor axon outgrowth (Kabashi et al., 2011). Importantly in *FUS* knockdown zebrafish, defects in escape response and motor axon outgrowth were rescued by co-injection of wildtype human *FUS* mRNA, but not with co-injection of mutant human *FUS* mRNA (Kabashi et al., 2011). Moreover, knockdown of *TARDBP* expression also resulted in defects in touch evoked escape response and motor axon outgrowth, and these deficits were rescued upon co-injection of wildtype *FUS* RNA (Kabashi et al., 2011). Conversely deficits caused by *FUS* knockdown could not be rescued by wildtype *TARDBP* expression, suggesting that FUS may act downstream of TDP-43 (Kabashi et al., 2011). Similarly,

expression of mutant *FUS*, or *FUS* knockdown in zebrafish, caused early swimming defects and defects in transmission across neuromuscular junctions (Armstrong and Drapeau, 2013). *FUS* knockdown phenotype in these zebrafish could be partially rescued by co-injection of wildtype *FUS*, but not by co-injection of mutant *FUS* (Armstrong and Drapeau, 2013). Despite *FUS* knockdown in zebrafish causing an early motor phenotype, stable CRISPR mediated ablation of *FUS* in zebrafish did not result in defects in motor axon formation or touch evoked escape responses at 2-3 dpf (Lebedeva et al., 2017).

ALS causative mutations have also been identified in the *SQSTM1* gene which encodes the autophagy associated p62 protein (Fecto et al., 2011). Morpholino-mediated knockdown of the zebrafish orthologue of *SQSTM1* resulted in defective axonal outgrowth and reduced total swimming distance when measured at 2 dpf (Lattante et al., 2015). Co-injection of wildtype *SQSTM1* RNA, but not mutant *SQSTM1* RNA, was sufficient to partially rescue both axonal outgrowth and swimming defects observed upon *SQSTM1* knockdown (Lattante et al., 2015).

To date, multiple gain of function and loss of function zebrafish models of *C9orf72*-ALS have been generated. Loss of function of endogenous zebrafish C9orf72 gene products was sufficient to induce a motor phenotype in two out of three reported models (Ciura et al., 2013). See section 1.7.1 for a full outline of evidence supporting loss of function toxicity in *C9orf72*-ALS, including the contributions of zebrafish models. Zebrafish models of gain of function toxicity have identified that C9orf72 expansion expression is sufficient to induce apoptosis and result in reduced survival of zebrafish (Lee et al., 2013, Ohki et al., 2017). Additionally, gain of function C9orf72 zebrafish models have shown that although DPR species are likely more toxic than RNA species, both RNA and DPR species may contribute to toxicity via independent mechanisms (Ohki et al., 2017, Swinnen et al., 2018). See sections 1.7.2 and 1.8 for a full outline of evidence supporting gain of function toxicity in *C9orf72*-ALS, including the contributions of zebrafish models. In summary, zebrafish models of multiple ALS disease genes have successfully recapitulated features of human ALS. Although, current C9orf72 zebrafish models have mainly focussed on elucidating the pathomechanisms or specific toxic species which underlie *C9orf72*-ALS, rather than utilising zebrafish as screening tools for potential therapeutic compounds.

1.14. Aims

Zebrafish models represent an excellent tool for studying neurodegenerative diseases. At the time of the initiation of this project, no stable genetic *C9orf72* zebrafish models had been created. We proposed to take full advantage of the unique characteristics of the zebrafish

which make it useful as a vertebrate drug screening tool. Additionally, investigation of common proteinopathies in ALS (outlined in section 1.5) has identified abnormal protein species which are observed across multiple ALS sub-types, therefore representing a potential therapeutic target which could benefit multiple ALS sub-types simultaneously. Taking the above points into consideration, the primary aims of the current study are detailed below:

1. To generate stable transgenic C9orf72 expansion model zebrafish lines, and to characterise their molecular and phenotypic hallmarks in order to assess whether they recapitulate features of human *C9orf72*-ALS.
2. To develop C9orf72 expansion model zebrafish lines into a validated drug screening paradigm.
3. To discover common pathological species between *C9orf72*-ALS and sALS by examining neuropathological tissue.

2. Chapter 2: Materials and methods

2.1. Materials

2.1.1. Antibodies

Antibody	Manufacturer/Product Code	Clonality	Application		
			Western blot	IHC	AR
Ms α GFP	Roche (11814460001)	Monoclonal	1:5000	1:500	no retrieval
Rb α Poly(PR)	Proteintech (23979-1-AP)	Polyclonal	1:500	1:200	Acid or alkali
Rat α Poly(PA)	EMD Millipore (MABN1790)	Monoclonal	1:1000	1:200	Acid or alkali
Rb α Poly(GP)	Proteintech (24494-1-AP)	Polyclonal	1:1000	1:200	Acid or alkali
Ms α Poly(GA)	Dieter Edbauer labs (clone 5F2)	Monoclonal	1:500	n/a	n/a
Rb α DsRed	Clontech (632496)	Polyclonal	1:1000	n/a	n/a
Ms α HSP70	Thermofisher (C92F3A-5)	Monoclonal	1:1000	n/a	n/a
Ms α Tubulin	Abcam (ab7291)	Monoclonal	1:10,000	n/a	n/a
Ms α Cleaved-PARP1	Cell signalling (9541)	Monoclonal	1:1000	n/a	n/a
Rb α TDP1	Abcam (ab4166)	Polyclonal	1:1000	n/a	n/a
Ms α GAPDH	Abcam (ab8245)	Monoclonal	1:10,000	n/a	n/a
Ms α SOD1	Abcam (ab52714)	Monoclonal	1:1000	n/a	n/a
Rb α C-terminal TDP-43	Proteintech (12892-1-AP)	Polyclonal	1:1000	n/a	n/a
Rb α H3K9me3	AbcAm (ab8898)	Polyclonal	1:1000	n/a	n/a
Ms α PAR	Santa Cruz (sc-56198)	Monoclonal	1:250	n/a	n/a

Table 2.1: List of antibody information

The antibodies used in the work throughout this thesis are listed in table 2.1.

IHC: immunohistochemistry, AR: antigen retrieval

2.1.2. RNA probes

To detect sense foci (in 5.3 zebrafish lines) RNA probes with sequence $(CCCCGG)_3$ were utilised. To detect antisense foci (2.2 zebrafish lines) RNA probes with sequence $(GGGGCC)_3$ were utilised. Both RNA probes have a 5' conjugation to the fluorophore TYE563 and are locked nucleic acids. Both probes were obtained custom made from Exiqon, and were used at a final concentration of 40nM.

2.1.3. Home office approval and zebrafish husbandry

Project license approval was obtained from the home office under the title "Aquatic models of human neurological disease" (PPL No. 70/8058). A zebrafish specific personal license was also obtained from the home office (PIL No. ICCB97EFC). Adult, larval and embryonic zebrafish were maintained at the University of Sheffield zebrafish facility at 28°C and bred according to established procedures (Westerfield, 2000). All animal experiments were conducted in accordance with the Animals (Scientific Procedures) Act 1986.

2.1.4. Human samples

Ethical approval for use of human frozen brain samples and human fixed brain samples was reviewed by the Sheffield Brain Tissue Bank Management Board, and approval to release tissue under REC 08/MRE00/103 was granted. See table 2.2 for a summary of patient information from frozen and fixed human samples used. In C9-ALS patients, the presence of >30 G_4C_2 expansions was confirmed by repeat primed PCR as part of the clinical protocol. However, the clinical protocol did not include estimation of expansion length by southern blotting, therefore expansion length data was not available for patients listed in table 2.2.

Diagnosis	Gender	Age	Use	Post mortem interval	Cause of death	Patient code
C9-ALS	M	48	IB	23	ALS	6
C9-ALS	F	69	PC/IB	55	ALS	7
C9-ALS	M	72	PC/IB	96	ALS	8
C9-ALS	F	62	PC/IB	63	ALS	9
C9-ALS	F	64	PC/IB	24	ALS	10
C9-ALS	F	67	PC/IB	24	ALS	17
C9-ALS	F	53	PC	48	ALS	n/a
C9-ALS	F	58	PC	2	ALS	n/a
C9-ALS	F	59	PC	28	ALS	n/a
C9-ALS	F	64	PC	7	ALS	n/a
C9-ALS	F	66	PC	10	ALS	n/a
C9-ALS	M	46	PC	4	ALS	n/a
C9-ALS	M	57	PC	-	ALS	n/a
C9-ALS	M	59	PC	72	ALS	n/a
C9-ALS	M	64	PC	48	ALS	n/a
C9-ALS	M	67	PC	38	ALS	n/a
C9-ALS	M	68	PC	31	ALS	n/a
Control	M	84	IB	72	wm damage	1
Control	F	63	PC/IB	22	-	2
Control	M	63	PC/IB	-	Ischaemic heart disease	3
Control	M	67	IB	63	Hepatocellular carcinoma	4
Control	M	58	IB	24	pontine haemorrhage	5
Control	F	59	PC/IB	5	Pneumonia	15
Control	M	47	IB	15	-	16
Control	M	46	IB	20	Ischaemic heart disease	20
Control	F	61	PC	-	-	n/a
Control	F	76	PC	-	-	n/a
Control	F	82	PC	-	-	n/a
Control	F	87	PC	14	myocardial infarct	n/a

Control	M	40	PC	-	-	n/a
Control	M	51	PC	-	-	n/a
Control	M	70	PC	-	-	n/a
Control	M	72	PC	31	-	n/a
sALS	F	64	IB	27	ALS	11
sALS	M	79	IB	48	ALS	12
sALS	F	69	IB	12	ALS	13
sALS	M	40	IB	96	ALS	14
sALS	M	60	IB	9	ALS	18
sALS	F	46	IB	20	ALS	19

Table 2.2: Human CNS tissue sample information

Patient codes are also listed next to the relevant sample in western blot and dot blot figures so that sample pathology may be linked with patient information. Patient codes are not listed for samples only used in Purkinje cell counts. F: female, M: male, PC: Purkinje cell counting, IB: immuno blotting, - indicates unknown information

2.2.   Methods

2.2.1. Generation and sequencing of GC rich DNA constructs (performed by Adrian Higginbottom and Tennore Ramesh respectively)

The GGGGCC hexanucleotide repeat gene was generated by first annealing two primers with sequence: forward, 5'-TCGAC(GGGGCC)$_{10}$-3' and reverse 5'-TCGA(CCCCGG)$_{10}$-3'. Due to the high propensity of GC rich DNA to form secondary structures, a long annealing protocol was required in which the annealing temperature is reduced from 95°C to 40°C, in a stepwise manner at 5°C/hour. The annealed primers were then ligated overnight at 16°C with standard T4 DNA ligase (New England Biolabs). The ligation mix was then run on a 1% agarose gel and multiple bands representing varying lengths of polymerised primers were cut out and purified. The maximum hexanucleotide repeat length obtained by this method was 99 repeats. The 99 hexanucleotide repeats were then ligated, at both 5' and 3' ends, to adaptors containing a BsrGI restriction endonuclease site. BsrGI sites were then endonuclease cleaved using the BsrGI restriction enzyme (New England Biolabs) and ligated into a construct containing the remaining transgene sequence. As BsrGI sites were present at both 5' and 3' DNA ends, this would allow the GC rich insert to ligate in either sense (G$_4$C$_2$) or antisense (C$_4$G$_2$) orientations. DNA constructs were then transformed in chemically competent E.coli cells (5-Alpha Competent E. coli, New England Biolabs) by incubating for 30 seconds at 32°C. Bacterial colonies were then grown overnight on carbenicillin treated agar plates at 37°C. A few colonies were picked for each DNA construct and grown overnight in 3ml of LB broth at 37°C. Plasmid DNA constructs were then purified from the E. coli using a Miniprep Kit (Qiagen). GC rich DNA expansions are known to be unstable in E. coli so both sense and antisense constructs were sequenced to confirm the presence of the hexanucleotide expansion. In the construct containing the sense

orientation, the full 99 G_4C_2 repeats were confirmed to be present by DNA sequencing (**Figure 3.1a;** see **Appendix B** for full transgene sequence). However, in the antisense orientation DNA sequencing was not able to read through the entire GC rich region, therefore only 89 C_4G_2 repeats were confirmed to be present (**Figure 4.1a;** see **Appendix C** for full transgene sequence).

2.2.2. Transgene construct

The expression of GFP, a hexanucleotide repeat expansion (HRE) and an auxin inducible degron (AID) gene was driven by a zebrafish ubiquitin promoter (Mosimann et al., 2011), a kind gift from Henry Roehl. In the case of the sense orientation HRE (5.3 zebrafish lines) a stop codon was formed after GFP and before the hexanucleotide repeats (**Figure 3.1a**). In the case of the antisense orientation HRE (2.2 zebrafish lines) a stop codon was formed after the HRE but before the AID gene (**Figure 4.1a**). The protein product of the AID gene becomes ubiquitinated and therefore targeted for proteasomal degradation selectively in the presence of auxin. We planned to exploit this system to selectively degrade the GFP-DPR fusion protein, however due to technical difficulties of working with an HRE, the *AID* gene could not be cloned in frame with the GFP and HRE in any of our constructs. Further 3' of the *AID* gene, is a *DsRed* gene driven by a heat shock protein 70 (*HSP70*) promoter, functioning as a fluorescent reporter of generalised cellular stress (Ramesh et al., 2010). Flanking both 5' and 3' of the transgene regions described above, are I-SceI meganuclease restriction sites.

2.2.3. Zebrafish strains used

All transgenic zebrafish used in the present study, and all non-transgenic zebrafish used as controls, were from the AB zebrafish strain. All subsequent generations of transgenic zebrafish were generated by breeding with zebrafish from the AB strain.

2.2.4. Generating transgenic zebrafish (performed by Tennore Ramesh)

DNA plasmid was digested with ISceI and the transgene containing fragment was purified and placed in solution with 1X concentration digestion buffer, 1 U/ml of ISceI, Phenol Red and 1X injection buffer (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 100 mM NaCl, 30 µM spermine and 70 µM spermidine) to a final concentration of DNA ranging from 80-100 ng/µl. DNA solution was then injected into early one-cell stage embryos in a 1nl volume (within 5 minutes of fertilisation). At 48 hpf the injected embryos were placed into 96 well PCR plates and heat shocked (30 minutes at 37°C followed by 30 minutes at 28°C, for three cycles) in a standard thermocycler. Heat shocking robustly increases *hsp70* promoter mediated

production of DsRed, thus allowing easy identification of transgenic zebrafish. Heat shocked embryos were examined for DsRed florescence at 72 hpf under a fluorescence microscope and highly chimeric embryos showing DsRed expression in a large proportion of their tissues were kept and grown to adulthood. Chimerism was determined by visually examining which tissue types expressed the transgene according to visible DsRed fluorescence (GFP florescence was also used if visible), zebrafish which showed transgene expression in at least muscle, heart and eye tissue were considered to be highly chimeric. The chimeric adult zebrafish (F0s) were outcrossed with AB zebrafish and their resulting offspring were screened for transgene expression using the same heat shocking-DsRed expression method described above. Any full transgenic offspring (F1s) identified during the screens were separated from their siblings and grown to adulthood. F1 zebrafish could then be outcrossed to AB zebrafish to give rise to an individual transgenic line.

2.2.5. In Situ Hybridisation (with assistance from Adrian Higginbottom)

At 10 dpf, zebrafish were terminally anaesthetised and then fixed in 4% paraformaldehyde in phosphate buffered saline (PBS) overnight at 4°C. Fixed zebrafish were then washed three times in PBS for 5 minutes per wash. Zebrafish were processed for paraffin embedding in a tissue processor and then embedded for longitudinal sectioning in paraffin. Paraffin embedded tissue blocks were sectioned at 10µm and dewaxed by 5 minute sequential immersions in each of the following solutions - xylene 100%, xylene 100%, ethanol 100%, ethanol 95%, ethanol 70% and dH_2O. Dewaxed sections were then incubated in 150µl of pre-hyb buffer (50% formamide, 300mM NaCl, 30mM trisodium citrate dihydrate, 10% dextran sulphate, 20mM monobasic sodium phosphate and 30mM dibasic sodium phosphate) for 1 hour at 66°C. A 5'TYE563 conjugated oligonucleotide-probe (for antisense foci: 5'-GGGGCCGGGGCCGGGG-3'; for sense foci: 5'-CCCCGGCCCCGGCCCC-3', Exiqon) was diluted in pre-hyb buffer (40nM probe) and the sections were incubated with 150µl of the probe-hyb buffer over night at 66°C. Sections were then washed for 10 minutes at room temperature with the first wash buffer (300mM NaCl, 30mM trisodium citrate dihydrate and 0.1% Tween20), then washed three times for 10 minutes each at 66°C in the second wash buffer (15mM NaCl and 1.5mM trisodium citrate dihydrate). Slides were then nuclei stained with Hoechst, mounted in Vectashield and imaged using confocal microscopy (TCS SP5II, Leica Microsystems) as described previously (Cooper-Knock et al., 2014).

2.2.6. Immunohistochemistry

Zebrafish were terminally anaesthetised in 4% tricaine solution and fixed in 4% paraformaldehyde at 4°C ON. Tissue was then embedded in paraffin blocks and sectioned

at 5μm using a microtome (RM2245, Leica Microsystems). Antigen retrieval was carried out by gradually heating the sections in a pressure cooker to 125°C while immersed in either Access Revelation buffer (pH 6.4, A.Menarini diagnostics) or Access Super RTU buffer (pH 9.5, A.Menarini diagnostics). Sections were then washed three times in PBDT (1% dimethyl sulfoxide, 0.5% triton X100, 1% bovine serum albumin in phosphate buffered saline) for 15 minutes per wash, at RT. Following this, normal goat serum (NGS) was added to PBDT (final concentration 0.1% NGS) and the tissue sections were blocked in this solution for one hour. Sections were then incubated in primary antibody at 4°C ON (see materials section for antibody manufacturers and dilutions). The following day, embryos were washed six times in PBDT for 15 minutes per wash and incubated in secondary antibody at 4°C ON. The next day, embryos were washed six times in PBDT for 15 minutes per wash, nuclei stained with Hoechst and mounted in Vectashield. Sections were then imaged by confocal microscopy (TCS SP5II, Leica Microsystems).

2.2.7. Western blots

Human cerebellum or motor cortex samples were ground to powder in liquid nitrogen prior to lysing. For adult zebrafish tissue, brain and spinal cord were dissected from deeply anaesthetised and decapitated zebrafish and flash frozen in liquid nitrogen. Laemmli buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.00004% bromophenol blue and 0.125M Tris-HCl pH6.8) was added to either human or zebrafish tissue at a ratio of 10μl:1mg of tissue. The samples were then sonicated at 25% power for three rounds of 10s each, with 30s cooling on ice between each round (Vibra-Cell, Sonics). Samples were then boiled at 95°C for 10 minutes, no centrifugation was performed.

To obtain protein lysates from embryonic (<5.2 dpf) zebrafish, embryos were first terminally anaesthetized in 4% tricaine solution, then all media was carefully removed and Laemmli buffer was added at a volume of 2μL per embryo. Samples were then lysed and stored as described above.

The equivalent of 1mg tissue (frozen tissue mass) of protein lysate was separated on a 10% tris SDS-polyacrylamide gel for one hour at 110V, protein was then transferred to a poly-vinylidene difluoride membrane at 110V for 1.5 hours. The membrane was then blocked for one hour using 5% milk in tris buffered saline with 0.1% Tween20 (TBST), before primary antibody diluted in blocking buffer was added overnight at 4°C (see Table 2.1 for antibody manufacturers and dilutions). Following this, the membranes were washed three times in TBST for 10 minutes per wash, before a secondary HRP conjugated antibody diluted in blocking buffer was added for 90 minutes. Membranes were now washed a further three times in TBST for 10 minutes per wash. Membranes were imaged by incubating with 2ml of

EZ-ECL chemiluminescent HRP enzyme substrate (biological industries) for one minute before being captured as a sub-saturating image using a G:box imaging system (Syngene). Membranes were re-probed using a different species primary antibody by first incubating in 0.01% sodium azide in blocking buffer for 1 hour, and then proceeding with addition of primary antibody as described above.

2.2.8. Embryonic spontaneous locomotor behaviour

At 5 dpf zebrafish were placed into individual wells of a 96well plate in 200μl aquarium water. The zebrafish were then habituated in in the dark for 10 minutes before a light stimulus was turned on for a duration of 10 minutes. Following this, the light stimulus was turned off and zebrafish activity was recorded for 10 minutes under dark conditions, this light –dark cycle was repeated once more for 2.2-7 zebrafish experiments, and twice more for 5.3 zebrafish experiments. Only activity under dark conditions was used in the analysis. All dark activity periods were combined together before calculating average speed and total distance moved values. Time spent swimming at different speeds was also calculated using the flowing speed thresholds: slow ($0<x<5$mm/sec), intermediate ($5<x<15$mm/sec) and fast ($x>15$mm/sec). Recordings were carried out using ZebraBox software (ViewPoint Behaviour Technologies).

2.2.9. Embryonic centre avoidance behaviour

At 5 dpf zebrafish were placed into a 6 well plate at a density of 30 zebrafish per well, in a volume of 3ml aquarium water. After a 30 minute habituation period with the lights on, the lights were turned off the moment recording began. Lights were off for 5 minutes then on for 5 minutes for 6 cycles. Beginning at 30 seconds after the lights were turned on, a frame was pulled out every minute using the Imagegrab tool, and this was repeated for each of the 6 lights on periods. Using ImageJ, circles of the same size were placed around the outside of every well so that only the centre of the well was visible, the % of zebrafish present in the centre of the well was then blind counted for every image and the average per well was calculated.

2.2.10. Developmental analysis of transgenic zebrafish

To obtain fertilised zebrafish eggs, adult zebrafish pair mates were placed into breeding tanks separated by a divider and left overnight. The following morning the dividers were removed and the eggs were collected 2-3 hours later, thus ensuring that all of the zebrafish included in the assay were of a similar developmental stage. In order to standardise water quality conditions a water change was performed in the morning immediately following

collection of eggs, another water change was performed in the afternoon and again the following morning. At 5 dpf zebrafish were placed into 2L of fresh water and a partial water change of roughly 350ml was carried out every day until 13 dpf at which point a constant flow of fresh water was applied to the tank. The zebrafish were genotyped and counted at 24hpf. Any zebrafish embryos which were malformed or had died before this point were not included in the assay. The zebrafish were again genotyped and counted at 5 dpf and 15 dpf.

2.2.11. Weighing zebrafish

Adult zebrafish were deeply anaesthetised in 4% tricaine solution, before being gently dried using a paper towel, zebrafish were then weighed in groups of 5 using a fine lab balance.

2.2.12. Swim tunnel and adult spontaneous locomotor behaviour

Zebrafish were not fed the day of the swim tunnel test and were allowed to acclimatise to the testing room for one hour. Zebrafish were individually placed into a 1" diameter swim tunnel chamber initially with no water flow. The water flow was then increased in 2L/min increments every 5 minutes (for 2-3 month old 5.3-9 zebrafish flow was increased at 1L/min every 10 minutes) until the maximum flow rate of 11.6L/min was achieved. For experiments using 2.2 zebrafish a wider 1.5" swimming chamber was used due to a lack of availability of the swimming chamber used in 5.3 zebrafish experiments. Zebrafish that stopped swimming during the test were allowed to briefly recover at 4L/min with the timer paused. Once the zebrafish began to swim again the flow was again increased to the rate at which the zebrafish previously failed, and the timer was started again once the flow rate had returned. Once a zebrafish either stopped swimming twice or successfully swam at 11.6L/min for 5 minutes, it was removed from the chamber and placed into a recovery tank. After 5 minutes in the recovery tank each zebrafish was placed into a behaviour monitoring tank and spontaneous swimming behaviour was monitored for 30 minutes using a camera linked to ZebraLab software (ViewPoint Behaviour Technologies). Average speed and total distance moved were then calculated for each individual zebrafish. Time spent swimming at different speeds was also calculated using the following speed thresholds: slow (x<60mm/sec), intermediate (60<x<120mm/sec) and fast (x>120mm/sec).

2.2.13. Motor neuron counts and myotome measurements

Zebrafish body segments just anterior to the pelvic fin were first fixed in 4% paraformaldehyde in PBS overnight at 4°C. Body segments were then washed three times in PBS for 5 minutes per wash. Zebrafish body segments were then decalcified by

incubating in 0.5M EDTA solution for 1 week. Body segments were then processed for paraffin embedding before being embedded for transverse sectioning in paraffin. Body segments were cut at 10µm thickness and dewaxed by 5 minute sequential immersions in each of the following solutions - xylene 100%, xylene 100%, ethanol 100%, ethanol 95%, ethanol 70% and dH$_2$O. Following this, sections were stained with haematoxylin and eosin (H&E) by immersion in haematoxylin (1 minute), tap water rinse, 1% acid alcohol (1 minute), tap water rinse, Scott's tap water (30 seconds), tap water rinse, eosin (5 minutes) and a final tap water rinse. Sections were then dehydrated by short immersions in ethanol 70%, ethanol 95% and ethanol 100% before being cleared in xylene and mounted in DPX. Cells with a soma size >75µm^2 and within 25,000µm^2 proximity of the central canal were designated as motor neurons. Three sections were counted per animal, by two independent blinded investigators. The two independent scores were averaged to give the final score for each animal.

The area of each individual body muscle myotome was measured by a blinded investigator from 6 images per animal. All muscle images were obtained from the epaxial muscle region just lateral from the dorsal spinal bone. Any myotome which was incomplete due to being partially out of frame was not included in the analysis.

2.2.14. Heat shock cell stress drug screening assay

At 2 dpf, transgenic zebrafish were removed from the chorion and placed into a 96 well plate in 200µl of drug or DMSO containing E3 zebrafish media. No water changes or additional dosing was performed. At 5 dpf zebrafish embryos were individually placed into the well of a V-bottom 96 well plate in 50µl of 4% tricaine solution. Each well was sonicated for 10s at 25% power, before the plate was centrifuged at 3000rpm for 10 minutes, 20µl of supernatant was then added to a 384 well plate. The 384 well plate was measured in a microplate reader (PHERAstar FSX, BMG Labtech), and both DsRed fluorescence (540nm excitation/590nm emission; 1399 gain) and GFP fluorescence (485nm excitation/520nm emission; 1865 gain) were quantified. This method was also used in 5 dpf zebrafish to obtain GFP and DsRed levels for comparison.

2.2.15. Embryonic drug injections

Immediately following fertilisation zebrafish embryos were placed into an injection mould in a small volume of E3 zebrafish media. Injections were carried out using a thinly pulled hollow glass pipette (make) attached to a Pneumatic PicoPump (World Precision Instruments). Approximately 1-2nl of drug or DMSO (diluted to appropriate concentration in E3 zebrafish medium with 1:1000 Texas Red), was injected directly into the embryonic yolk sac.

2.2.16. Dot blots

Human motor cortex or cerebellum tissue was ground to powder in liquid nitrogen, and lysis buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.00004% bromophenol blue and 0.125M Tris-HCl pH6.8) was added at a ratio of 10µl:1mg of tissue before the samples were briefly homogenised with an electric homogeniser. 20µl of each sample was vacuum drawn onto a PVDF membrane and immunoblotted for tubulin as a loading control.   A second dot blot was then prepared using the previous tubulin loading control to adjust the required volume of each sample to ensure equal loading. Each sample was diluted to a total volume of 400µl in PBS, before being divided into two 200µl aliquots immediately before being vacuum drawn onto a PVDF membrane. By splitting samples into two aliquots, duplicate membranes were generated to allow immunoblotting with two different sets of antibodies. Immunoblotting was carried out according to the same protocol descried for western blotting in section 2.2.7. The first duplicate membrane was immunoblotted with an anti-tubulin antibody and after treatment with sodium azide (to avoid signal cross over between antibodies of different species), was then immunoblotted for poly(PR). The second duplicate membrane was immunoblotted with an anti-poly(GA) antibody followed by sodium azide treatment and immunoblotting for anti-poly(GP). Sub-saturating images were captured using a G:box imaging system (Syngene). As the membranes used were duplicates (derived from the same sample preparation), the same tubulin image was used as a loading control for poly(PR), poly(GA) and poly(GP) images.

2.2.17. Larval drug dosing

At 1 dpf zebrafish were genotyped and placed into a 6 well plate containing either drug or DMSO in E3 zebrafish medium (30 zebrafish per well). Daily changes of 50% drug or DMSO containing E3 media were carried out. At 5 dpf, zebrafish were transferred to a 1.2L tank in 100ml of drug/DMSO media. An equal volume of drug/DMSO media was added to the tank daily. Once the water level in the tank had reached 800ml, the media was strained until 200ml was remaining and then equal volumes were added daily again. At 15 dpf zebrafish were removed from the tank and terminally anaesthetised in 4% tricaine solution, and then frozen at -80°C to later be used for western blotting. Any zebrafish which became sick during the course of the experiment were removed from the tank and terminally anaesthetised and the death was recorded.

2.2.18. Purkinje cell counts

Human cerebellum was formalin fixed and processed for paraffin embedding before being sectioned coronally at a thickness of 5μm. To ensure a similar anatomical region was used in each patient, only sections containing a portion of the dentate gyrus were used. Dentate sections were then dewaxed by 5 minute sequential immersions in each of the following solutions - xylene 100%, xylene 100%, ethanol 100%, ethanol 95%, ethanol 70% and dH$_2$O. Following this, sections were stained with haematoxylin and eosin (H&E) by immersion in haematoxylin (1 minute), tap water rinse, 1% acid alcohol (1 minute), tap water rinse, Scott's tap water (30 seconds), tap water rinse, eosin (5 minutes) and a final tap water rinse. Sections were then dehydrated by short immersions in ethanol 70%, ethanol 95% and ethanol 100% before being cleared in xylene and mounted in DPX. Whole mounted H&E stained slides were then automatically imaged at 40X zoom using a NanoZoomer Digital slide scanner (Hamamatsu). Images were visualised using the NDP view software (Hamamatsu), where the Purkinje cell layer was traced until either 200 Purkinje cells or 1 entire slide had been counted. The number of cells counted was then divided by the distance of Purkinje layer over which the cells were distributed, to give an average value of Purkinje cells per mm for each patient. Only Purkinje cells where the nucleolus was visible were counted, to ensure that cell would not influence the dataset.

2.2.19. Statistical analysis and definition of n number used

For adult and embryonic zebrafish behaviour experiments, each individual zebrafish was considered as one n number. Zebrafish in each group were derived from a minimum of three independent clutches unless otherwise stated.

For continuous datasets comparing two groups, normality was assessed using the Shapiro-Wilk test. Non-normally distributed data were compared using the Mann-Whitney U test. Normally distributed data were further assessed for homogeneity of variance using the F test. When variance was equal, datasets were compared using the unpaired t test. However, when variance was not equal unpaired t test with Welch's correction was used.

For continuous datasets comparing a single factor across more than two groups, normality was assessed using the Shapiro-Wilk test. Normally distributed data were compared using a one-way ANOVA. Non-normally distributed data were compared using the Kruskal-Wallace test. For continuous datasets comparing two factors across two or more groups a two-way ANOVA was used. Correlation between two datasets was analysed using Pearson correlation coefficient. Survival type data was analysed using the log-rank test.

Discrete data were analysed using a Chi-squared test for trend. For all figures in this document significance values are denoted as *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001.

3. Chapter 3: RNA-only zebrafish exhibit no ALS phenotype

3.1. Introduction

Although previous studies in zebrafish and drosophila have convincingly demonstrated that DPR proteins are generally more toxic than RNA foci, the contribution of each to cellular stress has thus far only been determined by generic readouts such as neurodegeneration and reduced viability (Lee et al., 2013, Mizielinska et al., 2014, Tran et al., 2015, Ohki et al., 2017). It is also important to accurately quantify the amount of cellular stress through simple molecular readouts, as this can allow early detection of the cell stress which precedes neurodegeneration or death (McGown et al., 2013). Furthermore, although neurodegeneration and loss of viability are good quality readouts, simple molecular readouts of cell stress can allow for higher throughput screening of potential neuroprotective compounds.

3.2. Aims

In order to address the above points, we aimed to generate two sets of transgenic zebrafish. One set of zebrafish would express only RNA foci (without ATG-driven DPRs) and the other set of fish would express both RNA foci and ATG-driven DPR proteins. Both sets of zebrafish would also have a tandem drug screening fluorescence reporter readout as used previously to successfully screen drugs in a *SOD1*-ALS zebrafish model (McGown et al., 2013). The two sets of zebrafish could therefore be used to identify the relative toxicity of RNA foci and DPR, and also screen for drugs which aid in alleviating this toxicity. Chapter 3 focuses on the generation and characterisation of RNA-only zebrafish.

3.3. C9orf72 expansion RNA-only transgenic zebrafish express hallmark RNA foci

At the single cell stage wildtype zebrafish of the AB strain were nuclear injected with a C9orf72 expansion containing construct (**figure 3.1a**). Expression of GFP and C9orf72 expansion RNA is driven by a zebrafish ubiquitin promoter. The C9orf72 expansions used in this construct were interrupted $(GGGGCC)_{99}$ hexanucleotides (strings of G_4C_2 repeats interrupted by TCGAC linkers). Specifically, the C9orf72 expansion sequence used was: $(G_4C_2)_{12}$-TCGAC-$(G_4C_2)_{17}$-TCGAC-$(G_4C_2)_{10}$-TCGAC-$(G_4C_2)_{22}$-TCGAC-$(G_4C_2)_{17}$-TCGAC-$(G_4C_2)_9$-TCGAC-$(G_4C_2)_{12}$ (see **Appendix B** for the full transgene sequence). The presence and size of the expansion was confirmed by DNA sequencing prior to injection. Further 3' of the C9orf72 expansion is the sequence of an auxin inducible degron (*AID*) gene. The AID gene was initially included in this construct as a control for its expression in DPR producing transgene constructs discussed later (see section 4.3 for a full explanation of the uses of the AID gene and production of DPR producing zebrafish). A start codon was placed at the

a

Disease Gene | Drug Screening Readout

| Ubiquitin Promotor | GFP | (GGGGCC)$_{99}$ interrupted repeats | AID | | Heat Shock Promotor 70 | DsRed |

Start Stop   Start Stop

b

| DAPI | RNA foci | Merge |

5.3-9
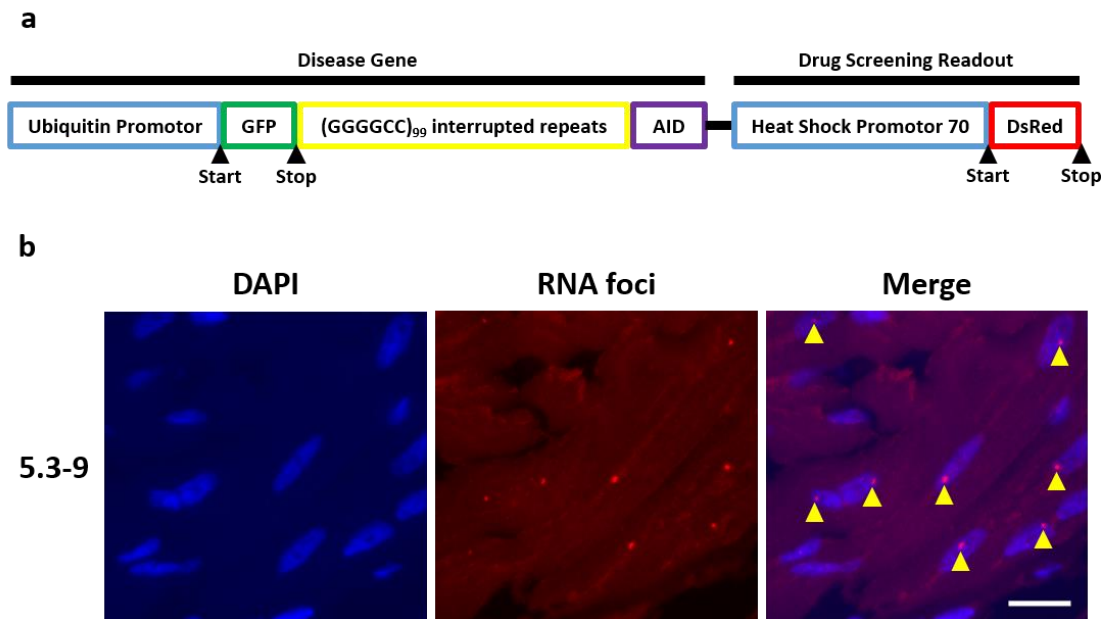
Figure 3.1: RNA-only 5.3 model zebrafish display RNA foci

(**a**) Schematic representation of the transgene inserted into 5.3 zebrafish. A zebrafish ubiquitin promotor drives GFP expression, but a stop codon prevents translation of all sequence downstream from GFP. A hsp70 promotor then drives DsRed production as a read out of cellular stress. (**b**) In situ hybridisation of paraffin embedded sections of 10dpf 5.3-9 zebrafish using a Cy3-conjugated (red) GC probe (CCCCGG)x4 showed that RNA foci are present in the nuclei of muscle cells. Yellow arrow heads denote RNA foci. Images are representative of 3 sections assessed from 3 individual zebrafish per group (not quantified). Scale bar = 10µm.

beginning of the GFP gene and a stop codon was placed at the end of the GFP gene, thus ensuring that no ATG-driven DPRs are produced (all potential frames are not amenable to RAN-translation in this construct due to the presence of stop codons or lack of ATG-like initiation codons). Thus, the ubiquitin promoter drives expression of GFP-*C9orf72* expansions-AID RNA, but only GFP is translated into protein, this is the disease gene component of the transgene construct. Further 3' of the disease gene is a hsp70 promoter driving expression of a DsRed gene, forming the drug screening portion of the DNA construct. NTG clutchmate zebrafish were used as controls for transgenics in all experiments as they provide the closest possible match for tank conditions, age and background genetics. NTG zebrafish do not express non-disease gene elements of the transgene expressed in transgenic zebrafish (e.g. GFP and DsRed). However, these genes have been widely used in zebrafish research and no toxicity has been reported through to

adult expression (Ju et al., 1999, Zhu and Zon, 2004, Ramesh et al., 2010, Blackburn et al., 2011).

In total approximately 500 zebrafish were injected with the RNA-only transgene construct. These zebrafish were referred to by the arbitrary term '5.3' - these zebrafish will be henceforth be referred to as 5.3 zebrafish lines. From those injected, 25 highly chimeric 5.3 zebrafish were identified. Chimerism was determined by visually examining which tissue types expressed the transgene (according to visible GFP or DsRed fluorescence). Zebrafish which showed transgene expression in at least muscle, heart and eye tissue (indicative of early transgene integration into embryos) were considered to be highly chimeric. These highly chimeric 5.3 zebrafish were grown to adulthood, then bred, and their offspring were screened for transgene expression by first heat shocking (see section 2.2.4 for details) and then examined for DsRed and GFP fluorescence under a fluorescence microscope. Of the 25 highly chimeric zebrafish, 7 gave rise to a proportion of full transgenic offspring. However, 2/7 exhibited extremely high transgene expression levels, and the offspring died within 5-7 dpf, thus preventing the establishment of a full transgenic line. The remaining 5/7 highly chimeric zebrafish gave rise to full transgenic offspring which were grown to adulthood and used to establish a full transgenic line. From each of the 5 full transgenic 5.3 zebrafish lines, a few embryos were examined by eye under a fluorescence microscope to judge the strength of their fluorescence relative to one another. High, middle and low expressing lines were selected for further examination, these lines were named 5.3-9, 5.3-5 and 5.3-6 respectively. As 5.3-9 demonstrated the highest transgene expression, this line was used for further characterisation. Transgenic 5.3-9 embryos and their NTG clutchmates were grown to 10 dpf and then processed for in situ hybridisation to test whether RNA foci from the *C9orf72* expansion could be observed. Numerous GGGGCC orientation RNA foci were observed in muscle cells of the 5.3-9 zebrafish, but not in their NTG clutchmates (**Figure 3.1b**). A limitation of the RNA foci data is that zebrafish were obtained from a single clutch, ideally 3 or more zebrafish clutches should be studied so that variation in background genetics may be accounted for and statistical comparison may be applied.

3.4.   5.3 zebrafish lines express GFP and DsRed

It is important to determine how strongly the transgene is expressed in each line of transgenic fish as it is likely that the highest expressing zebrafish would be most likely to develop a phenotype, if indeed the transgene is conferring toxicity to the zebrafish. The lines 5.3-9, 5.3-5 and 5.3-6 had already been identified as high, middle and low transgene expressers by eye. However, a more quantitative measure of this was required. For this reason, 5 dpf embryos from each of the aforementioned lines were lysed and individual GFP

65

and DsRed fluorescence values of each lysate were determined using a fluorescence plate reader. GFP fluorescence was significantly higher in the 5.3-9 line as compared to both 5.3-5 and 5.3-6 zebrafish lines (1.7×10^5 vs. 1.3×10^5 and 1.7×10^5 vs. 0.5×10^5 A.U. respectively; **Figure 3.2a**). Additionally, 5.3-5 GFP fluorescence was significantly higher than that of 5.3-6. GFP fluorescence was used as a convenient readout for transgene expression at the protein level. However, expression of the transgene was not examined at the DNA level (transgene copy number analysis) or RNA level (mRNA expression level analysis). Similar patterns were observed when DsRed florescence was measured. 5.3-9 DsRed fluorescence was significantly higher than both 5.3-5 and 5.3-6 (1.3×10^5 vs. 1.2×10^5 and 1.3×10^5 vs. 0.1×10^5 A.U. respectively; **Figure 3.2b**). Additionally, 5.3-5 DsRed fluorescence was significantly higher than that of 5.3-6. Although raw fluorescence values are an important measure of transgene expression levels, a better readout of average cell stress levels is given by the ratio of DsRed:GFP. This is because DsRed protein levels can be influenced by the number of copies of the transgene present in each individual line, as well as by cell stress driving the hsp70 promoter. Therefore, by normalising DsRed fluorescence levels to the GFP florescence levels of the same zebrafish, the influence of transgene copy number on DsRed fluorescence will be diminished, giving a more accurate representation of cell stress levels. The DsRed:GFP ratio of 5.3-9 zebrafish was significantly higher than both 5.3-5 and 5.3-6 zebrafish (0.78 vs. 0.23 and 0.78 vs. 0.12 respectively; **Figure 3.2c**). Interestingly, despite higher raw fluorescence values, the DsRed:GFP ratio of 5.3-5 zebrafish was not significantly higher than that of 5.3-6 zebrafish.

As 5.3-9 zebrafish showed significantly higher GFP, DsRed and DsRed:GFP ratio than the other 5.3 lines, it was considered that 5.3-9 zebrafish would be the most likely to develop a phenotype. Therefore, the majority of subsequent phenotypic characterisation was conducted in the 5.3-9 zebrafish line. In order to confirm that 5.3-9 zebrafish were producing GFP protein alone (i.e. no DPR species were being translated from the C9orf72 expansion), 5 dpf embryos from the 5.3-9 zebrafish were lysed and western blotted. Immunoblotting of 5.3-9 embryonic lysates with an antibody against GFP revealed a single band at 25KDa, the exact expected size of the GFP protein (**Figure 3.2d**). Both GFP and DsRed were also detectable by confocal imaging of endogenous fluorescence in 15dpf cryosectioned 5.3-9 zebrafish (**Figure 3.2e**). Although transgene driven expression of GFP was observed in muscle tissue of 15dpf cryosectioned 5.3-9 zebrafish, it was not clear whether GFP expression was also present in CNS tissues of 5.3-9 zebrafish.
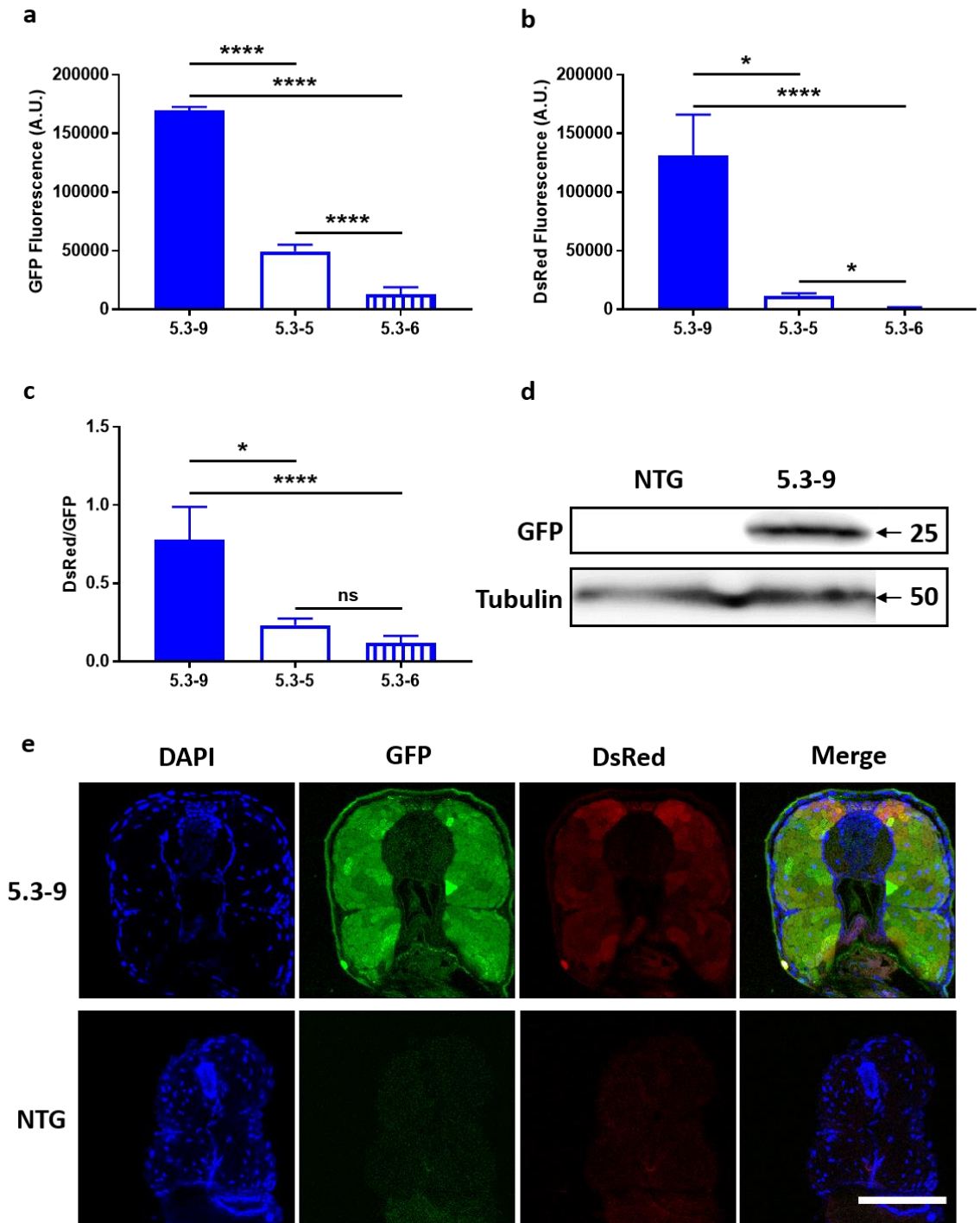
Figure 3.2: RNA-only 5.3 zebrafish lines express GFP and DsRed

(**a**) At 5dpf, 5.3-9 zebrafish have significantly higher GFP fluorescence than both 5.3-5 and 5.3-6 lines. Additionally, 5.3-5 zebrafish have significantly higher GFP fluorescence than 5.3-6 zebrafish. N=9 5.3-9, 10 5.3-5 and 9 5.3-6 zebrafish, derived from a single clutch. Statistical comparisons were carried out using a one-way ANOVA with Tukey's post-hoc test. (**b**) At 5dpf, 5.3-9 zebrafish have significantly higher DsRed fluorescence than both 5.3-5 and 5.3-6 lines. Additionally, 5.3-5 zebrafish have significantly higher DsRed fluorescence than 5.3-6 zebrafish. N=9 5.3-9, 10 5.3-5 and 9 5.3-6 zebrafish, derived from a single clutch. This dataset (**legend continues on next page**)

was not normally distributed according to the Shapiro-Wilk test for normality, and was therefore statistically compared with the Kruskal-Wallis test with Dunn's post-hoc test. (**c**) At 5dpf, 5.3-9 zebrafish have significantly higher DsRed:GFP ratio than both 5.3-5 and 5.3-6 lines. N=9 5.3-9, 10 5.3-5 and 9 5.3-6 zebrafish, derived from a single clutch. This dataset was not normally distributed according to the Shapiro-Wilk test for normality, and was therefore statistically compared with the Kruskal-Wallis test with Dunn's post-hoc test. (**d**) At 5dpf, whole 5.3-9 zebrafish protein lysates produce a single GFP immunoreactive band at 25KDa (the exact size of the GFP protein). (**e**) At 15 dpf, 5.3-9 zebrafish produce GFP and DsRed. The NTG zebrafish shown for comparison is 5 dpf. Transverse body sections. Scale bar = 100µm. Images are representative of 9 zebrafish imaged (not quantified). All data are shown as mean +/- standard deviation; *P < 0.05 and ****P < 0.0001.

3.5. GFP but not DPR proteins are expressed in adult 5.3-9 zebrafish brains

ALS is primarily a neurological disease, it is therefore critical to determine whether expression of the RNA-only transgene is present in the zebrafish CNS. For this reason, we lysed adult 5.3-9 zebrafish brains and western blotted them for GFP. A GFP immunoreactive band was detected at the predicted molecular weight of GFP in the 5.3-9 adult brain lysates, but not in those of NTG controls brain lysates (**Figure 3.3a**, top panel). GFP levels detected in the 5.3-9 zebrafish brain lysates were statistically significantly higher than those detected in the NTG controls (**Figure 3.3b**). This demonstrates that the C9orf72 expansion containing transgene is expressed in the CNS, despite the lack of an ALS-like phenotype being observed. As shown previously at 5 dpf, GFP was detected at 25KDa, the expected size of the GFP protein, thus demonstrating that translation of the C9orf72 expansion is prevented by the stop codon placed at the end of the GFP gene. However, DPR species can be produced from C9orf72 expansions even in the absence of a start codon due to the phenomenon of RAN-translation. Additionally, translation of RAN-translated DPR species would be initiated in close proximity to the C9orf72 expansion and therefore would not allow for translation of GFP. This means that RAN-translated DPRs would not be likely to be detected using a GFP antibody. In order to assess whether RAN-translated DPR proteins were being produced from the C9orf72 expansion, we probed 5.3-9 adult brain lysates with antibodies against poly(GA) and poly(PA) DPRs. Neither poly(GA) nor poly(PA) antibodies detected any specific protein bands in the 5.3-9 zebrafish brain lysates (**Figure 3.3a**, middle two panels).

Previous validation of the poly(GA) antibody used here, identified that the poly(GA) antibody binds both human and recombinant poly(GA) protein, and showed minimal cross reactivity with all other DPR species (Gendron et al., 2015). Similarly, the binding specificity of the
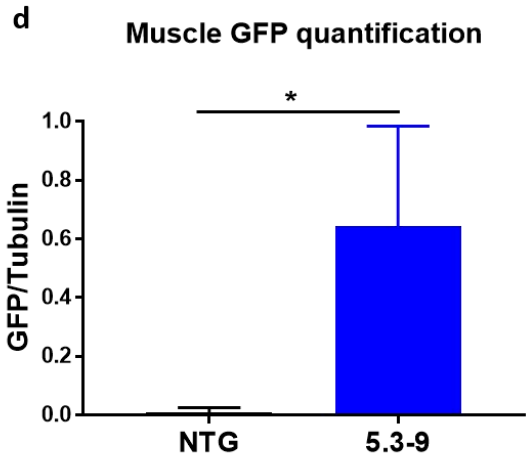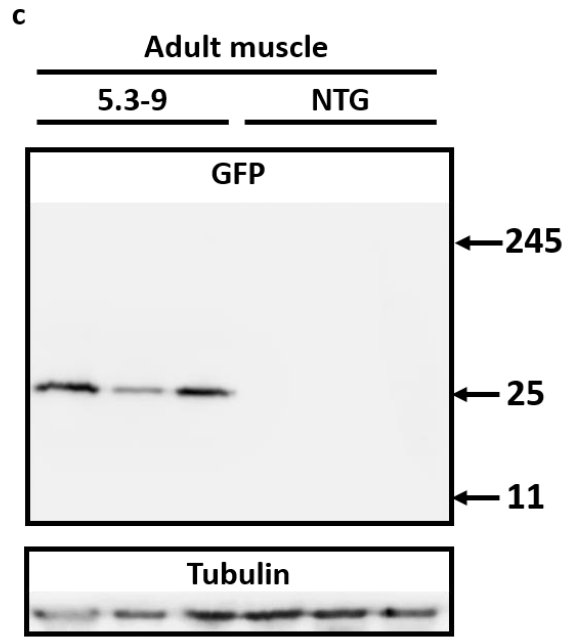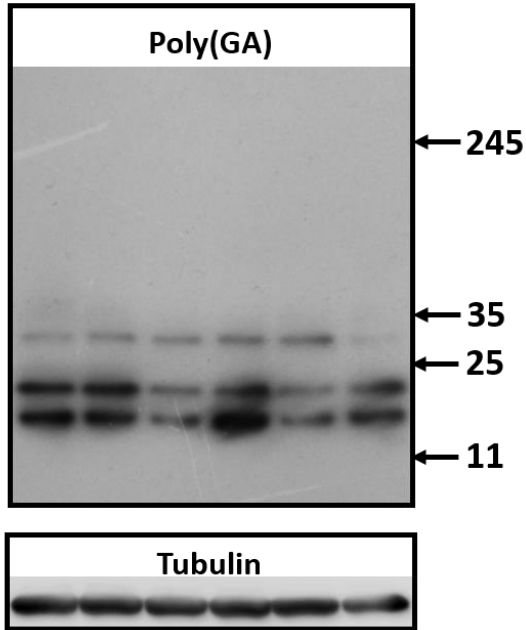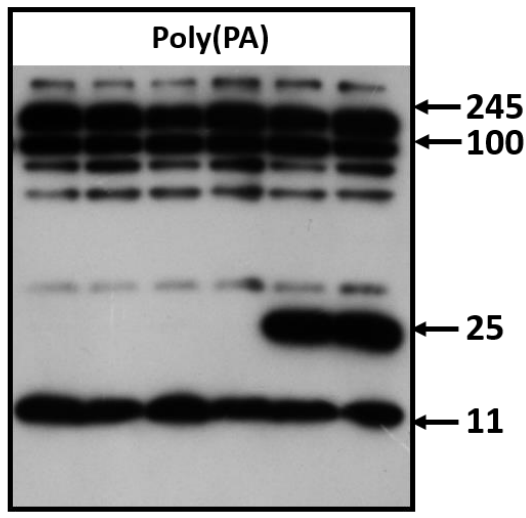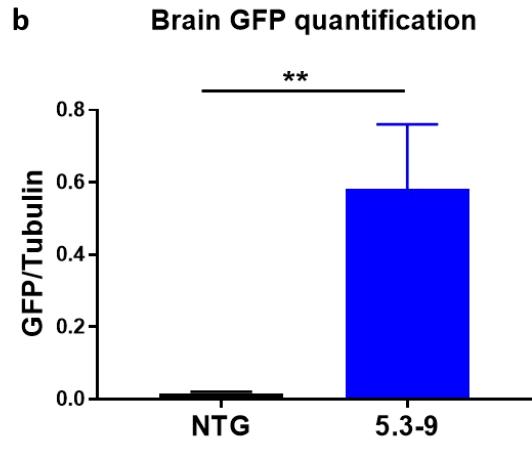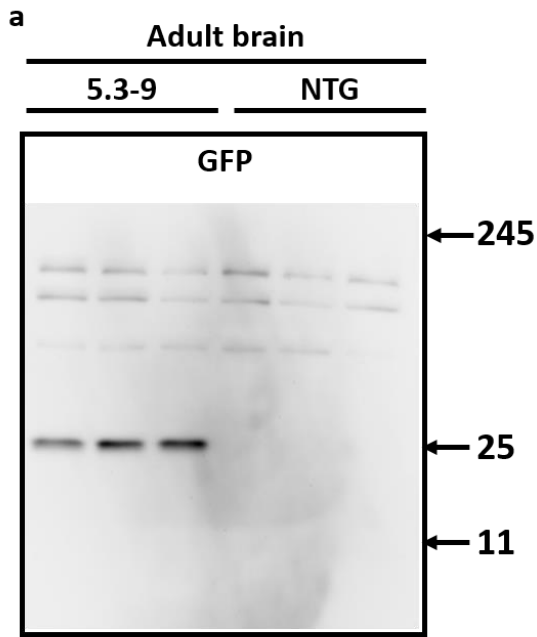
Figure 3.3: Adult 5.3-9 zebrafish produce GFP, but not poly(GA) or poly(PA) DPRs

(**a**) Adult 26 month old 5.3-9 zebrafish brains produced an immunoreactive band at 25KDa that is not present in any NTG lane, this is the predicted weight of the GFP protein. No 5.3-9 specific bands could be detected with either poly(GA) or poly(PA) antibodies, all the bands shown for these antibodies are non-specific bands and are not DPRs. Poly(GA) and poly(PA) blots underwent long exposures to help identify any DPR proteins which may have been expressed at low levels. (**b**) Quantification of GFP/tubulin levels in adult 5.3-9 zebrafish brains. Statistical comparison was carried out with an unpaired t-test. (**c**) Adult 26 month old 5.3-9 zebrafish muscle produces an immunoreactive band at 25KDa that is not present in any NTG lane, this is the predicted weight of the GFP protein. (**d**) Quantification of GFP/tubulin levels in adult 5.3-9 zebrafish muscle. Statistical comparison was carried out with an unpaired t-test.

poly(PA) antibody used here has been validated previously in a western blot which showed that poly(PA) immunoreactive bands were detected only in lysates from poly(PA) expressing cells, and not in lysates of cells expressing any other DPR species (May et al., 2014). Hence, a lack of poly(GA) or poly(PA) antibody specificity is not likely to be the cause of the absence of 5.3-9 specific bands detected in adult brain lysates. Thus, despite transgene expression being detectable in adult 5.3-9 brain tissue as shown by GFP production, no poly(PA) or poly(GA) containing DPRs are produced at detectable levels through conventional translation or RAN-translation. Furthermore, muscle tissue from the same zebrafish was also western blotted and probed for GFP. A GFP immunoreactive band was also detected at the predicted molecular weight of GFP protein in 5.3-9 adult zebrafish muscle tissue, but not in NTG controls (**Figure 3.3c**). GFP levels in adult muscle tissue of 5.3-9 zebrafish were statistically significantly higher than those in NTG control adult muscle tissue (**Figure 3.3d**). No other GFP immunoreactive bands were detected in adult muscle zebrafish tissue, indicating that only GFP protein, and not GFP tagged DPRs are produced in this tissue.

3.6. Embryonic 5.3-9 zebrafish show a mild hyperactivity phenotype

In order to determine whether ALS-like locomotor defects were present in 5.3-9 embryonic zebrafish, spontaneous locomotor activity was monitored in 5 dpf embryos. Swimming activity was monitored under dark conditions for 30 minutes, and average speed and total distance swam was calculated for each individual zebrafish. Despite average speed being 46% higher in 5.3-9 compared with NTG zebrafish, this increase did not reach statistical significance (**Figure 3.4a**). Similarly, 5.3-9 zebrafish swam on average 45% further than
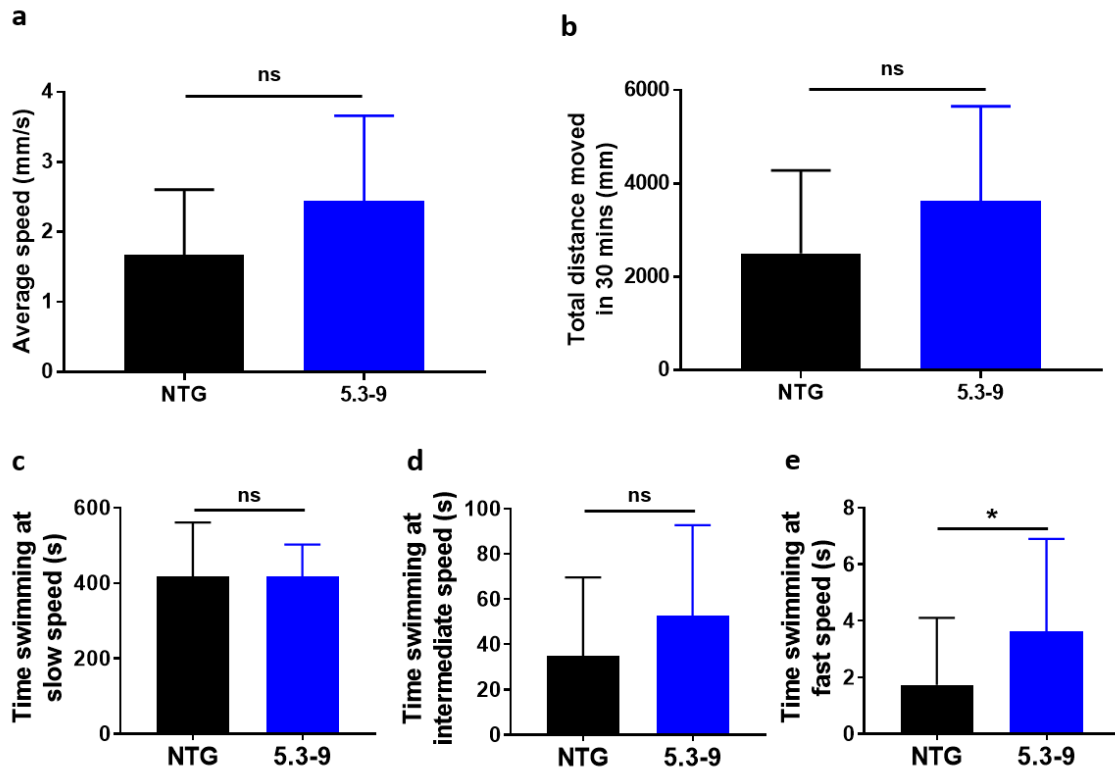
Figure 3.4: 5.3-9 zebrafish spend more time swimming at fast speeds

(**a**) No difference is observed in the average speed of the 5.3-9 and NTG zebrafish over a 30 minute recording period. Statistical comparison was carried out using an unpaired t-test. (**b**) No difference is observed in the average speed of the 5.3-9 and NTG zebrafish over a 30 minute recording period. Statistical comparison was carried out using an unpaired t-test. (**c**) No difference was observed in the time spent swimming at slow speed between 5.3-9 and NTG zebrafish over a 30 minute recording period. This dataset was not normally distributed according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. (**d**) No difference was observed in the time spent swimming at intermediate speed between 5.3-9 and NTG zebrafish over a 30 minute recording period. This dataset was not normally distributed according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. (**e**) 5.3-9 zebrafish spent significantly more time swimming at fast speeds in comparison to NTG zebrafish over a 30 minute recording period. This dataset was not normally distributed according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. N=16 individual fish per genotype for all comparisons, zebrafish were derived from a single clutch. Speed thresholds used were slow ($0 < x < 5mm/sec$), intermediate ($5 < x < 15mm/sec$) and fast ($x > 15mm/sec$). All data were recorded under dark conditions and are shown as mean +/- standard deviation; ns: not significant and $*P < 0.05$.

NTG zebrafish, however this difference also did not reach statistical significance (**Figure 3.4b**).

In order to analyse swimming behaviour in more detail, zebrafish swimming data were subdivided into time spent swimming at slow, intermediate and fast speeds. No significant difference was observed in the amount of time 5.3-9 and NTG zebrafish spent swimming at slow or intermediate speeds (**Figure 3.4c+d**). However, 5.3-9 zebrafish spent significantly more time swimming at fast speeds in comparison to NTG zebrafish (3.6 vs 1.7 seconds for 5.3-9 and NTG zebrafish respectively; **Figure 3.4e**).

Behavioural studies in embryonic zebrafish are known to generate highly variable data, this is at least in part due to the propensity of some embryonic zebrafish to not move for long periods of time (Liu et al., 2017). For this reason, large numbers of zebrafish from multiple clutches are typically favoured in embryonic zebrafish behavioural experiments. An important limitation of the behavioural data shown in figure 3.4, is that the zebrafish used were derived from a single clutch.

3.7. Survival of 5.3-9 zebrafish is normal through to adulthood

Production of DsRed in 5.3-9 zebrafish indicates activation of the *hsp70* promoter. Activation of the *hsp70* promoter suggests that RNA foci may be driving low level toxicity in 5.3-9 zebrafish. To assess whether the C9orf72 expansion and/or RNA foci were impacting upon the viability of zebrafish, we monitored survival of 5.3-9 zebrafish up to the early larval stage. At both 5 dpf and 15 dpf no significant difference in survival was observed between 5.3-9 zebrafish (mix of heterozygous and homozygous resulting from a 5.3-9 X 5.3-9 cross) and their NTG clutchmates (84 vs. 86% survival at 15dpf for 5.3-9 and NTG respectively; **Figure 3.5a**). In order to assess whether toxicity derived from RNA foci may manifest over a long period of time, heterozygous 5.3-9 zebrafish survival was monitored into adulthood. At 26 months old, the survival of 5.3-9 zebrafish was slightly higher than that of their NTG clutchmates (55 vs. 53% survival for heterozygous 5.3-9 and NTG adults respectively; **Figure 3.5b**). A limitation of these data is that all 3 clutches of fish were maintained in the same tank from 1dpf until 26 months old. Therefore, rather than 3 separate values for 26 months survival (1 value from each clutch), only a single survival value could be calculated (1 value from all three clutches combined). For this reason statistical comparison of 5.3-9 and NTG survival at 26 months was not possible. Although there is no overt loss of viability in RNA foci expressing zebrafish up to 26 months, it is important to assess whether RNA foci may cause a sub-lethal ALS-like phenotype such as impairments in neuromuscular physiology.
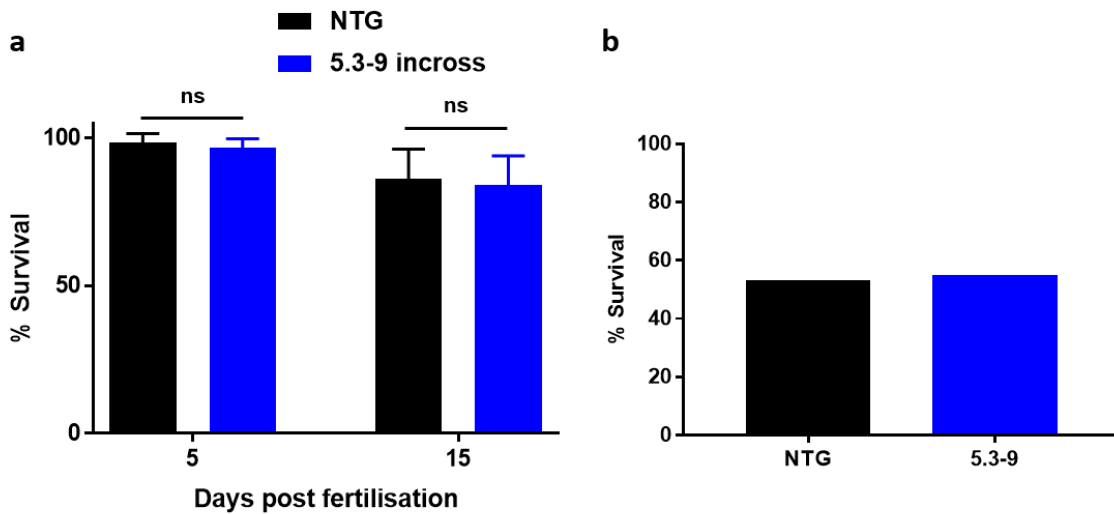
Figure 3.5: 5.3-9 zebrafish have normal survival at 26 months old

(**a**) Survival of 5.3-9 zebrafish is not significantly different from NTG clutch mates at 5dpf or 15 dpf. N=4 clutches for each group. Data were statistically compared using a two-way ANOVA with Sidak's post hoc test. (**b**) Survival of 5.3-9 zebrafish is similar to that of NTG clutchmates at 26 months old. N=60 fish of each genotype at 1 dpf. Data could not be compared statistically as survival was only assessed at a single time point, i.e. of the 60 fish in each group at 1dpf, how many were still alive at 26 months old. All data are shown as mean +/- standard deviation; ns: not significant.

3.8. Adult 5.3-9 zebrafish exhibit normal swimming endurance

Of the 5.3 zebrafish, the highest transgene expressing zebrafish (5.3-9) exhibit RNA foci pathology but do not show any overt reduction in survival when monitored well into adulthood. It is therefore important to determine whether the 5.3-9 zebrafish exhibit a more subtle ALS-like phenotype such as impaired motor function. To assess this, the swimming endurance of 5.3-9 zebrafish was measured using a swim tunnel setup. At 2-3 months old, the swimming endurance of 5.3-9 zebrafish showed a small but non-significant reduction compared with their NTG clutchmates (98 vs. 100 minutes median swimming time for 5.3-9 and NTG respectively; **Figure 3.6a**). Only 4 fish of each genotype were tested at 2-3 months old, before it was decided to wait until the 5.3-9 zebrafish were older and were more likely to display a clear phenotype. During this time the protocol used in the swim tunnel was optimised to allow faster exhaustion of the zebrafish being tested. For this reason the swim tunnel performance of the two age groups of 5.3-9 zebrafish tested are not directly comparable. At 26 months old, 5.3-9 zebrafish were re-tested using the new swim tunnel
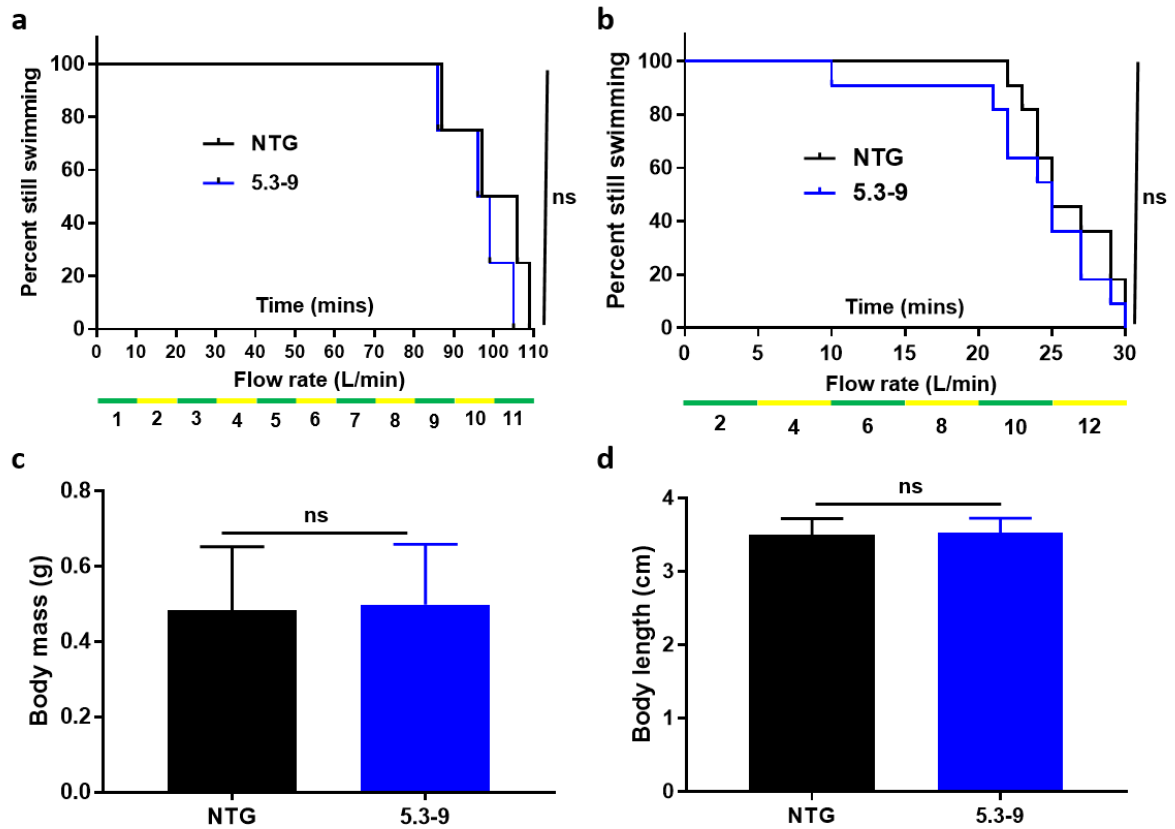
Figure 3.6: 5.3-9 zebrafish have normal swimming ability at 26 months olds

(**a**) At 2-3 months old, swimming ability of 5.3-9 zebrafish was not significantly different from that of NTG clutchmates. N=4 fish per genotype. Statistical comparison was carried out using the log-rank test. (**b**) At 26 months old, swimming ability of 5.3-9 zebrafish was not significantly different from that of NTG clutchmates. N=11 fish per genotype. Statistical comparison was carried out using the log-rank test. (**c**) At 26 months old, body mass of 5.3-9 zebrafish was not significantly different compared to their NTG clutchmates. N=11 fish per genotype. Statistical comparison was carried out using an unpaired t-test. (**d**) At 26 months old, body length of 5.3-9 zebrafish was not significantly different compared to their NTG clutchmates. N=11 fish per genotype. Statistical comparison was carried out using an unpaired t-test. All data are shown as mean +/- standard deviation; ns: not significant

protocol, but still showed no significant difference in swimming performance as compared with their NTG clutchmates (25 vs. 25 minutes median swimming time; **Figure 3.6b**). To ensure that the swimming ability of zebrafish was not being influenced by body mass or body length, these factors were measured after removal from the swim tunnel. Neither body mass nor body length was significantly different between 5.3-9 zebrafish and their NTG clutchmates (**Figure 3.6c+d**).

3.9. The G-quadruplex unfolding drug TMPYP$_4$ does not influence heat shock response activation

Despite a lack of detectable DPR species, 5.3-9 zebrafish exhibit DsRed production as early as 5 dpf, as seen in figure 3.2. One potential explanation for this is that RNA foci may directly or indirectly activate the hsp70 promoter and therefore drive DsRed production. The expression of RNA foci was confirmed in 10 dpf 5.3-9 zebrafish (see figure 3.1), but not in adult 5.3-9 zebrafish. GGGGCC RNA foci are formed by the GC-rich RNA molecules adopting G-quadruplex structures (Reddy et al., 2013). To test whether G-quadruplex RNA foci formation can drive the heat shock response, we injected zebrafish with TMPYP$_4$ which has previously been demonstrated to unfold G-quadruplex structures (Morris et al., 2012). First, preliminary injections were performed in NTG zebrafish immediately after fertilisation, in order to identify an appropriate injection dose. Although some degree of toxicity was observed with every dose of TMPYP$_4$ injected, 730 µM was selected as the dose to use in the injections of 5.3-9 zebrafish embryos, as it was the maximum dose which did not cause substantial death and abnormalities (30% death with 730µM injection vs 70% death with 7300µM injection; **Figure 3.7a**). Injections were then carried out in 5.3-9 zebrafish, where immediately after fertilisation embryos were injected with ultra-pure water (1nl), 730µM TMPYP$_4$ (single dose, 1nl), 730µM TMPYP$_4$ (double dose, 2nl) or left uninjected. No significant difference was observed between any of the groups (**Figure 3.7b**). It is important to note that we did not test whether TMPYP$_4$ injection was sufficient to prevent RNA foci formation or block typical RNA foci interactions, in the 5.3-9 zebrafish.

3.10. Discussion

RNA-only zebrafish were generated through nuclear microinjection of transgenic DNA constructs. The DNA constructs used did not contain transposable elements. Microinjections of this kind are often considered to be very laborious due to their predicted low rate of transgenesis. However, a combination of spermine and spermidine containing injection buffer (keeps DNA in compact conformation) and pre-screening (by eye under a fluorescence microscope) for highly chimeric embryos, led to 28% (7/25) of chimeric
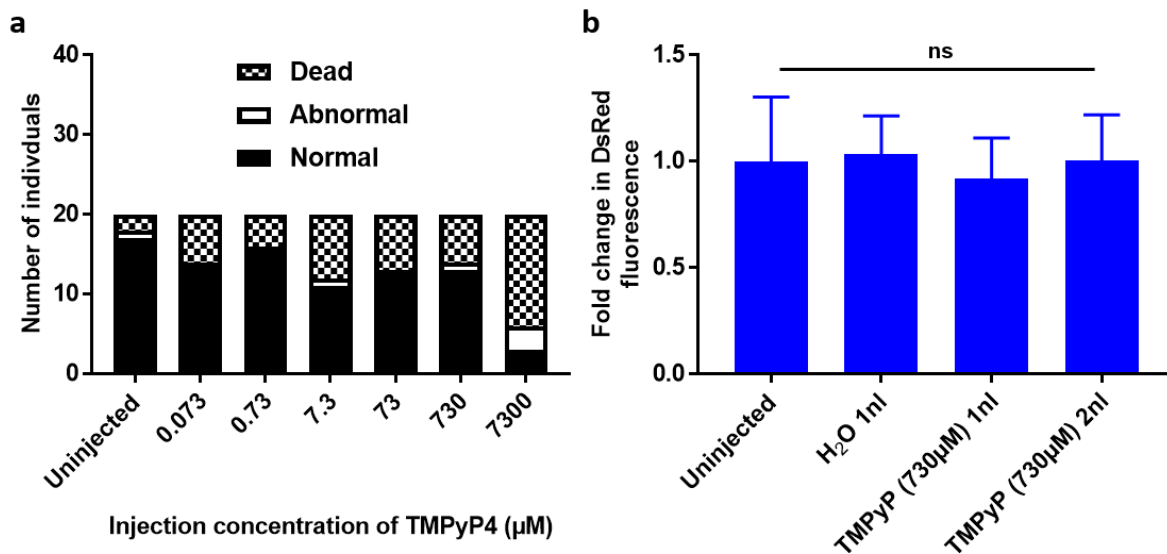
Figure 3.7: The G-quadruplex unfolding drug TMPYP4 does not influence heat shock response activation

(**a**) When a 1nl volume was injected into the yolk sac of NTG zebrafish immediately after fertilisation, every concentration of TMPYP4 caused an increase in death compared with uninjected zebrafish. The increase in death rates was markedly higher when 7300µM TMPYP4 was injected. 730µM was the highest dose which did not induce high death rates, so this was selected to be used in future injections with 5.3-9 zebrafish. (**b**) No significant change in DsRed fluorescence was observed in 5.3-9 zebrafish when injected with H_2O, or a single or double injection of 730µM TMPYP4, as compared to uninjected 5.3-9 zebrafish. Statistical comparison was carried out using a one-way ANOVA with Tukey's post hoc test.

zebrafish grown to adulthood giving rise to full transgenic offspring. Not only does this method offer reasonable rates of transgenesis without the need to clone transposons into DNA constructs, but it also obviates the lengthy process of backcrossing zebrafish which is necessitated when transposable elements have caused transgene integration in more than one locus. 2/7 highly chimeric zebrafish showed severely reduced fertility and the few fertilised offspring died within the first few days. The reason why these offspring show an embryonic lethal phenotype whereas 5.3-9 zebrafish have no detectable phenotype even during adulthood is not fully understood. It has however, been reported previously that

injection of *C9orf72* expansion containing RNA can lead to toxicity even in the absence of DPR proteins (Swinnen et al., 2018). Therefore there may be a threshold at which C9-

expansion RNA will trigger an early lethal phenotype, but RNA levels below this threshold may be manageable over the long term.

The 5.3-9 zebrafish line showed higher expression levels of GFP and DsRed, as compared to 5.3-6 and 5.3-5. If GFP and DsRed expression levels were solely determined by transgene copy number then GFP and DsRed expression should increase in linear proportion to one another, and therefore GFP:DsRed ratios should be comparable across all 5.3 zebrafish lines. However, the DsRed:GFP ratio was higher in 5.3-9 zebrafish, indicating that DsRed production may be influenced by C9orf72 expansion expression levels. Additionally, 5.3-5 zebrafish showed significantly higher GFP and DsRed levels as compared to 5.3-6, though no significant difference was detected in the DsRed:GFP ratio. This suggests that there may be an expression threshold required before *C9orf72* expansions will result in *hsp70* promoter activation and production of additional DsRed protein.

No significant locomotor phenotype could be detected in embryonic zebrafish as measured by average speed or total distance covered, despite an approximate 45% increase in both measurements in 5.3-9 zebrafish. However, a mild hyperactivity phenotype was detected in 5.3-9 zebrafish when swimming at fast speeds. All locomotor data obtained were highly variable, this is likely due to a combination of the relatively small sample size used and innate variability in zebrafish behaviour at this developmental stage (Liu et al., 2017).   A further limitation of these data is that the zebrafish used were obtained from a single clutch. Therefore, future experiments should repeat locomotor analysis with a further 2 clutches of zebrafish, ideally using a larger sample size, which may help reduce variation in the data.

The 5.3-9 zebrafish did not show any early or late onset reduction in viability/survival. Additionally, 5.3-9 zebrafish did not show any underlying motor impairment as measured by swim tunnel testing during adulthood. These 5.3-9 zebrafish showed GFP expression in CNS and muscle tissues during adulthood, but no detectable DPR species. This is consistent with previous reports suggesting that a stop codon in close proximity to the C9orf72 expansion, combined with a lack of ATG-like codons upstream of the repeats, will effectively prevent the formation of RAN-translation products (Todd et al., 2013, Mizielinska et al., 2014, Green et al., 2017). This suggests that expression of *C9orf72* expansion RNA alone is not sufficient to induce toxicity at this age. However, age is a known risk factor for ALS (Armon, 2003). Survival, swim tunnel and DPR expression data were obtained from 5.3-9 zebrafish at 26 months old. Zebrafish mean life span typically ranges between 36 and 42 months old depending on the strain (Gerhard et al., 2002). Therefore, on average 5.3-9 zebrafish may be expected to live 10-16 months longer, and should continue to be monitored

during this time to assess the role of aging on C9orf72 expansion pathogenicity in this zebrafish model.

Early injection of 5.3-9 zebrafish with the proposed G-quadruplex unfolding molecule TMPyP$_4$ (Morris et al., 2012), did not result in amelioration of the heat shock stress response. It is important to note that it was not assessed whether TMPyP4 was able to enter the cell/nucleus, or whether indeed the molecule was able to unfold the secondary structure of RNA foci. Accurately measuring these two factors in future experiments will be important to determine whether TMPyP4 can unfold the G-quadruplex structures formed by RNA foci and whether this may help reduce cellular stress.

It will also be important to re-analyse tissue from any 5.3-9 zebrafish which develops a phenotype in the future, to test for the presence of DPR proteins so that the relative contribution of toxicity from RNA and DPR may be further assessed. Finally, determining whether RNA foci or perhaps minute undetectable levels of DPRs are causing the heat shock mediated production of DsRed in the 5.3-9 zebrafish will help provide clues to the relative contributions to toxicity from these two species.

4.   Chapter 4: RNA+DPR zebrafish show ALS/FTD features

4.1.   Introduction

Reports from drosophila models indicate that RNA foci expression in isolation results in no neurodegenerative phenotype (Mizielinska et al., 2014, Tran et al., 2015). Zebrafish models expressing RNA foci have been reported to show a mild phenotype which is not indicative of ALS (cardiac oedema) (Ohki et al., 2017). These findings are somewhat corroborated by data presented in chapter 3 of this thesis, showing that zebrafish expressing RNA foci do not develop any measurable ALS-like phenotype when studied well into adulthood.

On the other hand, when DPR proteins are expressed, a previously reported zebrafish model shows a more severe form of the cardiac oedema phenotype and are no longer viable (Ohki et al., 2017). Additionally, drosophila models expressing DPR proteins show a neurodegenerative phenotype (Mizielinska et al., 2014, Tran et al., 2015). Whilst these studies are informative about the relative toxicities of RNA foci and DPR species, they do not accurately recapitulate an ALS phenotype. Furthermore, stable BAC mouse models which do recapitulate an ALS-like phenotype are not amenable to high throuput drug screening (Jiang et al., 2016, Liu et al., 2016). Therefore, there is currently a need for an *in vivo* model which can accurately recapitulate the ALS symptoms observed in human *C9orf72* patients, and which can be used to efficiently test potential therapeutic compounds.

4.2.   Aims

In order to address the above points, we aimed to generate C9orf72 expansion zebrafish expressing both RNA foci and DPR proteins. These RNA+DPR zebrafish could then be used to both study the molecular pathology generated by DPR species, and identify therapeutic compounds which may alleviate DPR toxicity. This chapter describes the generation and characterisation of RNA+DPR expressing zebrafish.

4.3.   C9orf72 expansion RNA+DPR transgenic zebrafish express hallmark RNA foci

At the single cell stage wildtype zebrafish of the AB strain were nuclear injected with a C9orf72 expansion containing construct (**figure 4.1a**). The constructs begins with a ubiquitin promotor which drives expression of a GFP gene and C9orf72 expansions (encoding a GFP-DPR fusion protein). The C9orf72 expansions used in this construct were interrupted $(CCCCGG)_{89}$ hexanucleotides (strings of C_4G_2 repeats interrupted by GTCGA linkers). 89 hexanucleotide repeats were counted, however more may be present as sequencing reads were not able to traverse the entire C9orf72 expansion. Specifically, the C9orf72 expansion sequenced was: $(C_4G_2)_{12}$-GTCGA-$(C_4G_2)_9$-GTCGA-$(C_4G_2)_{17}$-$(C_4G_2)_{22}$-GTCGA-$(C_4G_2)_{17}$-
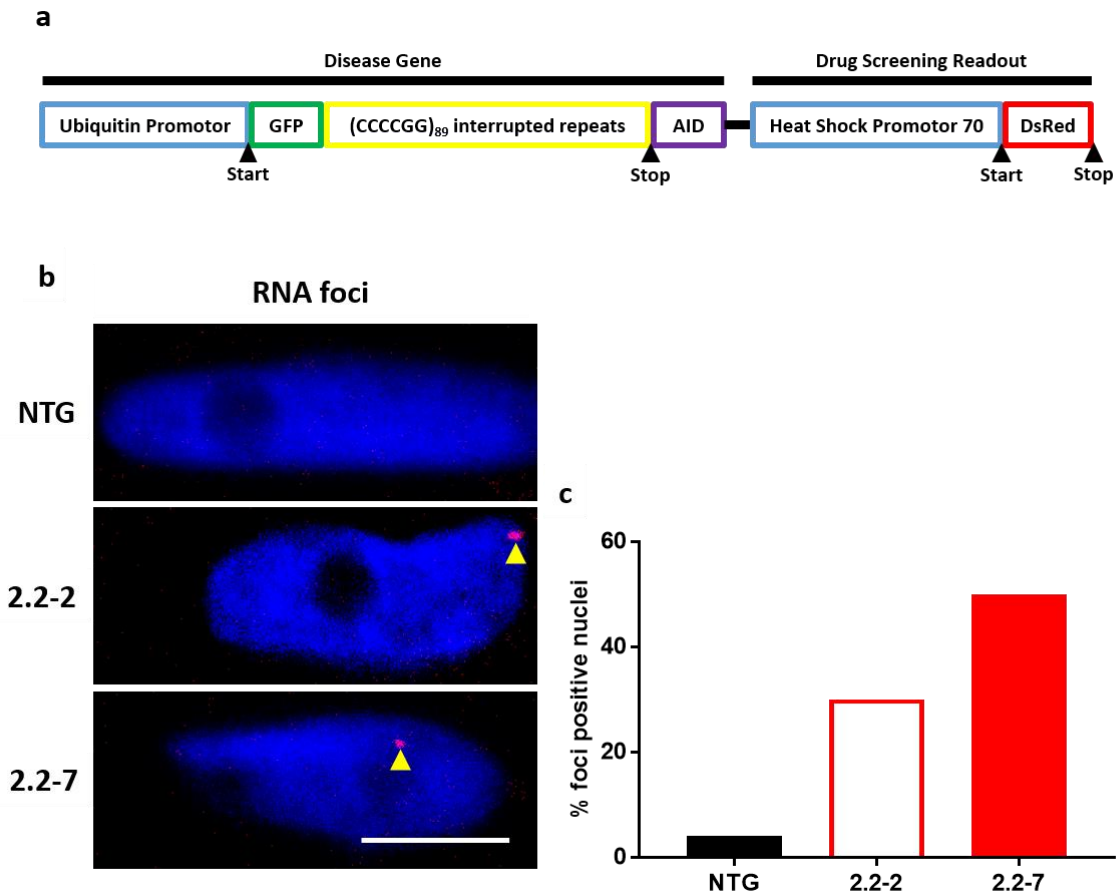
Figure 4.1: RNA+DPR 2.2 model zebrafish display RNA foci

(**a**) Schematic representation of the transgene inserted into 2.2 zebrafish. A zebrafish ubiquitin promotor drives GFP-DPR expression. The AID gene is out of frame and a stop codon prevents it from being translated into protein. A hsp70 promotor then drives DsRed production as a read out of cellular stress. (**b**) In situ hybridisation of paraffin embedded sections of 10dpf 2.2-7 zebrafish using a Cy3-conjugated (red) GC probe (GGGGCC)x4 showed that RNA foci are present in the nuclei of muscle cells. Yellow arrow heads denote RNA foci. Scale bar = 5μm. Images are representative of 6 body muscle sections assessed from 3 individual zebrafish per group. (**c**) Blinded quantification of RNA foci in 2.2 lines and NTG zebrafish muscle. Quantified images were taken from 3 zebrafish (6 sections were assessed for each group) from a single clutch per condition.

GTCGA-($C_4G_2)_{12}$ (see **Appendix C** for full transgene sequence). The presence of the C9orf72 expansion was confirmed by DNA sequencing prior to injection. Further 3' of the C9orf72 expansion is an AID gene. Upon co-expression of the TIR1 protein and exposure to the plant hormone auxin, the AID protein can be utilised to target itself, and other proteins with which it is fused, for proteasomal degradation (Nishimura et al., 2009). The AID gene

had been planned to encode the final portion of a GFP-DPR-AID fusion protein, which we hypothesised may allow for efficient proteasomal degradation of DPR species. However, due to unforeseen cloning difficulties, the AID gene was out of frame with respect to the start codon for GFP in this construct, this generated a premature stop codon, which resulted in the entire AID gene not being translated into protein. Thus, only a GFP-DPR fusion protein was encoded by this construct. As the C9orf72 expansion prevents polymerase processivity, frame correction of the AID gene with standard cloning techniques was not possible. It is possible that due to RAN-translation of all frames, some RAN-translated DPR-AID fusion protein may be produced, however this is likely a small percentage of the total DPR protein, due to RAN-translation being significantly less efficient than ATG-mediated translation (Kearse et al., 2016). Hence, AID mediated degradation of RAN-translated DPR-AID proteins was not considered a worthwhile avenue to pursue further. A start codon was placed at the beginning of the GFP gene and a stop codon was present at the end of the C9orf72 expansion. Thus, from this construct GFP and the C9orf72 expansion will be translated into a GFP-DPR fusion protein. The ubiquitin promoter driving GFP, C9orf72 expansions and AID (not translated) is the disease gene component of the construct.

Further 3' of the disease gene is a hsp70 promoter driving expression of a DsRed gene, this forms the drug screening portion of the DNA construct. For clarity, the 3 differences between the RNA-only construct used to generate the zebrafish described in chapter 3 and the RNA+DPR construct described here are: 1) C9orf72 expansions are of opposite orientation, G_4C_2 for RNA-only and C_4G_2 for RNA+DPR; 2) C9orf72 expansion length, 99 confirmed repeats for RNA-only and 89 confirmed repeats for RNA+DPR; 3) The position of disease gene stop codon, after GFP for RNA-only (codes for GFP protein only) and after C9orf72 expansion for RNA+DPR (codes for GFP-DPR fusion protein). The orientation of the repeats was reversed in RNA+DPR constructs due to the G_4C_2 expansions being too unstable in this context. NTG clutchmate zebrafish were used as controls for transgenics in all experiments as they provide the closest possible match for tank conditions, age and background genetics. NTG zebrafish do not express non-disease gene elements of the transgene expressed in transgenic zebrafish (e.g. GFP and DsRed) however these genes have been widely used in zebrafish research and no toxicity has been reported even when expressed throughout adulthood (Ju et al., 1999, Zhu and Zon, 2004, Ramesh et al., 2010, Blackburn et al., 2011). Furthermore, 5.3 zebrafish also express transgene elements such as GFP and DsRed and do not exhibit ALS-like phenotypic alterations. We therefore reason that it is highly unlikely that non-disease gene elements of the transgene contribute to the ALS-like phenotype, and NTG zebrafish may be used as a suitable control for further experimentation.

In total, approximately 500 zebrafish were injected with this RNA+DPR transgene construct. These zebrafish were referred to by the arbitrary term '2.2'. These zebrafish will be henceforth be referred to as 2.2 zebrafish lines. From those injected, 7 highly chimeric 2.2 zebrafish were identified. Chimerism was determined by visually examining which tissue types expressed the transgene (according to visible GFP or DsRed fluorescence). Zebrafish which showed transgene expression in at least muscle, heart and eye tissue (indicative of early transgene integration into embryos) were considered to be highly chimeric. These highly chimeric 2.2 zebrafish were grown to adulthood, then bred, and their offspring were screened for transgene expression by first heat shocking (see methods section for details) and then examined for DsRed and GFP fluorescence under a fluorescence microscope. Of the 7 highly chimeric zebrafish, 3 gave rise to a proportion of full transgenic offspring. However, 1/3 exhibited extremely high transgene expression levels, and the offspring died within 5 dpf, thus preventing the establishment of a full transgenic line. The remaining 2/3 highly chimeric zebrafish gave rise to full transgenic offspring which were grown to adulthood and used to establish a full transgenic line. The 2 separate zebrafish lines generated using the RNA+DPR transgene were arbitrarily named 2.2-2 and 2.2-7. A few embryos from 2.2-2 and 2.2-7 zebrafish lines were grown to 10 dpf and then processed for in situ hybridisation to test whether RNA foci from the C9orf72 expansion could be observed. In a blinded analysis, 50% (11/22) of muscle nuclei from 2.2-7 zebrafish contained foci, 30% (6/20) of 2.2-2 zebrafish nuclei contained foci and only 4% (1/25) of NTG zebrafish nuclei contained foci (**Figure 4.1b+c**). It is presumed, that the single focus observed in the NTG zebrafish was due to non-specific binding of the in situ probe. For 2.2-2 and 2.2-7 zebrafish all foci observed were localised to the nucleus and only a single focus per nucleus was observed. A limitation of the RNA foci data is that zebrafish were obtained from a single clutch, ideally 3 or more zebrafish clutches should be studied so that variation in background genetics may be accounted for and statistical comparison may be applied. As it had now been confirmed that 2.2 zebrafish express RNA foci, the next step was to test whether these zebrafish also produced DPR species.

4.4.　Multiple DPR species and stress response activation can be detected in embryonic 2.2 zebrafish

The various DPR species are known to have differential toxicity, with arginine rich species being considered the most toxic. To investigate whether there is a relationship between molecular weight (MW) and species toxicity, western blot was performed on 5 dpf 2.2-2 and 2.2-7 zebrafish lysates. The transgene construct expressed in both 2.2-zebrafish lines causes the production of GFP tagged DPR proteins via canonical ATG (start codon)

dependent translation, these GFP tagged DPRs are produced from C_4G_2 transcripts and can be detected at 5 dpf (**Figure 4.2a**, top panel). Interestingly, multiple GFP bands are detected in both the 2.2-2 and 2.2-7 zebrafish lines, and these bands were often unique to specific zebrafish lines. It was noted that, across multiple clutches of 5 dpf zebrafish embryos (>10 different clutches) that 2.2-2 zebrafish express 5 major GFP tagged DPR bands with 3 being common (3 lowest MW) and 2 being unique (2 highest MW), and 2.2-7 zebrafish express 6 major GFP tagged DPR bands with 3 being common and 3 unique (3 highest MW). The differential expression of DPRs between 2.2-2 and 2.2-7 zebrafish also holds true when probing for the DPR proteins directly (i.e. using antibodies which directly bind the dipeptide motifs). Poly(PA), poly(PR) and poly(GP) bands are each detectable at 5 dpf and are differentially expressed between 2.2-2 and 2.2-7 zebrafish lines (**Figure 4.2a**, middle three panels). Probing with poly(PA), poly(PR) and poly(GP) antibodies also revealed that some DPR bands detected did not co-localise with any of the ATG-dependent translation bands detected when using the GFP antibody, thus suggesting that these bands are likely to be produced via non-canonical RAN translation (**Figure 4.2a**, middle panels marked with asterix). Poly(GP) and poly(PR) antibodies were both purified from rabbit serum, however specific bands could be distinguished based on molecular weight.

In addition to the poly-PA, PR and GP DPRs produced from the antisense (C_4G_2) RNA transcripts, we were also able to detect poly(GA) DPR produced from the sense (G_4C_2) RNA transcript, however poly(GA) was only detected in the 2.2-7 zebrafish line (**Figure 4.2b**, top panel). Poly(GA) immunoblotting was carried on a separate PVDF membrane to GFP immunoblotting, in order to avoid cross reactivity (both poly(GA) and GFP antibodies are purified from mouse serum). The detection of poly(GA) indicates that bidirectional transcription of the C9orf72 expansion is occurring from our transgene. As the transcription of the RNA transcript containing the sense (G_4C_2) expansion is not driven by a conventional promoter region, this strongly suggests that poly(GA) protein is indeed produced via RAN translation.

The binding specificity of the poly(PR) and poly(GP) antibodies used here has previously been validated using an enzyme-linked immunosorbent assay to demonstrate that each antibody binds its target DPR antigen, and exhibits minimal cross-reactivity to all other DPR species (Davidson et al., 2016). Additionally, both poly(PR) and poly(GP) antibodies have been previously used to detect DPR species in *C9orf72*-ALS post mortem tissue samples (Davidson et al., 2016). Validation of the binding specificities of both poly(GA) and poly(PA) antibodies is discussed in section 3.5.
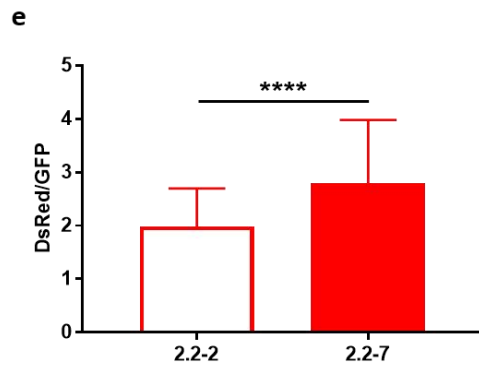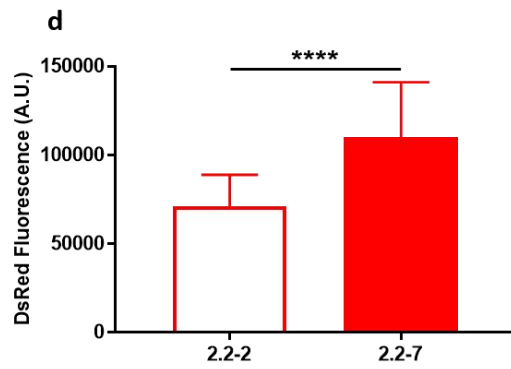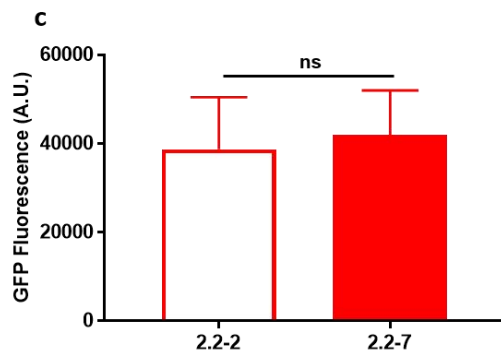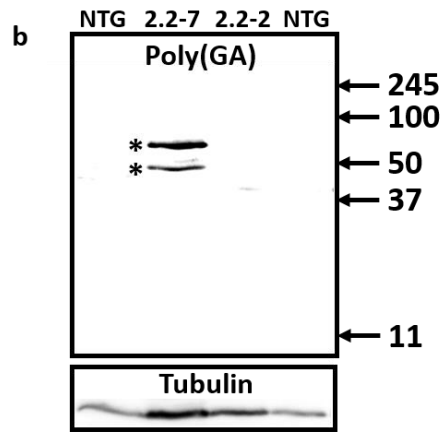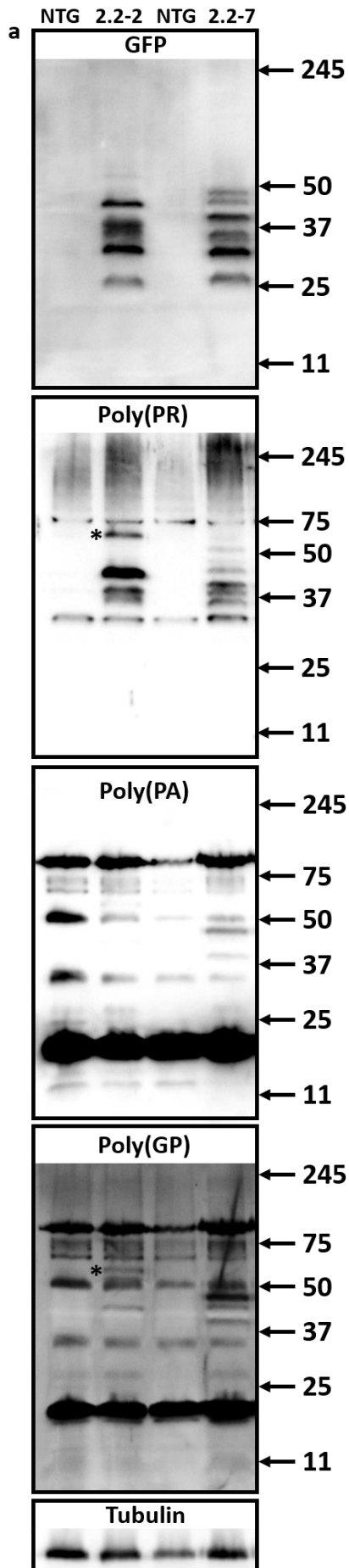
Figure 4.2: Both 2.2 zebrafish lines produce multiple DPR species and 2.2-7 zebrafish produces more DsRed

(**a**) Anti-sense DPRs (predominately ATG-driven) detected in 5dpf embryonic lysates. Asterix (*) denotes protein bands which are proposed to have been produced via RAN-translation. (**b**) Sense Poly(GA) DPRs (exclusively RAN-translation driven) detected in 5dpf embryonic lysate of 2.2-7 zebrafish. Asterix (*) denotes protein bands which are proposed to have been produced via RAN-translation. (**c**) GFP fluorescence at 5dpf does not differ between 2.2-2 and 2.2-7 zebrafish. Statistical comparison carried out using an unpaired t-test. (**d**) DsRed fluorescence at 5dpf is higher in 2.2-7 zebrafish compared with 2.2-2 zebrafish. This dataset was not normally distributed according to the Shapiro-Wilk test for normality, and was therefore statistically compared using the Mann-Whitney U test. (**e**) The ratio of DsRed/GFP at 5dpf is higher in 2.2-7 zebrafish compared with 2.2-2 zebrafish. This dataset was not normally distributed according to the Shapiro-Wilk test for normality, and was therefore statistically compared using the Mann-Whitney U test. N=75 2.2-2 and 76 2.2-7 individual zebrafish for all comparisons. All data are shown as mean +/- standard deviation; ****P < 0.0001 and ns: not significant.

Due to the presence of multiple bands of variable strength, quantifying transgene expression in each 2.2 zebrafish line by immunoblotting for GFP would have proven difficult. Instead individual embryos of each 2.2 zebrafish line were lysed and GFP and DsRed fluorescence were measured using a plate reader. The GFP fluorescence of 2.2-7 and 2.2-2 zebrafish was not significantly different ($4.2x10^4$ vs. $3.9x10^4$ A.U. respectively; **Figure 4.2c**), suggesting similar transgene expression levels in both 2.2 zebrafish lines. However, despite similar GFP expression levels in both 2.2 zebrafish, DsRed fluorescence of 2.2-7 zebrafish was significantly higher than that of the 2.2-2 zebrafish ($1.1x10^5$ vs. $0.71x10^5$ A.U. respectively; **Figure 4.2d**). Furthermore, the DsRed:GFP ratio of 2.2-7 zebrafish was significantly higher than that of 2.2-2 zebrafish (2.8 vs. 2.0 respectively; **Figure 4.2e**).

Embryonic expression of GFP and DsRed in 2.2 zebrafish, was also confirmed by confocal imaging of endogenous fluorescence in 5 dpf cryosections (**Figure 4.3**). GFP expression appeared to mainly localise to the nucleus of cells, although GFP and DAPI co-localisation was not quantified. In order to assess whether DPR/RNA foci expression leads to toxicity in 2.2 zebrafish, phenotypic characterisation of both 2.2 lines was next carried out.
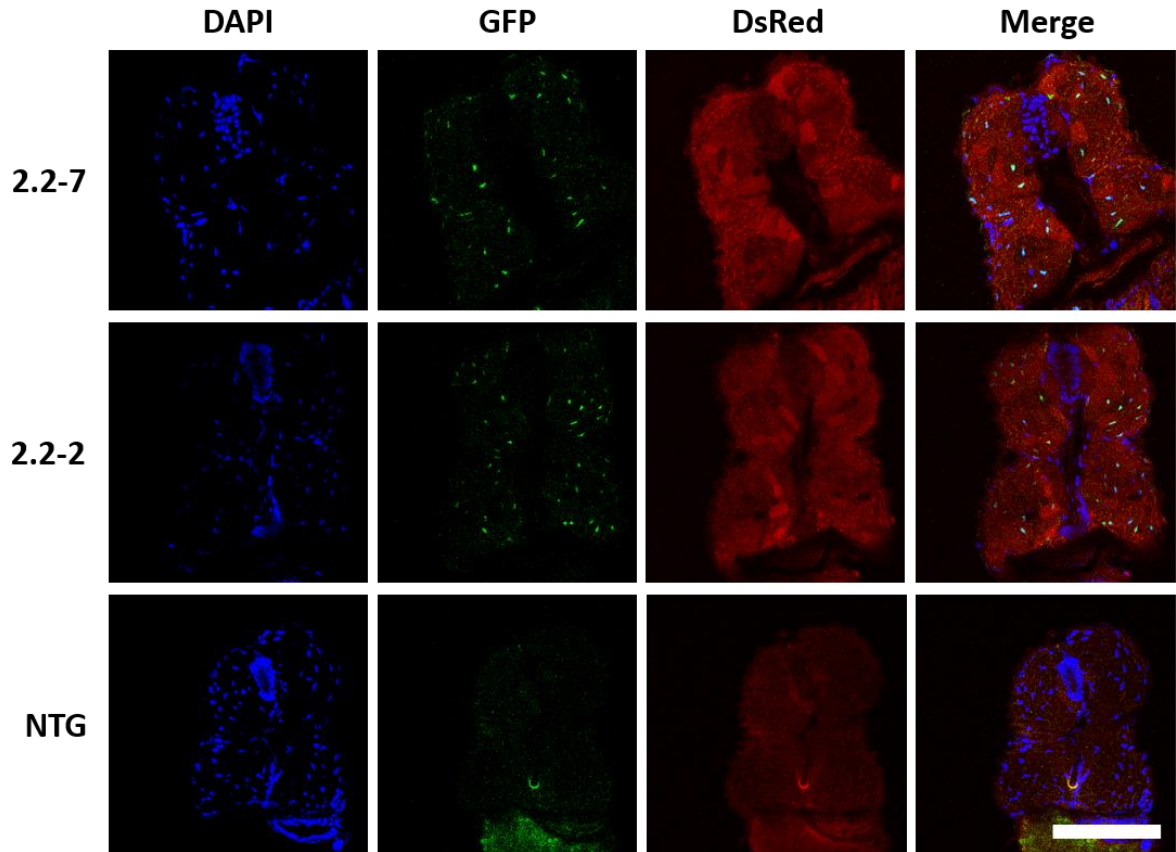
Figure 4.3: 2.2 zebrafish produce GFP and DsRed at 5dpf

At 5 dpf, both 2.2-7 and 2.2-2 zebrafish produce predominately nuclear localised GFP and diffusely expressed DsRed. Scale bar = 100µm. Transverse body sections. Images are representative of 9 zebrafish imaged (not quantified).

4.5.   Embryonic 2.2-7 zebrafish show increased spontaneous locomotor activity

Neither 2.2-zebrafish lines showed any overt morphological abnormalities during embryonic development (0- 5 dpf). To assess whether RNA foci and DPR pathology was causing any underlying motor deficits, the spontaneous locomotor activity of 2.2 zebrafish was monitored at 5 dpf. Embryonic 2.2 zebrafish were monitored in 96 well plates using the Viewpoint behaviour monitoring setup. Under dark conditions, the average speed of movement of 2.2-7 zebrafish was significantly higher than that of both 2.2-2 and NTG zebrafish, over the 20 minute recording period (2.1 vs 1.7 and 1.8 mm/s average speed for 2.2-7, 2.2-2 and NTG zebrafish respectively; **Figure 4.4a**). Similarly, under dark conditions the total distance moved by 2.2-7 zebrafish was significantly higher than that of both 2.2-2 and NTG zebrafish, over the 20 minute recording period (2440 vs 1950 and 1940 millimetres moved for 2.2-7, 2.2-2 and NTG zebrafish respectively; **Figure 4.4b**).
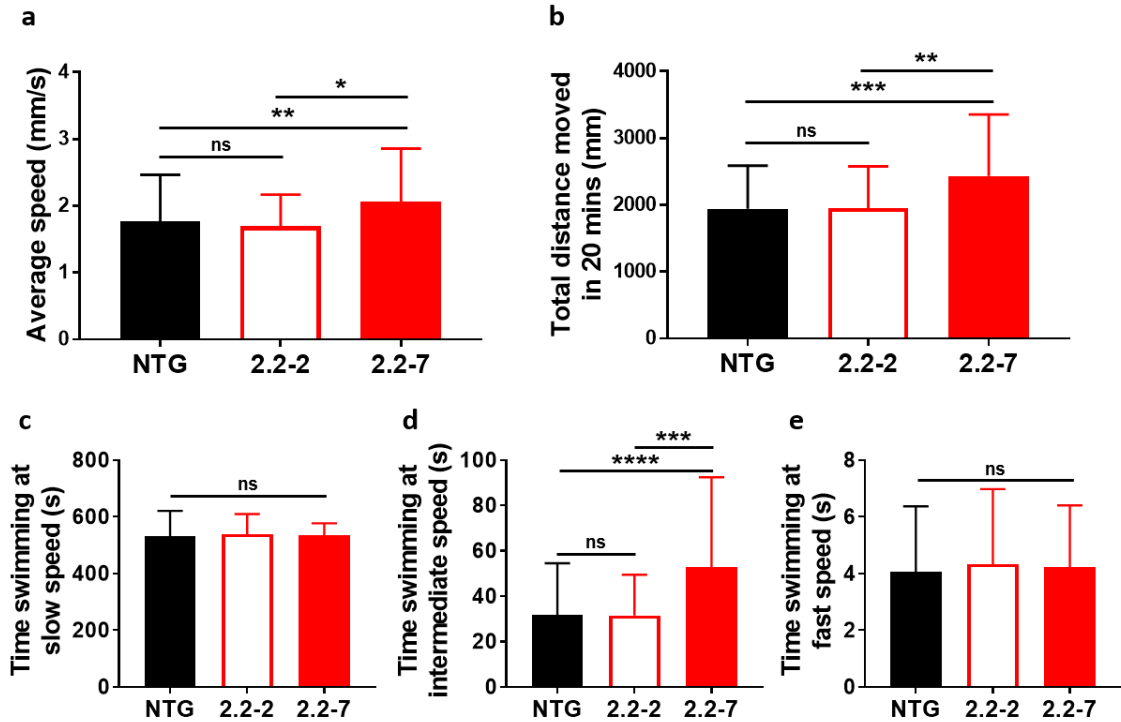
Figure 4.4: 2.2-7 zebrafish show increased average speed and total distance moved at 5dpf

(**a**) The average speed of 2.2-7 zebrafish is significantly higher than that of both 2.2-2 and NTG zebrafish. (**b**) The total distance moved over a 20 minute recording period is significantly higher in 2.2-7 zebrafish as compared to both 2.2-2 and NTG zebrafish. (**c**) There is no significant difference in the time spent swimming at slow speed between NTG, 2.2-2 and 2.2-7 zebrafish over a 20 minute recording period. (**d**) 2.2-7 zebrafish spent significantly more time swimming at intermediate speeds in comparison to both 2.2-2 and NTG zebrafish over a 20 minute recording period. (**e**) There is no significant difference in the time spent swimming at fast speed between NTG, 2.2-2 and 2.2-7 zebrafish over a 20 minute recording period. N=144 NTG, 71 2.2-2 and 60 2.2-7 individual fish per genotype for all comparisons, zebrafish were derived from three separate clutches. All datasets were not normally distributed according to the Shapiro-Wilk test for normality, and were therefore statistically compared using the Kruskal-Wallis test with Dunn's post-hoc test. Speed thresholds used were slow ($0 < x < 5$ mm/sec), intermediate ($5 < x < 15$ mm/sec) and fast ($x > 15$ mm/sec). All data were recorded under dark conditions and are shown as mean +/- standard deviation; ns: not significant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$.

In order to analyse swimming behaviour in more detail, zebrafish swimming data were subdivided into time spent swimming at slow, intermediate and fast speeds. No significant difference was detected in the time spent swimming at slow speed between NTG, 2.2-2 and 2.2-7 zebrafish (**Figure 4.4c**). However, 2.2-7 zebrafish spent significantly more time swimming at intermediate speeds in comparison to 2.2-2 and NTG zebrafish (53 vs 32 and 32 seconds for 2.2-7, 2.2-2 and NTG zebrafish respectively; **Figure 4.4d**). No significant difference was observed in the time spent swimming at fast speeds between 2.2-7, 2.2-2 and NTG zebrafish (**Figure 4.4e**). Thus, at the embryonic stage (5dpf) 2.2-7 zebrafish show signs of hyperactivity. We next set out to determine whether other behavioural or developmental abnormalities were present in either of the 2.2 zebrafish lines during embryonic and larval development.

4.6. Early behavioural abnormalities, reduced viability and reduced body mass are detected in 2.2-7 zebrafish

As expansions in human ALS cause a spectrum of both motor and cognitive deficits, we examined whether normal zebrafish behaviour was affected in 2.2-7 zebrafish at 5 dpf. Centre avoidance behaviour assays are a validated means of measuring willingness to explore in zebrafish (Schnorr et al., 2012), and are comparable to the open field test performed in mice. It was determined that 2.2-7 zebrafish were significantly less likely to venture into the centre of the well when compared to their NTG clutchmates (27% of 2.2-7 in centre vs. 34% of NTG zebrafish in centre; **Figure 4.5a+b**).

To determine if the early embryonic expression of RNA foci and DPR impacted upon the viability of the zebrafish, we carried out early (1-15 dpf) survival analysis. Heterozygous 2.2-2 zebrafish did not show any change in survival within 15 dpf as compared to NTG zebrafish (data from NTG clutchmates of all genotypes are pooled; **Figure 4.5c**). In contrast, heterozygous 2.2-7 zebrafish did show a significant decrease in survival within 15 dpf as compared to NTG zebrafish (79% vs. 92% survival for 2.2-7 and NTG respectively at 15dpf; **Figure 4.5c**).

It was noted that during early development, the 2.2-7 zebrafish appeared smaller than their NTG clutchmates. At 30 dpf there was a significant decrease in total body mass of 2.2-7 zebrafish compared to their NTG clutchmates (1.7mg (21%) vs. 9.9mg (100%) mean body mass for 2.2-7 and NTG respectively; **Figure 4.6a**). However, 2.2-2 zebrafish did not show a significant difference in body mass as compared to their own clutchmates at the same age (**Figure 4.6b**). In summary, 2.2-7 zebrafish but not 2.2-2 zebrafish, showed significant reduction in survival at 15 dpf, and reduction in body mass at 30 dpf. At 5 dpf, 2.2-7 zebrafish also displayed signs of atypical behaviour as measured by centre avoidance.
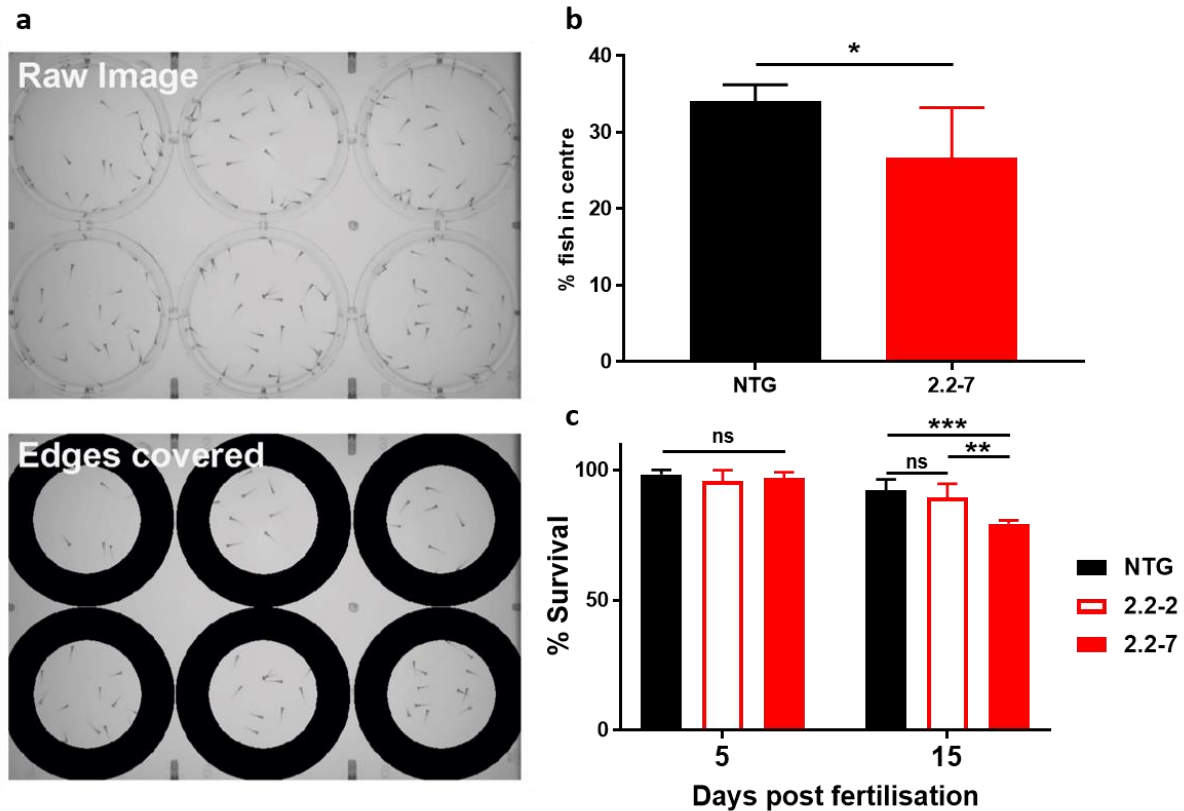
Figure 4.5: 2.2-7 zebrafish show early behavioural defects and reduced viability

(**a**) Representative images of the plate set-up used to monitor centre avoidance behaviour in zebrafish. 30 fish were placed in each well and one image every minute was analysed. Image shown as recorded (top) and then following removal of the region around the edge of the plate for analysis (bottom). In these representative images 2.2-7 zebrafish are placed across the top 3 wells and NTG zebrafish across the bottom 3 wells. (**b**) Quantification of centre avoidance behaviour showing 2.2-7 zebrafish are significantly less often found in the plate centre. N=6 clutches per genotype. Datasets were found to have unequal variance according to the F test, and were therefore statistically compared using a t-test with Welch's correction. (**c**) Survival of zebrafish did not change by 5dpf, however by 15dpf survival of the 2.2-7 line was significantly reduced compared to NTG and 2.2-2. N=4 clutches per genotype. Datasets were statistically compared using a two-way ANOVA with Sidak's post hoc test. All data are shown as mean +/- standard deviation; ns: not significant, $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$.
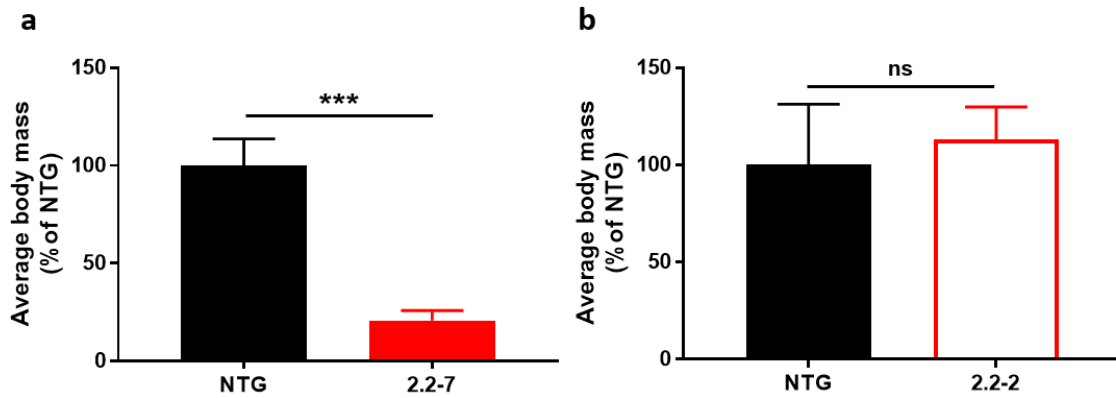
Figure 4.6: 2.2-7 zebrafish have reduced body mass at 30dpf

(**a**) At 30dpf 2.2-7 zebrafish have reduced average body mass in comparison to their NTG clutch mates. N=3 clutches per genotype (48 x 2.2-7 and 79 x NTG zebrafish in total). Statistical comparison was carried out using an unpaired t-test. (**b**) At 30dpf there was no difference in average body mass between 2.2-2 zebrafish and their NTG clutch mates. N=3 clutches per genotype (82 x 2.2-2 and 79 x NTG zebrafish in total). Statistical comparison was carried out using an unpaired t-test. All data are shown as mean +/- standard deviation; ns: not significant and ***P < 0.001.

The increased phenotypic severity of 2.2-7 zebrafish is consistent with their increased hsp70 promoter mediated DsRed production, as compared with 2.2-2 zebrafish. Transgene expression, DPR pathology and the motor/behavioural phenotype have now been characterised in developing zebrafish. The next step was therefore to characterise the expression of the transgene and associated phenotype in adult 2.2 zebrafish.

4.7.   Adult 2.2 zebrafish express nuclear localised DPR proteins

Previous reports indicate that DPR species often localise to the nucleus. This is particularly true for the arginine rich DPRs poly(PR) and poly(GR) (Mizielinska et al., 2014). To examine the cellular localisation of the DPR species produced in 2.2 zebrafish, immunofluorescence with DPR antibodies was carried out in adult zebrafish muscle. The majority of poly(GP), poly(PA) and poly(PR) DPR signal was found to be localised to the nucleus (**Figure 4.7**, **Figure 4.8** and **Figure 4.9**). The 2.2 zebrafish express interrupted C9orf72 expansions, meaning that runs of pure C_4G_2 are linked together with 5 non-GC base pairs. As there are 5 non-GC base pairs, this effectively shifts the reading frame for subsequent C_4G_2 expansions. Therefore, most individual DPR species produced in the 2.2 zebrafish will likely contain a mixture of multiple DPR species, as can be seen in figure 4.2a, in which the same protein band can be detected with multiple DPR antibodies. This is likely the reason why
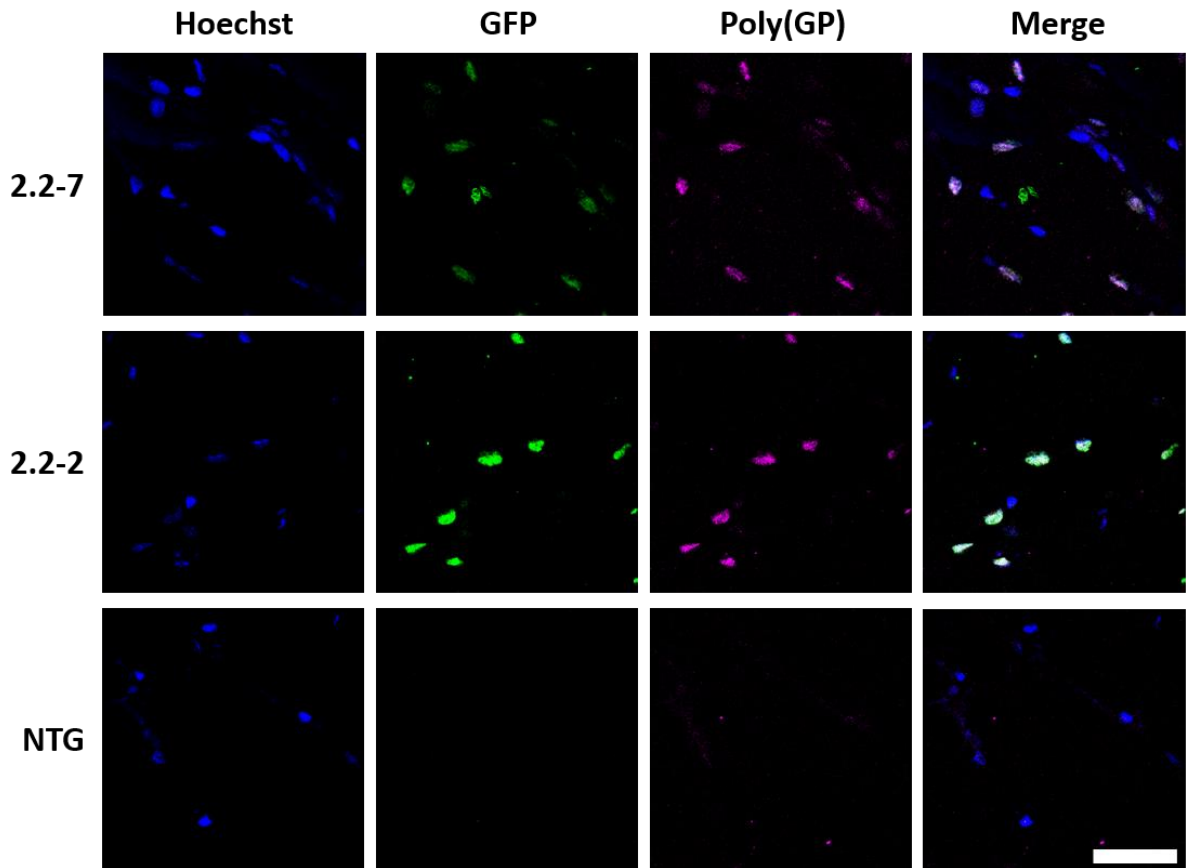
Figure 4.7: Both 2.2 model zebrafish express nuclear localised poly(GP) DPRs

GFP and poly(GP) immunostaining of adult (12 months old) zebrafish muscle tissue shows that GFP tagged DPRs and poly(GP) containing DPRs localise to the nucleus in 2.2-2 and 2.2-7 zebrafish. For all DPR images nuclei are stained with Hoechst (blue), GFP is stained with GFP antibody (green) and DPR proteins are stained with the poly(GP) antibody 24494-1-AP (purple). Scale bar = 25µm.
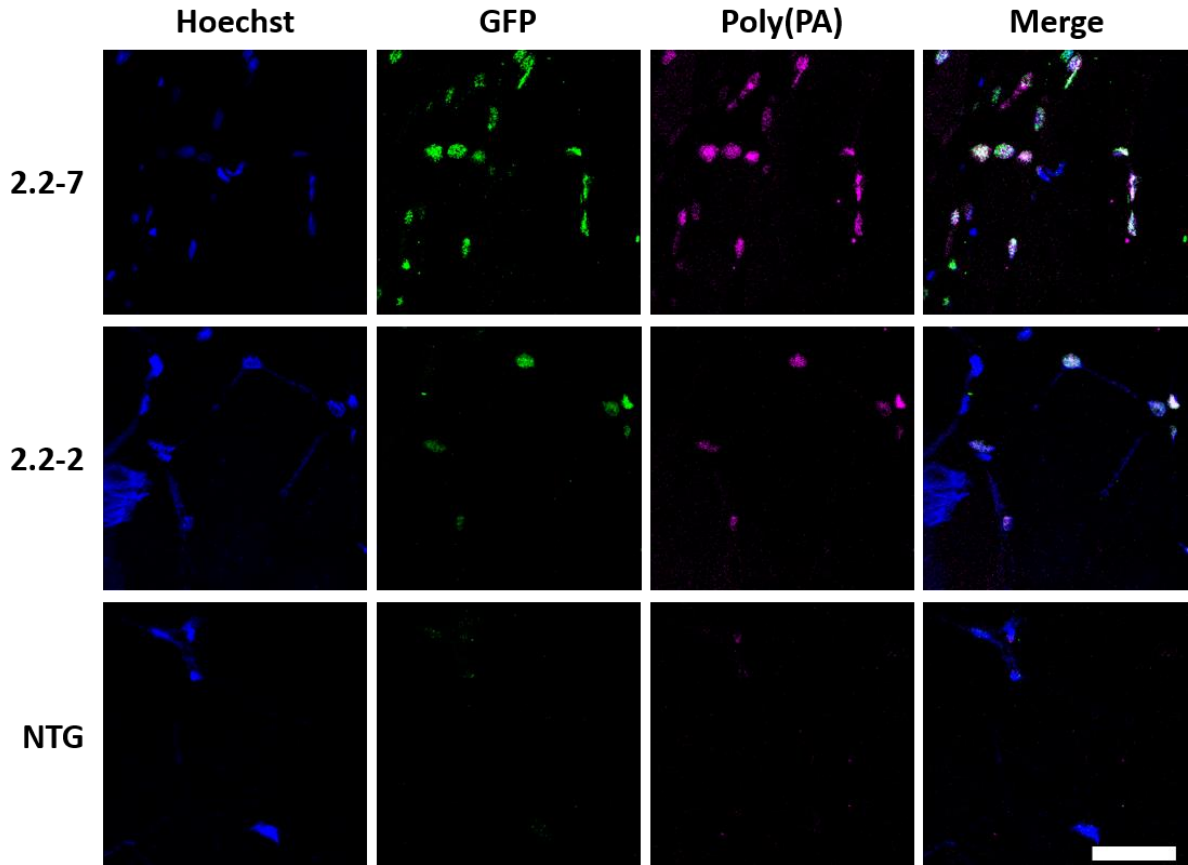
Figure 4.8: Both 2.2 model zebrafish express nuclear localised poly(PA) DPRs

GFP and poly(PA) immunostaining of adult (12 months old) zebrafish muscle tissue shows that GFP tagged DPRs and poly(PA) containing DPRs localise to the nucleus in 2.2-2 and 2.2-7 zebrafish. For all DPR images nuclei are stained with Hoechst (blue), GFP is stained with GFP antibody (green) and DPR proteins are stained with the poly(PA) antibody MABN1790 (purple). Scale bar = 25µm.
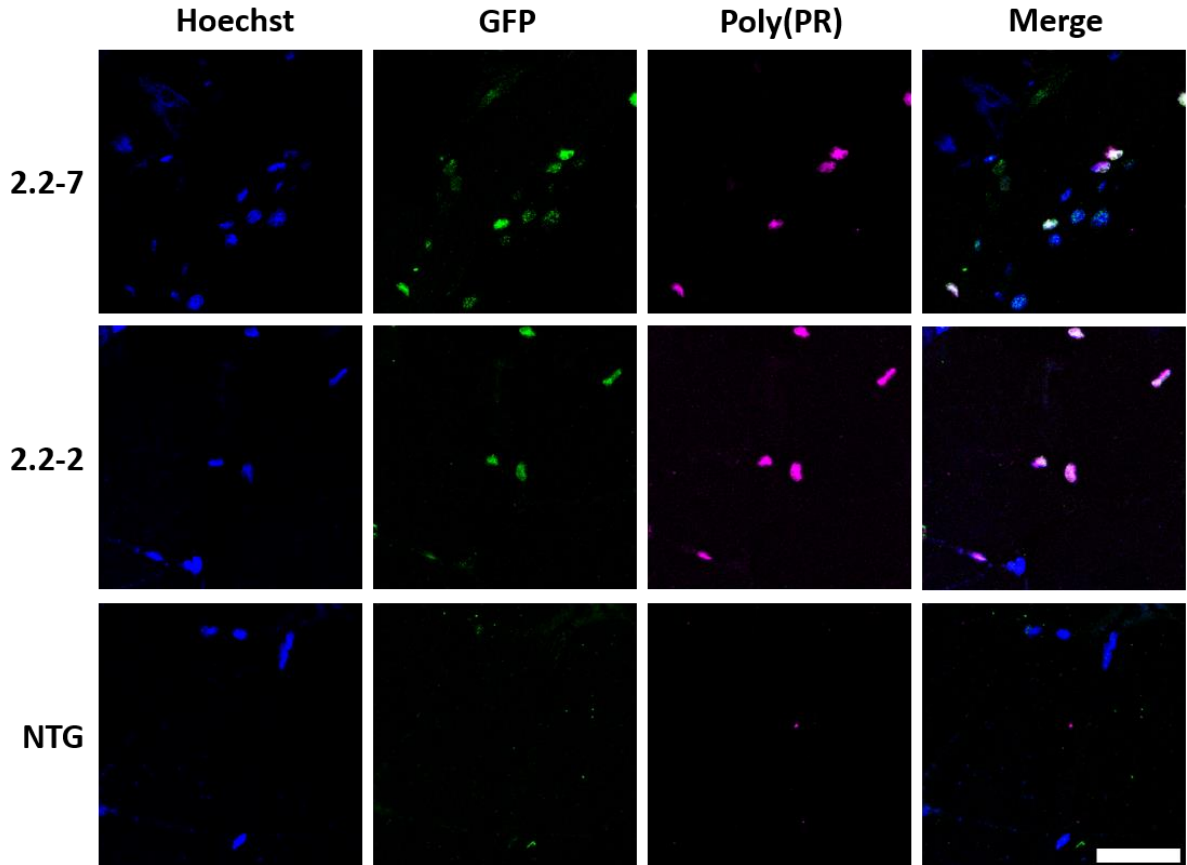
Figure 4.9: Both 2.2 model zebrafish express nuclear localised poly(PR) DPRs

GFP and poly(PR) immunostaining of adult (12 months old) zebrafish muscle tissue shows that GFP tagged DPRs and poly(PR) containing DPRs localise to the nucleus in 2.2-2 and 2.2-7 zebrafish. For all DPR images nuclei are stained with Hoechst (blue), GFP is stained with GFP antibody (green) and DPR proteins are stained with the poly(PR) antibody 23979-1-AP (purple). Scale bar = 25µm.

signal for multiple DPR species are all localised to the nucleus in 2.2 zebrafish. Based on previous reports, the nuclear localisation is most likely mediated by poly(PR) DPR proteins (Wen et al., 2014). DPR expression in muscle tissue from 2.2 zebrafish had now been confirmed. However, in ALS expression of C9orf72 expansions is primarily associated with neurodegeneration. Therefore, we next planned to test whether transgene expression in 2.2 zebrafish was also present in central nervous system tissues.

4.8.   DPR species are produced in 2.2 zebrafish CNS tissues

In human ALS, toxicity occurs primarily in cells of the central nervous system (CNS), and so it is essential to ascertain whether DPR species are also produced within the CNS of this zebrafish model. In adult spinal cord and brain of both 2.2-zebrafish lines, GFP-tagged DPR species and DPR species which were not immunoreactive with GFP antibodies, could be detected (**Figure 4.10a+b**). This suggests that both ATG-dependent translation and RAN translation of DPR species occurs within the CNS of the 2.2-zebrafish. DPR bands which are not immunoreactive for GFP and are therefore proposed to be produced via RAN-translation are marked with an asterix. In both 2.2-zebrafish lines, the band pattern of DPRs detected largely remains constant from 5 dpf until adulthood, although higher molecular weight (>50KDa) poly(PR) positive bands are more abundant in adult tissue. Of all the DPR species examined here, poly(PR) generally has the highest propensity to form high MW RAN-translation mediated bands. Poly(GP) and poly(GA) immunoblotting was not carried out due to the presence of multiple specific and non-specific bands already produced by antibodies derived from the same species. Poly(GP) and poly(PR) were both purified from rabbit serum, and poly(GA) and GFP antibodies were both purified from mouse serum. As multiple DPR species are produced in the CNS of adult 2.2 zebrafish, we next planned to investigate whether there was an adult onset phenotype in these zebrafish.

Figure 4.10: Both 2.2 zebrafish lines produce multiple DPR species in brain and spinal cord

(**a**) Anti-sense DPRs (predominately ATG-driven) detected in adult spinal cord lysates. Asterix (*) denotes protein bands which are proposed to have been produced via RAN-translation. (**b**) Anti-sense DPRs (predominately ATG-driven) detected in adult brain lysates. Asterix (*) denotes protein bands which are proposed to have been produced via RAN-translation.

4.9.   Reduced swimming endurance is observed in adult 2.2-7 zebrafish

To assess the neuro-muscular integrity of the 2.2-7 transgenic zebrafish, swimming endurance was tested using a swim tunnel, the aquatic equivalent to a treadmill. Due to equipment availability, the swim tunnel setup used previously to test the 5.3 zebrafish was not available when testing 2.2 zebrafish. Instead, this swim tunnel setup had a wider swimming chamber and so was not able to fully exhaust every zebrafish tested. Flow into the swimming chamber began at 2L/min and then was increased in 2L/min increments every 5 minutes until 11.6L/min (maximum flow rate for this setup) had been reached, at which point the experiment was halted. For this reason, direct comparisons cannot be drawn between 5.3 zebrafish and 2.2 zebrafish swim tunnel performance. Despite the swim tunnel not being able to fully exhaust every zebrafish, at 9 months of age 2.2-7 zebrafish showed significantly reduced swimming endurance as compared to their NTG clutchmates (38% vs. 77% still swimming at maximum flow rate for 2.2-7 and NTG respectively; **Figure 4.11a**). Zebrafish size can affect swimming ability, therefore the body length and body mass of each zebrafish used in the swim tunnel was measured. Despite decreased body mass during early development, neither variable was significantly different between 2.2-7 and NTG clutch mates at 9 months of age (**Figure 4.11b+c**). The swim tunnel test was repeated using the same experimental setup and the same cohort of zebrafish three months later (zebrafish were now 12 months old). Again, 2.2-7 zebrafish showed significantly reduced swimming endurance compared to their NTG clutchmates (23% vs. 85% still swimming at maximum flow rate for 2.2-7 and NTG respectively; **Figure 4.11d**). Furthermore, neither body length nor body mass was significantly different between transgenic and NTG groups at 12 months old (**Figure 4.11e+f**). Immediately, following removal from the swim tunnel zebrafish were transferred to a recovery tank where their spontaneous locomotor activity was monitored.

9 months old   **12 months old**
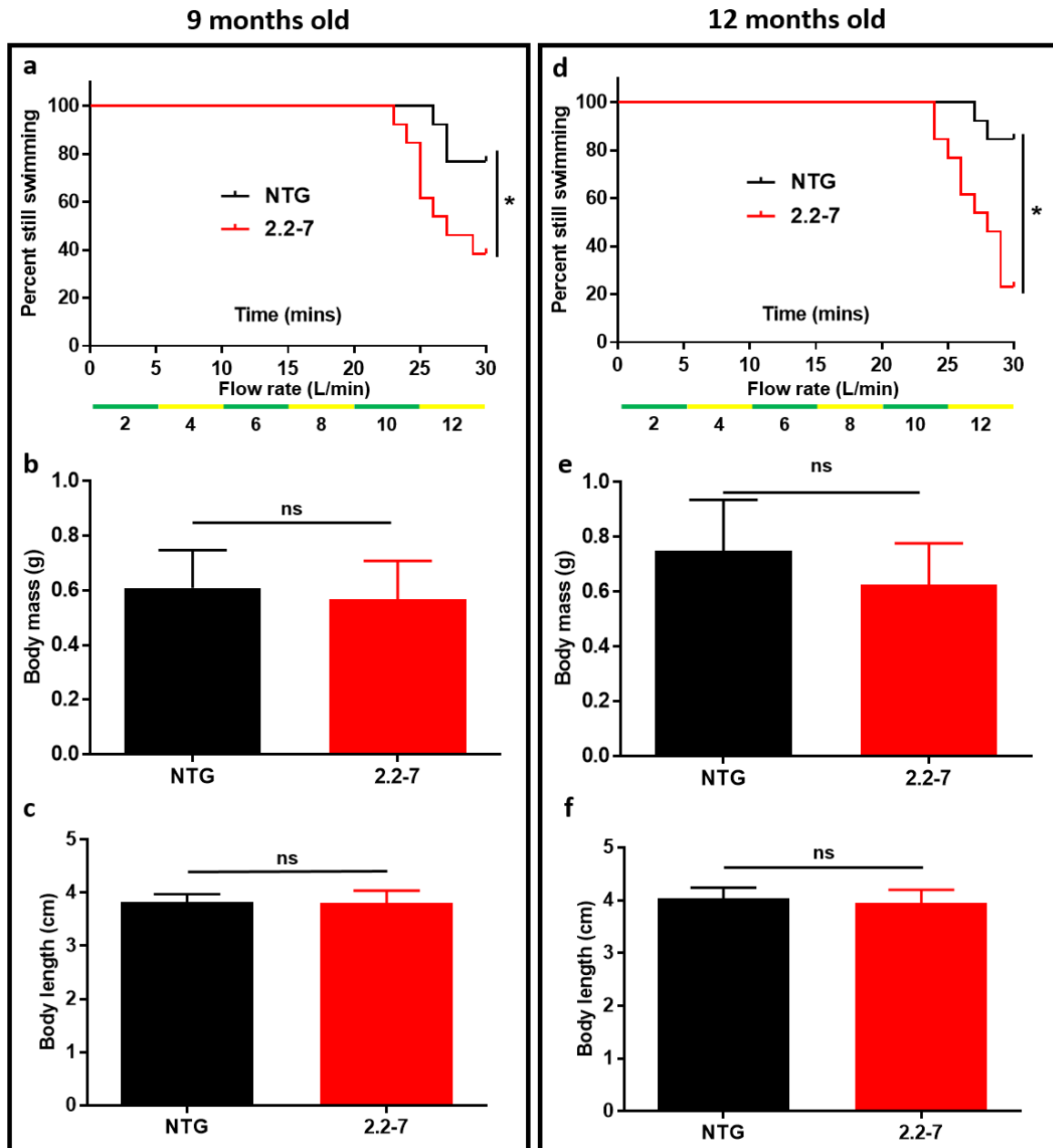
Figure 4.11: 2.2-7 zebrafish show adult onset swimming endurance deficits

(**a**) At nine months old, 2.2-7 transgenic zebrafish failed to continue swimming at earlier time points than their NTG clutch mates. N=13 fish per genotype. Statistical comparison was carried out using the Log-rank test. (**b**) At nine months old there was no difference in body mass between 2.2-7 zebrafish and their NTG clutchmates. N=13 fish per genotype. Statistical comparison was carried out using an unpaired t-test. (**c**) At nine months old there was no difference in body length between 2.2-7 zebrafish and their NTG clutchmates. N=13 fish per genotype. Statistical comparison was carried out using an unpaired t-test. (**d**) At twelve months old, 2.2-7 transgenic zebrafish failed to continue swimming at earlier time points than their NTG clutch (**legend continues on next page**)

mates. N=13 fish per genotype. Statistical comparison was carried out using the Log-rank test. (**e**) At twelve months old there was no difference in body mass between 2.2-7 zebrafish and their NTG clutchmates. N=13 fish per genotype. Statistical comparison was carried out using an unpaired t-test. (**f**) At twelve months old there is no difference in body length between 2.2-7 zebrafish and their NTG clutchmates. N=13 fish per genotype. Statistical comparison was carried out using an unpaired t-test. All body mass and body length data are shown as mean +/- standard deviation; ns: not significant and *P < 0.05.

4.10.   2.2-7 zebrafish show normal locomotor activity following swim tunnel testing

Zebrafish were removed from the swim tunnel after testing, placed into a recovery tank and allowed 5 minutes to recover before spontaneous locomotor activity recordings began. At 9 months of age there was no difference in average speed of movements made or the total distance moved by the zebrafish over a 30 minute recording period (**Figure 4.12a+b**). Similarly, at 12 months old there was no difference in average speed of movements made or the total distance moved by the zebrafish over a 30 minute recording period (**Figure 4.12c+d**).

In order to analyse swimming behaviour in more detail, zebrafish swimming data were subdivided into time spent swimming at slow, intermediate and fast speeds. No significant difference was observed in the time spent swimming at slow, intermediate or fast speed between 9 month old 2.2-7 and NTG zebrafish (**Figure 4.13a-c**). Similarly, no significant difference was observed in the time spent swimming at slow, intermediate or fast speeds between 12 month old 2.2-7 and NTG zebrafish (**Figure 4.13d-f**). Following the swim tunnel/locomotor testing of the 2.2-7 zebrafish, it was noted that some of the 2.2-7 zebrafish were rapidly losing body muscle (no zebrafish displaying this phenotype were included in the swim tunnel/locomotor experiments).

9 months

12 months
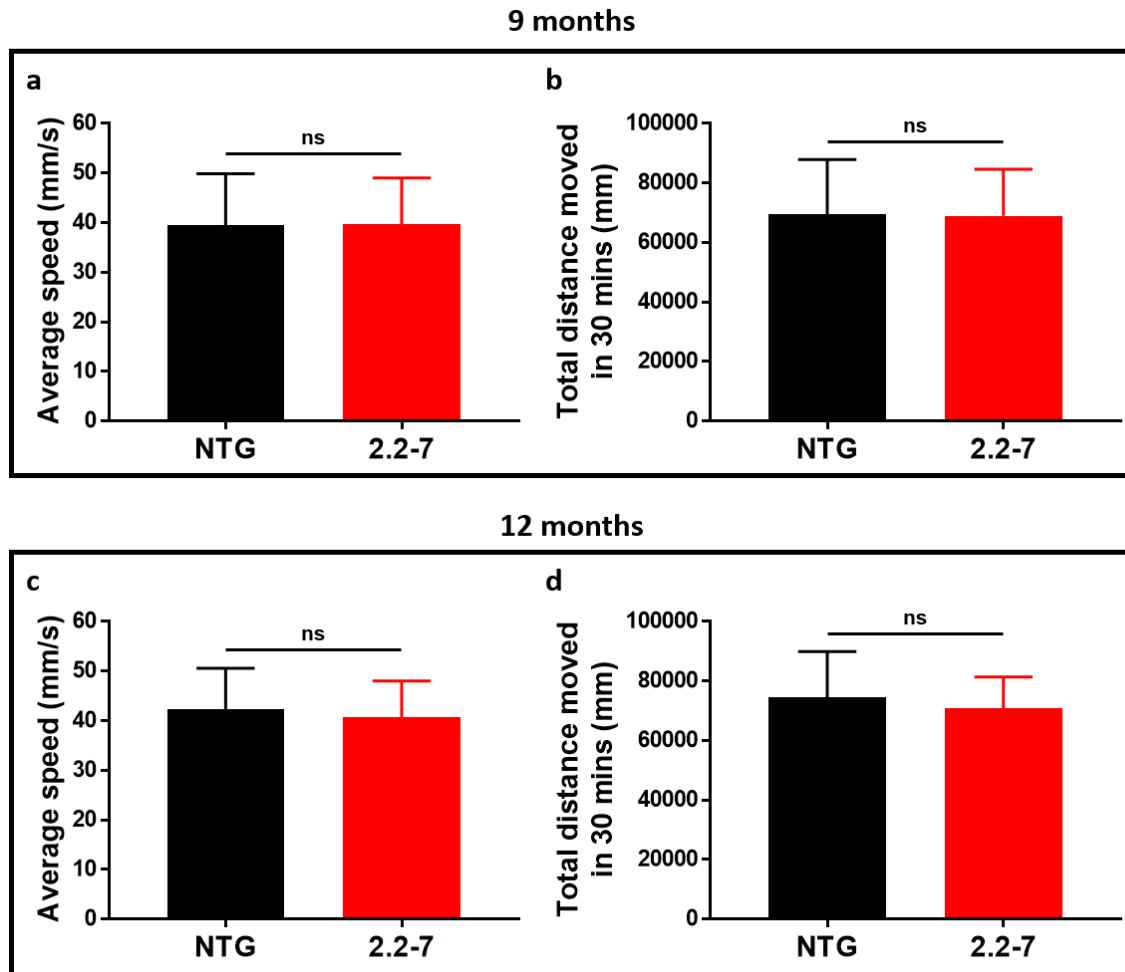
Figure 4.12: After swim tunnel, 2.2-7 zebrafish show normal swimming activity

(**a**) After being removed from the swim tunnel, 9 month old 2.2-7 fish did not show any significant difference in average speed of movements made over a 30 minute recording period in comparison to NTG zebrafish. N=12 fish per genotype. (**b**) After being removed from the swim tunnel, 9 month old 2.2-7 fish did not show any significant difference in total distance moved over a 30 minute recording period in comparison to NTG zebrafish. N=12 fish per genotype. (**c**) After being removed from the swim tunnel, 12 month old 2.2-7 fish did not show any significant difference in average speed of movements made over a 30 minute recording period in comparison to NTG zebrafish. N=13 fish per genotype. (**d**) After being removed from the swim tunnel, 12 month old 2.2-7 fish did not show any significant difference in total distance moved over a 30 minute recording period in comparison to NTG zebrafish. N=13 fish per genotype. All statistical comparisons were carried out using an unpaired t-test. All data are shown as mean +/- standard deviation; ns: not significant.
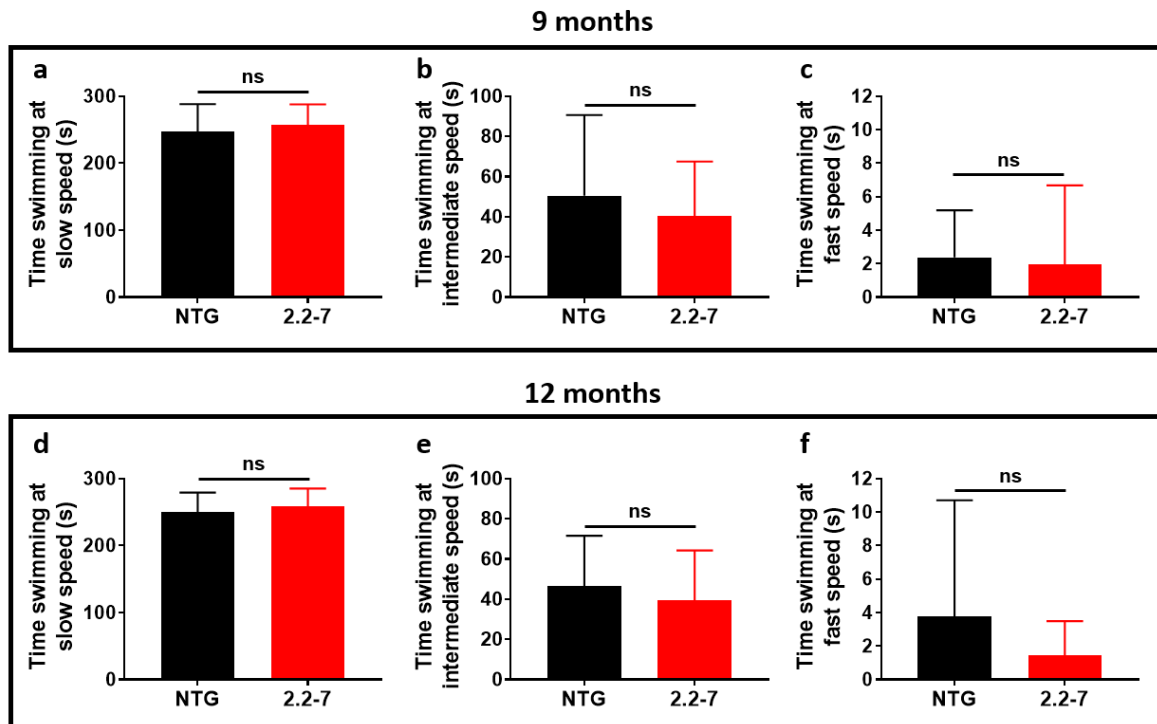
Figure 4.13: After swim tunnel, the time spent swimming at different speeds is not significantly different between 2.2-7 and NTG zebrafish

(**a**) There is no significant difference in the time spent swimming at slow speeds between 2.2-7 and NTG zebrafish at 9 months old. Statistical comparison was carried out using an unpaired t-test. N=12 fish per genotype. (**b**) There is no significant difference in the time spent swimming at intermediate speeds between 2.2-7 and NTG zebrafish at 9 months old. Statistical comparison was carried out using an unpaired t-test. N=12 fish per genotype. (**c**) There is no significant difference in the time spent swimming at fast speeds between 2.2-7 and NTG zebrafish at 9 months old. This dataset was not normally distributed according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. N=12 fish per genotype. (**d**) There is no significant difference in the time spent swimming at slow speeds between 2.2-7 and NTG zebrafish at 12 months old. Statistical comparison was carried out using an unpaired t-test. N=13 fish per genotype. (**e**) There is no significant difference in the time spent swimming at intermediate speeds between 2.2-7 and NTG zebrafish at 12 months old. Statistical comparison was carried out using an unpaired t-test. N=13 fish per genotype. (**f**) There is no significant difference in the time spent swimming at fast speeds between 2.2-7 and NTG zebrafish at 12 months old. This dataset was not normally distributed according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. N=13 fish per genotype. All data were recorded over a 30 minute period after zebrafish had been removed from the swim tunnel and allowed to rest for 5 minutes. Speed thresholds used were slow ($x<60$mm/sec), intermediate ($60<x<120$mm/sec) and fast ($x>120$mm/sec).Data are shown as mean +/- standard deviation; ns: not significant.

4.11.   Early mortality is observed in 2.2-7 zebrafish

While visually inspecting tanks containing 12 month old 2.2-7 zebrafish, it was noted that a small proportion of the 2.2-7 zebrafish showed marked loss of body muscle. These 2.2-7 zebrafish were then closely monitored, and it became clear that the loss of body mass was progressive and was beginning to interfere with the normal tank swimming behaviour of the 2.2-7 zebrafish. At this point specific criteria were decided upon in order to determine when a zebrafish had reached end-stage, and would need to be culled to prevent unnecessary suffering. A zebrafish was defined as having reached end-stage once it had lost the ability to maintain normal swimming (showing signs of paralysis) to the extent where it was no longer able to obtain food. End-stage 2.2-7 zebrafish displayed severe wasting in the body muscle region, had very poor locomotor skills and would often rest at the bottom of tank completely still (this is very abnormal for zebrafish who typically maintain almost constant swimming whilst awake). Over time more 2.2-7 zebrafish developed this progressive muscle wasting phenotype, and by 36 months there was a significant reduction in survival of the 2.2-7 zebrafish compared with that of NTG zebrafish (22% vs. 86% survival for 2.2-7 and NTG respectively at 36 months old; **Figure 4.14**). In contrast to 2.2-7, no NTG zebrafish displayed this muscle wasting phenotype. We next planned to assess adult muscle integrity of the 2.2 zebrafish using histological techniques.
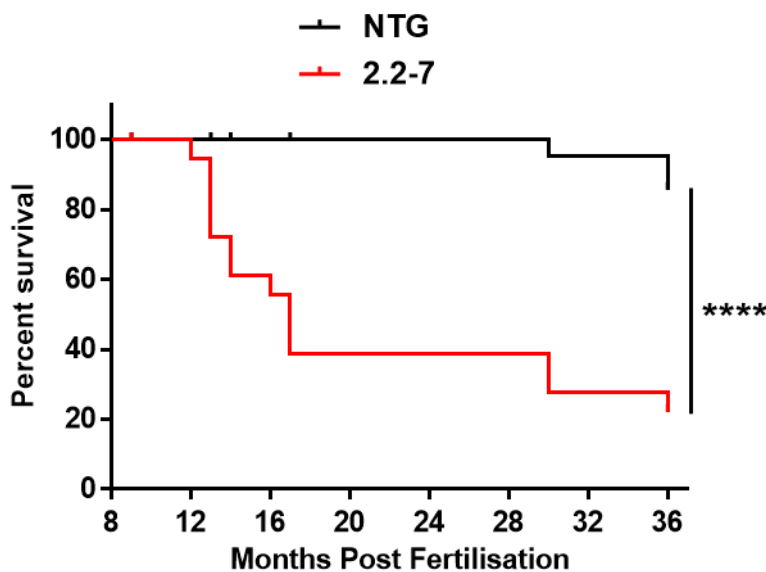
Figure 4.14: 2.2-7 zebrafish have reduced survival

Adult transgenic 2.2-7 zebrafish have reduced survival from 8-36 months in comparison to their NTG clutch mates which are housed in the same tank. N=17 2.2-7 and 27 NTG at 8 months. Statistical comparison was carried out using the Log-rank test. ****P < 0.0001.

4.12. Both 2.2 zebrafish show abnormal muscle histology

Progressive muscle atrophy is observed in all ALS patients and similar muscle atrophy was observed in end-stage 2.2-7 zebrafish. In order to investigate the nature and severity of this muscle atrophy, we H&E stained epaxial muscle sections from NTG, 2.2-2 and end-stage 2.2-7 zebrafish to examine their muscle structure. End-stage 2.2-7 zebrafish muscle displayed widespread severe atrophy, muscle fibres were sparse and disorganised (**Figure 4.15a**). The muscle of 2.2-2 zebrafish displayed more subtle changes. To assess 2.2-2 zebrafish, muscle integrity myotome size and number was counted from 2.2-2 and NTG epaxial muscle sections in a blinded analysis. Myotomes of 2.2-2 zebrafish muscle were significantly smaller and more numerous compared to NTG muscle myotomes (1.1 vs. 1.7mm^2 median myotome size for 2.2-2 and NTG respectively; **Figure 4.15a+b**). End-stage 2.2-7 zebrafish muscle fibre size and number was not quantified, as 2.2-7 myofibres were so severely disorganised that it was not possible to discern individual myotomes. It was now clear that end-stage 2.2-7 zebrafish have severely atrophied muscle tissue. We next planned to examine what the cause of muscle atrophy in these fish may have been.

4.13. 2.2-7 zebrafish show motor neuron loss

In ALS patients, the underlying molecular pathology ultimately leads to motor neuron death, which in turn leads to atrophy of denervated muscle. To investigate whether motor neuron loss underlies the muscle atrophy observed in 2.2-7 zebrafish, we counted motor neurons from the ventral spinal cord of NTG, 2.2-2 and end-stage 2.2-7 H&E stained zebrafish sections. Significant loss of motor neurons was observed in end-stage 2.2-7 zebrafish as compared with NTG controls (1.0 vs. 3.2 mean number of motor neurons per body section of 2.2-7 and NTG respectively; **Figure 4.16a+b**). A small, non-significant reduction in motor neurons was observed in 2.2-2 zebrafish as compared with NTG controls (**Figure 4.16a+b**).

As mentioned previously the 2.2 zebrafish lines were generated with the aim of producing an *in vivo* model which recapitulates key aspects of the ALS/FTD phenotype and can be used to screen potential neuroprotective compounds using hsp70 promoter driven DsRed as a cellular stress readout. As already described, the 2.2 zebrafish show behavioural and motor defects. Additionally, we have already shown that embryonic DsRed is higher in the phenotypically severe 2.2-7 zebrafish, compared with the phenotypically less severe 2.2-2 zebrafish. We next aimed to characterise how DsRed levels change over time in the CNS of 2.2-7 zebrafish as their phenotype becomes more severe.
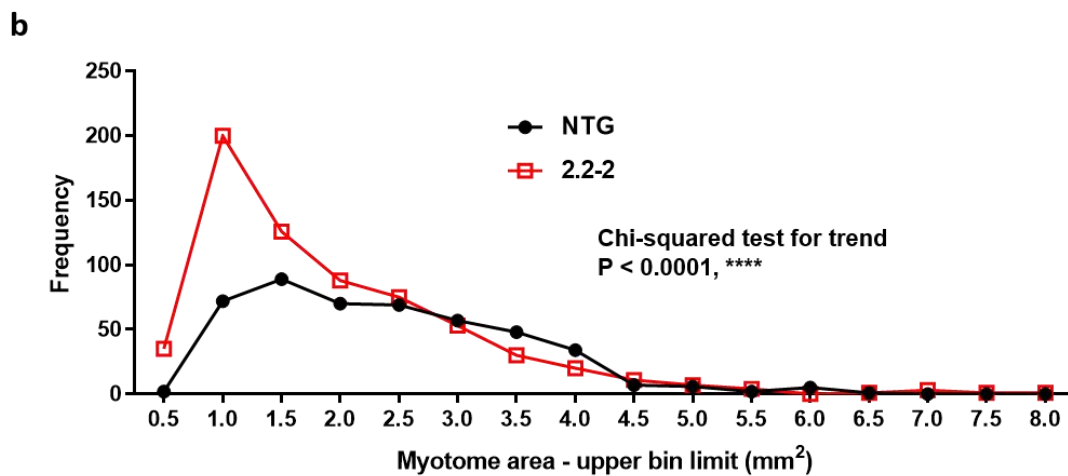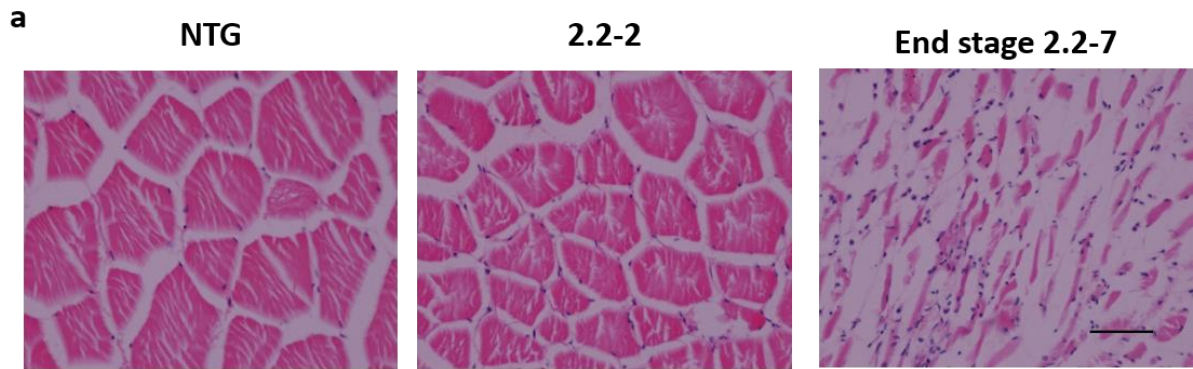
Figure 4.15: Both 2.2 zebrafish lines display abnormal muscle histology

(**a**) Representative H&E staining of zebrafish epaxial muscle (body muscle) myotomes. Scale bar = 50μm. (**b**) Frequency distribution of 2.2-2 and NTG myotome sizes. N=6 individual zebrafish per genotype, 6 sections were analysed for each individual zebrafish (36 sections per genotype in total). 2.2-7 zebrafish were culled once they had reached phenotypic end stage (age matched NTG zebrafish were culled alongside end stage 2.2-7 zebrafish), average age of culling was 12.6 months (8, 11, 13, 14, 15 and 15 months individually for both 2.2-7 and NTG zebrafish). 2.2-2 zebrafish were all culled at 12 months old. Statistical comparison was carried out using the Chi-squared test for trend. Myotome size data are shown as the frequency of myotome sizes binned into defined ranges. ****$P < 0.0001$.
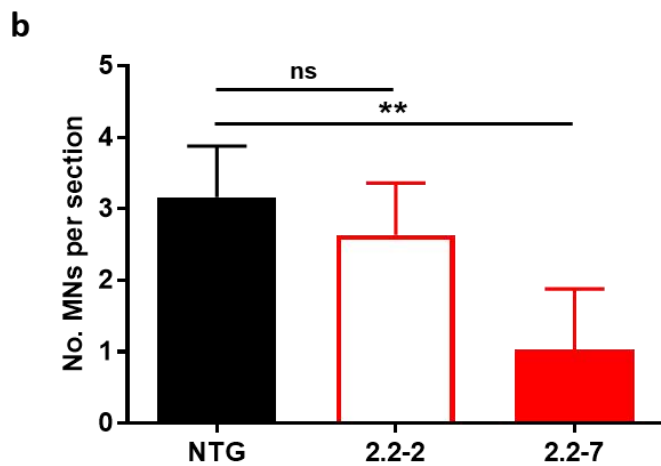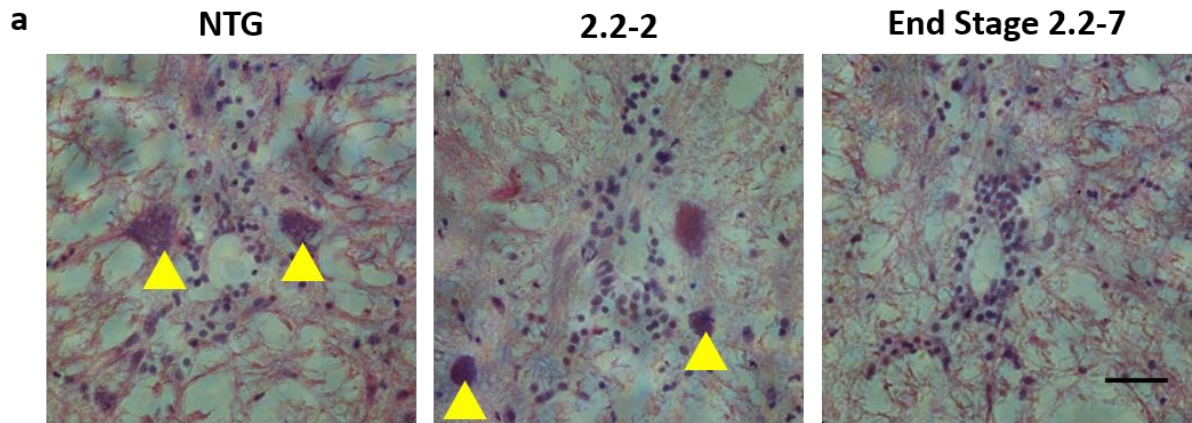
Figure 4.16: 2.2-7 zebrafish display motor neuron loss

(**a**) Representative H&E staining of zebrafish spinal cord sections in which motor neurons are denoted by yellow arrowheads. Scale bar = 25µm. (**b**) 2.2-7 zebrafish have significant motor neuron loss compared to NTG. N=6 individual fish per genotype, 3 sections were analysed for each individual zebrafish (18 sections per genotype in total). 2.2-7 zebrafish were culled once they had reached phenotypic end stage (age matched NTG zebrafish were culled alongside end stage 2.2-7 zebrafish), average age of culling was 12.6 months (8, 11, 13, 14, 15 and 15 months individually for both 2.2-7 and NTG zebrafish). 2.2-2 zebrafish were all culled at 12 months old. Statistical comparison was carried out using a one-way ANOVA with Tukey's post hoc test. Data are shown as mean +/- standard deviation; ns: not significant and **$P < 0.01$.

4.14. DsRed production correlates with disease severity in 2.2-7 zebrafish and heat shock activation occurs in human *C9orf72*-ALS

To examine whether CNS DsRed production changes as disease severity is increasing in 2.2-7 zebrafish, brains of end-stage 2.2-7 zebrafish (ages 15, 15 and 19 months), pre-symptomatic 2.2-7 zebrafish (all aged 7 months) and NTG zebrafish (age matched to end-stage) were lysed and western blotted. Pre-symptomatic was defined as fish which did not show any overt swimming or musculature abnormalities, and were of a younger age than the age at which the earliest swim tunnel abnormalities had been detected. GFP tagged DPRs were increased in the brains of end-stage zebrafish in comparison to the brains of pre-symptomatic zebrafish (1.0 vs. 0.17 A.U. for end-stage and pre-symptomatic respectively; **Figure 4.17a+b**). Similarly, DsRed was also increased in the brains of end-stage zebrafish in comparison to the brains of pre-symptomatic zebrafish (1.0 vs. 0.43 A.U. for end-stage and pre-symptomatic respectively; **Figure 4.17a+c**), suggesting that there is a correlation between DPR production, phenotypic severity and hsp70 promoter driven DsRed production.

Activation of the heat shock response due to cellular stress driven by aberrant protein expression is an already well characterised pathway (Dedmon et al., 2005, Bukau et al., 2006). However, as this system lies at the foundation of our proposed drug screening strategy in the *C9orf72* model zebrafish, we deemed it necessary to examine the role of the heat shock protein system in *C9orf72*-ALS patients. To assess whether the heat shock response is activated in human *C9orf72*-ALS, cerebellum grey matter from *C9orf72*-ALS patients or healthy controls was lysed and western blotted for HSP70 protein (the most ubiquitous and abundant heat shock protein). Indeed, HSP70 protein expression was significantly higher in *C9orf72*-ALS patients as compared to healthy controls (1.1 vs. 0.58 HSP70/tubulin for *C9orf72* patients and controls respectively; **Figure 4.17d+e**). Activation of the heat shock response in human ALS cases validates that identification of compounds which are able to alleviate heat shock activation may have neuroprotective effects in *C9orf72*-ALS. To further validate the heat shock activation drug screening system, we next planned to identify a positive control compound which can reduce DsRed production in the C9orf72 zebrafish model.
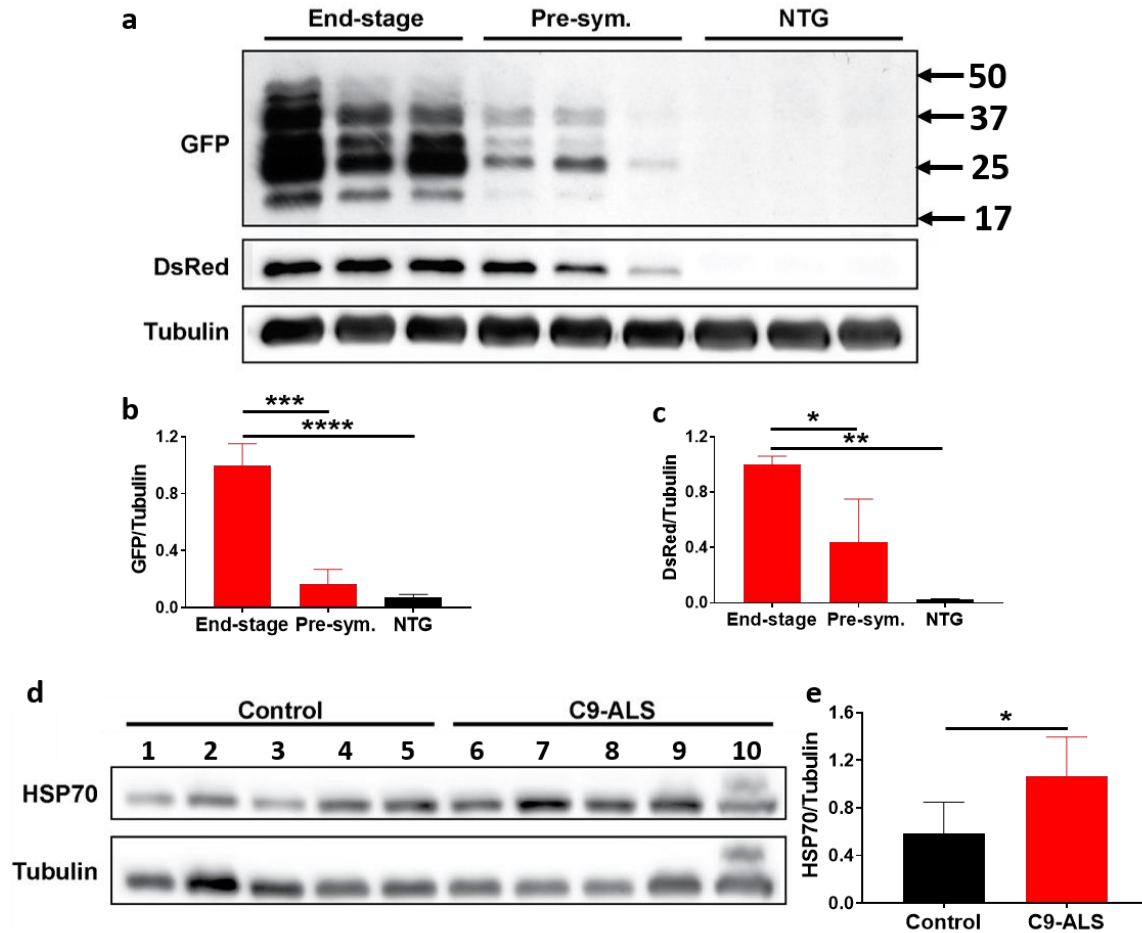
Figure 4.17: Heat shock stress response activation is increased by C9orf72 expansions

(**a**) In end-stage 2.2-7 zebrafish brains, levels of GFP tagged DPR and DsRed proteins are increased compared with pre-symptomatic 2.2-7 and NTG. (**b**) Quantification of GFP tagged DPR protein normalised to tubulin in adult zebrafish brains. N=3 adult brains per condition. Statistical comparison was carried out using a one-way ANOVA with Tukey's post hoc test. (**c**) Quantification of DsRed protein normalised to tubulin in adult zebrafish brains. N=3 adult brains per condition. Statistical comparison was carried out using a one-way ANOVA with Tukey's post hoc test. (**d**) In human cerebellum samples, HSP70 protein levels are higher in C9-ALS patients as compared to controls. N=5 samples per group. The patient code is listed above each sample (see table 2.2 for full patient information). (**e**) Quantification of HSP70 protein levels normalised to tubulin in human cerebellum. Statistical comparison was carried out using an unpaired t-test. All data are shown as mean +/- standard deviation; *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001.

4.15. Riluzole treatment reduces DsRed production in 2.2-7 zebrafish

The first choice to use as a positive control compound in order to test the validity of the DsRed drug screening system in zebrafish was riluzole. The reason for this, is that riluzole is known to have neuroprotective effects and is currently the only disease modifying treatment prescribed for ALS patients in the UK. In order to test the efficacy of riluzole in our DsRed drug screening paradigm, 2.2-7 zebrafish were treated with either 10µM riluzole or vehicle control (DMSO) from 2-5 dpf. At 5 dpf, zebrafish were lysed and their individual DsRed florescence levels were measured using a plate reader. Riluzole treatment resulted in a significant reduction in hsp70 promoter driven production of DsRed protein, as compared with DMSO treatment (0.87 vs. 1.0 fold change in DsRed for riluzole and DMSO treatments respectively; **Figure 4.18a**). As the DsRed readout had now been validated by demonstrating that it could identify riluzole as a hit, two compounds named ivermectin and selamectin were also tested using the same treatment paradigm. The reason for selecting ivermectin and selamectin was that they had both previously been identified as hits in a *SOD1*-ALS drug screening zebrafish model (Ramesh lab, unpublished data). Interestingly, 10µM riluzole, 1µM ivermectin and 10µM selamectin treatments all significantly reduced DsRed production in 2.2-7 zebrafish, as compared to DMSO treatment (0.68, 0.73, 0.62 and 1.0 fold change in DsRed for riluzole, ivermectin, selamectin and DMSO treatments respectively; **Figure 4.18b**). Elucidating the mechanism by which drugs are reducing the heat shock response activation is an important step in targeting future therapies. Ivermectin has previously been shown to inhibit α and β nuclear importins (Wagstaff et al., 2012). This led us to hypothesise that ivermectin may reduce cellular stress by blocking nuclear import of poly(PR) DPR species. To test this hypothesis, we treated zebrafish with the β-importin specific blocker importazole. However, neither 1µM or 5µM importazole treatments significantly changed DsRed production in 2.2-7 zebrafish, as compared to untreated 2.2-7 zebrafish (**Figure 4.18c**). Further, work is required to elucidate the mechanism by which ivermectin or any of the other compounds tested are reducing activation of the heat shock stress response. Andrographolide is a potent nuclear factor E2-related factor 2 (nfr2) activator (Mead et al., 2013), which reduces motor neuron toxicity mediated by co-culture with astrocytes derived from *C9orf72* patients (Ferraiuolo lab, unpublished data). However, treatment of 2.2-7 zebrafish with 10µM or 40µM andrographolide did not result in a significant change in DsRed production, as compared to untreated 2.2-7 zebrafish (**Figure 4.18d**). As the ALS/FTD and drug screening capability of the 2.2-7 zebrafish was now established, we next set out to identify a drug with known neuroprotective properties which may have therapeutic potential in treating *C9orf72*-ALS.
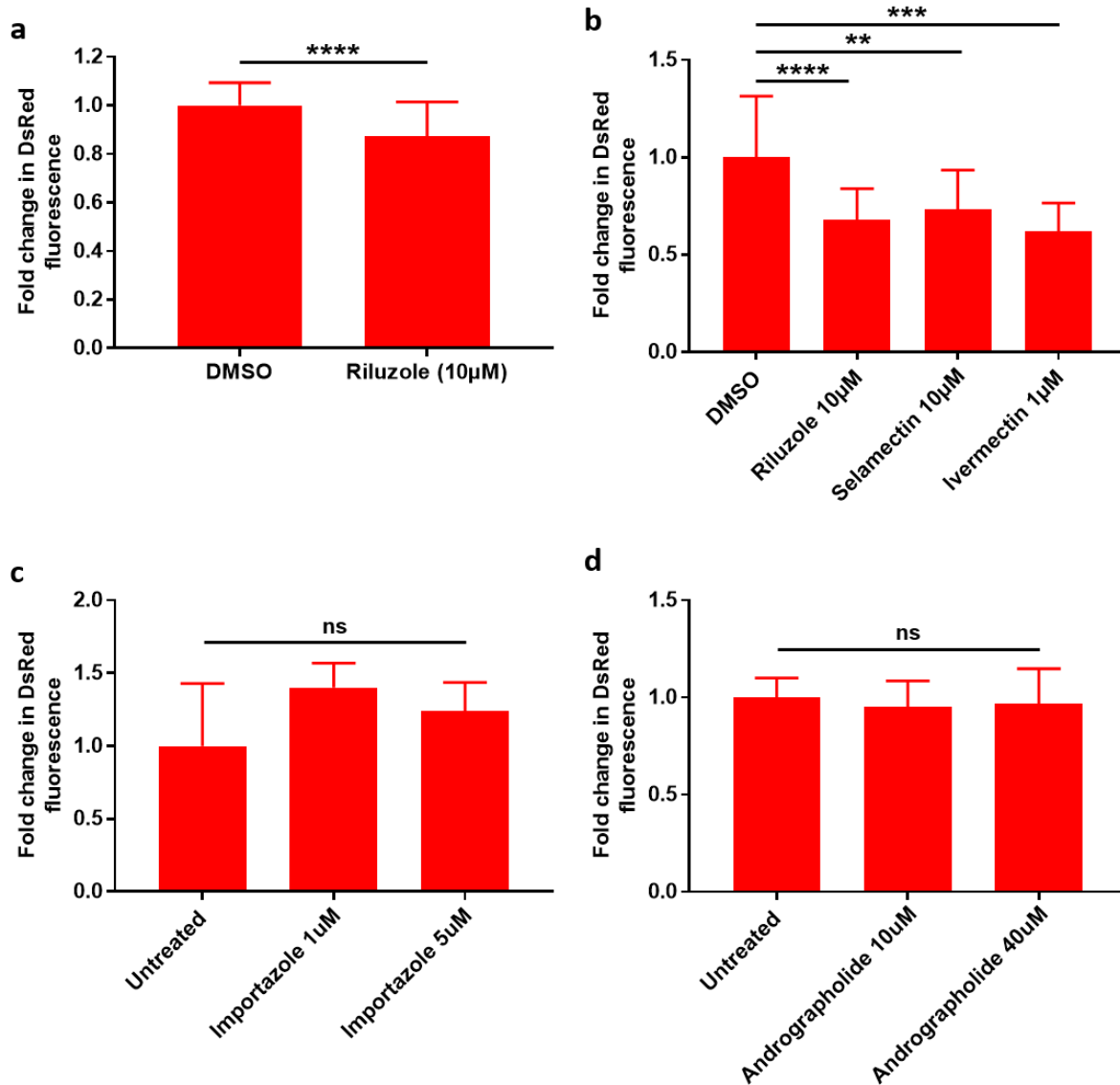
Figure 4.18: Riluzole, selamectin and ivermectin can all alleviate stress response activation in 2.2-7 zebrafish

(**a**) Treatment with 10µM riluzole from 2-5dpf significantly reduced the DsRed fluorescence in 2.2-7 zebrafish. N=36 DMSO treated and 27 riluzole treated individual zebrafish. This dataset was not normally distributed according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. (**b**) Treatment with 10µM riluzole, 10µM selamectin or 1µM ivermectin from 2-5 dpf significantly reduced the DsRed fluorescence in 2.2-7 zebrafish. N=33 DMSO treated, 34 riluzole treated, 34 selamectin treated and 34 ivermectin treated individual zebrafish. Statistical comparison was carried out using a one-way ANOVA with Tukey's post hoc test. (**c**) Treatment with either 1µM or 5µM importazole from 2-5 dpf did not significantly change DsRed production in 2.2-7 zebrafish, as compared to untreated 2.2-7   (**legend continues on next page**)

zebrafish. N=8 untreated, 3 importazole 1µM treated and 5 importazole 5µM treated, these zebrafish were derived from a single clutch. Statistical comparison was carried out using a one-way ANOVA with Dunnett's post hoc test. (**d**) Treatment with either 10µM or 40µM andrographolide from 2-5 dpf did not significantly change 2.2-7 DsRed production as compared to untreated 2.2-7 zebrafish. N=10 untreated, 10 andrographolide 10µM treated and 8 andrographolide 40µM treated individual zebrafish, these zebrafish were derived from a single clutch. Statistical comparison was carried out using a one-way ANOVA with Dunnett's post hoc test. All data are shown as mean +/- standard deviation; ns: not significant, **P < 0.01, ***P < 0.001 and ****P < 0.0001.

4.16. Over activation of PARP1 does not drive toxicity in 2.2-7 zebrafish

In a subset of patients with cerebellar ataxia caused by *XRCC1* mutations, neurodegeneration was discovered to be driven by over-activation of the DNA damage sensing protein PARP1 (Hoch et al., 2017). Additionally, deletion of the PARP1 gene in XRCC1 mutant mice ameliorated the loss of cerebellar neurons (Hoch et al., 2017). Given the current evidence suggesting chronic DNA damage as a mechanism of C9orf72 expansion mediated toxicity (Lopez-Gonzalez et al., 2016, Farg et al., 2017, Walker et al., 2017), we hypothesised that PARP1 over-activation may also contribute to *C9orf72*-ALS. In order to test this hypothesis, we treated zebrafish with the PARP1 specific inhibitor olaparib. To assess olaparib target engagement, we treated 2.2-7 zebrafish from 2-5 dpf with 10µM and 1µM olaparib. These zebrafish were then lysed and immunoblotted for PAR chains (the product of PARP1 enzymatic activity). Both 10µM and 1µM olaparib treatments reduced the amount of PAR chains produced in 2.2-7 zebrafish, however the 10µM treatment was markedly more effective (**Figure 4.19a**).

Next, to assess whether inhibiting PARP1 activity could ameliorate C9orf72 expansion driven toxicity, we treated zebrafish with olaparib from 2-15 dpf and assessed whether this would rescue the reduced survival observed in 2.2-7 zebrafish at 15 dpf. Unexpectedly, 10µM olaparib treatment caused clear toxic effects on both 2.2-7 and NTG zebrafish. At around 10 dpf the zebrafish became lethargic, and over the next three days of treatment the condition of the zebrafish worsened, with many zebrafish appearing morphologically abnormal. The 10µM olaparib treatment was stopped at 13 dpf and all zebrafish were culled. The 10µM olaparib treatment resulted in a significant reduction in survival in both 2.2-7 and NTG groups (28% vs 17% 2.2-7 survival and 85% vs 67% NTG survival, for DMSO and 10µM olaparib treatments respectively; **Figure 4.19b+f**). The dose of olaparib was then lowered to 1µM and the water change protocol was altered in attempt to reduce

water quality stress on the zebrafish, as even the DMSO treated 2.2-7 zebrafish showed poor survival in the previous experiment. However, at 10 dpf the same signs of toxicity were observed in the 2.2-7 zebrafish, and zebrafish were culled at 12 dpf. With 1µM olaparib treatment a significant reduction in the survival of 2.2-7 zebrafish was observed, however the reduction in the survival of NTG zebrafish was not significant (80% vs 62% 2.2-7 survival and 95% vs 90% NTG survival, for DMSO and 1µM olaparib treatments respectively; **Figure 4.19c+f**). The dose of olaparib was again lowered to 0.1µM, and the same water change protocol as previously described was used. However, at 10 dpf the same signs of toxicity were observed in the 2.2-7 zebrafish, and zebrafish were culled at 11 dpf. With 0.1µM olaparib treatment there was a strong trend towards a reduction in the survival of 2.2-7 zebrafish ($P = 0.0504$), however there was no change in survival of NTG zebrafish (90% vs 76% 2.2-7 survival and 95% vs 95% NTG survival, for DMSO and olaparib treatments respectively; **Figure 4.19d+f**). The effect of olaparib on the heat shock response was also assessed by treating 2.2-7 with 5µM olaparib from 2-5 dpf and then measuring DsRed production. Treatment with 5µM olaparib did not significantly change the heat shock stress response activation as measured by DsRed florescence (**Figure 4.19e**).

Heat shock response activation gives a general readout of cellular stress conditions. For this reason we next aimed to compare the level of heat shock response activation in 5.3 (RNA-only) and 2.2 (RNA+DPR) zebrafish lines.
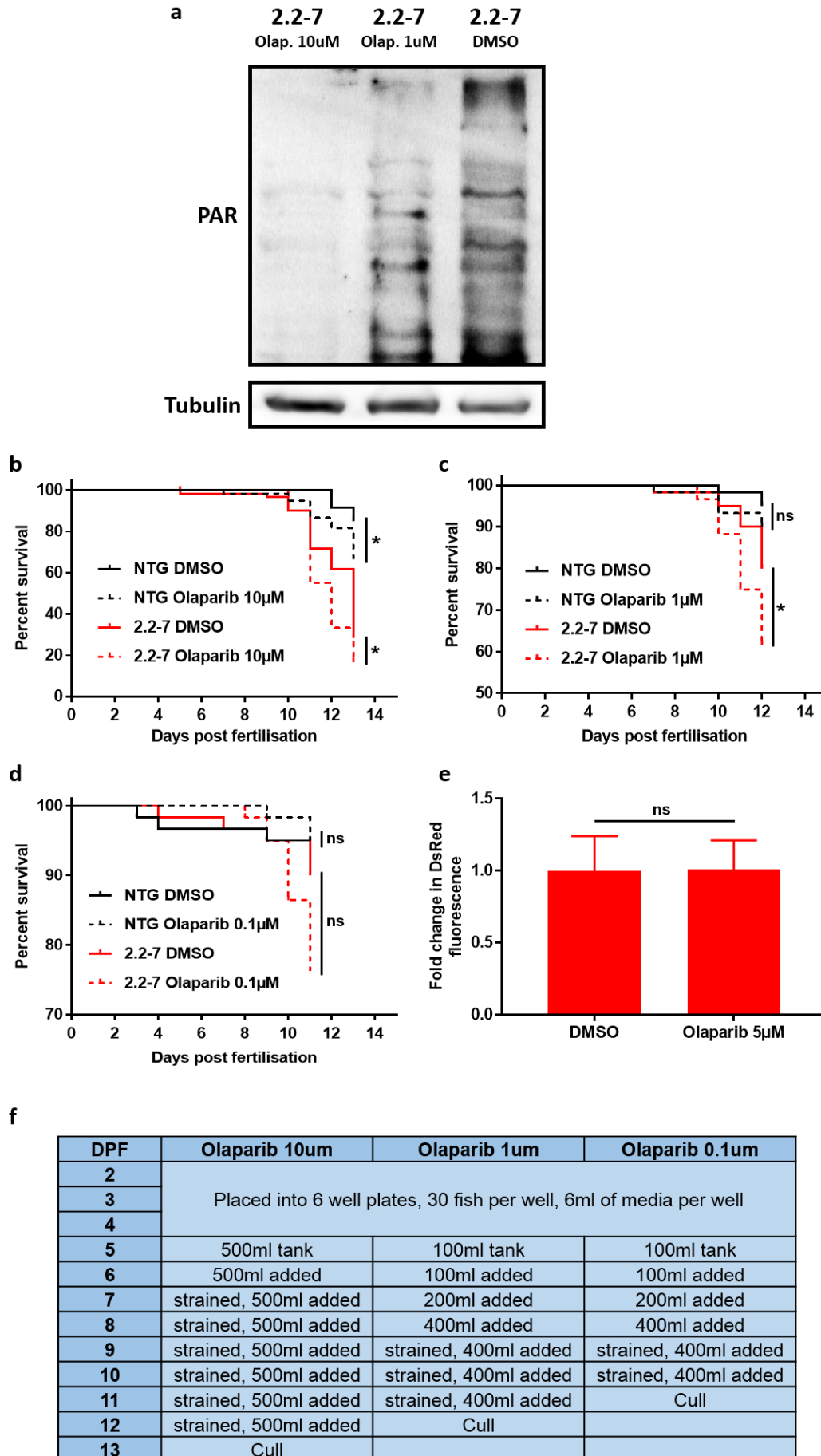
Figure 4.19: Olaparib treatment reduces PAR production but does not rescue reduced viability in 2.2-7 zebrafish

(**a**) Treatment of 2.2-7 zebrafish from 2-5 dpf with the PARP1 specific inhibitor olaparib markedly reduced the production of PARylated proteins at 10µM, and to a lesser extent at 1µM also. (**b**) Treatment with 10µM olaparib from 2-13 dpf was toxic, and caused a significant reduction in survival of 2.2-7 zebrafish and their NTG clutchmates, as compared to DMSO treated zebrafish of the same genotype. N=60 individual zebrafish per condition, zebrafish derived from 3 clutches. Statistical comparison was carried out using the Log-rank test. (**c**) Treatment with 1µM olaparib from 2-12dpf was also toxic and caused a significant reduction in survival in 2.2-7 zebrafish but not in NTG zebrafish, as compared with DMSO treated zebrafish of the same genotype. N=60 individual zebrafish per condition, zebrafish derived from 3 clutches. Statistical comparison was carried out using the Log-rank test. (**d**) Treatment with 0.1µM olaparib from 2-11 dpf caused a strong trend towards reduced survival in 2.2-7 zebrafish but did not have any effect on NTG survival, as compared to DMSO treated zebrafish of the same genotype. N=60 individual zebrafish per condition, zebrafish derived from 3 clutches. Statistical comparison was carried out using the Log-rank test. (**e**) Treatment with 5µM olaparib from 2-5 dpf did not significantly change DsRed production in 2.2-7 zebrafish, as compared to DMSO treated 2.2-7 zebrafish. N=26 DMSO treated and 29 olaparib treated individual zebrafish, derived from 3 clutches. Statistical comparison was carried out using an unpaired t-test. (**f**) Summary of the olaparib treatment and water change protocols used in each of the 3 survival experiments. Olaparib treatment was planned for 2-15 dpf, but was ended early in every experiment to protect the welfare of the zebrafish after observing signs of early toxicity. DsRed fold change data is shown as mean +/- standard deviation; ns: not significant, *P < 0.05.

4.17. Heat shock response activation is higher in 2.2 than 5.3 zebrafish lines

In both 5.3 (RNA-only) and 2.2 (RNA+DPR) zebrafish, activation of the *hsp70* promoter drives production of the DsRed protein. Due to this, it is possible to compare the relative expressions of DsRed by each line of zebrafish and use this as a generic readout of the amount of cellular stress being exhibited. Cell stress influences DsRed production by increasing the drive on the *hsp70* promoter. However, transgene copy number can also influence the amount of DsRed produced by each line of zebrafish. For this reason, in order to compare the relative stress levels of each zebrafish line, we normalised DsRed to GFP production, thus controlling for transgene expression levels within each zebrafish line. There was no significant difference in DsRed/GFP levels between any of the 5.3 zebrafish lines (**Figure 4.20**). However, 2.2-2 zebrafish had significantly higher DsRed/GFP levels than
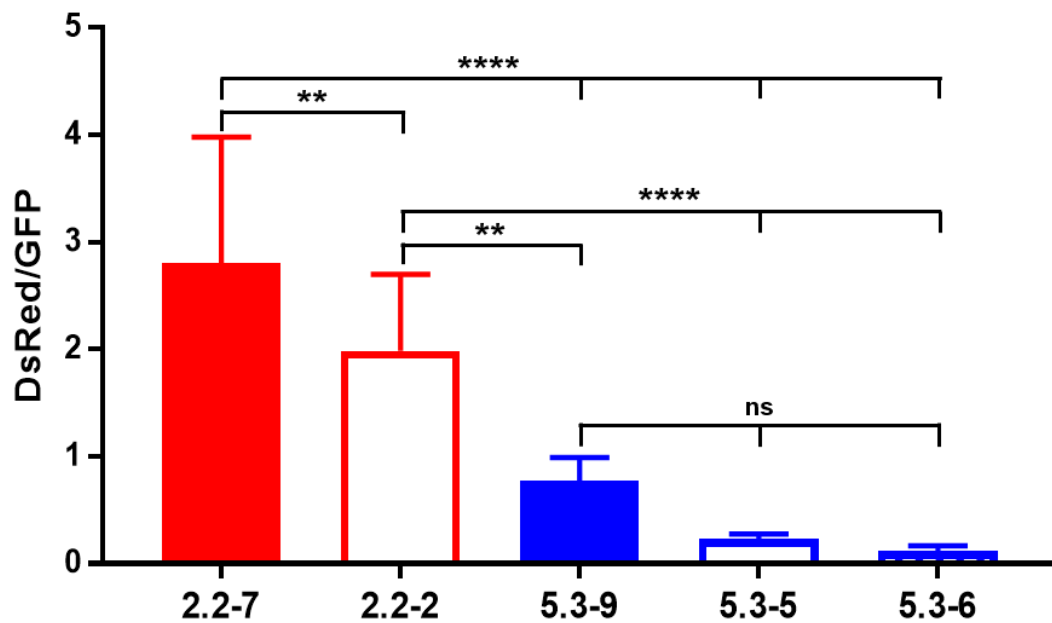
Figure 4.20: 2.2 (RNA+DPR) zebrafish lines exhibit higher heat shock response activation than 5.3 (RNA-only) zebrafish lines when normalised to GFP expression

The ratio of DsRed:GFP does not differ between any zebrafish in the 5.3 lines. Additionally, Both 2.2 zebrafish lines exhibit a significantly higher DsRed:GFP ratio than every 5.3 zebrafish line. Finally, 2.2-7 zebrafish also have a higher DsRed:GFP ratio than 2.2-2 zebrafish. N=76 x 2.2-7, 75 x 2.2-2, 9 x 5.3-9, 10 x 5.3-5 and 9 x 5.3-6 individual zebrafish, zebrafish from 5.3 lines were derived from a single clutch. This dataset was not normally distributed according to the Shapiro-Wilk test for normality, therefore statistical comparisons were carried out using the Kruskal-Wallis test with Dunn's post hoc test. All data are shown as mean +/- standard deviation; ns: not significant, **$P < 0.01$ and ****$P < 0.0001$.

all three 5.3 zebrafish lines (2.0 vs 0.78, 0.23 and 0.12 A.U., for 2.2-2, 5.3-9, 5.3-5 and 5.3-6 respectively; **Figure 4.20**). Additionally, 2.2-7 zebrafish had significantly higher DsRed/GFP levels than 2.2-2 and all three 5.3 zebrafish lines (2.8 vs 2.0, 0.78, 0.23 and 0.12 A.U., for 2.2-7, 2.2-2, 5.3-9, 5.3-5 and 5.3-6 respectively; **Figure 4.20**). Therefore, DsRed/GFP levels correlate with phenotypic severity. It is important to note however that the zebrafish from each of the 5.3 lines were all derived from a single clutch. Therefore, figure 4.20 should be considered as preliminary data until it has been repeated with a further 2 clutches for 5.3 zebrafish.

4.18. Discussion

As previously mentioned in discussion of RNA-only zebrafish injections, the microinjection/pre-screening method used here allows for highly efficient rates of transgenesis without the additional effort and potential pitfalls when using transposable elements. RNA+DPR zebrafish were generated with an even higher rate of transgenesis. Of the highly chimeric zebrafish grown to adulthood, 43% (3/7) gave rise to full transgenic offspring. This further highlights the efficiency of this system for producing transgenic zebrafish. One of the highly chimeric RNA+DPR zebrafish, the highest expresser as determined by eye, showed severely reduced fertility and produced offspring which died within the first few days of life. This phenotype is identical to the phenotype observed in high expresser RNA-only chimeric zebrafish. Thus, these data suggest that high enough expression of C9orf72 expansion RNA is sufficient to generate a toxic phenotype. However, it would be difficult to prove that RNA-only zebrafish are not also producing DPR species due to a lack of sample tissue caused by the low fertility of their chimeric founder (approximately 1 in 200 eggs is fertilised, and only a maximum of 50% of these are expected to be transgenic). Previous reports have also identified that high expression of C9orf72 RNA (with no detectable DPRs) is sufficient to generate an early lethal phenotype, albeit the phenotype was less severe and less frequently observed than the phenotype of zebrafish which expressed poly(GA) encoding DPR constructs under the same promoter (Ohki et al., 2017). These data clearly indicate that C9orf72 expansion RNA can contribute to early toxicity in stable transgenic animal models.

Of the two RNA+DPR zebrafish which did not show early lethality, the GFP expression levels of both zebrafish were not significantly different at 5 dpf. However, the DsRed level and DsRed:GFP ratio of the 2.2-7 zebrafish was significantly higher. These data suggest that different DPR species may have a differential ability to activate the heat shock stress response. This would be consistent with previous reports showing that different DPR species have variable levels of *in vivo* toxicity (Mizielinska et al., 2014, Tran et al., 2015). Interestingly, adult 2.2-7 zebrafish also show a more severe phenotype than 2.2-2 zebrafish, exhibiting reduced swimming endurance, reduction in body mass, motor neuron loss and advanced muscle atrophy. This suggests that the level of heat shock activation is a good predictor of phenotypic severity in 2.2 zebrafish. Furthermore, hsp70 promoter-mediated DsRed production was significantly higher in end-stage vs pre-symptomatic zebrafish, further demonstrating that within the same zebrafish line heat shock activation correlates with disease severity. Zebrafish of the 2.2-7 line also exhibited centre avoidance behaviour, decreased early viability and an adult onset reduction in survival. It will be important in the future to assess 2.2-2 zebrafish using these assays also so that it can be determined

whether the observed milder phenotype of the 2.2-2 zebrafish is present in all phenotypic readouts.

Reduced body mass was observed in the 2.2-7 zebrafish at 30 dpf, however the reduction had recovered to normal levels by adulthood. This suggests that the reduction in body mass was not caused by tissue degeneration, but rather by a developmental delay which can be compensated for given enough time.

Interestingly, not only did DsRed production increase from pre-symptomatic to end-stage in zebrafish brains, but levels of GFP-tagged DPRs were also more abundant in end-stage tissue. The reason for this increase in DPR abundance is not currently known. One possibility is that a reduction in the efficiency of protein clearance mechanisms leads to DPR accumulation. Certainly impairment of protein clearance pathways is a well characterised feature of ALS /FTD pathogenesis (Shahheydari et al., 2017). Another potential explanation for the progressive increase in DPR abundance could be due to increased drive through the ubiquitin promoter with age. DPR species have previously been shown to bind ubiquitin in intracellular inclusions (Mackenzie et al., 2014, May et al., 2014, Zhang et al., 2014). This ubiquitin binding may cause increased transcriptional drive on the ubiquitin promoter thus producing more C9orf72-RNA and more DPR protein (C9orf72 expansion expression is driven by the ubiquitin promoter in 2.2 zebrafish), resulting in the formation of a positive feedback loop. These possibilities deserve careful exploration in the future.

DPR expression in the CNS, activation of the stress response in the CNS, centre avoidance behaviour and motor neuron loss all provide evidence of neurological dysfunction in 2.2-7 zebrafish. Muscle tissue in 2.2-7 zebrafish also expresses DPR species, and may therefore contribute to the adult onset motor neuron loss through cell to cell transmission of DPR species, or by loss of trophic support for motor neuron axon terminals due to muscle degeneration (Griesbeck et al., 1995, Westergard et al., 2016). It is important to note that, contribution to motor neuron pathology from muscle tissues has also been suggested in human ALS, as muscle tissue from ALS patients frequently contains abnormal protein aggregates such as p62 positive inclusions (Al-Sarraj et al., 2014). In contrast, centre avoidance is primarily a behavioural abnormality and as such is unlikely to be influenced by muscle expression of DPR species. The reason why abnormal centre avoidance behaviour is detectable at 5dpf, while reduced swimming endurance and motor neuron loss are not observed until adulthood, is not fully understood. It is possible that this may reflect the cell type specific toxicity which is frequently reported with DPR proteins (Mackenzie et al., 2013, Gendron et al., 2015, Schludi et al., 2015). In the future, careful time course analysis of

muscle and motor neuron pathology in 2.2-7 zebrafish would help to distinguish myogenic from neurogenic pathologies.

Overall, it is clear from the characterisation of the RNA+DPR zebrafish that the phenotype expressed is more severe than the phenotype observed in RNA-only zebrafish. This is discussed further in chapter 6.

5. Chapter 5: C9orf72 expansions induce upregulation of TDP1 and SOD1 proteinopathy

5.1. Introduction

Dysregulation of multiple seemingly unrelated pathways have now been reported in *C9orf72*-ALS cases. It is difficult to separate pathways which are driving ALS pathogenesis from those which are consequences of neurodegeneration, unrelated to disease progression. These problems are further complicated by a lack of translation of animal model data. For example, it is now clear that in animal models DPR drive toxicity to a larger extent than RNA foci (Tran et al., 2015, Ohki et al., 2017). However, in human pathological studies, DPR expression poorly correlates with disease severity (Mackenzie et al., 2013). In particular, most human pathological studies seem to agree that in the motor cortex, DPR expression is low and disease severity is high, whereas in the cerebellum the reverse is true (Mackenzie et al., 2013, Gendron et al., 2015, Mackenzie et al., 2015, Schludi et al., 2015). In order to further test the argument that DPRs do not correlate with clinical phenotype, it is important to carefully quantify DPR levels in disease affected areas using robust techniques, and investigate whether early toxicity can be observed in the cerebellum.

5.2. Aims

We aimed to first assess whether DPR proteins can be reliably detected in disease relevant areas of human post-mortem samples using a quantitative immunoblotting technique. Following this, we aimed to characterise in detail whether the cerebellum shows any early signs of toxicity. Finally, we aimed to further interrogate the types of pathology induced by C9orf72 expansions in human post-mortem tissue.

5.3. Multiple DPR species are detectable in cerebellum but only poly(GA) is detectable in motor cortex

Post-mortem grey matter samples from the cerebellum and white matter samples from the motor cortex (white matter motor cortex was selected due to lack of grey matter motor cortex patient tissue availability at the time of the experiment) were lysed and drawn through a vacuum onto a PVDF membrane which was then immunoblotted. To obtain loading controls, each lysed sample was divided into two tubes immediately before being vacuum drawn onto the PVDF membrane. Thus, two duplicate membranes were obtained, one membrane was immuno blotted with tubulin (raised in mouse) followed by poly(PR) (raised in rabbit), while the other membrane was immunoblotted with poly(GA) (raised in mouse) followed by poly(GP) (raised in rabbit). Tubulin images were used as a loading control for all

DPR antibodies, as both membranes contained duplicates of the same samples. For more details see methods section 2.2.16.

Poly(GA) protein levels were significantly higher in motor cortex and cerebellum samples from *C9orf72* patients as compared with equivalent samples from non-neurological disease controls (1.1 vs 0.022 poly(GA)/tubulin for motor cortex samples from *C9orf72* and control patients respectively; **Figure 5.1a-c**; 5.1 vs 0.0043 poly(GA)/tubulin for cerebellum samples from *C9orf72* and control patients respectively; **Figure 5.1d-f**). In both motor cortex and cerebellum samples, poly(GA) protein levels in sALS patients were not significantly different compared to those in non-neurological disease controls (**Figure 5.1a-f**).

Poly(GP) protein levels were not significantly different in white matter motor cortex samples from *C9orf72* patients or sALS patients when compared with non-neurological disease controls (**Figure 5.2a-c**). However, poly(GP) protein levels in grey matter cerebellum samples from *C9orf72* patients were significantly higher compared with non-neurological controls (4.3 vs 0.043 for *C9orf72* and control samples respectively; **Figure 5.2d-f**). No difference in poly(GP) levels was observed in grey matter samples from sALS patients as compared to non-neurological controls.

No significant difference was observed in poly(PR) protein levels in white matter motor cortex samples from *C9orf72* patients as compared with non-neurological controls (**Figure 5.3a-c**). Unexpectedly, poly(PR) protein levels from motor cortex were higher in sALS samples as compared to non-neurological disease controls (0.44 vs 0.10 poly(PR)/tubulin for sALS and control samples respectively; **Figure 5.3a-c**). This difference was primarily driven by a single sample suggesting that this patient could potentially harbour a poly(PR) encoding expansion in another genetic locus. Finally, poly(PR) protein levels were significantly higher in grey matter cerebellum samples from *C9orf72* patients as compared to non-neurological disease controls (8.7 vs 0.15 poly(PR)/tubulin for cerebellum samples from *C9orf72* patients and controls respectively; **Figure 5.3d-f**).

There are some limitations to the dot blots described above. As the same tubulin image was used as a loading control for both duplicate membranes used, this assumes that protein levels bound to each membrane were equal. Additionally, only one protein concentration of each sample was used, rather than a dilution series. Furthermore, unlike western blots, dot blots do not allow for the separation of specific and non-specific immunoreactive protein species by molecular weight. Therefore, the precise proportion of signal which is contributed by actual DPR proteins in C9-ALS cases cannot be determined.
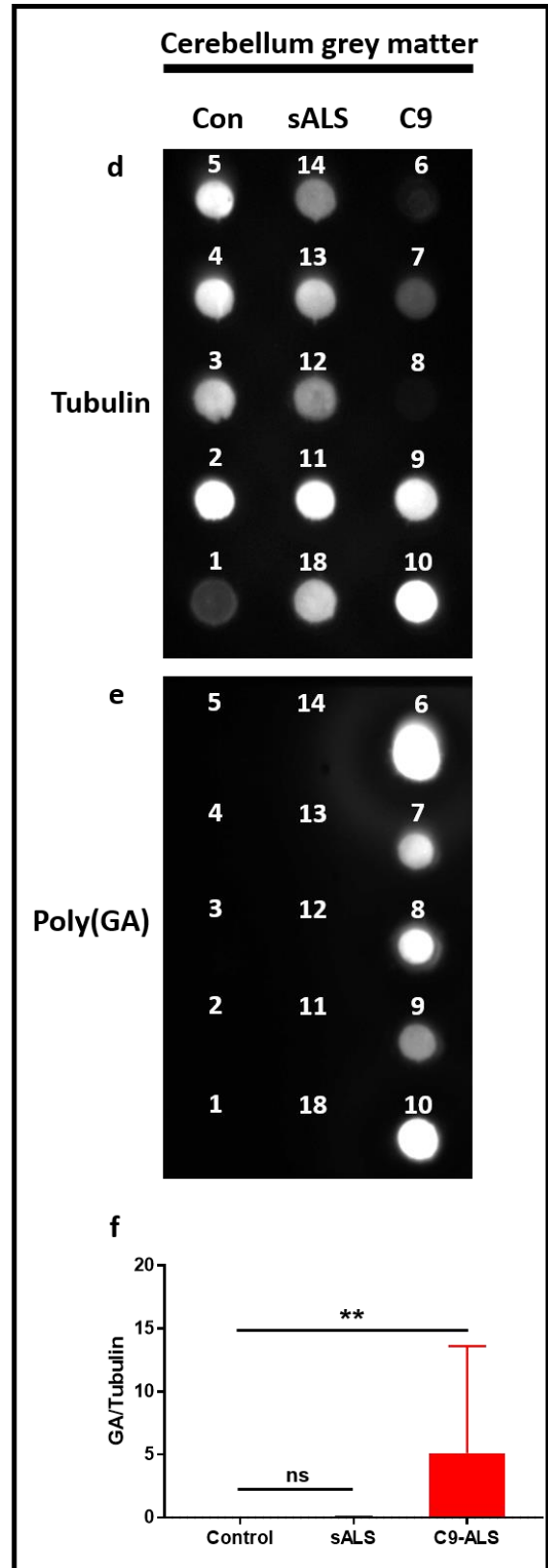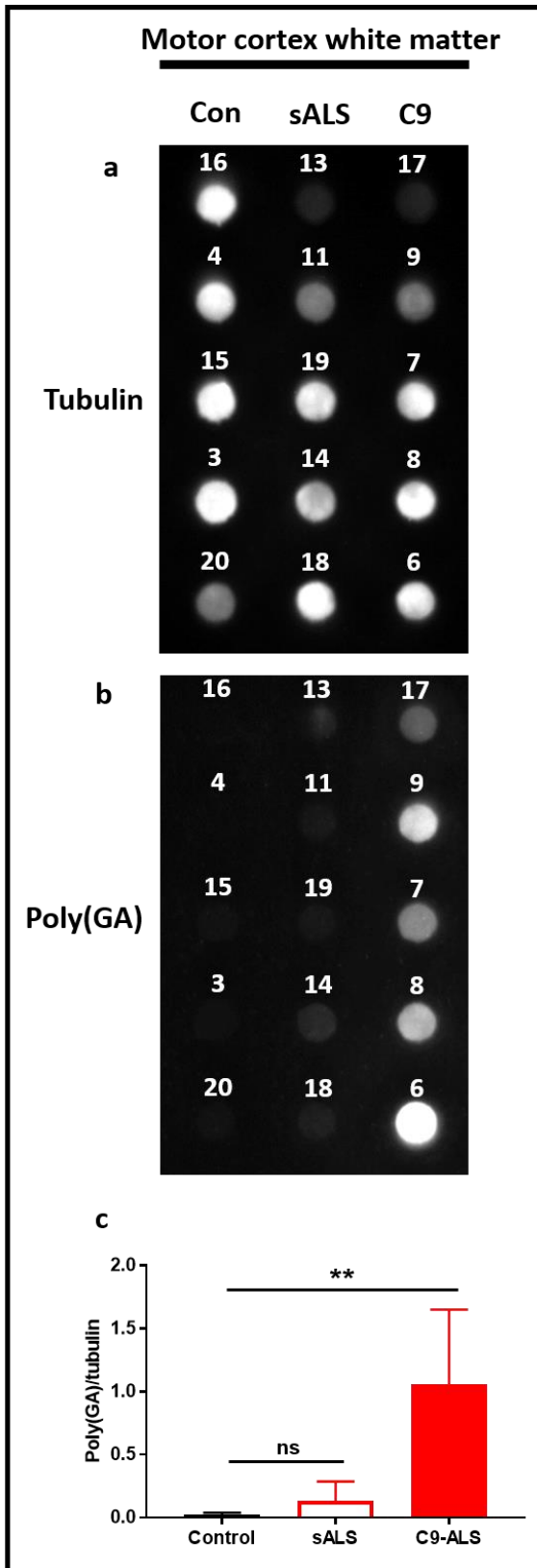
Figure 5.1: Poly(GA) DPR proteins are produced in both motor cortex and cerebellum of C9orf72 ALS patients

(**a**) Dot blots of motor cortex white matter protein show similar loading in all groups, as measured by tubulin signal. (**b**) Dot blots of motor cortex white matter proteins show a markedly higher poly(GA) signal in samples from C9-ALS patients. (**c**) Quantification showing that in motor cortex white matter, significantly higher poly(GA) signal is detected in C9-ALS samples in comparison to control samples, when normalised to tubulin. As this dataset was not normally distributed according to the Shapiro-Wilk test for normality, statistical comparison was carried out using the Kruskal-Wallis test with Dunn's post hoc test. (**d**) Dot blots of cerebellum grey matter protein show similar loading in all groups, as measured by tubulin signal. (**e**) Dot blots of cerebellum grey matter proteins show a markedly higher poly(GA) signal in samples from C9-ALS patients. (**f**) Quantification showing that in cerebellum grey matter, significantly higher poly(GA) signal is detected in C9-ALS samples in comparison to control samples, when normalised to tubulin. As this dataset was not normally distributed according to the Shapiro-Wilk test for normality, statistical comparison was carried out using the Kruskal-Wallis test with Dunn's post hoc test. For all samples, the patient code is listed above each sample (see table 2.2 for full patient information). Con: Control, sALS: sporadic-ALS, C9: C9-ALS. All data are shown as mean +/- standard deviation; ns: not significant and **$P < 0.01$.

Motor cortex white matter

Con sALS C9

a

Tubulin	16	13	17
	4	11	9
	15	19	7
	3	14	8
	20	18	6

b

Poly(GP)	16	13	17
	4	11	9
	15	19	7
	3	14	8
	20	18	6

c

Cerebellum grey matter

Con sALS C9

d

Tubulin	5	14	6
	4	13	7
	3	12	8
	2	11	9
	1	18	10

e

Poly(GP)	5	14	6
	4	13	7
	3	12	8
	2	11	9
	1	18	10

f

Figure 5.2 : Poly(GP) DPR proteins are produced in cerebellum but not white matter motor cortex of C9orf72 ALS patients

(**a**) Dot blots of motor cortex white matter protein show similar loading in all groups, as measured by tubulin signal. (**b**) Dot blots of motor cortex white matter protein show similar levels of poly(GP) signal in all groups. (**c**) Quantification showing that in motor cortex white matter, no significant difference in poly(GP) is observed between any group, when normalised to tubulin. Statistical comparisons were carried out using a one-way ANOVA with Dunnett's post hoc test. (**d**) Dot blots of cerebellum grey matter protein show similar loading in all groups, as measured by tubulin signal. (**e**) Dot blots of cerebellum grey matter proteins show a markedly higher poly(GP) signal in samples from C9-ALS patients. (**f**) Quantification showing that in cerebellum grey matter, significantly higher poly(GP) signal is detected in C9-ALS samples in comparison to control samples, when normalised to tubulin. As this dataset was not normally distributed according to the Shapiro-Wilk test for normality, statistical comparison was carried out using the Kruskal-Wallis test with Dunn's post hoc test. For all samples, the patient code is listed above each sample (see table 2.2 for full patient information). Con: Control, sALS: sporadic-ALS, C9: C9-ALS. All data are shown as mean +/- standard deviation; ns: not significant and **$P < 0.01$.
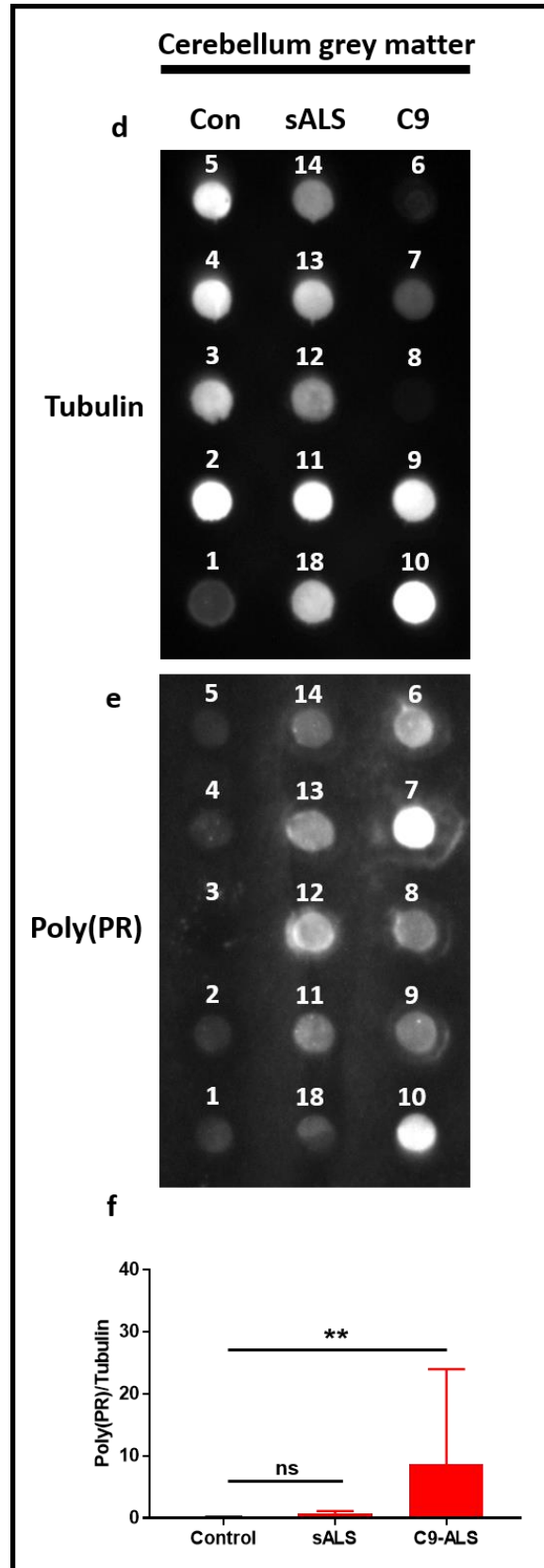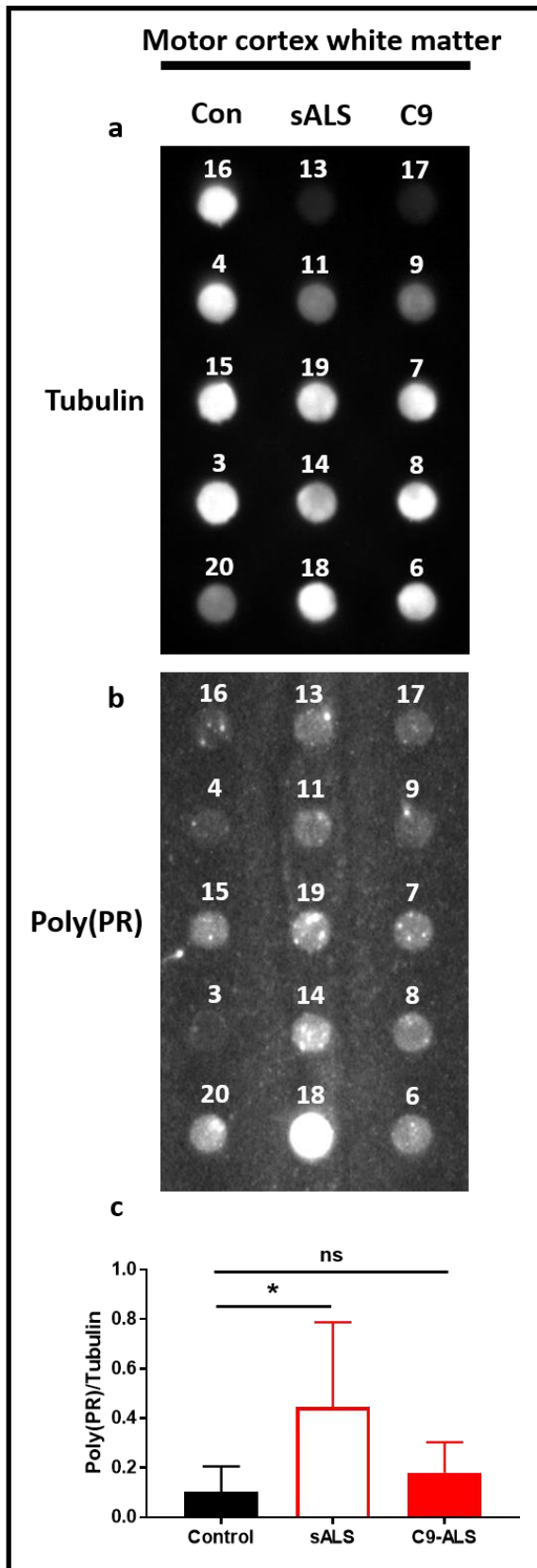
Figure 5.3: Poly(PR) DPR proteins are produced in cerebellum but not white matter motor cortex of C9orf72 ALS patients

(**a**) Dot blots of motor cortex white matter protein show similar loading in all groups, as measured by tubulin signal. (**b**) Dot blots of motor cortex white matter proteins show a markedly higher poly(PR) signal in samples from sALS patients, but not C9-ALS patients. (**c**) Quantification showing that in motor cortex white matter, a significantly higher poly(PR) signal is observed in the sALS samples in comparison to control samples, when normalised to tubulin. As this dataset was not normally distributed according to the Shapiro-Wilk test for normality, statistical comparison was carried out using the Kruskal-Wallis test with Dunn's post hoc test. (**d**) Dot blots of cerebellum grey matter protein show similar loading in all groups, as measured by tubulin signal. (**e**) Dot blots of cerebellum grey matter proteins show a markedly higher poly(PR) signal in samples from C9-ALS patients. (**f**) Quantification showing that in cerebellum grey matter, significantly higher poly(PR) signal is detected in C9-ALS samples in comparison to control samples, when normalised to tubulin. As this dataset was not normally distributed according to the Shapiro-Wilk test for normality, statistical comparison was carried out using the Kruskal-Wallis test with Dunn's post hoc test. For all samples, the patient code is listed above each sample (see table 2.2 for full patient information). Con: Control, sALS: sporadic-ALS, C9: C9-ALS. All data are shown as mean +/- standard deviation; ns: not significant, *P < 0.05 and **P < 0.01.

These data show that a diverse range of DPR proteins are readily detectable in cerebellum tissue samples from *C9orf72* patients. Given the reported toxicity of DPR proteins, we next aimed to assess whether sub-clinical neurodegeneration may be detected in *C9orf72* cerebellum samples.

5.4. Purkinje cell loss is not a feature of C9orf72-ALS

In order to assess whether DPR load in the cerebellum of *C9orf72*-ALS patients is capable of causing sub-clinical neurodegeneration, we counted Purkinje neurons from sections of the cerebellum adjacent to the dentate gyrus. Beginning at the bottom of each slide, the contour of the Purkinje cell layer was followed until 200 Purkinje cells had been counted or the top of the slide had been reached (**Figure 5.4a**). To control for the effect of any potential cell shrinkage, only Purkinje neurons in which the nucleolus was visible were counted during the blinded quantification (**Figure 5.4b**). Surprisingly, despite the presence of DPR proteins in *C9orf72* patient cerebellum samples, there was no significant difference between Purkinje cell counts of *C9orf72* patients and those from non-neurological disease controls (**Figure 5.4c**). As there was no detectable neurodegeneration of Purkinje cells, we next focussed on identifying whether molecular pathologies could be detected in the cerebellum.

a

b
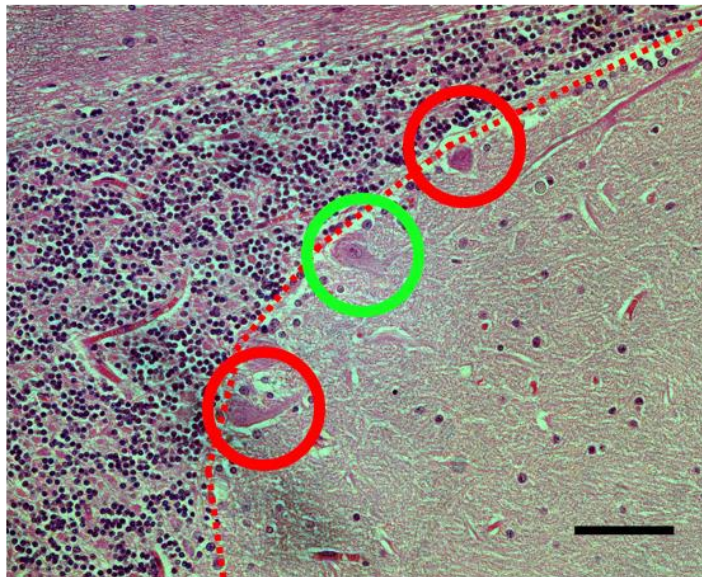
c
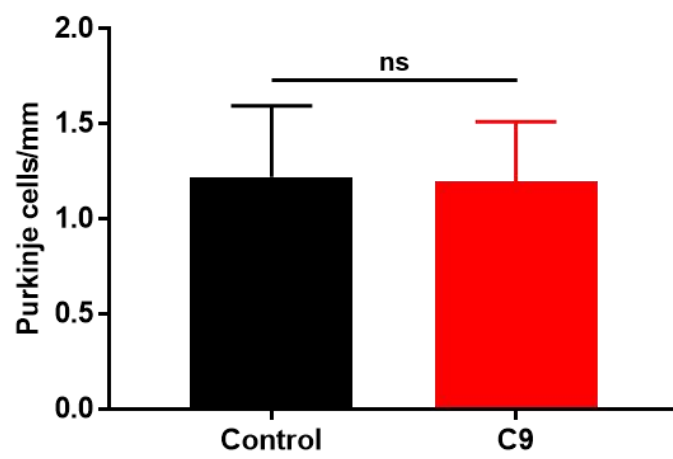
Figure 5.4: C9orf72 ALS patients have normal cerebellar Purkinje cell counts

(**a**) Representative image of cerebellar folia adjacent to the dentate gyrus. The red dashed line indicates the Purkinje cell layer, from which Purkinje cells were counted. Scale bar = 2mm. (**b**) Representative image showing three Purkinje cells in the Purkinje cell layer (red dashed line). The red circles indicate Purkinje cells that would not be counted in the analysis as the nucleolus is not visible. Whereas, the green circle indicates a Purkinje cell which would be included in the analysis as the nucleolus is visible. Scale bar = 250μm. (**c**) In the Purkinje cell layer adjacent to the dentate gyrus of the cerebellum, there is no significant difference in the number of Purkinje cells per mm in C9orf72 ALS patients as compared with non-ALS controls. In order to control for potential cell shrinkage, only Purkinje neurons in which the nucleolus was visible were included in the analysis. N=11 controls and 16 C9-ALS patients. Statistical comparison was carried out using an unpaired t-test. Data are shown as mean +/- standard deviation; ns: not significant.

5.5. The DNA repair factor TDP1 is upregulated in *C9orf72*-ALS and sALS

Previous reports indicate that C9orf72 expansions cause DNA damage and inhibit DNA repair pathways (Lopez-Gonzalez et al., 2016, Farg et al., 2017, Walker et al., 2017). Loss of function of the DNA repair factor Tyrosyl-DNA phosphodiesterase 1 (TDP1) has been identified as a cause of neurodegeneration in patients suffering from SCAN-1 (El-Khamisy et al., 2005). We set out to assess whether TDP1 is among the multitude of DNA repair factors inhibited by C9orf72 expansions and may therefore contribute to C9orf72 expansion driven neurodegeneration. We lysed post-mortem tissue samples from the grey and white matter of the motor cortex and the grey matter of the cerebellum, separated the proteins by SDS-PAGE and then immuno blotted for TDP1. In the grey matter of the cerebellum, TDP1 protein levels were significantly higher in *C9orf72* patients samples as compared to non-neurological disease control samples (0.45 vs 0.017 TDP1/tubulin for *C9orf72* patient and control samples respectively; **Figure 5.5a+d**). TDP1 protein levels in grey matter cerebellum samples from sALS patients were not significantly different compared to non-neurological control samples.

Interestingly, in the white matter of the motor cortex TDP1 levels in sALS patients were significantly higher than in non-neurological controls (0.71 vs 0.27 TDP1/GAPDH, for sALS and controls samples respectively; **Figure 5.5b+e**). However, in white matter of the motor cortex TDP1 levels in *C9orf72* patient samples were not significantly different than in non-
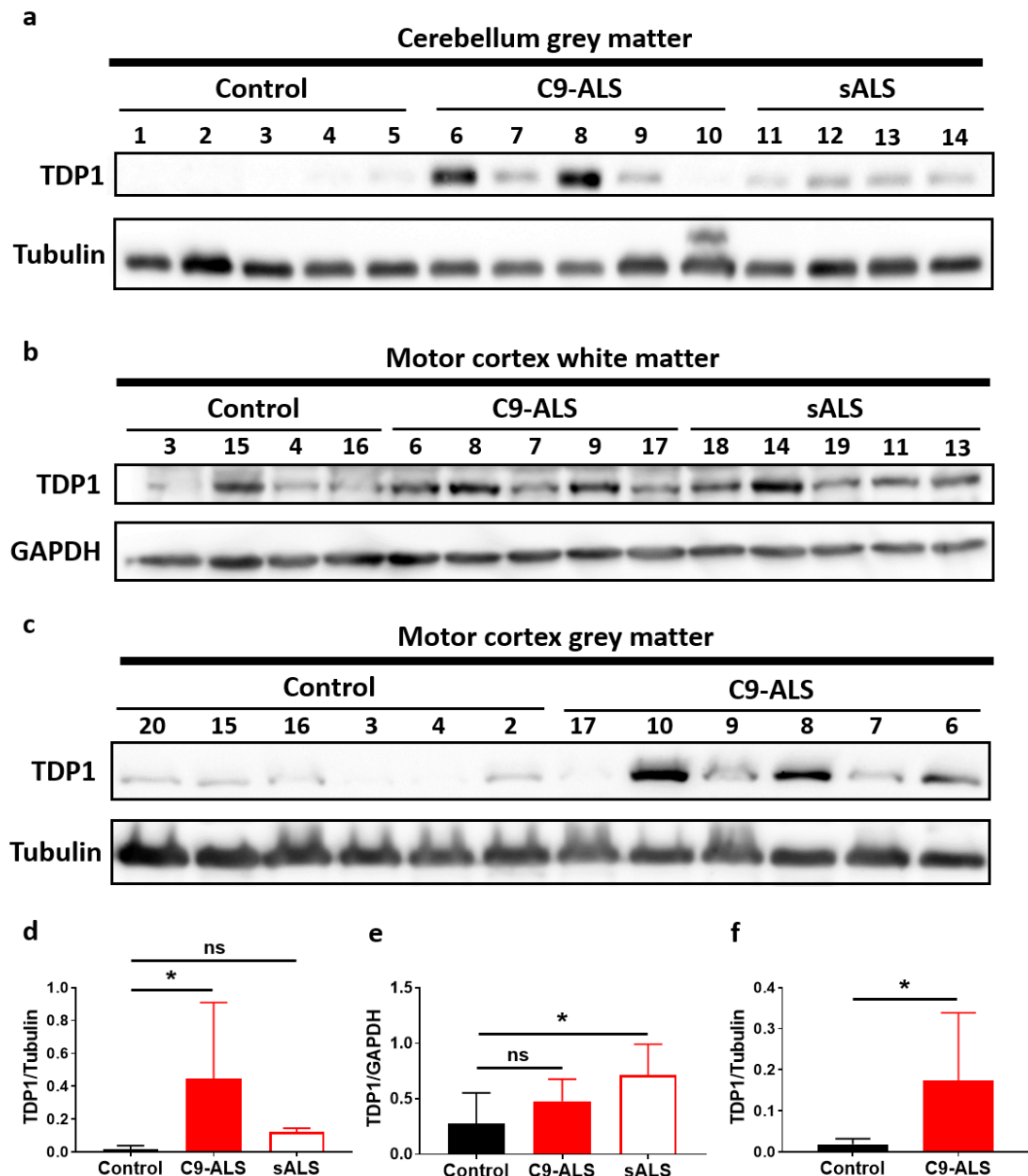
Figure 5.5: TDP1 protein is upregulated in sALS and C9-ALS patient brains

(**a**) In cerebellum grey matter, TDP1 protein is markedly increased in C9-ALS patient samples. (**b**) In motor cortex white matter, TDP1 protein is markedly increased in sALS patient samples. (**c**) In motor cortex grey matter, TDP1 protein is markedly increased in C9-ALS samples. (**d**) Quantification shows that cerebellar grey matter TDP1 levels are significantly higher in C9-ALS samples compared to controls, when normalised to tubulin. This dataset was not normally distributes according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Kruskal-Wallis test followed by Dunn's post hoc test. (**e**) Quantification shows that motor cortex white matter TDP1 levels are significantly higher in sALS samples compared (**Legend continues on next page**)

to controls, when normalised to tubulin. Statistical comparisons were carried out using a one-way ANOVA with Dunnett's post hoc test. (**f**) Quantification shows that motor cortex grey matter TDP1 levels are significantly higher in C9-ALS samples compared to controls, when normalised to tubulin. This dataset was not normally distributes according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. For all samples, the patient code is listed above each sample (see table 2.2 for full patient information). All data are shown as mean +/- standard deviation; ns: not significant and *P < 0.05.

neurological controls. Finally, in the grey matter from the motor cortex, TDP1 levels in *C9orf72* patients were significantly higher than in non-neurological disease controls (0.17 vs 0.019 TDP1/tubulin, for *C9orf72* patient samples and control samples respectively; **Figure 5.5c+f**).

Taken together, these data indicate that protein expression of the DNA repair protein TDP1 is not inhibited by C9orf72 expansions, but is actually upregulated in *C9orf72*-ALS and sALS, depending on tissue type. As it seems that loss of TDP1 function does not drive neurodegeneration in *C9orf72*-ALS, we next set out to investigate whether wtSOD1 proteinopathy may be implicated in C9orf72 expansion pathology.

5.6.   C9orf72 expansions induce wtSOD1 and TDP-43 proteinopathy in disease relevant areas

To date, the role of wtSOD1 pathology in ALS has typically been studied using misfolded-SOD1 specific antibodies in histological CNS sections, which thus far have suggested that misfolded-wtSOD1 is not a component of pathology (Da Cruz et al., 2017). To examine wtSOD1 pathology in ALS patients, we probed total protein lysates from brain samples with a conventional SOD1 antibody (not misfolded-SOD1 specific) raised against amino acids 1-100 of the wtSOD1 protein. We identified that a high molecular weight (HMW) form of SOD1 detected at just above 35KDa, was significantly more abundant in white matter from motor cortex of ALS patients, as compared to comparable regions in non-neurological-disease control samples (**Figure 5.6a+b**). Fragmentation of TDP-43 is a common feature observed in the majority of ALS cases, this primarily results in the formation of 35KDa and 25KDa TDP-43 fragments (Inukai et al., 2008, Tsuji et al., 2012, Kametani et al., 2016). Here, we focussed on the 35KDa fragment of TDP-43 (TDP-35) as this was more abundant in our samples. We identified that TDP-35 was also significantly more abundant in white matter from motor cortex of *C9orf72*-ALS patients, as compared to comparable regions in non-

a Motor cortex white matter
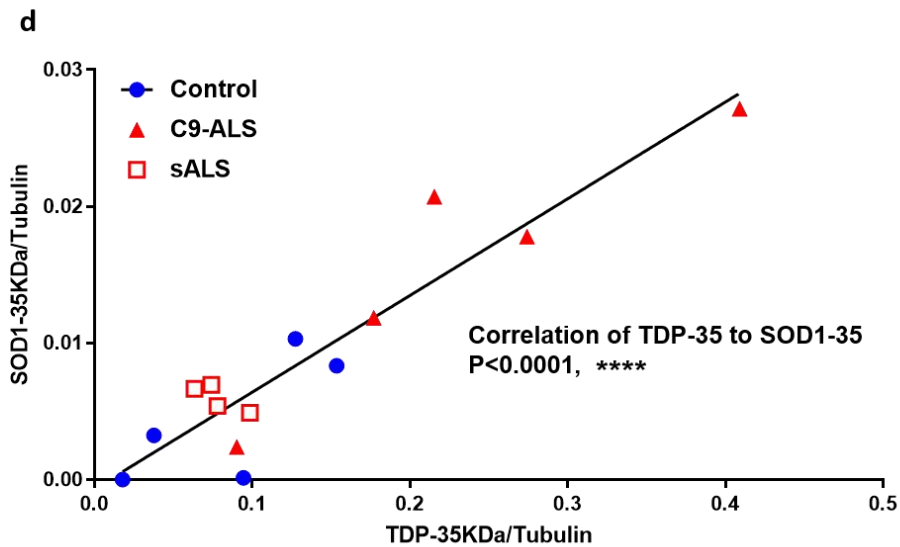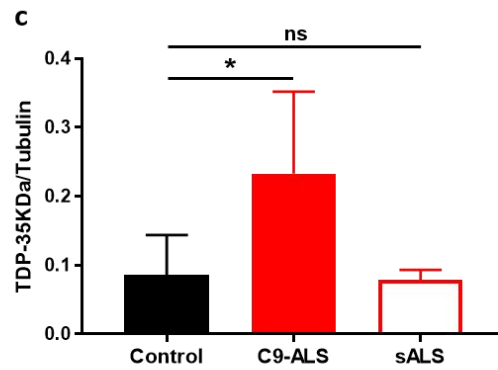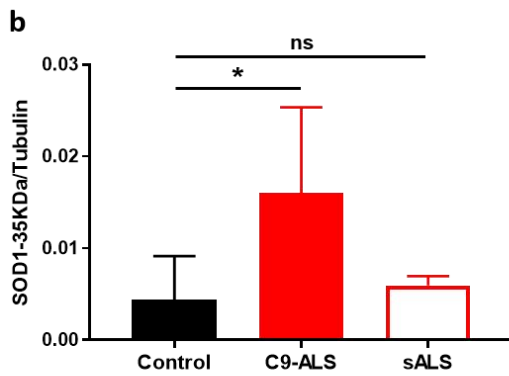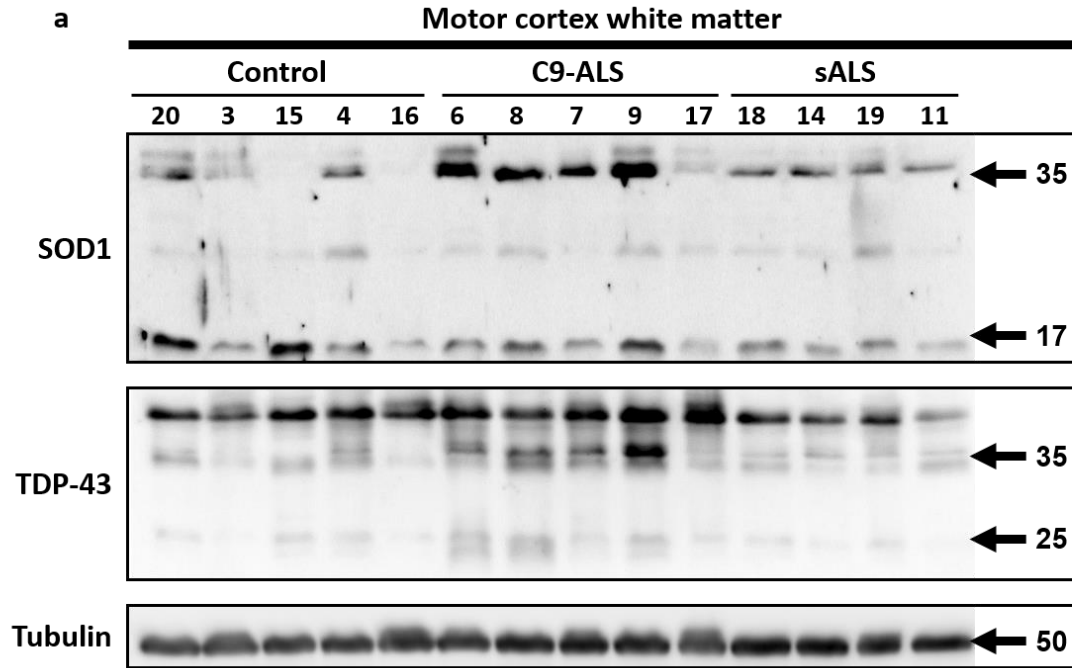
Figure 5.6: HMW-SOD1 and TDP-35 are increased in C9-ALS motor cortex white matter

(**a**) A marked increase in 35KDa SOD1 and 35KDa TDP-43 immunoreactive bands is observed in C9-ALS samples. (**b**) Quantification shows that the signal of 35KDa SOD1 immunoreactive bands is significantly higher in C9-ALS patient samples compared with controls, when normalised to tubulin. Statistical comparisons were carried out using a one-way ANOVA with Dunnett's post hoc test. (**c**) Quantification shows that the signal of 35KDa TDP-43 immunoreactive bands is significantly higher in C9-ALS patient samples compared with controls, when normalised to tubulin. Statistical comparisons were carried out using a one-way ANOVA with Dunnett's post hoc test. (**d**) There is a significant correlation between normalised TDP-35KDa and SOD1-35KDa irrespective of genotype. Correlation was analysed using the Pearson correlation coefficient. For all samples, the patient code is listed above each sample (see table 2.2 for full patient information). Western blot quantification data are shown as mean +/- standard deviation; ns: not significant, $*P < 0.05$ and $****P < 0.0001$.

neurological-disease control samples (**Figure 5.6a+c**). Intriguingly, we also observed that those patients who showed abundant TDP-35 also showed abundant HMW-SOD1. Indeed, there was a strong positive correlation between the abundance of TDP-35 and the abundance of HMW-SOD1 across all white matter motor cortex samples (**Figure 5.6d**).

Furthermore, in grey matter samples from the motor cortex, HMW-SOD1 protein was again significantly more abundant in *C9orf72* patient samples as compared with non-neurological disease control samples (**Figure 5.7a+b**). Likewise, in grey matter samples from the motor cortex fragmented TDP-43 was also significantly more abundant in *C9orf72* patient samples in comparison to non-neurological disease control samples (**Figure 5.7a+c**). Similarly to findings from white matter, samples from grey matter of the motor cortex also showed a strong positive correlation between the abundance of TDP-35 and the abundance of HMW-SOD1 (**Figure 5.7d**).

a

Motor cortex grey matter

Control | C9-ALS

20 15 16 3 4 2 | 17 10 9 8 7 6

SOD1 — 35

TDP-43 — 35 / 25

Tubulin — 50

b

c

d

Correlation of TDP-35 to SOD1-35
P<0.0001, ****
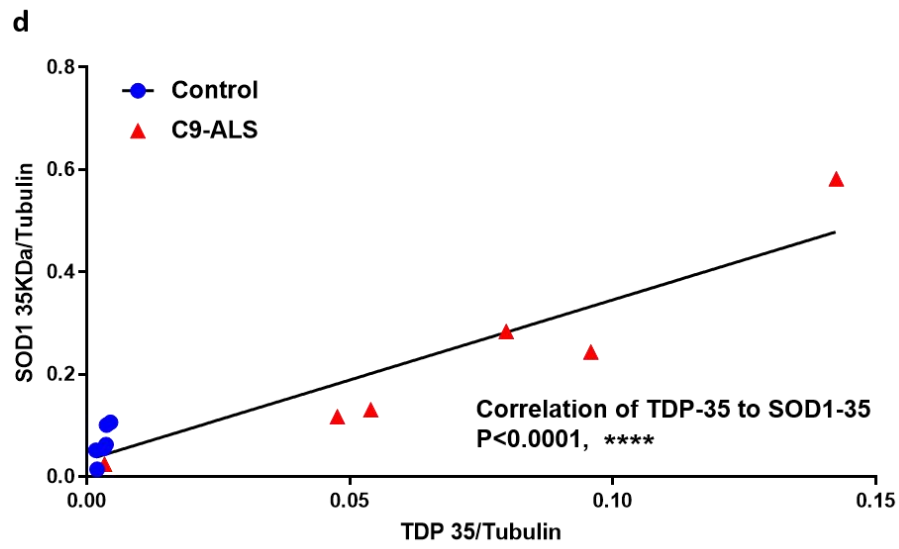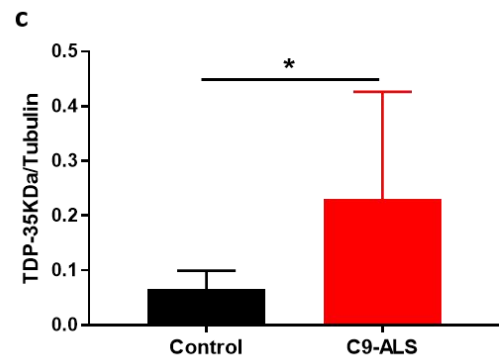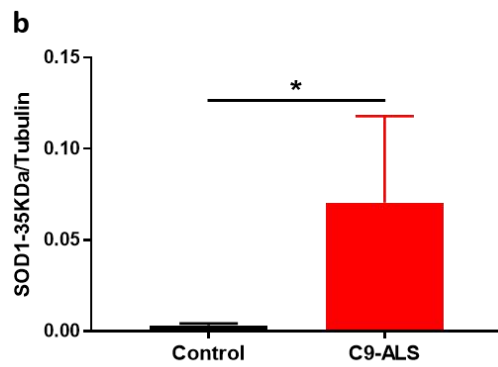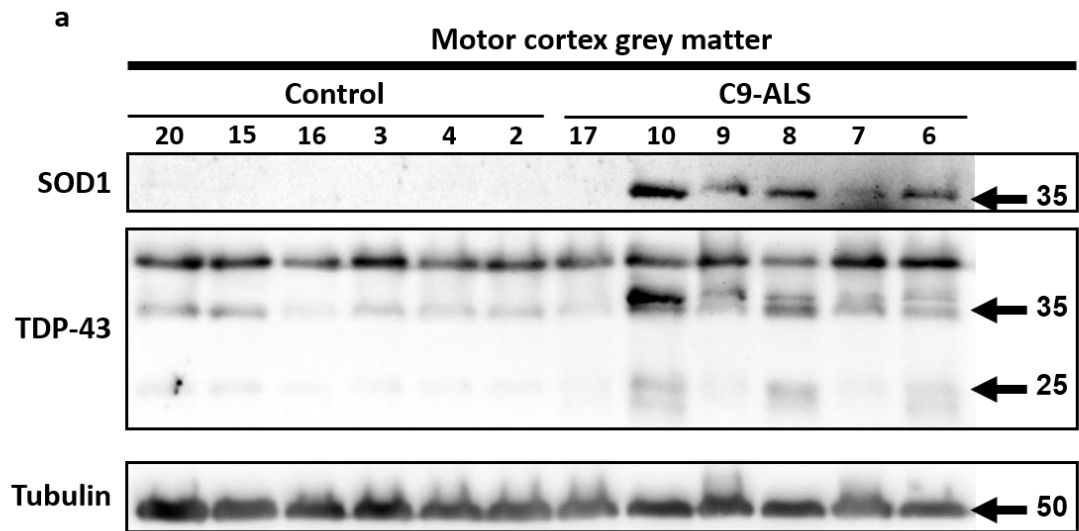
Figure 5.7: HMW-SOD1 and TDP-35 are increased in C9-ALS motor cortex grey matter

(**a**) A marked increase in 35KDa SOD1 and 35KDa TDP-43 immunoreactive bands is observed in C9-ALS samples. (**b**) Quantification shows that the signal of 35KDa SOD1 immunoreactive bands is significantly higher in C9-ALS patient samples compared with controls, when normalised to tubulin. This dataset was not normally distributed according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. (**c**) Quantification shows that the signal of 35KDa TDP-43 immunoreactive bands is significantly higher in C9-ALS patient samples compared with controls, when normalised to tubulin. This dataset was not normally distributes according to the Shapiro-Wilk test for normality and was therefore statistically compared using the Mann-Whitney U test. (**d**) There is a significant correlation between normalised TDP-35KDa and SOD1-35KDa irrespective of genotype. Correlation was analysed using the Pearson correlation coefficient. For all samples, the patient code is listed above each sample (see table 2.2 for full patient information). Western blot quantification data are shown as mean +/- standard deviation; *P < 0.05 and ****P < 0.0001.

In order to identify whether HMW-SOD1 and TDP-35 pathology is specific to disease affected areas, we also examined the abundance of these molecular species in cerebellum tissue. The abundance of HMW-SOD1 was not significantly different between either sALS or *C9orf72*-ALS patients when compared to non-neurological-disease controls, in grey matter samples from the cerebellum (**Figure 5.8a+b**). Similarly, the abundance of TDP-35 was not significantly different between either sALS or *C9orf72*-ALS patients and non-neurological-disease controls, when examined in grey matter samples from the cerebellum (**Figure 5.8a+c**). Interestingly, despite no change in the abundance of HMW-SOD1 and TDP-35 in cerebellum samples, there was still a strong positive correlation between the two molecular species (**Figure 5.8d**). This suggests that there may be a fundamental association between TDP-35 and HMW-SOD1, even under non-pathological conditions. This human pathological tissue data suggests that there may be a link between C9orf72 and SOD1 ALS pathologies.

a

Cerebellum grey matter

Control | C9-ALS | sALS

1 2 3 4 5 6 7 8 9 10 11 12 13 14

SOD1
35
17

TDP-43
35
25

Tubulin
50

b

SOD1-35KDa/Tubulin

ns

Control C9-ALS sALS

c

TDP-35KDa/Tubulin

ns

Control C9-ALS sALS

d

SOD1-35KDa/Tubulin

- Control
- C9-ALS
- sALS

Correlation of TDP-35 to SOD1-35
P<0.0001, ****

TDP-35KDa/Tubulin
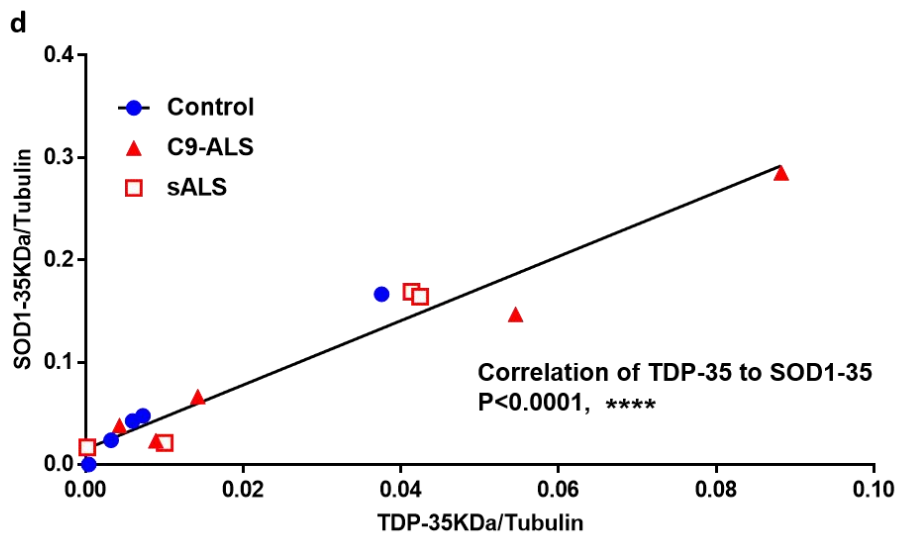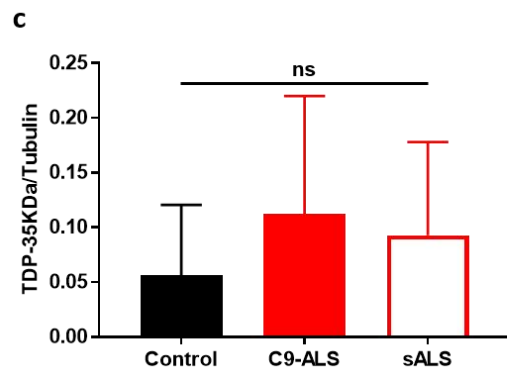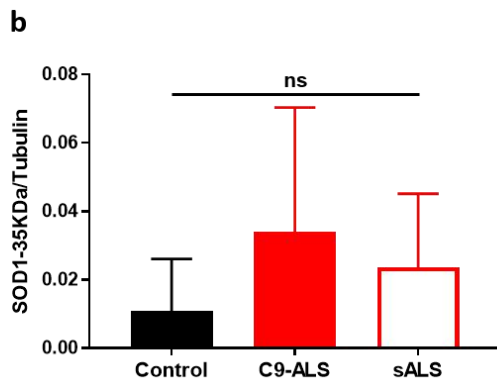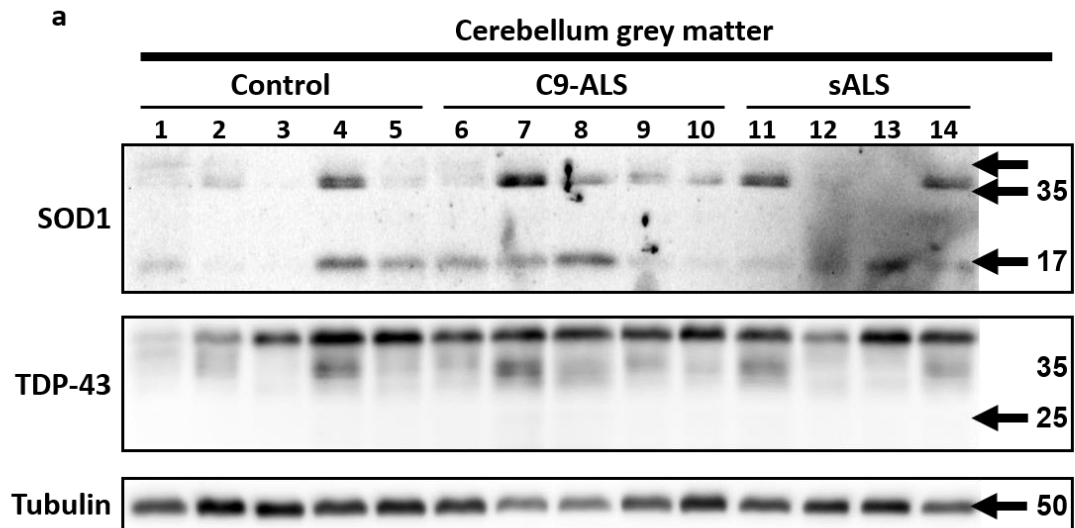
Figure 5.8: No difference in HMW-SOD1 and TDP-35 in C9-ALS cerebellum grey matter

(**a**) Similar levels of 35KDa SOD1 and 35KDa TDP-43 immunoreactive bands are observed across all three groups. (**b**) Quantification shows that the signal of 35KDa SOD1 immunoreactive bands is not significantly different between all three groups, when normalised to tubulin. Statistical comparison were carried out using a one-way ANOVA with Dunnett's post hoc test. (**c**) Quantification shows that the signal of 35KDa TDP-43 immunoreactive bands is not significantly different between all three groups, when normalised to tubulin. Statistical comparison were carried out using a one-way ANOVA with Dunnett's post hoc test. (**d**) There is a significant correlation between normalised TDP-35KDa and SOD1-35KDa irrespective of genotype. Correlation was analysed using the Pearson correlation coefficient. For all samples, the patient code is listed above each sample (see table 2.2 for full patient information). Western blot quantification data are shown as mean +/- standard deviation; ns: not significant and ****$P < 0.0001$.

5.7. Discussion

Unexpectedly, poly(PR) levels were significantly higher in sALS patient motor cortex white matter samples, although this increase was predominately driven by a single sample. There are exactly 24 hexanucleotide sequences which could potentially encode poly(PR) proteins, G_4C_2 is only 1 of these. Therefore it may be beneficial to screen sALS patients for additional DNA repeat expansions in other genetic loci which may also be causative of ALS/FTD. It is noteworthy that the nature of DNA repeat expansions made *C9orf72* hexanucleotides difficult for researchers to detect with traditional techniques and therefore allowed them to remain hidden for many years.

A wider variety of DPR species were detectable in cerebellum grey matter samples from *C9orf72* patients (poly(GA+GP+PR) detectable) vs motor cortex white matter samples from *C9orf72* patients (only poly(GA) detectable). Additionally, poly(GA) levels in cerebellum (300 fold higher than sALS cerebellum samples) were markedly higher than levels in motor cortex (8 fold higher than sALS cerebellum samples). This is consistent with other reports of higher DPR expression in the cerebellum compared with other brain regions (Mackenzie et al., 2013, Mann et al., 2013, Baborie et al., 2015, Davidson et al., 2016).

Despite substantial DPR load, the number of cerebellar Purkinje neurons was not decreased in *C9orf72* patient brains. Purkinje cells were originally selected as a readout of cerebellar neurodegeneration as they are the largest neurons in the cerebellum, and therefore may be

more susceptible to DPR mediated dysregulation. The theory that large cell types may have increased susceptibility to certain pathologies has been suggested in an attempt to explain why motor neurons are selectively lost in ALS cases. However, in the case of the *C9orf72*-ALS patients studied presently, large Purkinje neurons suffered no decrease in cell number despite DPR loads in local regions. Previous studies indicate that multiple DPR species are detectable in molecular, granular and Purkinje cells of the cerebellum of C9orf72 patients (Mackenzie et al., 2013, Mackenzie et al., 2015, Saberi et al., 2017). This may suggest that resistance of cerebellar neurons to DPR mediated pathology may be conferred due to epigenetic differences in gene regulation in these types of neurons in comparison with ALS/FTD affected motor neurons and pyramidal neurons. The precise gene expression profile which may confer DPR resistance to cerebellar neurons may be investigated in the future by using laser capture followed by single cell RNA sequencing to determine transcriptomic profile differences between Purkinje cells and motor neurons. Additionally, counting of cerebellar granule or molecular cell layer neurons may be informative as to whether sub-populations of neurons in the cerebellum are differentially susceptible to C9orf72 expansion mediated toxicity.

Upregulation of HMW-SOD1 and fragmented TDP-43 was also discovered in motor cortex grey and white matter from *C9orf72*-ALS patients. The nature of these species is discussed at length in chapter 6. Most interestingly, the abundance of HMW-SOD1 strongly correlated with levels of fragmented TDP-43 in all tissue types tested, including non-neurological control tissue and tissue from disease unaffected areas. This correlation therefore appears to be fundamental and occurs regardless of other pathological features. This suggests that fragmented TDP-43 may induce expression of HMW-SOD1 or vice versa (also discussed further in chapter 6). However, it is important at this point not to rule out that there may be a third, currently unknown species, which drives expression of both fragmented TDP-43 and HMW-SOD1 simultaneously.

6.   Chapter 6: Discussion

6.1.   RNA foci vs DPR toxicity

We have generated two distinct types of C9orf72 zebrafish, these are RNA-only and RNA+DPR zebrafish models. Our RNA-only zebrafish express RNA foci and GFP protein (GFP detectable in embryonic and adult tissues). Despite sustained transgene expression, no DPRs could be detected in RNA-only zebrafish at adulthood. In this respect, RNA-only zebrafish were successful in segregating RNA foci and DPR expression. Additionally, the highest expressing RNA-only zebrafish did not show any reduction in survival or underlying reduction in swimming endurance at 26 months (although a mild hyper activity phenotype was detected at 5dpf). These data are consistent with previous reports indicating that RNA foci expression alone is not sufficient to produce a toxic phenotype in drosophila models (Mizielinska et al., 2014, Tran et al., 2015). Conversely, another report identified RNA foci mediated toxicity in zebrafish even when the absence of DPRs was confirmed (Swinnen et al., 2018). These data suggest that C9orf72 expansion RNA may have to be expressed above a certain threshold for toxicity to be observed.

The second type of zebrafish generated were RNA+DPR zebrafish. These zebrafish showed RNA foci and RAN-translated DPR expression. RNA+DPR zebrafish also exhibited increased average speed of movement and increased total distance moved at 5dpf in comparison to NTG zebrafish. The identification of a hyperactivity phenotype in both RNA-only and RNA+DPR zebrafish at 5dpf was unexpected. The cause of this early hyperactivity phenotype remains unclear, although motor neuron hyperexcitability has been previously reported in the early stages of ALS and is thought to contribute to the fasciculations characteristic of ALS onset in many patients (Iwai et al., 2016). RNA+DPR zebrafish also exhibited centre avoidance behaviour at 5dpf. This suggests an unwillingness to explore the environment and is consistent with similar results from open field tests conducted in RNA foci and DPR expressing *C9orf72* BAC transgenic mice (Liu et al., 2016). These features then progressed to reduced body mass, reduced swimming endurance, motor neuron loss, muscle atrophy and early mortality. These data are consistent with the motor neuron loss and early mortality reported in a BAC transgenic mouse model expressing RNA foci and DPR species (Liu et al., 2016), as well as the motor deficits reported in two independently generated virally transduced mouse models expressing RNA foci and DPR species (Chew et al., 2015, Herranz-Martin et al., 2017).

It is important to highlight that RNA-only and RNA+DPR zebrafish have different C9orf72 expansion orientations (G_4C_2 and C_4G_2 respectively) and different expansion lengths (99 repeats and 89 confirmed repeats respectively). The presence of uncontrolled variables makes direct comparison of RNA-only and RNA+DPR zebrafish more difficult. For example,

comparison of RNA foci abundance and distribution between the two lines was not possible as two different in situ hybridisation RNA probes with potentially different affinities/specificities were required. For these reasons it is difficult to make precise comparisons between the molecular pathologies of the RNA-only and RNA+DPR zebrafish reported here. However, it is reasonable to conclude that within our zebrafish models, expression of RNA foci + DPR species cause a more severe phenotype than expression of RNA foci alone.

Other *in vivo* model data supports the conclusion that RNA foci + DPR expression is more potently toxic than RNA foci alone (Mizielinska et al., 2014, Tran et al., 2015, Ohki et al., 2017). It is possible that RNA foci contribute towards toxicity to a lesser extent than DPRs, or that RNA foci and DPRs act additively/synergistically to generate toxicity (2 hits theory). Certainly, previous data in zebrafish models suggest that RNA foci and DPR may both contribute towards toxicity via independent mechanisms (Swinnen et al., 2018). It is important to note that use of alternate codons have allowed DPR species to be encoded without use of G_4C_2 C9orf72 RNA (Mizielinska et al., 2014). However, alternate codons encoding DPR species still require expression of repetitive exogenous RNA species, and therefore do not fully eliminate the potential of RNA mediated toxicity. Therefore, data presented here and previously, indicate that DPR and RNA foci are both toxic in C9orf72 *in vivo* models, with DPR presence usually correlating with a more severe phenotype.

6.2.   Zebrafish DPR expression profiles

Western blotting of RNA+DPR zebrafish lysates revealed that multiple DPR species were present, producing a laddered appearance. Both sense and antisense DPRs were detected, and DPRs could be produced via both conventional ATG-driven translation and RAN-translation. Detection of species of varying MW has also been reported during RAN-translation of CAG repeats (Zu et al., 2011), and during RAN-translation of GGGGCC in C9-ALS patients (Zu et al., 2013). In both previously mentioned cases the RAN-translation products formed a smear rather than distinct laddered protein bands as in our C9-zebrafish. The reason for this discrepancy may be due to the interruptions in the CCCCGG sequence used to generate our model causing the translation machinery to uncouple, thus consistently producing protein species of the same size. It is also possible that due to instability of the repeat DNA, multiple concatemerised transgene copies may contain variable repeat lengths and could therefore be producing multiple RNA transcripts of variable length, and thus multiple DPRs of variable length. Certainly, this reasoning may also explain why different DPR band expression profiles are observed between 2.2-2 and 2.2-7 zebrafish.

More RAN-translation mediated bands were detected in adult 2.2-7 zebrafish CNS compared to adult 2.2-2 zebrafish CNS. Additionally, poly(GA) expression was only detected in 5dpf 2.2-7 zebrafish, suggesting that bidirectional transcription is only occurring in 2.2-7 zebrafish. It is not currently known whether bi-directional transcription and increased formation of RAN-translation products during adulthood may be partially responsible for the increased toxicity observed in the 2.2-7 zebrafish line. Activation of cellular stress pathways has been shown to potentiate RAN-translation, potentially forming a feedforward loop (Green et al., 2017, Cheng et al., 2018). Given that heat shock stress activation is higher in 2.2-7 zebrafish, this may potentiate RAN-translation thus resulting in the observed increased number of bi-directionally transcribed and RAN-translated products, in comparison to 2.2-2 zebrafish. Whether activation of the heat shock response influences RAN-translation efficiency in 2.2 zebrafish is an interesting line of investigation for the future. Interestingly, poly(PR) species were detected at higher MWs than other DPR species. In the future, it may be important to investigate whether poly(PR)s tendency to form high MW species is related to its potent *in vivo* toxicity.

6.3. Use of interrupted C9orf72 expansions

Both RNA+DPR model zebrafish lines contained an interrupted C_4G_2 transgene construct. The advantages in using this construct are: **1)** Only anti-sense RNA foci have been found to co-localise with TDP-43 mislocalisation (a common molecular hallmark of neurodegeneration in ALS) (Cooper-Knock et al., 2015b). **2)** Anti-sense repeats will produce a greater proportion of the highly toxic proline-arginine repeat peptide (Mizielinska et al., 2014) and **3)** Interruptions in the expansion allow for largely equal expression levels of all three anti-sense DPR proteins.

There are also potential disadvantages of using interrupted C9orf72 expansions. Interruptions may increase the biochemical stability of DNA/RNA in comparison to pure repeat expansion DNA/RNA. Although, interrupted C9orf72 expansions and pure repeat expansions are reported to produce similar levels of RNA foci, it has also been noted that larger expansions are more stable when interruptions are present (Mizielinska et al., 2014). This suggests that interruptions in C9orf72 expansions may alter DNA/RNA structure, which could potentially influence the affinity with which RNA foci sequester RNA-binding proteins. Additionally, there is conflicting evidence as to whether interrupted expansions increase or decrease genomic instability once incorporated into genomic DNA in a cellular setting, which further suggests that there are fundamental differences in the biochemical properties of interrupted and pure repeat expansions (Ananda et al., 2014, Landrian et al., 2017). A further disadvantage of the use of interrupted repeats is that frameshifting interruptions will

result in all three DPR species (either 3 sense orientation or 3 antisense orientation DPRs) being encoded into a single protein. DPR species have distinct biochemical properties. In drosophila models where DPR proteins are individually expressed, poly(PR) localises to the nucleus, whereas poly(PA) and poly(GP) primarily localise to the cytoplasm (Mizielinska et al., 2013, Tran et al., 2015). In contrast, in the 2.2 zebrafish reported in this thesis, all three antisense DPR species are expressed in a single protein, and all three antisense DPRs are primarily localised to the nucleus (likely due to the biochemical nature of the poly(PR) containing regions of this protein). This is an example of how frameshifting interruptions can result in the production of protein species which have distinct characteristics from protein species encoded by pure repeat expansions.

Due to the technical difficulty of sequencing large DNA repeat expansions, it is currently unknown whether C9orf72 expansion bearing patients possess interrupted repeats or pure repeats. Notably, other repeat expansion disorders such as spinocerebellar ataxia are known to contain interruptions in some cases (Chung et al., 1993). Ultimately, improvements in DNA expansion sequencing technologies which allow the presence or absence of interruptions in *C9orf72* patient repeats to be confirmed will likely inform whether future C9orf72 expansion models should contain interrupted or pure C9orf72 expansions.

6.4. Gain of function or loss of function mechanisms as drivers of *C9orf72-ALS/FTD*

Both RNA foci and DPR protein production are gain of function mechanisms of potential toxicity in C9orf72 expansion ALS. However, an additional toxic mechanism proposed for C9orf72 expansion ALS/FTD, is loss of function of C9orf72 endogenous gene products leading to haploinsufficiency. This mechanism was first proposed after the observation that mRNA encoding C9orf72 endogenous protein products was reduced in expansion bearing patients (DeJesus-Hernandez et al., 2011). Furthermore, recent mass spectrometry analysis revealed decreased expression of the long isoform of the C9orf72 protein in the frontal cortex of C9orf72 expansion harbouring FTD patients (Viode et al., 2018). Loss of C9orf72 protein function in two transient zebrafish models, and a stable *C9orf72*-knockout C. elegans model, all resulted in motor neuron pathologies (Ciura et al., 2013, Therrien et al., 2013, Yeh et al., 2018). Conversely, early data from stable *C9orf72* knockout zebrafish suggest that no neurodegenerative or motor phenotype is observed (Stepto et al., 2014). Indeed, it has been reported that the transient knockdown phenotype observed in zebrafish is often not consistent with the phenotype exhibited upon stable knockout of the same gene (Kok et al., 2015). Furthermore, multiple *C9orf72* loss of function mouse models did not display any ALS/FTD phenotype (Koppers et al., 2015, Atanasio et al., 2016, Jiang et al.,

2016, O'Rourke et al., 2016, Sudria-Lopez et al., 2016, Ji et al., 2017). Moreover, multiple C9orf72 expansion gain of function mouse models have accurately recapitulated key disease features of ALS/FTD (O'Rourke et al., 2015, Liu et al., 2016, Herranz-Martin et al., 2017, Walker et al., 2017). Taken together, the conflicting data from *in vivo* models of *C9orf72* loss of function, and the success of gain of function mouse models in recapitulating key features of ALS/FTD have caused haploinsufficiency of the C9orf72 protein to fall out of favour as a proposed driving mechanism of ALS/FTD.

Some researchers argue that, although haploinsufficiency of the C9orf72 protein is unlikely to drive ALS/FTD progression, it may be a modifier of RNA/DPR toxicity (Webster et al., 2016b, Nassif et al., 2017, Shi et al., 2018). The C9orf72 protein itself has been identified as being crucial in the initiation of autophagy (Farg et al., 2014, Sellier et al., 2016, Webster et al., 2016a). Importantly, *C9orf72* knockout mice have been reported to show impaired autophagy functioning (Ji et al., 2017), and impaired autophagy has been associated with ALS causative genetic mutations in *SQSTM1* and *UBQLN2* genes (Goode et al., 2016, Osaka et al., 2016). Taken together, these data form the basis of the argument that haploinsufficiency of the C9orf72 protein may cause defects in normal autophagy processing which can potentially augment RNA/DPR mediated toxicity. In line with this, induced motor neurons from *C9orf72* patients, exhibit reduced autophagy-mediated clearing of DPR aggregates, which has been identified as contributing toward neurodegeneration (Shi et al., 2018). Moreover, knockout of the mouse orthologue of the *C9orf72* gene in the background of a previously described BAC transgenic mouse expressing human C9orf72 expansions with flanking regulatory regions (Liu et al., 2016), caused a more severe motor phenotype in these mice (Shao et al., 2019). However, most *C9orf72* patients are heterozygous for the hexanucleotide expansion, and therefore retain a single wildtype functional copy of the *C9orf72* gene. Therefore, patients homozygous for C9orf72 hexanucleotide expansions may be expected to show an even greater reduction of C9orf72 endogenous protein and therefore more severe autophagy defects, thus modifying RNA/DPR toxicity to produce more severe clinical outcomes. However, a case report of a single patient with homozygous C9orf72 expansions displayed a clinical profile which was within the expected pattern of disease progression of heterozygous patients, thus suggesting that homozygous reduction of C9orf72 protein production did not further augment RNA/DPR toxicity in this case (Fratta et al., 2013). A possible experiment to resolve this issue, would be to measure C9orf72 protein/mRNA levels in disease affected areas of post-mortem ALS/FTD patient CNS tissue, and test whether C9orf72 expression levels correlate with severity of clinical outcome.

6.5. Clinicopathological correlations of DPRs in ALS/FTD

A significant obstacle in our understanding of the driving pathomechanisms of *C9orf72*-ALS/FTD is the relative lack of DPR pathology in disease affected regions. *In vivo* models of *C9orf72*-ALS/FTD, including the present study, have clearly identified that DPR species are more toxic than RNA foci alone (Mizielinska et al., 2014, Tran et al., 2015, Ohki et al., 2017). However, multiple neuropathological studies have identified that DPR burden in disease affected areas (spinal cord, motor cortex, frontal cortex and temporal cortex) is typically lower in comparison to disease unaffected areas (basal ganglia and cerebellum) of *C9orf72*-ALS/FTD patients (Mackenzie et al., 2013, Mann et al., 2013, Baborie et al., 2015, Davidson et al., 2016). A potential explanation is that neurons exhibiting substantial DPR burden in disease affected areas may have completely degenerated when end-stage ALS/FTD is reached, and are therefore not detectable in post-mortem tissues. This argument however does not explain why areas such as the cerebellum have a substantial DPR load, yet in the present study and a previous study, cerebellar DPR load did not correlate with significant neurodegeneration (Mackenzie et al., 2013). As discussed previously, this may indicate a difference in the ability of different neuronal sub-populations to resist DPR-mediated toxicity. Investigation of the transcriptomic and epigenetic differences between affected and unaffected CNS regions in ALS/FTD patients is an attractive option to highlight pathways for potential therapeutic modulation. More recently, poly(GR) DPR species have been identified to correlate with neurodegeneration in C9-ALS patients (Saberi et al., 2017, Sakae et al., 2018). Additionally, the majority of poly(GR) DPR inclusions were also positive for TDP-43 (Saberi et al., 2017). TDP-43 positive inclusions have also previously been shown to correlate with the extent of neurodegeneration in *C9orf72*-ALS/FTD post-mortem tissue (Mackenzie et al., 2013). Taken together, these data indicate that most DPR species show poor spatial correlation with neurodegeneration and TDP-43 proteinopathy, with the exception of poly(GR). This suggests that further studies should be directed towards studying poly(GR) specific mechanisms of toxicity.

In order to compare the spatial distribution of various DPR species, most neuropathological studies rely heavily on immunohistochemical staining and occasionally western blotting using standard protocols (Mackenzie et al., 2013, Mann et al., 2013, Baborie et al., 2015, Mackenzie et al., 2015, Davidson et al., 2016, Saberi et al., 2017). Whilst analysing CNS DPR levels in the present study, both immunohistochemical staining and western blotting were attempted, but both techniques resulted in difficult to interpret results in which control samples showed apparently high levels of DPR-like inclusions and multiple protein bands. Eventually, dot blotting tissue lysates under a vacuum was utilised as this method was able to detect significantly higher signal in C9orf72 patient cerebellum samples compared to

sALS or non-neurological control samples. This may be because dot blotting removes the opportunity for DPR species to become trapped in western blot gels during electrophoresis. However, dot blotting of samples also has limitations due to an inability to separate specific and non-specific protein species bound by DPR antibodies (see section 5.7 for a full discussion of limitations of the dot blot technique). The limitations of current DPR detection techniques highlights the need for the development of more sensitive DPR detection methods. More sensitive detection methods will help reveal whether disease affected areas such as spinal cord may be expressing low levels of DPR species. Additionally, multiple lysis buffers were tested when optimising DPR extraction from both zebrafish and human tissues. Conventional RIPA or even high triton X-100 buffers were found to be inadequate when attempting to extract DPR species. Furthermore, DPR species were found to readily pellet out during centrifugation. For these reasons, tissue was lysed directly in loading buffer and no centrifugation step was used (see section 2.2.7 for lysis method details). These techniques allowed much more sensitive detection of both zebrafish and human DPRs and may therefore be worthwhile to be adopted in future studies in replacement of standard lysis protocols.

6.6. The heat shock response in ALS

The heat shock response involves the up-regulation of heat shock proteins in response to the presence of aberrant cellular proteins in an attempt to refold, or target these aberrant proteins for degradation (Dedmon et al., 2005, Bukau et al., 2006). Activation of the heat shock response was observed to some degree in both RNA-only and RNA+DPR zebrafish. Normalising for transgene expression levels by comparing ratios of DsRed:GFP between lines showed that both RNA+DPR zebrafish had significantly higher hsp70 promoter mediated production of DsRed, as compared with all RNA-only zebrafish lines. These data suggest that DPRs drive heat shock activation more potently than RNA foci alone. Additionally, activation of the heat shock response was also detected in post-mortem cerebellum samples from *C9orf72*-ALS/FTD patients. This is in line with previous reports, which identified activation of the heat shock response at the RNA level, in both cerebellum and frontal cortex of *C9orf72*-ALS/FTD patients (Prudencio et al., 2015, Mordes et al., 2018). Furthermore, the ALS causative G93R mutant SOD1 protein causes activation of the heat shock response in zebrafish (McGown et al., 2013).

The G93R *SOD1*-ALS model zebrafish described previously also contained the same hsp70:DsRed drug screening construct used in the C9orf72 zebrafish reported in this thesis (Ramesh et al., 2010). In the *SOD1* zebrafish, treatment with riluzole (a disease modifying ALS treatment) ameliorated the extent of heat shock response activation (McGown et al.,

2013). Therefore, in *SOD1* zebrafish, heat shock promoter mediated DsRed production was used as a disease biomarker against which to screen potentially therapeutic compounds. Approximately 2000 compounds have now been screened using these *SOD1*-ALS zebrafish models, and two of the top hits identified were the compounds ivermectin and selamectin (Ramesh lab, unpublished data). Taking this into consideration, riluzole, ivermectin and selamectin were also tested in C9orf72 model zebrafish, and were shown to ameliorate heat shock promoter mediated DsRed production. Ivermectin has previously been administered to *SOD1*-ALS transgenic mouse models by an independent lab, and was found to increase survival in these mice (Andries et al., 2007). These data further demonstrate that compounds identified by hsp70:DsRed screening in zebrafish show genuine potential as prospective ALS treatments. Ivermectin has been demonstrated to both ameliorate excitotoxicity and block specific nuclear importers, although the precise mechanism of the therapeutic properties of ivermectin in ALS models is not currently known (Andries et al., 2007, Wagstaff et al., 2012). Determining the mechanism of action of both selamectin and ivermectin in ALS models will be an important next step, which would allow additional compounds acting via similar mechanisms to be assessed in ALS models.

However, there are also limitations to the use of the heat shock response as a drug screening readout for ALS. As the drug screening readout relies upon hsp70 promoter mediated production of DsRed, it is possible that hits identified in each drug screen may mediate their effect by directly blocking heat shock response activation (through blocking transcription factor binding at the hsp70 promoter or by other means), rather than by reducing cellular stress. Similarly, it is also possible that increased heat shock activation in ALS models and patient tissue is caused by disease associated mechanisms which are not directly related to disease progression. In this case, drugs which reduce heat shock response activation would be unlikely to have significant therapeutic benefit to ALS patients. Both of the above mentioned limitations of the hsp70:DsRed drug screening system may be addressed by secondary validation of identified drugs using a phenotypic readout. A simple paradigm which may address the limitations of the hsp70:DsRed system, would be to conduct higher throughput screening of compounds using the hsp70:DsRed system in zebrafish, and then to validate the top hits from these screens using survival, or time until onset of motor symptoms in ALS model zebrafish or mice as a phenotypic readout.

In summary, activation of the heat shock response in *C9orf72*-ALS patient post-mortem tissue has been reported by two independent sources, and amelioration of heat shock activation by treatment with riluzole, ivermectin or selamectin can be achieved in *in vivo* models of two distinct genetic sub-types of ALS. These data suggest that heat shock response activation may be a robust readout for identifying hits from compound libraries in

C9orf72 and SOD1-ALS zebrafish, before secondary validation of identified hits by phenotypic readouts.

6.7.   DNA damage and neurodegeneration in ALS

The DNA damage response associated proteins PARP1 and TDP1 were identified as potential targets for investigation in *C9orf72*-ALS/FTD pathology due to their known role in other neurodegenerative disorders (El-Khamisy et al., 2005, Hoch et al., 2017), and the known involvement of DNA damage in *C9orf72*-ALS/FTD (Lopez-Gonzalez et al., 2016, Farg et al., 2017, Walker et al., 2017). We did not find any evidence for over-activation of PARP1 driving *C9orf72*-ALS/FTD pathogenesis. Unexpectedly, PARP1 inhibition with olaparib had a detrimental effect on the survival of 2.2-7 C9orf72 zebrafish, and seemed to be selectively toxic to 2.2-7 zebrafish and not NTG zebrafish at lower doses. These data indicate that PARP1's crucial role in recruiting DNA repair factors to local DNA lesions was likely protective in C9orf72 zebrafish rather than detrimental as previously thought (Ciccia and Elledge, 2010). Moreover, we did not find evidence of loss of TDP1 function, the reported cause of neurodegeneration in SCAN-1 patients (El-Khamisy et al., 2005). Rather, TDP1 protein was found to be upregulated in grey matter samples of cerebellum and motor cortex in *C9orf72* patients and in white matter samples from motor cortex of sALS patients. Taken together, these data support the notion that C9orf72 expansions cause DNA damage, and also suggest that a component of DNA damage may be present in sALS patient tissue. Finally, over activation of PARP1 or loss of TDP1 function are not likely candidates as drivers of ALS/FTD pathogenesis.

6.8.   Relationship between C9orf72 expansions, TDP-43 and SOD1

A contentious question that is raised within ALS research, is whether a common mechanism exists which accounts for the pathogenesis of multiple ALS-subtypes. Nearly all sporadic and familial (except *SOD1*-ALS and *FUS*-ALS) ALS cases show ubiquitinated cytoplasmic TDP-43 positive aggregates which co-localise with disease affected areas (Arai et al., 2006, Neumann et al., 2006, Mackenzie et al., 2007, King et al., 2015). However, the role of SOD1 in non-*SOD1* ALS is more controversial, with multiple studies reporting misfolded/aggregated SOD1 in spinal cord tissue from sALS patients (Bosco et al., 2010, Forsberg et al., 2010, Pokrishevsky et al., 2012, Grad et al., 2014), and multiple other studies refuting this finding (Kerman et al., 2010, Brotherton et al., 2012, Liu et al., 2012, Ayers et al., 2014, Da Cruz et al., 2017).

In the present study, HMW-SOD1 protein was found to be significantly more abundant in post-mortem white and grey matter motor cortex of *C9orf72*-ALS patients. This is in line with

a recent study which identified misfolded SOD1 staining in the CNS of 15 C9orf72-ALS/FTD cases (10/15 of these cases showed positive misfolded SOD1 staining in the motor cortex) (Forsberg et al., 2019). The reason why SOD1 forms HMW species in C9orf72 patient motor cortex is not known. Although, formation of high molecular weight species has also been reported in cell models expressing multiple variants of mutant SOD1 protein (Brown and Borchelt, 2014). HMW-SOD1 species also showed a non-significant increase in sALS patient samples from motor cortex white matter. Given that the increase in HMW-SOD1 in *C9orf72* patient samples was much clearer in grey matter motor cortex, it will be vital to examine HMW-SOD1 expression in this region of sALS patients also. Additionally, we confirmed a previous report of increased abundance of fragmented TDP-43 in the frontal cortex of *C9orf72*-ALS patients (Tsuji et al., 2012). Unexpectedly, fragmented TDP-43 was not increased in white matter motor cortex samples from sALS patients, and future investigation of grey matter samples from these patients will be of vital importance. Importantly, both HMW-SOD1 and fragmented TDP-43 pathology were only more abundant in the motor cortex of *C9orf72* patients, and not in the cerebellum, indicating that both proteinopathies were specific to disease affected areas of *C9orf72*-ALS patients.

Interestingly, in all analysed post-mortem tissues, there was a strong positive correlation between fragmented TDP-43 and HMW-SOD1, thus suggesting that there may be a fundamental association between the two species. It has previously been reported that TDP-43 proteinopathy can induce wtSOD1 pathology in cultured cells (Pokrishevsky et al., 2012, Pokrishevsky et al., 2016). Certainly, this mechanism may explain why TDP-43 pathology is rarely observed in fALS patients harbouring a mutation in the *SOD1* gene (Mackenzie et al., 2007). As a *SOD1* mutation may be sufficient to alter SOD1 protein conformation, TDP-43 proteinopathy may no longer be a prerequisite for the initiation of SOD1 proteinopathy. It is important to note that, the presence of wtSOD1 pathology in ALS remains highly controversial, and whether TDP-43 proteinopathy may induce wtSOD1 pathology in ALS patients would require much further investigation of the temporal relationship between the two proteinopathies. Furthermore, correlation of fragmented TDP-43 and HMW SOD1 protein levels does not prove that one species induces the expression of the other, as it is also possible that another currently unknown process may drive expression of both fragmented TDP-43 and HMW SOD1 simultaneously. The finding of SOD1 pathology in *C9orf72* patients is of particular importance, due to the ongoing *SOD1* antisense oligonucleotide clinical trials being conducted in ALS patients with *SOD1* mutations (clinical trial identifier: NCT03070119). It remains to be seen whether manipulation of SOD1 protein levels in patients with *C9orf72*-ALS may confer therapeutic benefit.

A final note on SOD1 and TDP-43 proteinopathies in ALS, is that previous studies have relied very heavily upon the use of immunohistochemical staining and misfolded SOD1 specific antibodies. In the present study we used a conventional SOD1 antibody (not misfolded-SOD1 specific) raised against amino acids 1-100 of the wtSOD1 protein to detect pathological HMW-SOD1. Previous studies have indicated that use of misfolded SOD1 specific antibodies would be unlikely to detect these HMW-SOD1 species, even when native gels are utilised to maintain endogenous protein structure (Brown and Borchelt, 2014). Additionally, western blotting of protein samples is a much more efficient and quantitative method to use when analysing the extent of TDP-43 proteinopathy. Furthermore, use of misfolded SOD1 specific antibodies has led to the publication of conflicting data. For these reasons, future studies could also include western blotting with conventional antibodies as a potentially more quantitative method of assessing both SOD1 and TDP-43 proteinopathy.

6.9.   Conclusions

The body of *in vivo* and *in vitro* evidence from various C9orf72 expansion models is now clearly pointing toward DPRs being the most toxic species in *C9orf72*-ALS/FTD. RNA foci formation almost certainly contribute to toxicity in *C9orf72* models also, but the presence of DPR species are often required for toxicity to be observed, suggesting a second hit may be required in order to overwhelm the cells compensatory mechanisms. Future therapies such as antisense oligonucleotides against RNA foci are a very attractive option, as they would allow for the simultaneous reduction of both RNA and DPR burden. Haploinsufficiency of the C9orf72 protein seems unlikely as a disease driving mechanism and does not represent an easily druggable target. As such therapies addressing loss of function of the C9orf72 protein may be a less valuable prospect. The obstacle of DPR protein expression not correlating with disease affected CNS regions is slowly being overcome, as careful analysis of poly(GR) has highlighted its correlation with disease affected regions and TDP-43 proteinopathy. The development of more sensitive detection methods for DPR species will be required before low level expression of DPRs in disease affected areas can be entirely ruled out.

Activation of the heat shock response is a feature of both *SOD1* and *C9orf72*-ALS. Additionally, heat shock activation has successfully been used to identify riluzole as a hit in both *SOD1* and C9orf72 zebrafish models, thus validating heat shock activation as a good quality readout against which potential therapeutic compounds can be screened.

C9orf72 expansions drive fragmentation of TDP-43 and SOD1 proteinopathy, and increased abundance of these species was detected in motor cortex, but not in cerebellum of *C9orf72*-ALS patients. Although the study of TDP-43 and SOD1 proteinopathy as a common ALS

mechanism seems to have fallen out of favour recently, future studies should carefully consider the roles of these pathological species. If indeed it exists, finding a common mechanism for ALS pathogenesis may allow for the development of neuroprotective therapies targeted against multiple sub-types of ALS/FTD.

8. Bibliography

Abe, K., Aoki, M., Tsuji, S., et al. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol,* 16**,** 505-512. (2017)

Al-Sarraj, S., King, A., Cleveland, M., et al. Mitochondrial abnormalities and low grade inflammation are present in the skeletal muscle of a minority of patients with amyotrophic lateral sclerosis; an observational myopathology study. *Acta Neuropathol Commun,* 2**,** 165. (2014)

Al-Sarraj, S., King, A., Troakes, C., et al. 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 Neuropathol,* 122**,** 691-702. (2011)

Almeida, S., Gascon, E., Tran, H., et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathologica,* 126**,** 385-399. (2013)

Ananda, G., Hile, S. E., Breski, A., et al. Microsatellite interruptions stabilize primate genomes and exist as population-specific single nucleotide polymorphisms within individual human genomes. *PLoS Genet,* 10**,** e1004498. (2014)

Andries, M., Van Damme, P., Robberecht, W., et al. Ivermectin inhibits AMPA receptor-mediated excitotoxicity in cultured motor neurons and extends the life span of a transgenic mouse model of amyotrophic lateral sclerosis. *Neurobiol Dis,* 25**,** 8-16. (2007)

Anichtchik, O., Diekmann, H., Fleming, A., et al. Loss of PINK1 function affects development and results in neurodegeneration in zebrafish. *J Neurosci,* 28**,** 8199-207. (2008)

Anzai, I., Tokuda, E., Mukaiyama, A., et al. A misfolded dimer of Cu/Zn-superoxide dismutase leading to pathological oligomerization in amyotrophic lateral sclerosis. *Protein Sci,* 26**,** 484-496. (2017)

Arai, T., Hasegawa, M., Akiyama, H., et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun,* 351**,** 602-11. (2006)

Arai, T., Nonaka, T., Hasegawa, M., et al. Neuronal and glial inclusions in frontotemporal dementia with or without motor neuron disease are immunopositive for p62. *Neurosci Lett,* 342**,** 41-4. (2003)

Armon, C. An evidence-based medicine approach to the evaluation of the role of exogenous risk factors in sporadic amyotrophic lateral sclerosis. *Neuroepidemiology,* 22**,** 217-28. (2003)

Armstrong, G. A. & Drapeau, P. Loss and gain of FUS function impair neuromuscular synaptic transmission in a genetic model of ALS. *Hum Mol Genet,* 22**,** 4282-92. (2013)

Ascherio, A., Weisskopf, M. G., O'Reilly E, J., et al. Vitamin E intake and risk of amyotrophic lateral sclerosis. *Ann Neurol,* 57**,** 104-10. (2005)

Ash, P. E., Bieniek, K. F., Gendron, T. F., et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron,* 77**,** 639-46. (2013)

Atanasio, A., Decman, V., White, D., et al. C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. *Sci Rep,* 6**,** 23204. (2016)

Auer, T. O. & Del Bene, F. CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods,* 69**,** 142-50. (2014)

Ayers, J. I., Fromholt, S., Koch, M., et al. Experimental transmissibility of mutant SOD1 motor neuron disease. *Acta Neuropathol,* 128**,** 791-803. (2014)

Baborie, A., Griffiths, T. D., Jaros, E., et al. 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**,** 601-12. (2015)

Baker, M., Mackenzie, I. R., Pickering-Brown, S. M., et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature,* 442**,** 916-9. (2006)

Baldwin, K. R., Godena, V. K., Hewitt, V. L., et al. Axonal transport defects are a common phenotype in Drosophila models of ALS. *Hum Mol Genet,* 25**,** 2378-2392. (2016)

Bannwarth, S., Ait-El-Mkadem, S., Chaussenot, A., et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain,* 137**,** 2329-45. (2014)

Beard, J. D. & Kamel, F. Military service, deployments, and exposures in relation to amyotrophic lateral sclerosis etiology and survival. *Epidemiol Rev,* 37**,** 55-70. (2015)

Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development,* 105**,** 733-7. (1989)

Bennett, C. L., Dastidar, S. G., Ling, S. C., et al. Senataxin mutations elicit motor neuron degeneration phenotypes and yield TDP-43 mislocalization in ALS4 mice and human patients. *Acta Neuropathol,* 136**,** 425-443. (2018)

Bennion Callister, J., Ryan, S., Sim, J., et al. Modelling C9orf72 dipeptide repeat proteins of a physiologically relevant size. *Hum Mol Genet,* 25**,** 5069-5082. (2016)

Bensimon, G., Lacomblez, L. & Meininger, V. A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. *N Engl J Med,* 330**,** 585-91. (1994)

Bertrandy, S., Burlet, P., Clermont, O., et al. The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. *Hum Mol Genet,* 8**,** 775-82. (1999)

Bigio, E. H., Weintraub, S., Rademakers, R., et al. Frontotemporal lobar degeneration with TDP-43 proteinopathy and chromosome 9p repeat expansion in C9ORF72: clinicopathologic correlation. *Neuropathology,* 33**,** 122-33. (2013)

Bjorkoy, G., Lamark, T. & Johansen, T. p62/SQSTM1: a missing link between protein aggregates and the autophagy machinery. *Autophagy,* 2**,** 138-9. (2006)

Blackburn, J. S., Liu, S., Raimondi, A. R., et al. High-throughput imaging of adult fluorescent zebrafish with an LED fluorescence macroscope. *Nat Protoc,* 6**,** 229-41. (2011)

Blennow, K., de Leon, M. J. & Zetterberg, H. Alzheimer's disease. *Lancet,* 368**,** 387-403. (2006)

Boeynaems, S., Bogaert, E., Kovacs, D., et al. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol Cell,* 65**,** 1044-1055.e5. (2017)

Boon, K. L., Xiao, S., McWhorter, M. L., et al. Zebrafish survival motor neuron mutants exhibit presynaptic neuromuscular junction defects. *Hum Mol Genet,* 18**,** 3615-25. (2009)

Borchelt, D. R., Lee, M. K., Slunt, H. S., et al. Superoxide dismutase 1 with mutations linked to familial amyotrophic lateral sclerosis possesses significant activity. *Proc Natl Acad Sci U S A,* 91**,** 8292-6. (1994)

Borroni, B., Bonvicini, C., Alberici, A., et al. Mutation within TARDBP leads to frontotemporal dementia without motor neuron disease. *Hum Mutat,* 30**,** E974-83. (2009)

Bosco, D. A., Morfini, G., Karabacak, N. M., et al. Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. *Nat Neurosci,* 13**,** 1396-403. (2010)

Braak, H. & Del Tredici, K. Where, when, and in what form does sporadic Alzheimer's disease begin? *Curr Opin Neurol,* 25**,** 708-14. (2012)

Bradford, Y. M., Toro, S., Ramachandran, S., et al. Zebrafish Models of Human Disease: Gaining Insight into Human Disease at ZFIN. *Ilar j,* 58**,** 4-16. (2017)

Bradley, A., Evans, M., Kaufman, M. H., et al. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature,* 309**,** 255-6. (1984)

Branchu, J., Boutry, M., Sourd, L., et al. Loss of spatacsin function alters lysosomal lipid clearance leading to upper and lower motor neuron degeneration. *Neurobiol Dis,* 102**,** 21-37. (2017)

Bretaud, S., Allen, C., Ingham, P. W., et al. p53-dependent neuronal cell death in a DJ-1-deficient zebrafish model of Parkinson's disease. *J Neurochem,* 100**,** 1626-35. (2007)

Brettschneider, J., Toledo, J. B., Van Deerlin, V. M., et al. Microglial activation correlates with disease progression and upper motor neuron clinical symptoms in amyotrophic lateral sclerosis. *PLoS One,* 7**,** e39216. (2012)

Brotherton, T. E., Li, Y., Cooper, D., et al. Localization of a toxic form of superoxide dismutase 1 protein to pathologically affected tissues in familial ALS. *Proc Natl Acad Sci U S A,* 109**,** 5505-10. (2012)

Brown, H. H. & Borchelt, D. R. Analysis of mutant SOD1 electrophoretic mobility by Blue Native gel electrophoresis; evidence for soluble multimeric assemblies. *PLoS One,* 9**,** e104583. (2014)

Brown, R. H. & Al-Chalabi, A. Amyotrophic Lateral Sclerosis. *N Engl J Med,* 377**,** 162-172. (2017)

Brzustowicz, L. M., Lehner, T., Castilla, L. H., et al. Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3. *Nature,* 344**,** 540-1. (1990)

Bukau, B., Weissman, J. & Horwich, A. Molecular chaperones and protein quality control. *Cell,* 125**,** 443-51. (2006)

Bunina, T. L. [On intracellular inclusions in familial amyotrophic lateral sclerosis]. *Zh Nevropatol Psikhiatr Im S S Korsakova,* 62**,** 1293-9. (1962)

Burberry, A., Suzuki, N., Wang, J. Y., et al. Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci Transl Med,* 8**,** 347ra93. (2016)

Byrne, S., Heverin, M., Elamin, M., et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: a population-based case-control cohort study of familial and sporadic amyotrophic lateral sclerosis. *Ann Neurol,* 74**,** 699-708. (2013)

Byrne, S., Walsh, C., Lynch, C., et al. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry,* 82**,** 623-7. (2011)

Cautain, B., Hill, R., de Pedro, N., et al. Components and regulation of nuclear transport processes. *Febs j,* 282**,** 445-62. (2015)

Cedarbaum, J. M., Stambler, N., Malta, E., et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *J Neurol Sci,* 169**,** 13-21. (1999)

Chen, T., Huang, B., Shi, X., et al. Mutant UBQLN2(P497H) in motor neurons leads to ALS-like phenotypes and defective autophagy in rats. *Acta Neuropathol Commun,* 6**,** 122. (2018)

Chen, Y. Z., Bennett, C. L., Huynh, H. M., et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am J Hum Genet,* 74**,** 1128-35. (2004)

Cheng, W., Wang, S., Mestre, A. A., et al. C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation. *Nat Commun,* 9**,** 51. (2018)

Chew, J., Gendron, T. F., Prudencio, M., et al. Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science,* 348**,** 1151-1154. (2015)

Chió, A., Calvo, A., Mazzini, L., et al. Extensive genetics of ALS: A population-based study in Italy. *Neurology,* 79. (2012)

Chio, A., Logroscino, G., Traynor, B. J., et al. Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. *Neuroepidemiology,* 41**,** 118-30. (2013)

Chitramuthu, B. P., Baranowski, D. C., Kay, D. G., et al. Progranulin modulates zebrafish motoneuron development in vivo and rescues truncation defects associated with knockdown of Survival motor neuron 1. *Mol Neurodegener,* 5**,** 41. (2010)

Chitramuthu, B. P., Kay, D. G., Bateman, A., et al. Neurotrophic effects of progranulin in vivo in reversing motor neuron defects caused by over or under expression of TDP-43 or FUS. *PLoS One,* 12**,** e0174784. (2017)

Choi, S. Y., Lopez-Gonzalez, R., Krishnan, G., et al. C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. *Nat Neurosci,* Epub ahead of print. (2019)

Chow, C. Y., Landers, J. E., Bergren, S. K., et al. Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. *Am J Hum Genet,* 84**,** 85-8. (2009)

Chung, M. Y., Ranum, L. P., Duvick, L. A., et al. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet,* 5**,** 254-258. (1993)

Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol Cell,* 40**,** 179-204. (2010)

Cirulli, E. T., Lasseigne, B. N., Petrovski, S., et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science,* 347**,** 1436-41. (2015)

Ciura, S., Lattante, S., Le Ber, I., et al. Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Annals of Neurology,* 74**,** 180-187. (2013)

Cooper-Knock, J., Bury, J. J., Heath, P. R., et al. C9ORF72 GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis. *PLoS ONE,* 10**,** e0127376. (2015a)

Cooper-Knock, J., Higginbottom, A., Stopford, M. J., et al. Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathologica,* 130**,** 63-75. (2015b)

Cooper-Knock, J., Moll, T., Ramesh, T., et al. Mutations in the Glycosyltransferase Domain of GLT8D1 Are Associated with Familial Amyotrophic Lateral Sclerosis. *Cell Rep,* 26**,** 2298-2306.e5. (2019)

Cooper-Knock, J., Walsh, M. J., Higginbottom, A., et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain,* 137**,** 2040-51. (2014)

Couthouis, J., Hart, M. P., Erion, R., et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Hum Mol Genet,* 21**,** 2899-2911. (2012)

Couthouis, J., Hart, M. P., Shorter, J., et al. A yeast functional screen predicts new candidate ALS disease genes. *Proc Natl Acad Sci U S A,* 108**,** 20881-90. (2011)

Cox, L. E., Ferraiuolo, L., Goodall, E. F., et al. Mutations in CHMP2B in lower motor neuron predominant amyotrophic lateral sclerosis (ALS). *PLoS One,* 5**,** e9872. (2010)

Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., et al. Proteomic analysis of the mammalian nuclear pore complex. *J Cell Biol,* 158**,** 915-27. (2002)

Cruts, M., Gijselinck, I., van der Zee, J., et al. Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature,* 442**,** 920-924. (2006)

Da Costa, M. M., Allen, C. E., Higginbottom, A., et al. A new zebrafish model produced by TILLING of SOD1-related amyotrophic lateral sclerosis replicates key features of the disease and represents a tool for in vivo therapeutic screening. *Dis Model Mech,* 7**,** 73-81. (2014)

Da Cruz, S., Bui, A., Saberi, S., et al. Misfolded SOD1 is not a primary component of sporadic ALS. *Acta Neuropathol,* 134**,** 97-111. (2017)

Dadon-Nachum, M., Melamed, E. & Offen, D. The "dying-back" phenomenon of motor neurons in ALS. *J Mol Neurosci,* 43**,** 470-477. (2011)

Dantzer, F., Ame, J. C., Schreiber, V., et al. Poly(ADP-ribose) polymerase-1 activation during DNA damage and repair. *Methods Enzymol,* 409**,** 493-510. (2006)

Dantzer, F., de La Rubia, G., Menissier-De Murcia, J., et al. Base excision repair is impaired in mammalian cells lacking Poly(ADP-ribose) polymerase-1. *Biochemistry,* 39**,** 7559-69. (2000)

Davidson, Y., Robinson, A. C., Liu, X., et al. 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**,** 242-54. (2016)

De Strooper, B., Saftig, P., Craessaerts, K., et al. Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature,* 391**,** 387-90. (1998)

Dedmon, M. M., Christodoulou, J., Wilson, M. R., et al. Heat shock protein 70 inhibits alpha-synuclein fibril formation via preferential binding to prefibrillar species. *J Biol Chem,* 280**,** 14733-14740. (2005)

DeJesus-Hernandez, M., Finch, N. A., Wang, X., et al. In-depth clinico-pathological examination of RNA foci in a large cohort of C9ORF72 expansion carriers. *Acta Neuropathol,* 134**,** 255-269. (2017)

DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron,* 72**,** 245-56. (2011)

Deng, H. X., Chen, W., Hong, S. T., et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature,* 477**,** 211-5. (2011)

Deng, J., Yang, M., Chen, Y., et al. FUS Interacts with HSP60 to Promote Mitochondrial Damage. *PLoS Genet,* 11**,** e1005357. (2015)

Doetschman, T., Gregg, R. G., Maeda, N., et al. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature,* 330**,** 576-8. (1987)

Dormann, D., Rodde, R., Edbauer, D., et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *EMBO J,* 29**,** 2841-57. (2010)

El-Khamisy, S. F., Saifi, G. M., Weinfeld, M., et al. Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature,* 434**,** 108-13. (2005)

Elden, A. C., Kim, H. J., Hart, M. P., et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature,* 466**,** 1069-75. (2010)

Elshafey, A., Lanyon, W. G. & Connor, J. M. Identification of a new missense point mutation in exon 4 of the Cu/Zn superoxide dismutase (SOD-1) gene in a family with amyotrophic lateral sclerosis. *Hum Mol Genet,* 3**,** 363-4. (1994)

Factor-Litvak, P., Al-Chalabi, A., Ascherio, A., et al. Current pathways for epidemiological research in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener,* 14 Suppl 1**,** 33-43. (2013)

Fallini, C., Bassell, G. J. & Rossoll, W. Spinal muscular atrophy: the role of SMN in axonal mRNA regulation. *Brain Res,* 1462**,** 81-92. (2012)

Farg, M. A., Konopka, A., Soo, K. Y., et al. The DNA damage response (DDR) is induced by the C9orf72 repeat expansion in amyotrophic lateral sclerosis. *Hum Mol Genet,* 26**,** 2882-2896. (2017)

Farg, M. A., Sundaramoorthy, V., Sultana, J. M., et al. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol Genet,* 23**,** 3579-95. (2014)

Farrar, M. A. & Kiernan, M. C. The Genetics of Spinal Muscular Atrophy: Progress and Challenges. *Neurotherapeutics,* 12**,** 290-302. (2015)

Fasana, E., Fossati, M., Ruggiano, A., et al. A VAPB mutant linked to amyotrophic lateral sclerosis generates a novel form of organized smooth endoplasmic reticulum. *FASEB J,* 24**,** 1419-30. (2010)

Fecto, F., Yan, J., Vemula, S. P., et al. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol,* 68**,** 1440-1446. (2011)

Ferraiuolo, L., Meyer, K., Sherwood, T. W., et al. Oligodendrocytes contribute to motor neuron death in ALS via SOD1-dependent mechanism. *Proc Natl Acad Sci U S A,* 113**,** E6496-e6505. (2016)

Ferrari, R., Kapogiannis, D., Huey, E. D., et al. FTD and ALS: a tale of two diseases. *Curr Alzheimer Res,* 8**,** 273-94. (2011)

Flinn, L., Mortiboys, H., Volkmann, K., et al. Complex I deficiency and dopaminergic neuronal cell loss in parkin-deficient zebrafish (Danio rerio). *Brain,* 132**,** 1613-1623. (2009)

Flinn, L. J., Keatinge, M., Bretaud, S., et al. TigarB causes mitochondrial dysfunction and neuronal loss in PINK1 deficiency. *Ann Neurol,* 74**,** 837-847. (2013)

Force, A., Lynch, M., Pickett, F. B., et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics,* 151**,** 1531-45. (1999)

Forsberg, K., Graffmo, K., Pakkenberg, B., et al. Misfolded SOD1 inclusions in patients with mutations in C9orf72 and other ALS/FTD-associated genes. *J Neurol Neurosurg Psychiatry*. (2019)

Forsberg, K., Jonsson, P. A., Andersen, P. M., et al. Novel antibodies reveal inclusions containing non-native SOD1 in sporadic ALS patients. *PLoS One,* 5**,** e11552. (2010)

Forstl, H. & Kurz, A. Clinical features of Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci,* 249**,** 288-90. (1999)

Fraher, D., Sanigorski, A., Mellett, N. A., et al. Zebrafish Embryonic Lipidomic Analysis Reveals that the Yolk Cell Is Metabolically Active in Processing Lipid. *Cell Rep,* 14**,** 1317-1329. (2016)

Fratta, P., Mizielinska, S., Nicoll, A. J., et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep,* 2**,** 1016. (2012)

Fratta, P., Poulter, M., Lashley, T., et al. Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathol,* 126**,** 401-409. (2013)

Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature,* 525**,** 129-33. (2015)

Freischmidt, A., Wieland, T., Richter, B., et al. Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nat Neurosci,* 18**,** 631-6. (2015)

Fukunaga, K., Shinoda, Y. & Tagashira, H. The role of SIGMAR1 gene mutation and mitochondrial dysfunction in amyotrophic lateral sclerosis. *J Pharmacol Sci,* 127**,** 36-41. (2015)

Garate, Z., Davis, B. R., Quintana-Bustamante, O., et al. New frontier in regenerative medicine: site-specific gene correction in patient-specific induced pluripotent stem cells. *Hum Gene Ther,* 24**,** 571-83. (2013)

Gendron, T. F., van Blitterswijk, M., Bieniek, K. F., et al. Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. *Acta Neuropathol,* 130**,** 559-73. (2015)

Gerhard, G. S., Kauffman, E. J., Wang, X., et al. Life spans and senescent phenotypes in two strains of Zebrafish (Danio rerio). *Exp Gerontol,* 37**,** 1055-68. (2002)

Gill, C., Phelan, J. P., Hatzipetros, T., et al. SOD1-positive aggregate accumulation in the CNS predicts slower disease progression and increased longevity in a mutant SOD1 mouse model of ALS. *Sci Rep,* 9**,** 6724. (2019)

Goedert, M., Spillantini, M. G., Del Tredici, K., et al. 100 years of Lewy pathology. *Nat Rev Neurol,* 9**,** 13-24. (2013)

Goedert, M., Wischik, C. M., Crowther, R. A., et al. Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau. *Proc Natl Acad Sci U S A,* 85**,** 4051-4055. (1988)

Goff, S. P. & Berg, P. Construction of hybrid viruses containing SV40 and lambda phage DNA segments and their propagation in cultured monkey cells. *Cell,* 9**,** 695-705. (1976)

Goode, A., Butler, K., Long, J., et al. Defective recognition of LC3B by mutant SQSTM1/p62 implicates impairment of autophagy as a pathogenic mechanism in ALS-FTLD. *Autophagy,* 12**,** 1094-1104. (2016)

Gordon, P. H. Amyotrophic Lateral Sclerosis: An update for 2013 Clinical Features, Pathophysiology, Management and Therapeutic Trials. *Aging and Disease,* 4**,** 295-310. (2013)

Grad, L. I., Pokrishevsky, E., Silverman, J. M., et al. Exosome-dependent and independent mechanisms are involved in prion-like transmission of propagated Cu/Zn superoxide dismutase misfolding. *Prion,* 8**,** 331-335. (2014)

Graf, M., Ecker, D., Horowski, R., et al. High dose vitamin E therapy in amyotrophic lateral sclerosis as add-on therapy to riluzole: results of a placebo-controlled double-blind study. *J Neural Transm (Vienna),* 112**,** 649-60. (2005)

Granieri, E., Carreras, M., Tola, R., et al. Motor neuron disease in the province of Ferrara, Italy, in 1964-1982. *Neurology,* 38**,** 1604-1608. (1988)

Green, K. M., Glineburg, M. R., Kearse, M. G., et al. RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat Commun,* 8**,** 2005. (2017)

Greenway, M. J., Andersen, P. M., Russ, C., et al. ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nat Genet,* 38**,** 411-3. (2006)

Griesbeck, O., Parsadanian, A. S., Sendtner, M., et al. Expression of neurotrophins in skeletal muscle: quantitative comparison and significance for motoneuron survival and maintenance of function. *J Neurosci Res,* 42**,** 21-33. (1995)

Haas, R. H., Nasirian, F., Nakano, K., et al. Low platelet mitochondrial complex I and complex II/III activity in early untreated Parkinson's disease. *Ann Neurol,* 37**,** 714-22. (1995)

Hadano, S., Hand, C. K., Osuga, H., et al. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nat Genet,* 29**,** 166-73. (2001)

Haeusler, A. R., Donnelly, C. J., Periz, G., et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature,* 507**,** 195-200. (2014)

Hammer, R. P., Jr., Tomiyasu, U. & Scheibel, A. B. Degeneration of the human Betz cell due to amyotrophic lateral sclerosis. *Exp Neurol,* 63**,** 336-46. (1979)

Hao le, T., Burghes, A. H. & Beattie, C. E. Generation and Characterization of a genetic zebrafish model of SMA carrying the human SMN2 gene. *Mol Neurodegener,* 6**,** 24. (2011)

Hao le, T., Wolman, M., Granato, M., et al. Survival motor neuron affects plastin 3 protein levels leading to motor defects. *J Neurosci,* 32**,** 5074-84. (2012)

Hardiman, O., Al-Chalabi, A., Chio, A., et al. Amyotrophic lateral sclerosis. *Nat Rev Dis Primers,* 3**,** 17071. (2017)

Hardiman, O. & van den Berg, L. H. Edaravone: a new treatment for ALS on the horizon? *Lancet Neurol,* 16**,** 490-491. (2017)

Hautbergue, G. M., Castelli, L. M., Ferraiuolo, L., et al. SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nat Commun,* 8**,** 16063. (2017)

He, Y., Vogelstein, B., Velculescu, V. E., et al. The antisense transcriptomes of human cells. *Science,* 322**,** 1855-1857. (2008)

Herranz-Martin, S., Chandran, J., Lewis, K., et al. Viral delivery of C9orf72 hexanucleotide repeat expansions in mice leads to repeat-length-dependent neuropathology and behavioural deficits. *Dis Model Mech,* 10**,** 859-868. (2017)

Hewamadduma, C. A., Grierson, A. J., Ma, T. P., et al. Tardbpl splicing rescues motor neuron and axonal development in a mutant tardbp zebrafish. *Hum Mol Genet,* 22**,** 2376-86. (2013)

Higelin, J., Catanese, A., Semelink-Sedlacek, L. L., et al. NEK1 loss-of-function mutation induces DNA damage accumulation in ALS patient-derived motoneurons. *Stem Cell Res,* 30**,** 150-162. (2018)

Higgins, C. M., Jung, C. & Xu, Z. ALS-associated mutant SOD1G93A causes mitochondrial vacuolation by expansion of the intermembrane space and by involvement of SOD1 aggregation and peroxisomes. *BMC Neurosci,* 4**,** 16. (2003)

Highley, J. R., Kirby, J., Jansweijer, J. A., et al. Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones. *Neuropathol Appl Neurobiol,* 40**,** 670-85. (2014)

Hoch, N. C., Hanzlikova, H., Rulten, S. L., et al. XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature,* 541**,** 87-91. (2017)

Hofmann, A., Kessler, B., Ewerling, S., et al. Efficient transgenesis in farm animals by lentiviral vectors. *EMBO Rep,* 4**,** 1054-1060. (2003)

Howe, K., Clark, M. D., Torroja, C. F., et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature,* 496**,** 498-503. (2013)

Humphreys, P. N., Bellamy, D., Stevenson, A., et al. A comparison of the breeding success of two strains of laboratory rats in relation to age at mating. *J Reprod Fertil,* 48**,** 421-422. (1976)

Hyman, A. A., Weber, C. A. & Julicher, F. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol,* 30**,** 39-58. (2014)

Ibrahim, S., Moatter, T. & Saleem, A. F. Spinal muscular atrophy: clinical spectrum and genetic mutations in Pakistani children. *Neurol India,* 60**,** 294-8. (2012)

Inukai, Y., Nonaka, T., Arai, T., et al. Abnormal phosphorylation of Ser409/410 of TDP-43 in FTLD-U and ALS. *FEBS Lett,* 582**,** 2899-2904. (2008)

Ivics, Z., Mates, L., Yau, T. Y., et al. Germline transgenesis in rodents by pronuclear microinjection of Sleeping Beauty transposons. 9, 773-93. (2014)

Iwai, Y., Shibuya, K., Misawa, S., et al. Axonal Dysfunction Precedes Motor Neuronal Death in Amyotrophic Lateral Sclerosis. *PLoS One,* 11, e0158596. (2016)

Jain, S., Wheeler, J. R., Walters, R. W., et al. ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell,* 164, 487-98. (2016)

Jedrzejowska, M., Milewski, M., Zimowski, J., et al. Phenotype modifiers of spinal muscular atrophy: the number of SMN2 gene copies, deletion in the NAIP gene and probably gender influence the course of the disease. *Acta Biochim Pol,* 56, 103-108. (2009)

Jensen, L., Jorgensen, L. H., Bech, R. D., et al. Skeletal Muscle Remodelling as a Function of Disease Progression in Amyotrophic Lateral Sclerosis. *Biomed Res Int,* 2016, 5930621. (2016)

Ji, Y., Ugolino, J., Brady, N. R., et al. Systemic Deregulation of Autophagy Upon Loss of ALS- and FTD-linked C9orf72. *Autophagy,* 13, 1254-1255. (2017)

Jiang, J., Zhu, Q., Gendron, T. F., et al. Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. *Neuron,* 90, 535-50. (2016)

Jinek, M., Chylinski, K., Fonfara, I., et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337, 816-21. (2012)

Johnson, J. O., Mandrioli, J., Benatar, M., et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron,* 68, 857-64. (2010)

Johnson, J. O., Pioro, E. P., Boehringer, A., et al. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat Neurosci,* 17, 664-666. (2014)

Jovicic, A., Mertens, J., Boeynaems, S., et al. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci,* 18, 1226-1229. (2015)

Ju, B., Xu, Y., He, J., et al. Faithful expression of green fluorescent protein (GFP) in transgenic zebrafish embryos under control of zebrafish gene promoters. *Dev Genet,* 25, 158-67. (1999)

Kabashi, E., Bercier, V., Lissouba, A., et al. FUS and TARDBP but not SOD1 interact in genetic models of amyotrophic lateral sclerosis. *PLoS Genet,* 7, e1002214. (2011)

Kabashi, E., Lin, L., Tradewell, M. L., et al. Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Hum Mol Genet,* 19, 671-83. (2010)

Kabashi, E., Valdmanis, P. N., Dion, P., et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet,* 40, 572-574. (2008)

Kalia, L. V. & Lang, A. E. Parkinson's disease. *Lancet,* 386, 896-912. (2015)

Kamel, F., Umbach, D. M., Bedlack, R. S., et al. Pesticide exposure and amyotrophic lateral sclerosis. *Neurotoxicology,* 33, 457-62. (2012)

Kamel, F., Umbach, D. M., Lehman, T. A., et al. Amyotrophic lateral sclerosis, lead, and genetic susceptibility: polymorphisms in the delta-aminolevulinic acid dehydratase and vitamin D receptor genes. *Environ Health Perspect,* 111**,** 1335-1339. (2003)

Kametani, F., Obi, T., Shishido, T., et al. Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral sclerosis brains. *Sci Rep,* 6**,** 23281. (2016)

Kapeli, K., Martinez, F. J. & Yeo, G. W. Genetic mutations in RNA-binding proteins and their roles in ALS. *Hum Genet,* 136**,** 1193-1214. (2017)

Kato, M., Han, T. W., Xie, S., et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell,* 149**,** 753-67. (2012)

Kawamata, T., Akiyama, H., Yamada, T., et al. Immunologic reactions in amyotrophic lateral sclerosis brain and spinal cord tissue. *Am J Pathol,* 140**,** 691-707. (1992)

Kawasaki, T., Saito, K., Mitsui, K., et al. Introduction of a Foreign Gene into Zebrafish and Medaka Cells Using Adenoviral Vectors. *Zebrafish,* 6**,** 253-258. (2009)

Kearse, M. G., Green, K. M., Krans, A., et al. CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell,* 62**,** 314-322. (2016)

Kenna, K. P., McLaughlin, R. L., Byrne, S., et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. *J Med Genet,* 50**,** 776-83. (2013)

Kenna, K. P., van Doormaal, P. T., Dekker, A. M., et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet,* 48**,** 1037-42. (2016)

Kerman, A., Liu, H. N., Croul, S., et al. Amyotrophic lateral sclerosis is a non-amyloid disease in which extensive misfolding of SOD1 is unique to the familial form. *Acta Neuropathol,* 119**,** 335-44. (2010)

Kim, H. J., Kim, N. C., Wang, Y. D., et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature,* 495**,** 467-73. (2013a)

Kim, H. J. & Taylor, J. P. Lost in Transportation: Nucleocytoplasmic Transport Defects in ALS and Other Neurodegenerative Diseases. *Neuron,* 96**,** 285-297. (2017)

Kim, N. C., Tresse, E., Kolaitis, R. M., et al. VCP is essential for mitochondrial quality control by PINK1/Parkin and this function is impaired by VCP mutations. *Neuron,* 78**,** 65-80. (2013b)

Kim, S. H., Stiles, S. G., Feichtmeier, J. M., et al. Mutation-dependent aggregation and toxicity in a Drosophila model for UBQLN2-associated ALS. *Hum Mol Genet,* 27**,** 322-337. (2018)

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., et al. Stages of embryonic development of the zebrafish. *Developmental Dynamics,* 203**,** 253-310. (1995)

Kimmel, C. B., Sepich, D. S. & Trevarrow, B. Development of segmentation in zebrafish. *Development,* 104**,** 197-207. (1988)

Kimura, Y., Hisano, Y., Kawahara, A., et al. Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci Rep,* 4**,** 6545. (2014)

King, A., Troakes, C., Smith, B., et al. ALS-FUS pathology revisited: singleton FUS mutations and an unusual case with both a FUS and TARDBP mutation. *Acta Neuropathol Commun,* 3**,** 62. (2015)

Kok, F. O., Shin, M., Ni, C. W., et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell,* 32**,** 97-108. (2015)

Koppers, M., Blokhuis, A. M., Westeneng, H. J., et al. C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Ann Neurol,* 78**,** 426-38. (2015)

Kraft, A. D., Resch, J. M., Johnson, D. A., et al. Activation of the Nrf2-ARE pathway in muscle and spinal cord during ALS-like pathology in mice expressing mutant SOD1. *Exp Neurol,* 207**,** 107-17. (2007)

Kwiatkowski, T. J., Jr., Bosco, D. A., Leclerc, A. L., et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science,* 323**,** 1205-1208. (2009)

Kwon, I., Xiang, S., Kato, M., et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science,* 345**,** 1139-45. (2014)

Lacorte, E., Ferrigno, L., Leoncini, E., et al. Physical activity, and physical activity related to sports, leisure and occupational activity as risk factors for ALS: A systematic review. *Neurosci Biobehav Rev,* 66**,** 61-79. (2016)

Lagier-Tourenne, C. & Cleveland, D. W. Rethinking ALS: the FUS about TDP-43. *Cell,* 136**,** 1001-1004. (2009)

Laird, A. S., Van Hoecke, A., De Muynck, L., et al. Progranulin is neurotrophic in vivo and protects against a mutant TDP-43 induced axonopathy. *PLoS One,* 5**,** e13368. (2010)

Lambert, R. Breeding Strategies for Maintaining Colonies of Laboratory Mice. *Jackson Laboratory Resource Manual.* (2009)

Landrian, I., McFarland, K. N., Liu, J., et al. Inheritance patterns of ATCCT repeat interruptions in spinocerebellar ataxia type 10 (SCA10) expansions. *PLoS One,* 12**,** e0175958. (2017)

Largaespada, D. A. Generating and manipulating transgenic animals using transposable elements. *Reprod Biol Endocrinol,* 1**,** 80. (2003)

Lattante, S., de Calbiac, H., Le Ber, I., et al. Sqstm1 knock-down causes a locomotor phenotype ameliorated by rapamycin in a zebrafish model of ALS/FTLD. *Hum Mol Genet,* 24**,** 1682-90. (2015)

Lawrence, C., Adatto, I., Best, J., et al. Generation time of zebrafish (Danio rerio) and medakas (Oryzias latipes) housed in the same aquaculture facility. *Lab Anim (NY),* 41**,** 158-65. (2012)

Lebedeva, S., de Jesus Domingues, A. M., Butter, F., et al. Characterization of genetic loss-of-function of Fus in zebrafish. *RNA Biol,* 14**,** 29-35. (2017)

Lee, K. H., Zhang, P., Kim, H. J., et al. C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell,* 167**,** 774-788.e17. (2016)

Lee, Y. B., Chen, H. J., Peres, J. N., et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep,* 5**,** 1178-86. (2013)

Lefebvre, S., Burlet, P., Liu, Q., et al. Correlation between severity and SMN protein level in spinal muscular atrophy. *Nat Genet,* 16**,** 265-269. (1997)

Leigh, P. N., Anderton, B. H., Dodson, A., et al. Ubiquitin deposits in anterior horn cells in motor neurone disease. *Neurosci Lett,* 93**,** 197-203. (1988)

Leigh, P. N., Whitwell, H., Garofalo, O., et al. Ubiquitin-immunoreactive intraneuronal inclusions in amyotrophic lateral sclerosis. Morphology, distribution, and specificity. *Brain,* 114 (Pt 2)**,** 775-88. (1991)

Lemmens, R., Van Hoecke, A., Hersmus, N., et al. Overexpression of mutant superoxide dismutase 1 causes a motor axonopathy in the zebrafish. *Hum Mol Genet,* 16**,** 2359-2365. (2007)

Levine, T. P., Daniels, R. D., Gatta, A. T., et al. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. *Bioinformatics,* 29**,** 499-503. (2013)

Li, M. & Yu, X. Function of BRCA1 in the DNA damage response is mediated by ADP-ribosylation. *Cancer Cell,* 23**,** 693-704. (2013)

Lin, W. C., Sanchez, H. B., Deerinck, T., et al. Aberrant development of motor axons and neuromuscular synapses in erbB2-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 1299-1304. (2000)

Lin, Y., Mori, E., Kato, M., et al. Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell,* 167**,** 789-802.e12. (2016)

Liscic, R. M. Molecular basis of ALS and FTD: implications for translational studies. *Arh Hig Rada Toksikol,* 66**,** 285-90. (2015)

Liu, H. N., Tjostheim, S., Dasilva, K., et al. Targeting of monomer/misfolded SOD1 as a therapeutic strategy for amyotrophic lateral sclerosis. *J Neurosci,* 32**,** 8791-8799. (2012)

Liu, S. & Leach, S. D. Zebrafish models for cancer. *Annu Rev Pathol,* 6**,** 71-93. (2011)

Liu, Y., Ma, P., Cassidy, P. A., et al. Statistical Analysis of Zebrafish Locomotor Behaviour by Generalized Linear Mixed Models. *Sci Rep,* 7**,** 2937. (2017)

Liu, Y., Pattamatta, A., Zu, T., et al. C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. *Neuron,* 90**,** 521-34. (2016)

Logroscino, G., Traynor, B. J., Hardiman, O., et al. Descriptive epidemiology of amyotrophic lateral sclerosis: new evidence and unsolved issues. *J Neurol Neurosurg Psychiatry,* 79**,** 6-11. (2008)

Lopez-Gonzalez, R., Lu, Y., Gendron, T. F., et al. Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. *Neuron,* 92**,** 383-391. (2016)

Lopez, K. L. R., Simpson, J. E., Watson, L. C., et al. TIGAR inclusion pathology is specific for Lewy body diseases. *Brain Res,* 1706**,** 218-223. (2019)

Lowe, J., Lennox, G., Jefferson, D., et al. A filamentous inclusion body within anterior horn neurones in motor neurone disease defined by immunocytochemical localisation of ubiquitin. *Neurosci Lett,* 94**,** 203-10. (1988)

Lunn, M. R. & Wang, C. H. Spinal muscular atrophy. *Lancet,* 371**,** 2120-33. (2008)

Luty, A. A., Kwok, J. B., Dobson-Stone, C., et al. Sigma nonopioid intracellular receptor 1 mutations cause frontotemporal lobar degeneration-motor neuron disease. *Ann Neurol,* 68**,** 639-49. (2010)

Mackenzie, I. R., Arzberger, T., Kremmer, E., et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol,* 126**,** 859-79. (2013)

Mackenzie, I. R., Bigio, E. H., Ince, P. G., et al. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Ann Neurol,* 61**,** 427-34. (2007)

Mackenzie, I. R., Frick, P., Grasser, F. A., et al. Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol,* 130**,** 845-61. (2015)

Mackenzie, I. R., Frick, P. & Neumann, M. The neuropathology associated with repeat expansions in the C9ORF72 gene. *Acta Neuropathol,* 127**,** 347-57. (2014)

Magrane, J., Cortez, C., Gan, W. B., et al. Abnormal mitochondrial transport and morphology are common pathological denominators in SOD1 and TDP43 ALS mouse models. *Hum Mol Genet,* 23**,** 1413-24. (2014)

Maier, M., Welt, T., Wirth, F., et al. A human-derived antibody targets misfolded SOD1 and ameliorates motor symptoms in mouse models of amyotrophic lateral sclerosis. *Sci Transl Med,* 10. (2018)

Mailman, M. D., Heinz, J. W., Papp, A. C., et al. Molecular analysis of spinal muscular atrophy and modification of the phenotype by SMN2. *Genet Med,* 4**,** 20-26. (2002)

Mann, D. M., Rollinson, S., Robinson, A., et al. Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathol Commun,* 1**,** 68. (2013)

Markovinovic, A., Cimbro, R., Ljutic, T., et al. Optineurin in amyotrophic lateral sclerosis: Multifunctional adaptor protein at the crossroads of different neuroprotective mechanisms. *Prog Neurobiol,* 154**,** 1-20. (2017)

Martire, S., Mosca, L. & d'Erme, M. PARP-1 involvement in neurodegeneration: A focus on Alzheimer's and Parkinson's diseases. *Mech Ageing Dev,* 146-148**,** 53-64. (2015)

Maruyama, H., Morino, H., Ito, H., et al. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature,* 465**,** 223-226. (2010)

May, S., Hornburg, D., Schludi, M. H., et al. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol,* 128**,** 485-503. (2014)

McClintock, B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A,* 36**,** 344-355. (1950)

McGown, A., McDearmid, J. R., Panagiotaki, N., et al. Early interneuron dysfunction in ALS: insights from a mutant sod1 zebrafish model. *Ann Neurol,* 73**,** 246-58. (2013)

McWhorter, M. L., Monani, U. R., Burghes, A. H., et al. Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *J Cell Biol,* 162**,** 919-31. (2003)

Mead, R. J., Higginbottom, A., Allen, S. P., et al. S[+] Apomorphine is a CNS penetrating activator of the Nrf2-ARE pathway with activity in mouse and patient fibroblast models of amyotrophic lateral sclerosis. *Free Radic Biol Med,* 61**,** 438-52. (2013)

Meerang, M., Ritz, D., Paliwal, S., et al. The ubiquitin-selective segregase VCP/p97 orchestrates the response to DNA double-strand breaks. *Nat Cell Biol,* 13**,** 1376-82. (2011)

Meyer, A. & Schartl, M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol,* 11**,** 699-704. (1999)

Mirkin, S. M. Expandable DNA repeats and human disease. *Nature,* 447**,** 932-40. (2007)

Mizielinska, S., Gronke, S., Niccoli, T., et al. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science,* 345**,** 1192-1194. (2014)

Mizielinska, S., Lashley, T., Norona, F. E., et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol,* 126**,** 845-57. (2013)

Mizielinska, S., Ridler, C. E., Balendra, R., et al. Bidirectional nucleolar dysfunction in C9orf72 frontotemporal lobar degeneration. *Acta Neuropathol Commun,* 5**,** 29. (2017)

Mizuno, Y., Amari, M., Takatama, M., et al. Immunoreactivities of p62, an ubiquitin-binding protein, in the spinal anterior horn cells of patients with amyotrophic lateral sclerosis. *J Neurol Sci,* 249**,** 13-18. (2006a)

Mizuno, Y., Amari, M., Takatama, M., et al. Transferrin localizes in Bunina bodies in amyotrophic lateral sclerosis. *Acta Neuropathol,* 112**,** 597-603. (2006b)

Molliex, A., Temirov, J., Lee, J., et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell,* 163**,** 123-33. (2015)

Mordes, D. A., Prudencio, M., Goodman, L. D., et al. Dipeptide repeat proteins activate a heat shock response found in C9ORF72-ALS/FTLD patients. *Acta Neuropathol Commun,* 6**,** 55. (2018)

Moreno, C. B., Hernandez-Beltran, N., Munevar, D., et al. Central neuropathic pain in Parkinson's disease. *Neurologia,* 27**,** 500-503. (2012)

Mori, K., Weng, S. M., Arzberger, T., et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science,* 339**,** 1335-1338. (2013)

Morris, M. J., Wingate, K. L., Silwal, J., et al. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res,* 40**,** 4137-45. (2012)

Mosimann, C., Kaufman, C. K., Li, P., et al. Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development,* 138**,** 169-77. (2011)

Nassif, M., Woehlbier, U. & Manque, P. A. The Enigmatic Role of C9ORF72 in Autophagy. *Front Neurosci,* 11**,** 442. (2017)

Naumann, M., Pal, A., Goswami, A., et al. Impaired DNA damage response signaling by FUS-NLS mutations leads to neurodegeneration and FUS aggregate formation. *Nat Commun,* 9**,** 335. (2018)

Nery, L. R., Silva, N. E., Fonseca, R., et al. Presenilin-1 Targeted Morpholino Induces Cognitive Deficits, Increased Brain Abeta1-42 and Decreased Synaptic Marker PSD-95 in Zebrafish Larvae. *Neurochem Res,* 42**,** 2959-2967. (2017)

Neumann, M., Sampathu, D. M., Kwong, L. K., et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science,* 314**,** 130-133. (2006)

Nicolas, A., Kenna, K. P., Renton, A. E., et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron,* 97**,** 1268-1283.e6. (2018)

Nihei, K., McKee, A. C. & Kowall, N. W. Patterns of neuronal degeneration in the motor cortex of amyotrophic lateral sclerosis patients. *Acta Neuropathol,* 86**,** 55-64. (1993)

Nishimura, A. L., Mitne-Neto, M., Silva, H. C., et al. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am J Hum Genet,* 75**,** 822-31. (2004)

Nishimura, K., Fukagawa, T., Takisawa, H., et al. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods,* 6**,** 917-22. (2009)

Nornes, S., Groth, C., Camp, E., et al. Developmental control of Presenilin1 expression, endoproteolysis, and interaction in zebrafish embryos. *Exp Cell Res,* 289**,** 124-32. (2003)

Nornes, S., Newman, M., Wells, S., et al. Independent and cooperative action of Psen2 with Psen1 in zebrafish embryos. *Exp Cell Res,* 315**,** 2791-2801. (2009)

O'Rourke, J. G., Bogdanik, L., Muhammad, A. K., et al. C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. *Neuron,* 88**,** 892-901. (2015)

O'Rourke, J. G., Bogdanik, L., Yanez, A., et al. C9orf72 is required for proper macrophage and microglial function in mice. *Science,* 351**,** 1324-1329. (2016)

Oakes, J. A., Davies, M. C. & Collins, M. O. TBK1: a new player in ALS linking autophagy and neuroinflammation. *Mol Brain,* 10**,** 5. (2017)

Ohki, Y., Wenninger-Weinzierl, A., Hruscha, A., et al. Glycine-alanine dipeptide repeat protein contributes to toxicity in a zebrafish model of C9orf72 associated neurodegeneration. *Mol Neurodegener,* 12**,** 6. (2017)

Okamoto, K., Hirai, S., Amari, M., et al. Bunina bodies in amyotrophic lateral sclerosis immunostained with rabbit anti-cystatin C serum. *Neurosci Lett,* 162**,** 125-128. (1993)

Okamoto, K., Mizuno, Y. & Fujita, Y. Bunina bodies in amyotrophic lateral sclerosis. *Neuropathology,* 28**,** 109-15. (2008)

Orlacchio, A., Babalini, C., Borreca, A., et al. SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. *Brain,* 133**,** 591-8. (2010)

Osaka, M., Ito, D. & Suzuki, N. Disturbance of proteasomal and autophagic protein degradation pathways by amyotrophic lateral sclerosis-linked mutations in ubiquilin 2. *Biochem Biophys Res Commun,* 472**,** 324-31. (2016)

Paquet, D., Bhat, R., Sydow, A., et al. A zebrafish model of tauopathy allows in vivo imaging of neuronal cell death and drug evaluation. *Journal of Clinical Investigation,* 119**,** 1382-1395. (2009)

Parkinson, N., Ince, P. G., Smith, M. O., et al. ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). *Neurology,* 67**,** 1074-1077. (2006)

Pasinelli, P., Belford, M. E., Lennon, N., et al. Amyotrophic lateral sclerosis-associated SOD1 mutant proteins bind and aggregate with Bcl-2 in spinal cord mitochondria. *Neuron,* 43**,** 19-30. (2004)

Pederson, T. The nucleolus. *Cold Spring Harb Perspect Biol,* 3**,** a000638. (2011)

Pedrini, S., Sau, D., Guareschi, S., et al. ALS-linked mutant SOD1 damages mitochondria by promoting conformational changes in Bcl-2. *Hum Mol Genet,* 19**,** 2974-86. (2010)

Peters, O. M., Cabrera, G. T., Tran, H., et al. Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. *Neuron,* 88**,** 902-909. (2015)

Phukan, J., Elamin, M., Bede, P., et al. The syndrome of cognitive impairment in amyotrophic lateral sclerosis: a population-based study. *J Neurol Neurosurg Psychiatry,* 83**,** 102-108. (2012)

Piao, Y. S., Wakabayashi, K., Kakita, A., et al. Neuropathology with clinical correlations of sporadic amyotrophic lateral sclerosis: 102 autopsy cases examined between 1962 and 2000. *Brain Pathol,* 13**,** 10-22. (2003)

Pokrishevsky, E., Grad, L. I. & Cashman, N. R. TDP-43 or FUS-induced misfolded human wild-type SOD1 can propagate intercellularly in a prion-like fashion. *Sci Rep,* 6**,** 22155. (2016)

Pokrishevsky, E., Grad, L. I., Yousefi, M., et al. Aberrant localization of FUS and TDP43 is associated with misfolding of SOD1 in amyotrophic lateral sclerosis. *PLoS One, 7,* e35050. (2012)

Poulopoulos, M., Levy, O. A. & Alcalay, R. N. The neuropathology of genetic Parkinson's disease. *Mov Disord,* 27, 831-42. (2012)

Powis, R. A., Karyka, E., Boyd, P., et al. Systemic restoration of UBA1 ameliorates disease in spinal muscular atrophy. *JCI Insight,* 1, e87908. (2016)

Prabhudesai, S., Bensabeur, F. Z., Abdullah, R., et al. LRRK2 knockdown in zebrafish causes developmental defects, neuronal loss, and synuclein aggregation. *J Neurosci Res,* 94, 717-35. (2016)

Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol,* 26, 668-679. (2016)

Prudencio, M., Belzil, V. V., Batra, R., et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci,* 18, 1175-82. (2015)

Pu, Y. Z., Liang, L., Fu, A. L., et al. Generation of Alzheimer's Disease Transgenic Zebrafish Expressing Human APP Mutation Under Control of Zebrafish appb Promotor. *Curr Alzheimer Res,* 14, 668-679. (2017)

Ramesh, T., Lyon, A. N., Pineda, R. H., et al. A genetic model of amyotrophic lateral sclerosis in zebrafish displays phenotypic hallmarks of motoneuron disease. *Dis Model Mech,* 3, 652-62. (2010)

Ratovitski, T., Corson, L. B., Strain, J., et al. Variation in the biochemical/biophysical properties of mutant superoxide dismutase 1 enzymes and the rate of disease progression in familial amyotrophic lateral sclerosis kindreds. *Hum Mol Genet,* 8, 1451-60. (1999)

Reaume, A. G., Elliott, J. L., Hoffman, E. K., et al. Motor neurons in Cu/Zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury. *Nat Genet,* 13, 43-47. (1996)

Reddy, K., Zamiri, B., Stanley, S. Y., et al. 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. (2013)

Reimer, M. M., Kuscha, V., Wyatt, C., et al. Sonic hedgehog is a polarized signal for motor neuron regeneration in adult zebrafish. *J Neurosci,* 29, 15073-15082. (2009)

Renton, A. E., Majounie, E., Waite, A., et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron,* 72, 257-68. (2011)

Richard, G. F., Kerrest, A. & Dujon, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev,* 72, 686-727. (2008)

Rink, E. & Wullimann, M. F. Connections of the ventral telencephalon and tyrosine hydroxylase distribution in the zebrafish brain (Danio rerio) lead to identification of an ascending dopaminergic system in a teleost. *Brain Res Bull,* 57, 385-387. (2002)

Robinson, K. J., Yuan, K. C., Don, E. K., et al. Motor Neuron Abnormalities Correlate with Impaired Movement in Zebrafish that Express Mutant Superoxide Dismutase 1. *Zebrafish,* 16. (2018)

Robinson, M., Lee, B. Y. & Hane, F. T. Recent Progress in Alzheimer's Disease Research, Part 2: Genetics and Epidemiology. *J Alzheimers Dis,* 57**,** 317-330. (2017)

Rochette, C. F., Gilbert, N. & Simard, L. R. SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to Homo sapiens. *Hum Genet,* 108**,** 255-66. (2001)

Roos, W. P. & Kaina, B. DNA damage-induced cell death by apoptosis. *Trends Mol Med,* 12**,** 440-50. (2006)

Rosen, D. R., Siddique, T., Patterson, D., et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature,* 362**,** 59-62. (1993)

Rotunno, M. S. & Bosco, D. A. An emerging role for misfolded wild-type SOD1 in sporadic ALS pathogenesis. *Front Cell Neurosci,* 7**,** 253. (2013)

Ryan, C. L., Baranowski, D. C., Chitramuthu, B. P., et al. Progranulin is expressed within motor neurons and promotes neuronal cell survival. *BMC Neurosci,* 10**,** 130. (2009)

Saberi, S., Stauffer, J. E., Jiang, J., et al. 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. (2017)

Saberi, S., Stauffer, J. E., Schulte, D. J., et al. Neuropathology of Amyotrophic Lateral Sclerosis and Its Variants. *Neurol Clin,* 33**,** 855-76. (2015)

Sakae, N., Bieniek, K. F., Zhang, Y. J., et al. Poly-GR dipeptide repeat polymers correlate with neurodegeneration and Clinicopathological subtypes in C9ORF72-related brain disease. *Acta Neuropathol Commun,* 6**,** 63. (2018)

Sakowski, S. A., Lunn, J. S., Busta, A. S., et al. Neuromuscular effects of G93A-SOD1 expression in zebrafish. *Mol Neurodegener,* 7**,** 44. (2012)

Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-55. (2014)

Sareen, D., O'Rourke, J. G., Meera, P., et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med,* 5**,** 208ra149. (2013)

Sasaki, S. & Iwata, M. Mitochondrial alterations in the spinal cord of patients with sporadic amyotrophic lateral sclerosis. *J Neuropathol Exp Neurol,* 66**,** 10-16. (2007)

Sato, K., Otomo, A., Ueda, M. T., et al. Altered oligomeric states in pathogenic ALS2 variants associated with juvenile motor neuron diseases cause loss of ALS2-mediated endosomal function. *J Biol Chem,* 293**,** 17135-17153. (2018)

Schapira, A. H., Mann, V. M., Cooper, J. M., et al. Anatomic and disease specificity of NADH CoQ1 reductase (complex I) deficiency in Parkinson's disease. *J Neurochem,* 55**,** 2142-2145. (1990)

Scheuner, D., Eckman, C., Jensen, M., et al. Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat Med,* 2**,** 864-70. (1996)

Schiffer, D., Cordera, S., Cavalla, P., et al. Reactive astrogliosis of the spinal cord in amyotrophic lateral sclerosis. *J Neurol Sci,* 139 Suppl**,** 27-33. (1996)

Schludi, M. H., May, S., Grasser, F. A., et al. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathol,* 130**,** 537-555. (2015)

Schmid, B., Hruscha, A., Hogl, S., et al. Loss of ALS-associated TDP-43 in zebrafish causes muscle degeneration, vascular dysfunction, and reduced motor neuron axon outgrowth. *Proc Natl Acad Sci U S A,* 110**,** 4986-91. (2013)

Schnorr, S. J., Steenbergen, P. J., Richardson, M. K., et al. Measuring thigmotaxis in larval zebrafish. *Behav Brain Res,* 228**,** 367-74. (2012)

Sellier, C., Campanari, M. L., Julie Corbier, C., et al. Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. *EMBO J,* 35**,** 1276-97. (2016)

Shahheydari, H., Ragagnin, A., Walker, A. K., et al. Protein Quality Control and the Amyotrophic Lateral Sclerosis/Frontotemporal Dementia Continuum. *Front Mol Neurosci,* 10**,** 119. (2017)

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

Shao, Q., Liang, C., Chang, Q., et al. C9orf72 deficiency promotes motor deficits of a C9ALS/FTD mouse model in a dose-dependent manner. *Acta Neuropathol Commun,* 7**,** 32. (2019)

Shi, Y., Lin, S., Staats, K. A., et al. Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat Med,* 24**,** 313-325. (2018)

Smith, B. N., Ticozzi, N., Fallini, C., et al. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron,* 84**,** 324-31. (2014)

Soman, S., Keatinge, M., Moein, M., et al. Inhibition of the mitochondrial calcium uniporter rescues dopaminergic neurons in pink1(-/-) zebrafish. *Eur J Neurosci,* 45**,** 528-535. (2017)

Soraru, G., D'Ascenzo, C., Nicolao, P., et al. Muscle histopathology in upper motor neuron-dominant amyotrophic lateral sclerosis. *Amyotroph Lateral Scler,* 9**,** 287-93. (2008)

Soraru, G., Orsetti, V., Buratti, E., et al. TDP-43 in skeletal muscle of patients affected with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler,* 11**,** 240-243. (2010)

Spillantini, M. G., Schmidt, M. L., Lee, V. M., et al. Alpha-synuclein in Lewy bodies. *Nature,* 388**,** 839-40. (1997)

Sreedharan, J., Blair, I. P., Tripathi, V. B., et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science,* 319**,** 1668-72. (2008)

Stephens, B., Guiloff, R. J., Navarrete, R., et al. Widespread loss of neuronal populations in the spinal ventral horn in sporadic motor neuron disease. A morphometric study. *J Neurol Sci,* 244**,** 41-58. (2006)

Stepto, A., Gallo, J. M., Shaw, C. E., et al. Modelling C9ORF72 hexanucleotide repeat expansion in amyotrophic lateral sclerosis and frontotemporal dementia. *Acta Neuropathol,* 127**,** 377-89. (2014)

Stopford, M. J., Higginbottom, A., Hautbergue, G. M., et al. C9ORF72 hexanucleotide repeat exerts toxicity in a stable, inducible motor neuronal cell model, which is rescued by partial depletion of Pten. *Hum Mol Genet,* 26**,** 1133-1145. (2017)

Streisinger, G., Walker, C., Dower, N., et al. Production of clones of homozygous diploid zebra fish (Brachydanio rerio). *Nature,* 291**,** 293-6. (1981)

Sudria-Lopez, E., Koppers, M., de Wit, M., et al. Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. *Acta Neuropathol,* 132**,** 145-147. (2016)

Sveinbjornsdottir, S. The clinical symptoms of Parkinson's disease. *J Neurochem,* 139 Suppl 1**,** 318-324. (2016)

Swaminathan, A., Bouffard, M., Liao, M., et al. Expression of C9orf72-related dipeptides impairs motor function in a vertebrate model. *Hum Mol Genet,* 27**,** 1754-1762. (2018)

Swinnen, B., Bento-Abreu, A., Gendron, T. F., et al. A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta Neuropathol,* 135**,** 427-443. (2018)

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., et al. The double-strand-break repair model for recombination. *Cell,* 33**,** 25-35. (1983)

Takahashi, Y., Fukuda, Y., Yoshimura, J., et al. ERBB4 mutations that disrupt the neuregulin-ErbB4 pathway cause amyotrophic lateral sclerosis type 19. *Am J Hum Genet,* 93**,** 900-905. (2013)

Tao, Z., Wang, H., Xia, Q., et al. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet,* 24**,** 2426-41. (2015)

Taylor, J. S., Braasch, I., Frickey, T., et al. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res,* 13**,** 382-90. (2003)

Therrien, M., Rouleau, G. A., Dion, P. A., et al. Deletion of C9ORF72 results in motor neuron degeneration and stress sensitivity in C. elegans. *PLoS One,* 8**,** e83450. (2013)

Thiyagarajan, N., Ferguson, R., Subramanian, V., et al. Structural and molecular insights into the mechanism of action of human angiogenin-ALS variants in neurons. *Nat Commun,* 3**,** 1121. (2012)

Thomason, L., Court, D. L., Bubunenko, M., et al. Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol,* Chapter 1**,** Unit 1.16. (2007)

Thompson, S., Clarke, A. R., Pow, A. M., et al. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell,* 56**,** 313-21. (1989)

Ticozzi, N., Vance, C., Leclerc, A. L., et al. Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. *Am J Med Genet B Neuropsychiatr Genet,* 156b**,** 285-90. (2011)

Todd, P. K., Oh, S. Y., Krans, A., et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron,* 78**,** 440-55. (2013)

Tran, H., Almeida, S., Moore, J., et al. Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron,* 87**,** 1207-14. (2015)

Troost, D., Van den Oord, J. J. & Vianney de Jong, J. M. Immunohistochemical characterization of the inflammatory infiltrate in amyotrophic lateral sclerosis. *Neuropathol Appl Neurobiol,* 16**,** 401-10. (1990)

Tsuji, H., Arai, T., Kametani, F., et al. Molecular analysis and biochemical classification of TDP-43 proteinopathy. *Brain,* 135**,** 3380-91. (2012)

Turner, M. R. Increased premorbid physical activity and amyotrophic lateral sclerosis: born to run rather than run to death, or a seductive myth? *J Neurol Neurosurg Psychiatry,* 84**,** 947. (2013)

van den Heuvel, D. M., Harschnitz, O., van den Berg, L. H., et al. Taking a risk: a therapeutic focus on ataxin-2 in amyotrophic lateral sclerosis? *Trends Mol Med,* 20**,** 25-35. (2014)

Vance, C., Rogelj, B., Hortobagyi, T., et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science,* 323**,** 1208-1211. (2009)

Veldink, J. H., Kalmijn, S., Groeneveld, G. J., et al. Intake of polyunsaturated fatty acids and vitamin E reduces the risk of developing amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry,* 78**,** 367-71. (2007)

Viode, A., Fournier, C., Camuzat, A., et al. New Antibody-Free Mass Spectrometry-Based Quantification Reveals That C9ORF72 Long Protein Isoform Is Reduced in the Frontal Cortex of Hexanucleotide-Repeat Expansion Carriers. *Front Neurosci,* 12**,** 589. (2018)

Wagstaff, K. M., Sivakumaran, H., Heaton, S. M., et al. Ivermectin is a specific inhibitor of importin alpha/beta-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. *Biochem J,* 443**,** 851-856. (2012)

Walker, C., Herranz-Martin, S., Karyka, E., et al. C9orf72 expansion disrupts ATM-mediated chromosomal break repair. *Nat Neurosci,* 20**,** 1225-1235. (2017)

Wang, H., Guo, W., Mitra, J., et al. Mutant FUS causes DNA ligation defects to inhibit oxidative damage repair in Amyotrophic Lateral Sclerosis. *Nat Commun,* 9**,** 3683. (2018)

Wang, M. D., Little, J., Gomes, J., et al. Identification of risk factors associated with onset and progression of amyotrophic lateral sclerosis using systematic review and meta-analysis. *Neurotoxicology,* 61**,** 101-130. (2017)

Wang, W. Y., Pan, L., Su, S. C., et al. Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. *Nat Neurosci,* 16**,** 1383-91. (2013)

Webster, C. P., Smith, E. F., Bauer, C. S., et al. The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *EMBO J,* 35**,** 1656-76. (2016a)

Webster, C. P., Smith, E. F., Grierson, A. J., et al. C9orf72 plays a central role in Rab GTPase-dependent regulation of autophagy. *Small GTPases,* 9**,** 399-408. (2016b)

Wei, Q., Zhou, Q., Chen, Y., et al. Analysis of SOD1 mutations in a Chinese population with amyotrophic lateral sclerosis: a case-control study and literature review. *Sci Rep,* 7**,** 44606. (2017)

Wen, X. M., Tan, W. Z., Westergard, T., et al. 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. (2014)

Wenk, G. L. Neuropathologic changes in Alzheimer's disease. *J Clin Psychiatry,* 64 Suppl 9**,** 7-10. (2003)

Westerfield, M. *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio),* Eugene, Univ. of Oregon Press. (2000)

Westerfield, M., McMurray, J. V. & Eisen, J. S. Identified motoneurons and their innervation of axial muscles in the zebrafish. *J Neurosci,* 6**,** 2267-77. (1986)

Westergard, T., Jensen, B. K., Wen, X., et al. Cell-to-Cell Transmission of Dipeptide Repeat Proteins Linked to C9orf72-ALS/FTD. *Cell Rep,* 17**,** 645-652. (2016)

White, M. R., Mitrea, D. M., Zhang, P., et al. C9orf72 Poly(PR) Dipeptide Repeats Disturb Biomolecular Phase Separation and Disrupt Nucleolar Function. *Mol Cell,* 74**,** 1-16. (2019)

Wiedemann, F. R., Winkler, K., Kuznetsov, A. V., et al. Impairment of mitochondrial function in skeletal muscle of patients with amyotrophic lateral sclerosis. *J Neurol Sci,* 156**,** 65-72. (1998)

Wilke, C., Pomper, J. K., Biskup, S., et al. Atypical parkinsonism in C9orf72 expansions: a case report and systematic review of 45 cases from the literature. *J Neurol,* 263**,** 558-74. (2016)

Wilkins, B. J. & Pack, M. Zebrafish models of human liver development and disease. *Compr Physiol,* 3**,** 1213-30. (2013)

Wilkinson, R. N., Jopling, C. & van Eeden, F. J. Zebrafish as a model of cardiac disease. *Prog Mol Biol Transl Sci,* 124**,** 65-91. (2014)

Wilkinson, R. N. & van Eeden, F. J. The zebrafish as a model of vascular development and disease. *Prog Mol Biol Transl Sci,* 124**,** 93-122. (2014)

Williams, K. L., Topp, S., Yang, S., et al. CCNF mutations in amyotrophic lateral sclerosis and frontotemporal dementia. *Nat Commun,* 7**,** 11253. (2016)

Wu, C. H., Fallini, C., Ticozzi, N., et al. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature,* 488**,** 499-503. (2012)

Xi, Y., Ryan, J., Noble, S., et al. Impaired dopaminergic neuron development and locomotor function in zebrafish with loss of pink1 function. *Eur J Neurosci,* 31**,** 623-33. (2010)

Xu, Z., Poidevin, M., Li, X., et al. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc Natl Acad Sci U S A,* 110, 7778-83. (2013)

Yang, Y., Hentati, A., Deng, H. X., et al. The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nat Genet,* 29**,** 160-165. (2001)

Yeh, T. H., Liu, H. F., Li, Y. W., et al. C9orf72 is essential for neurodevelopment and motility mediated by Cyclin G1. *Exp Neurol,* 304**,** 114-124. (2018)

Yin, S., Lopez-Gonzalez, R., Kunz, R. C., et al. Evidence that C9ORF72 Dipeptide Repeat Proteins Associate with U2 snRNP to Cause Mis-splicing in ALS/FTD Patients. *Cell Rep,* 19**,** 2244-2256. (2017)

Zelko, I. N., Mariani, T. J. & Folz, R. J. Superoxide dismutase multigene family: a comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. *Free Radic Biol Med,* 33**,** 337-49. (2002)

Zhang, H., Duan, C. & Yang, H. Defective autophagy in Parkinson's disease: lessons from genetics. *Mol Neurobiol,* 51**,** 89-104. (2015a)

Zhang, K., Daigle, J. G., Cunningham, K. M., et al. Stress Granule Assembly Disrupts Nucleocytoplasmic Transport. *Cell,* 173, 958-971.e17. (2018)

Zhang, K., Donnelly, C. J., Haeusler, A. R., et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature,* 525**,** 56-61. (2015b)

Zhang, Y., Nguyen, D. T., Olzomer, E. M., et al. Rescue of Pink1 Deficiency by Stress-Dependent Activation of Autophagy. *Cell Chem Biol,* 24**,** 471-480.e4. (2017)

Zhang, Y. J., Gendron, T. F., Grima, J. C., et al. C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. 19**,** 668-77. (2016)

Zhang, Y. J., Jansen-West, K., Xu, Y. F., et al. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol,* 128**,** 505-24. (2014)

Zhou, B., Geng, Y., Liu, C., et al. Characterizations of distinct parallel and antiparallel G-quadruplexes formed by two-repeat ALS and FTD related GGGGCC sequence. *Sci Rep,* 8**,** 2366. (2018)

Zhu, H. & Zon, L. I. Use of the DsRed fluorescent reporter in zebrafish. *Methods Cell Biol,* 76**,** 3-12. (2004)

Zhu, S., Tai, C., Petkau, T. L., et al. Progranulin promotes activation of microglia/macrophage after pilocarpine-induced status epilepticus. *Brain Res,* 1530**,** 54-65. (2013)

Zu, T., Gibbens, B., Doty, N. S., et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A,* 108**,** 260-265. (2011)

Zu, T., Liu, Y., Banez-Coronel, M., et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A,* 110**,** E4968-77. (2013)

Appendix A   Table of *in vivo* C9orf72 models

Gain or loss of function	Type of model	Species	RNA Foci	Sense DPR	Antisense DPR	Motor or cognitive defects	Neuro-degen.	Expression type	Reference	Notes
Gain	Expansion containing BAC	Mouse	Yes	Yes	Not tested	No	No	Stable	(Peters et al., 2015)	Exons 1-6 of human *C9orf72* gene with 140.5 Kb of upstream sequence was expressed. G4C2 repeats lengths of both ~300 and ~500 were reported within the same mice.
Gain	Expansion containing BAC	Mouse	Yes	Yes	Not tested	No	No	Stable	(O'Rourke et al., 2015)	Full length human *C9orf72* gene was expressed. G4C2 repeat size ranged from ~100-1000 repeats
Gain	Expansion containing BAC	Mouse	Yes	Yes	Not tested	Motor and Cognitive	Yes	Stable	(Liu et al., 2016)	Full length human *C9orf72* gene was expressed with 52Kb upstream and 19 Kb downstream sequence. Multiple lines of mice were generated with the following G4C2 repeat sizes (500), (500 and 32), (36 and 29) and (37), all lines were phenotypic with the exception of (37). Mice displayed anxiety like behaviour, paralysis and motor neuron loss. Only BAC mouse model to report TDP-43 pathology.
Gain	Expansion containing BAC	Mouse	Yes	Yes	Not tested	Cognitive	Yes	Stable	(Jiang et al., 2016)	Exons 1-5 of the human *C9orf72* gene with 140Kb of upstream sequence was expressed. G4C2 repeat lengths of ~110 or ~450 were reported in different mouse lines. Mice expressing ~450 repeats showed anxiety like behaviour and impaired working memory.
Gain	Viral overexpression of G4C2 expansion	Mouse	Yes	Yes	Not tested	Motor and Cognitive	Yes	Viral expression	(Chew et al., 2015)	66 G4C2 repeats were virally overexpressed by intracerebroventricular injection of virus. Mice showed anxiety like behaviour and rotarod performance deficits. Cortical neuron loss was also observed

Gain	Viral overexpression of poly(GA) encoding construct	Mouse	NA	Yes	NA	Motor and Cognitive	Yes	Viral expression	(Zhang et al., 2016)	50 poly(GA) encoding repeats were virally overexpressed by intracerebroventricular injection of virus. Mice showed anxiety like behaviour and rotarod performance deficits.
Gain	Viral overexpression of G4C2 expansion	Mouse	Yes	Yes	No	Motor and Cognitive	No	Viral expression	(Herranz-Martin et al., 2017)	102 G4C2 repeats were virally overexpressed by cisterna magna injection of virus. Mice showed defects in gait analysis and object recognition tests. Increased cleaved-PARP1 was reported in cerebellar regions but no cell loss was observed.
Gain	Viral overexpression of poly(GA) encoding construct	Mouse	NA	Yes	NA	Motor	Yes	Viral expression	(Walker et al., 2017)	69 poly(GA) encoding repeats were virally overexpressed by cisterna magna injection of virus. Mice showed defects in gait analysis and neuron loss in the brainstem
Gain	DPR encoding construct	Mouse	NA	Yes	NA	Cognitive	Yes	Stable	(Choi et al., 2019)	Mice stably expressing 80 poly(GR) encoding repeats in the forebrain. Mice showed defects in tests of social interaction and increased anxiety in the elevated plus maze test.
Loss	*C9orf72* gene knockout	Mouse	NA	NA	NA	No	No	Stable	(Koppers et al., 2015, Atanasio et al., 2016, Sudria-Lopez et al., 2016, Jiang et al., 2016, O'Rourke et al., 2016, Ji et al., 2017)	6 independently generated models *C9orf72* knockout mice reported similar results
Both	Expansion containg BAC combined with *C9orf72* gene knockout	Mouse	Not tested	Not tested	Not tested	Motor	Not tested	Stable	(Shao et al., 2019)	Phenotypic BAC mice generated previously by Liu and other authors were crossed with C9orf72 knockout mice. The precise BAC expansions size was not reported for these mice. Knockout of one or both copies of mouse C9orf72 increased the severity of the motor phenotype observed in BAC expansion mice.

Gain	Expansion containing construct	Zebrafish	Yes	No	Not tested	Not tested	No	Stable	(Ohki et al., 2017)	80 G4C2 repeats without an upstream ATG start codon. Mild cardiac phenotype.
Gain	Poly(GA) producing expansion	Zebrafish	Yes	Yes	Not tested	Not tested	No	Stable	(Ohki et al., 2017)	80 G4C2 repeats with an upstream start codon in the frame of poly(GA). Severe cardiac phenotype.
Gain	Expansion containing construct	Zebrafish	Yes	Not tested	Not tested	Not tested	Not tested	Chimeric expression	(Lee et al., 2013)	72 G4C2 repeat containing constructs were injected into zebrafish at the one cell stage. An increase in apoptotic markers was reported but not specifically in neural tissue
Gain	Expansion containing construct (sense)	Zebrafish	Yes	No	Not tested	Not tested	Motor axonop-athy	Transient	(Swinnen et al., 2018)	0.844µM of 3, 4, 10, 35, 70 and 90 G4C2 repeat RNA was injected at the 1-2 cell stage. Motor axonopathy was observed with 35 repeats or more. Poly(GR) was not detected in any condition. Poly(GP) was present at just above the limit of detection in 90 repeat condition only in an immunoassay.
Gain	Expansion containing construct (antisense)	Zebrafish	Yes	No	No	Not tested	Motor axonop-athy	Transient	(Swinnen et al., 2018)	0.844µM of 35 and 70 C4G2 repeat RNA was injected at the 1-2 cell stage. Motor axon abnormal branching was observed in both conditions. Poly(PR) and poly(GP) were not detectable in any condition
Gain	DPR encoding constructs	Zebrafish	NA	Yes	Yes	Not tested	Motor axonop-athy	Transient	(Swinnen et al., 2018)	0.844µM of 50 codon optimised repeats of poly(GR), poly(PR), poly(GP), poly (PA) and poly(GA) repeat RNA was injected at the 1-2 cell stage. Motoraxonopathy was observed in poly(GR) and poly(PR) injected conditions, but not in poly(GA), poly(PA) or poly(GP) conditions
Gain	Expansion containing construct (RNA-only)	Zebrafish	See notes	See notes	See notes	Not tested	Motor axonop-athy	Transient	(Swinnen et al., 2018)	0.844µM of RNA-only RNA of repeat lengths 70 (sense), 108 (sense), 70 (antisense) and 108 (antisense) were injected at the 1-2 cell stage. Motor axonopathy was observed in all conditions. Although RNA foci and DPR production was not investigated in the present study, the presence of RNA foci and absence
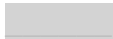
Loss	C9orf72 knockdown	Zebrafish	NA	NA	NA	Motor	Axonop-athy	Transient	(Ciura et al., 2013)	Morpholino mediated knockdown. Defects in spontaneous swimming and touch evoked escape responses were observed. Defects in early axonal outgrowth were also reported.
Loss	Non-functional C9orf72 expression	Zebrafish	NA	NA	NA	Motor	Yes	Transient	(Yeh et al., 2018)	Injection of non-functional C9orf72 RNA (lacking various DENN domain segments). Also confirmed with C9orf72 morpholino knockdown. Defects in touch evoked escape response and neuronal apoptosis observed
Loss	C9orf72 gene knockout	Zebrafish	NA	NA	NA	No	No	Stable	(Stepto et al., 2014)	Details of model are not yet published
Gain	Poly(PR) encoding construct	Drosophila	NA	NA	Yes	Not tested	Yes	Stable	(Wen et al., 2014)	50 poly(PR) encoding repeats resulted in neurodegeneration, however an equal number of poly(PA) and poly(GA) encoding repeats did not
Gain	Expansion containing construct	Drosophila	Yes	Yes	Yes	Not tested	Yes	Stable	(Mizielinska et al., 2014)	103 G4C2 repeats, produced poly(GP) and poly(GR) DPR proteins
Gain	Expansion containing construct (RNA-only)	Drosophila	Yes	No	No	Not tested	No	Stable	(Mizielinska et al., 2014)	108 RNA-only G4C2 repeats. RNA-only repeats were generated by inserting bi-directional multi-frame stop codons every 12 repeats
Gain	Poly(PR) encoding construct	Drosophila	NA	NA	Yes	Not tested	Yes	Stable	(Mizielinska et al., 2014)	100 poly(PR) encoding repeats. The same results were also obtained with 100 poly(GR) encoding repeats
Gain	Poly(GA) encoding construct	Drosophila	NA	Yes	NA	Not tested	No	Stable	(Mizielinska et al., 2014)	100 poly(GA) encoding repeats. A small reduction in lifespan was observed in this model. The same results were obtained with 100 poly(PA) encoding repeats, however no change in lifespan was observed in this case
Gain	Expansion containing construct	Drosophila	Yes	Yes	Not tested	No	No	Stable	(Tran et al., 2015)	160 G4C2 intronic repeats expressed between exon 1 and exon 3 of the human C9orf72 gene.

"of DPR when expressing these constructs in drosophila has been reported previously (Mizielinska et al., 2014)"

Gain	Expansion containing construct	Drosophila	Yes	Yes	Not tested	Not tested	Yes	Stable	(Tran et al., 2015)	A poly(A) tail was not present in this construct. Poly(GP) was detected, but not poly(GR). 34 G4C2 repeats expressed in the context of a poly(A) tail. Poly(GP) and poly(GR) were detected.
Loss	*C9orf72* gene knockout	C. elegans	NA	NA	NA	Motor	Yes	Stable	(Therrien et al., 2013)	Paralysis and motor neuron degeneration was observed

Neurodegen: Presence of neurodegeneration. Studies which utilised multiple distinct animal models have been separated into multiple table entries for clarity.

Appendix B   Sequence of 99 G$_4$C$_2$ containing transgene

▢ (grey)	**ISceI meganuclease restriction site**
▢ (cyan)	**Zebrafish ubiquitin promotor**
▢ (green)	**GFP**
▢ (blue)	**Start codon**
▢ (dark red)	**Stop codon**
▢ (yellow)	**(G$_4$C$_2$)$_{99}$ C9orf72 expansions**
▢ (magenta)	**Heat shock promotor 70**
▢ (red)	**DsRed**

Listed below is the entire DNA sequence contained in a single transgene copy -

cagggtaatagcttttaatctcaaaaaacattaaatgaaatgcatacaaggttttatcctgctttagaactgtttgtatttaatt
atcaaactataagacagacaatctaatgccagtacacgctactcaaagttgtaaaacctcagatttaacttcagtagaag
ctgattctcaaaattgttagtgtcaagcctagctcttttggggctgaaaagcaatcctgcagtgctgaaaagcctctcaca
ggcagccgatgcgggaagaggtgtattagtcttgatagagaggctgcaaatagcaggaaacgtgagcagagactcc
ctggtgtctgaaacacaggccagatgggtttaaacgaattcgtcgagaccagcaaagttctagaatttgtcgaaacattt
atgttatatatttcctgaaaaaaattctgagtaagttcttaagtgttattgccagcaacataaacaacagacggcaaaatga
ataaatgataacaaagcagtaggcttaaataaacctaattttttataggctgttctctacaaccctcaaacagtgattagtttt
gtacttataaacttgcccctttcattcatatttcaagaaaattggttcagaagatctggatattctagcagttgttcaagctcat
ggagggatcagtgacctgattccacaatgactaggcctaatccagaaattagatgactgtcaacataaaaaggcacag
cactcactagctgccctatatattttattatattttacatatattattttatttatttagctctgagtgctgtactttctggttaaaga
aaactgcttacaacagctaacctgtactacctcaggctcagggaatttggaacaggtttgtctggtttgtttctttaaccat
gcatgcttgttttcaactatggcaacacagtcacatgggacattacagaaatgatttgtcgatgacatgcgacttttctttaa
taaagcgcaaagatcccaaaaagcaaacttttaacaaaaatcatataattatattttcaatccagctttgtagcaactttgt
gctgctgttcactcagcaacagatagtcagtataaggtcagtgtgtctcaaagcagtgccatctgtttcacacattgcgtt
ctatatataagtgtgctggttgacacgacactgtataaggcctaggctaaaacacaaacaatgtagaatgacactgtgtt
tttttgtaaacaaatgttgtttttggttaaacatctttgtgaaaacatcctcctgtcatgtatttgctatattcaaatgttaaacc
cgtgcagaatagaacatatacaaaaaaaaacaacacaacacattttaaacattattaaatatcaagtattgctggcagtt
ctgtttctgttttacagtaccctttgccacagttctccgcttttcctggtccagattccacaagtctgattcaccaatagcaaa
gcgaataaacaaccaaagcagccaatcactgcttgtagactgtcctgcgagaccggcccattccagcacattctgga
aacttcctttatatgataattataaatacatttaaattattgatacaaaacatgtaattcctagaacataaccatagcaatcatt
agttttcagggtaattatgtattttttaggatttgactgcggaaagatctggtcatgtgacgtctcatgaacgtcacggccct
gggtttctataaatacagtaggactctcgaccatcggcagattttttcgaagaagaagatcagtttcaggagccgtactgtt
ccgtttcaacgcaaatattaacggtaagagcgaatttcctagtttgttttcatgccattctttaaaaccatagcgtattacttta
attatagtaaactttcgctttctttattacaagagacgtttgtgttgattctccgcggacatttctggtcagacaatcagaaa
atgaccgcggaggaccagtaacttgcattacacgtaagttaaatcttcgtgtattaaaatggttaggttgttaacgtcaaat
aggttaccgtgtttgcgtgtgatcaggttggttttgttagattttttgtcagtattttttaatttatttgttttagtttatttattttttttgct
gaatcatagtttgtgaacaaagaacccggatgttacatacagtacagccgccatgttacagagagttataacttaatcatt
ttaaaaataattttgccttacttttagtttgtcatgttgagaaatgaggaaatgttaaaatgaggaaatatccaattaatttaat
atatcaaaataatccatgattacaatgcactgaactggagaaaattaagatgttttctagtgtcatgaaacaaatgtaaga
gatgtacattgtagatgttttatgtcaagaattggctagttgatgcagcatactggcgatactcagttgtaataacagtaac

gttacatgttaatagactactgagtatgctgttctgtctatgtatgctctgtaagctacgagaaggactttttaaacagtaaa
gggtgcaatattttacaaattgaattaaataaaggctgtctattaagtaatatgcttgatattttcttacttgatcggaaataa
gaaaaatataaacgttgttgctctaaaaatcctagttcagtttagccaaccacaaataccttttgttcctccaacagttttt
ttttcttctctataatatttggcagtctatagtactccaaatgtttccccacagtctaactaattggtacagccaaaatcatga
cacttattgcaataataattttggttcattggcattgttgatagcctgtgccactaatatggtcgattgatcatgcttcaggaa
gaaaactatattgtttgatgtaagattattaaatcttcacctgcttccattacaaactattcccatcttattgaattctggtatgt
cttaaaggattagttcacttcccaaatcaaaatttacttagcctttttttcatccatgatccctttttttcatcattaatgaagaaat
tgtttttgaaaaagtttcaagatttttttctctatattgtggagcttgttagtttaaaattccaaaatgcaatatgtggcttcaaat
ggttctaaatgatcccagtcaaggaataacagtcttatctaatgaaaccattagacctttaaaaataaaaaaaataaaagt
atttatttttaaatgactgagtgattaagttgaatttcagcgtttccttactgtgtagaagtccttccttactggccccacccttt
ggttctctgccaatctgctacctaatgtaatgttgtggaacattattattctttatttcttaattttttatttttttattttaaaacaatgt
aaactgcacagatgtgcagtttgtttaaaatggccaatgctttggaaatgcatgacataattagatttcatgatgcacaaa
gccaaatctcagagcttgtgcaaaatgagctatcatttcactaggtaagaccctaaattttcatataggatcatttggacaa
ttttgctgcaggtaaaatgcattctatagtccactgtcagccattgtttttggatagtatttattttttctctacaagtatagtcaat
agttttctattattttaaaggtttgtaacatttaagggtgaccaaatgcaaagtaaaatttcattttcgggtgaactatctcgttt
aacatgggagaagtgcaaaacatacattattggctagaacattgtagtatttttaaatggaaatgtgtgattgctaatcttacttttgaatttgtttacagggatctataacttcgtatagtataccttatacgaagttatcgagctcataacttcgtataggatact
ttatacgaagttataacgtcgagataacttcgtatagcatacattatacgaagttataggatcaatggtgagcaagggcg
aggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtc
cggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgt
gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcag
cacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaacta
caagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaa
ggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaa
gcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgacc
actaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccg
ccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactct
cggcatggacgagctgtacaaaaaagcaggctttaaagggaggtagtgagtcgacggggccggggccggggccg
ggccggggccggggccggggccggggccggggccggggccggggccggggcctcgacggggccggggc
cggggccggggccggggccggggccggggccggggccggggccggggccggggccggggccggggccgg
ggccggggccggggccctcgacggggccggggccggggccggggccggggccggggccggggccggggcc
ggggccggggccggggcctcgacggggccggggccggggccggggccggggccggggccggggccggggg
ccggggccggggccggggccggggccggggccggggccggggccggggccggggccggggccggggccg
gggccggggccggggcctcgacggggccggggccggggccggggccggggccggggccggggccggggc
cggggccggggccggggccggggccggggccggggccggggcctcgacggggccggggccggggccg
gggccggggccggggccggggccggggccggggccggggcctcgacggggccggggccggggccg
gggccggggccggggccggggccggggccggggccggggccggggcctcgagatatctagaccca
gctttcttgtacatgtagaattcatgatgggcagtgtcgagctgaatctgagggagactgagctgtgtcttggtcttcccg
gtggagatacagtggctccggtaaccggaaacaagagagggttctcagagacggttgatctgaagctaaatctgaat
aatgagcctgcaaacaaggaaggatctacgactcatgacgtcgtgacttttgattccaaggagaagagtgcttgtccta
aagatccagccaaacctccggccaaggcacaagttgtgggatggccaccggtgagatcataccggaagaacgtgat
ggtttcctgccaaaaatcaagcggtggcccggaggcggcggcgttcgtgaaggtatcaatggacggagcaccgtac
ttgaggaaaatcgatttgaggatgtataaaagctacgatgagctttctaatgctttgtccaacatgttcagctcttttaccat
gggcaaacatggaggagaagaaggaatgatagacttcatgaatgagaggaaattgatggatttggtgaatagctggg
actatgttccctcttatgaagacaaagacggtgattggatgctcgtcggcgacgttccttggccaatgttcgtcgatacat
gcaagcgtttacgtctcatgaaaggatcggatgccattggtctcgctccgagggcgatggagaagtgcaagagcaga
gcttagcccggcggcatggacgagctgtacaagtaatacgtagatccagacatgataagatacattgatgagtttgga
caaaccacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagc
tgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcaggggggaggtgtgggaggttttttaattcgc
ggccaactttattatacatagttgataattcactggccgtcgttttacggtactacggtacggatccttcagggtggatcc
ttcaggggtgtcgcttggtgatttccaaaaatcaaattaattttattaaactattagaacgagcatgtttttgtctatatgctaca

gaagataaaaaataataggagttaacagttataaaacaacacactttgtttctattgattgttgaccacactggggtctcat
taagttagattaaagacacactaactgggtcaaatgcagcagattgatttcatggcaccagggtaaactttctaacacttt
tacggcaatcatatacatcaaaattaaatacaggccacgactgaacaaggaggatgatctccaaaattaaacaaagag
acttgtgcctatttctctgagggtaaacatgacctctcaagttagcaagttgtttttaacactacaaaaatagttaagactgc
aatcccagaataaagtattggttttaaccaatcaatatagtacagtaaacatccatttgtttgttgaaacgttaaacgaatc
tgaccaagctattagcttatataaaacaggtttgcctttatgtagctgaaaataccacaggcccgattttgctactgtgtaa
aacatttcagcaagatttttttatattgcattttttctactgaatcgttcaaacattttatcattttagtttgttcattcattgcaact
ggaaaaacaacacaccacacaaccgcacattttcagcaataagtacaataaaacactcaaataaaaaaaaacttttta
aatctctttgtattttgaccgctgtttcgcgtaatttcacggtaaaactctggaaatctccactacattcctctcagcggctc
ctctcaatgacagctgaagaagtgacgcggctgcctgctgtgtttgattggtcgaattcactggaggcttccagaaca
gtgtagagtctgaacgggtgcgcgctctgctgtatttaaagggcgaaagagagaccgcagagaaactcaaccgaag
agaagcgacttgacaaagaagaaaagagcagcctgacaggacttttccccgacgaggtgtttattcgctctatttaaga
atctactgtaaggtaagtctcaatatattgtactctagtggctaatcaaaattttatagagattatatgtacttaatgtcaaaaa
atctactttgtatatgtaatctttttacatgtggactgcctatgttcatcttattttaggtctactagaaaattatatttcccgttttc
acaataaggatttaaaaaaaaagcaatgaacagactggcatttactttatgttgctgacattattatatatgagcataataa
ccataaatactagcaaatgtcctaaatgaatttgtgttaatgttgtctacaaaagaaaagaaaattagcgttttacttgtaca
actaataataacttagttattaagagaatttcacttgttgactagaaaaatcctttcataatgaaacaattgcaccataaattg
tataaatataaaattaattctaattgtttttttttttttcctgcagtcgagcgccaccatggcctactccgagaaagtcatcaccg
agttcatgcgcttcaaggtgcgcatggagggcaccgtgaacggccacgagttcgagatcgagggcgagggcgagg
gccgcccctacgagggccacaacaacgtgaagctgaaggtgaccaagggcggccacctgcccttcgcctgggac
atcctgtcccaccagttccagtacggctccaaggtgtacgtgaagcaccccgccgacatcaccgactacaagaagct
gtcattccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggcgacagtgacccagga
ctcctccctgcaggacggctgcttcatctacaaggtgaagttcatcggagtgaacttcacctccgacggccccgtgatg
cagaagaagaccatgggctgggaggcctcaaccgagcgcatgtaccccccgcgacggcgtgctgaagggcgaga
cccacaaggccctgaagctgaaggacggcggccactacctggtggagttcaagtccatctacatggacaagaagcc
agtgcagctgaccggctactactacgtggacgccaagctggacatcacctcccacaacgaggactacaccatcgtg
gagcagtacgagcgcaccgagggccgccaccacctgttctaagcggccgcgactctagatcataatcagccatacc
acatttgtagaggttttacttgctttaaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttg
ttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcact
gcattctagttgtggtttgtccaaactcatcaatgtatcttaggtacctagggataa

Appendix C   Sequence of 89 C_4G_2 containing transgene

ISceI meganuclease restriction site

Zebrafish ubiquitin promotor

GFP

Start codon

Stop codon

$(C_4G_2)_{89}$ C9orf72 expansions

Heat shock promotor 70

DsRed

NB: Due to difficulties in sequencing GC rich DNA we were not able to obtain a sequencing read which spanned the entire C_4G_2 expansion, therefore expansion size is listed as 89 C_4G_2 repeats which is the minimum known size ascertained by comparing forward and reverse sequencing reads through the expansion. Additionally, as sequencing reads could not span the C_4G_2 expansion, the reading frame could not be determined and the precise location of the stop codon is not known.

Listed below is the entire DNA sequence contained in a single transgene copy -

cagggtaatagcttttaatctcaaaaaacattaaatgaaatgcatacaaggtttttatcctgctttagaactgtttgtatttaatt

atcaaactataagacagacaatctaatgccagtacacgctactcaaagttgtaaaacctcagatttaacttcagtagaag

ctgattctcaaaattgttagtgtcaagcctagctcttttgggggctgaaaagcaatcctgcagtgctgaaaagcctctcaca

ggcagccgatgcgggaagaggtgtattagtcttgatagagaggctgcaaatagcaggaaacgtgagcagagactcc

ctggtgtctgaaacacaggccagatgggtttaaacgaattcgtcgagaccagcaaagttctagaatttgtcgaaacattt

atgttatatatttcctgaaaaaaattctgagtaagttcttaagtgttattgccagcaacataaacaacagacggcaaaatga

ataaatgataacaaagcagtaggcttaaataaacctaatttttataggctgttctctacaaccctcaaacagtgattagtttt

gtacttataaacttgccctttcattcatatttcaagaaaattggttcagaagatctggatattctagcagttgttcaagctcat

ggagggatcagtgacctgattccacaatgactaggcctaatccagaaattagatgactgtcaacataaaaaggcacag

cactcactagctgccctatatattttattatatttttacatatattatttttatttatttagctctgagtgctgtactttctggttaaaga

aaactgcttacaacagctaacctgtactacctcaggctcagggaatttggaacaggtttgtctggtttgtttctttaaccat

gcatgcttgttttcaactatggcaacacagtcacatgggacattacagaaatgatttgtcgatgacatgcgacttttctttaa

taaagcgcaaagatcccaaaaagcaaacttttaacaaaaatcatataattatattttcaatccagctttgtagcaactttgt

gctgctgttcactcagcaacagatagtcagtataaggtcagtgtgtctcaaagcagtgccatctgtttcacacattgcgtt

ctatatataagtgtgctggttgacacgacactgtataaggcctaggctaaaacacaaacaatgtagaatgacactgtgtt
ttttttgtaaacaaatgttgtttttggttaaacatctttgtgaaaacatcctcctgtcatgtatttgctatattcaaatgttaaacc
cgtgcagaatagaacatatacaaaaaaaaacaacacaacacattttttaaacattattaaatatcaagtattgctggcagtt
ctgtttctgttttacagtaccctttgccacagttctccgcttttcctggtccagattccacaagtctgattcaccaatagcaaa
gcgaataaacaaccaaagcagccaatcactgcttgtagactgtcctgcgagaccggcccattccagcacattctgga
aacttcctttatatgataattataaatacatttaaattattgatacaaaacatgtaattcctagaacataaccatagcaatcatt
agttttcagggtaattatgtattttttaggatttgactgcggaaagatctggtcatgtgacgtctcatgaacgtcacggccct
gggtttctataaatacagtaggactctcgaccatcggcagattttttcgaagaagaagatcagtttcaggagccgtactgtt
ccgtttcaacgcaaatattaacggtaagagcgaatttcctagtttgttttcatgccattctttaaaaccatagcgtattacttta
attatagtaaactttcgctttctttattacaagagacgttttgtgttgattctccgcggacattttcggtcagacaatcagaaa
atgaccgcggaggaccagtaacttgcattacacgtaagttaaatcttcgtgtattaaaatggttaggttgttaacgtcaaat
aggttaccgtgtttgcgtgtgatcaggttggttttgttagattttttgtcagtatttttaatttatttgttttagtttatttatttttttttgct
gaatcatagtttgtgaacaaagaacccggatgttacatacagtacagccgccatgttacagagagttataacttaatcatt
ttaaaaataattttgccttactttttagtttgtcatgttgagaaatgaggaaatgttaaaatgaggaaatatccaattaatttaat
atatcaaaataatccatgattacaatgcactgaactggagaaaattaagatgtttttctagtgtcatgaaacaaatgtaaga
gatgtacattgtagatgtttttatgtcaagaattggctagttgatgcagcatactggcgatactcagttgtaataacagtaac
gttacatgttaatagactactgagtatgctgttctgtctatgtatgctctgtaagctacgagaaggactttttttaaacagtaaa
gggtgcaatattttttacaaattgaattaaataaaggctgtctattaagtaatatgcttgatattttttcttacttgatcggaaataa
gaaaaaatataaacgttgttgctctaaaaatcctagttcagtttagccaaccacaaataccttttttgttcctccaacagttttt
ttttcttctctataatatttggcagtctatagtactccaaatgtttccccacagtctaactaattggtacagccaaaatcatga
cacttattgcaataataattttggttcattggcattgttgatagcctgtgccactaatatggtcgattgatcatgcttcaggaa
gaaaactatattgtttgatgtaagattattaaatcttcacctgcttccattacaaactattcccatcttattgaattctggtatgt
cttaaaggattagttcacttcccaaatcaaaatttacttagcctttttttcatccatgatccctttttttttcatcattaatgaagaaat
tgtttttgaaaaagtttcaagatttttttctctatattgtggagcttgttagtttaaaattccaaaatgcaatatgtggcttcaaat
ggttctaaatgatcccagtcaaggaataacagtcttatctaatgaaaccattagacctttaaaaataaaaaaaataaaagt
atttatttttaaatgactgagtgattaagttgaatttcagcgtttccttactgtgtagaagtccttccttactggcccccacccttt
ggttctctgccaatctgctacctaatgtaatgttgtggaacattattattctttatttcttaattttttattttttattttaaaacaatgt
aaactgcacagatgtgcagtttgtttaaaatggccaatgctttggaaatgcatgacataattagatttcatgatgcacaaa
gccaaatctcagagcttgtgcaaaatgagctatcatttcactaggtaagaccctaaattttcatataggatcatttggacaa
ttttgctgcaggtaaaatgcattctatagtccactgtcagccattgtttttggatagtatttattttttctctacaagtatagtcaat
agttttctattattttaaaggtttgtaacatttaagggtgaccaaatgcaaagtaaaatttcattttcgggtgaactatctcgttt
aacatgggagaagtgcaaaacatacattattggctagaacattgtagtatttttttaaatggaaatgtgtgattgctaatctta
ctttgaatttgtttacagggatctataacttcgtatagtataccttatacgaagttatcgagctcataacttcgtataggatact
ttatacgaagttataacgtcgagataacttcgtatagcatacattatacgaagttataggatcaatggtgagcaagggcg
aggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtc
cggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgt
gccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcag

cacgacttcttcaagtccgccatgccccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaacta

caagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaa

ggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaa

gcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgacc

actaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccg

ccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactct

cggcatggacgagctgtacaagaaagctgggtctagatatctcgaggccccggccccggccccggccccggcccc

ggccccggccccggccccggccccggccccggccccggccccgtcgaggccccggccccggccccggcccc

ggccccggccccggccccggccccggccccgtcgaggccccggccccggccccggccccggccccggcccc

ggccccggccccggccccggccccggccccggccccggccccggccccggccccggccccggccccggcccc

cggccccggccccggccccggccccggccccggccccggccccggccccggccccggccccggccccggcc

ccggccccggccccggccccggccccggccccggccccggcccctcgaggccccggccc

cggccccggccccggccccggccccggccccggccccggccccggccccggccccggccccggccccggcc

ccggccccggccccggccccgtcgaggccccggccccggccccggccccggccccggccccggccccggcc

ccggccccggccccggccccggccccgtcgactcactacctccctttaaagcctgcttttttgtacatgtagaattcatg

atgggcagtgtcgagctgaatctgagggagactgagctgtgtcttggtcttcccggtggagatacagtggctccggta

accggaaacaagagagggttctcagagacggttgatctgaagctaaatctgaataatgagcctgcaaacaaggaag

gatctacgactcatgacgtcgtgactttttgattccaaggagaagagtgcttgtcctaaagatccagccaaacctccggc

caaggcacaagttgtgggatggccaccggtgagatcataccggaagaacgtgatggtttcctgccaaaaatcaagcg

gtggcccggaggcggcggcgttcgtgaaggtatcaatggacggagcaccgtacttgaggaaaatcgatttgaggat

gtataaaagctacgatgagctttctaatgctttgtccaacatgttcagctctttaccatgggcaaacatggaggagaaga

aggaatgatagacttcatgaatgagaggaaattgatggatttggtgaatagctgggactatgttccctcttatgaagaca

aagacggtgattggatgctcgtcggcgacgttccttggccaatgttcgtcgatacatgcaagcgtttacgtctcatgaaa

ggatcggatgccattggtctcgctccgagggcgatggagaagtgcaagagcagagcttagcccggcggcatggac

gagctgtacaagtaatacgtagatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagt

gaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaaca

acaattgcattcattttatgtttcaggttcaggggggaggtgtgggaggtttttttaattcgcggccaactttattatacatagtt

gataattcactggccgtcgttttacggtactacggtacggatccttcagggggt<mark>ggatccttcagggggtgtcgcttggtgat

ttccaaaaatcaaattaattttattaaactattagaacgagcatgtttgtctatatgctacagaagataaaaaataataggatt

gttaacagttataaaacaacacactttgtttctattgattgttgaccacactggggtctcattaagttagattaaagacacac

taactgggtcaaatgcagcagattgatttcatggcaccagggtaaactttctaacacttttacggcaatcatatacatcaa

aattaaatacaggccacgactgaacaaggaggatgatctccaaaattaaacaaagagacttgtgcctatttctctgagg

gtaaacatgacctctcaagttagcaagttgtttttaacactacaaaaatagttaagactgcaatcccagaataaagtattg

gttttaaccaatcaatatagtacagtaaacatccatttgtttgttgaaacgttaaacgaatctgaccaagctattagcttata

taaaacaggtttgccttttatgtagctgaaaataccacaggcccgattttgctactgtgtaaaacatttcagcaagattttttt

tatattgcattttttctactgaatcgttcaaacattttatcattttagtttgttcattcattgcaactggaaaaacaacacaccac

acaaccgcacattttttcagcaataagtacaataaaacactcaaataaaaaaaaaactttttaaatctctttgtattttttgaccg</mark>

ctgtttcgcgtaatttcacggtaaaactctggaaatctccactacattcctctcagcggctcctctcaatgacagctgaag

aagtgacgcggctgcctgctgtgttttgattggtcgaattcactggaggcttccagaacagtgtagagtctgaacgggt

gcgcgctctgctgtatttaaagggcgaaagagagaccgcagagaaactcaaccgaagagaagcgacttgacaaag

aagaaaagagcagcctgacaggacttttccccgacgaggtgtttattcgctctatttaagaatctactgtaaggtaagtct

caatatattgtactctagtggctaatcaaaattttatagagattatatgtacttaatgtcaaaaaatctactttgtatatgtaatc

tttttacatgtggactgcctatgttcatcttattttaggtctactagaaaattatatttcccgttttcacaataaggatttaaaaaa

aaagcaatgaacagactggcatttactttatgttgctgacattattatatatgagcataataaccataaatactagcaaatgt

cctaaatgaatttgtgttaatgttgtctacaaaagaaaagaaaattagcgtttacttgtacaactaataataacttagttatta

agagaatttcacttgttgactagaaaaatcctttcataatgaaacaattgcaccataaattgtataaatataaaattaattcta

attgttttttttttttcctgcagtcgagcgccacc**atg**gcctactccgagaaagtcatcaccgagttcatgcgcttcaaggtg

cgcatggagggcaccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgccccctacgagggcc

acaacaacgtgaagctgaaggtgaccaagggcggccacctgcccttcgcctgggacatcctgtcccaccagttcca

gtacggctccaaggtgtacgtgaagcaccccgccgacatcaccgactacaagaagctgtcattccccgagggcttca

agtgggagcgcgtgatgaacttcgaggacggcggcgtggcgacagtgacccaggactcctccctgcaggacggct

gcttcatctacaaggtgaagttcatcggagtgaacttcacctccgacggccccgtgatgcagaagaagaccatgggct

gggaggcctcaaccgagcgcatgtacccccgcgacggcgtgctgaagggcgagacccacaaggccctgaagctg

aaggacggcggccactacctggtggagttcaagtccatctacatggacaagaagccagtgcagctgaccggctact

actacgtggacgccaagctggacatcacctcccacaacgaggactacaccatcgtggagcagtacgagcgcaccg

agggccgccaccacctgttc**taa**gcggccgcgactctagatcataatcagccataccacatttgtagaggttttacttgc

tttaaaaaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagctt

ataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtcca

aactcatcaatgtatcttaggtacc**tagggataa**

Appendix D   PhD outputs

The data described in this thesis have been presented at various stages of my PhD in the form of poster presentations at the University of Sheffield research day and in the form of oral presentations in the SITraN seminar series.

Additionally, data generated during this thesis have been published in the journal Acta Neuropathologica Communications:

> Shaw MP, Higginbottom A, McGown A, Castelli LM, James E, Hauterbeurgue GM, Shaw PJ* and Ramesh TM* Stable transgenic C9orf72 zebrafish model key aspects of the ALS/FTD phenotype and reveal novel pathological features*. Acta Neuropathol Commun*. 6:125. (November 2018)

> * These authors contributed equally to this work

A further manuscript including the data on '5.3' RNA-only zebrafish models is currently in preparation.