MECHANISTIC INSIGHT INTO DISEASE PATHOGENESIS OF HEXANUCLEOTIDE REPEAT EXPANSION DISORDERS

A Dissertation
Presented to
The Academic Faculty

by

Zachary Thomas McEachin

In Partial Fulfillment
of the Requirements for the Degree
Biomedical Engineering in the
College of Engineering

Georgia Institute of Technology / Emory University
August 2019

MECHANISTIC INSIGHT INTO DISEASE PATHOGENESIS OF HEXANUCLEOTIDE REPEAT EXPANSION DISORDERS

Approved by:

Dr. Gary J. Bassell, Advisor
Department of Cell Biology
Emory University

Dr. Jonathan D. Glass
Department of Neurology
Emory University

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Thomas Kukar
Department of Pharmacology
Emory University

Dr. Philip J. Santangelo
School of Biomedical Engineering
Georgia Institute of Technology

Date Approved: [July 12, 2019]

This thesis is dedicated to my mother, my father and my children, Zachary and Zoe.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Gary Bassell. Your mentorship and guidance have made this project possible. The freedom to pursue a project that truly interests me has allowed me to maintain my excitement and passion for my work. My experience in your lab, under your mentorship, has allowed me to develop independently and confidently as a scientist. I am truly grateful for the opportunities you have given me over the years, and I appreciate your constant support in and outside of lab.

I would like to sincerely thank all members of my committee, Dr. Jonathan Glass, Dr. Phil Santangelo, Dr. Nicholas Hud, and Dr. Thomas Kukar. The guidance and support you have given me over the years have undoubtedly benefited my project. Whether it was something I learned in a class you taught, a brief chat in the hallway, or long discussions over lunch, I have valued every interaction.

I would like to thank all members of the Bassell lab, both past and present for their support, encouragement and insightful discussions over the years. To Anwesha, Arielle, and Ryan – I appreciate your support, conversations, and help. To Megan, I owe all my reagents to you! Thank you, you were a very important part of this journey and I appreciate your help and friendship over the years. Nisha, I am not sure what this PhD would have been without you. You have inspired me in more ways than you know. I wouldn't have wanted to navigate this journey with anyone else. Thank you for everything.

I would like to thank all those that I have collaborated with over the years. Wilfried Rossoll, thank you for the coffee, conversations, and collaborations. Nick Boulis, thank you for taking a chance on me at the beginning and the continued opportunities to grow as a scientist. Chad Hales, you have been a colleague, mentor, and friend. Malu Tansey, thank you for your support and for showing me that "you don't get what you don't ask for"! Jie Jiang, despite arriving later in my graduate career, you have been a great friend and collaborator. To all those in the CND and ALS clinic that I have interacted with over the years, thank you for your help.

To my family, I cannot begin to describe the love and support you have shown me throughout this journey. Marilyn, you have always been there for me no matter what. I have always looked up to you and I appreciate everything you have done for me. I love you. David, you have always been a role model to me. I hope that I can meet the standard you have set not only in my career, but as a brother, and as a father as well. I love you. Mark, I know that I can talk to you about anything first, good or bad. I appreciate your support over the years and fatherly advice. I love you. To my mom and dad, this would not have been possible without you. Throughout life, despite everything, you have shown me unconditional love and support. I know the struggles you have endured and sacrifices you have made over the years and there are no words that can capture my appreciation for everything you have done. I love you.

To my children, Zachary and Zoe, you are my blessings. You are my motivation in this life. As I finish up this last sentence of my thesis, I can only think of your future and the change and happiness that I know you will bring to this world. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AAV	Adeno-Associated Virus
ALS	Amyotrophic Lateral Sclerosis
BAC	Bacterial Artificial Chromosome
c9ALS/FTD	*C9orf72*-linked Amyotrophic Lateral Sclerosis/Frontotemporal Dementia
C9orf72	Chromosome 9 open reading frame 72
cDPR	Chimeric Dipeptide Repeat
DEG	Differentially Expressed Genes
DENN	Differentially Expressed in Normal and Neoplastic cells
DM1	Myotonic Dystrophy Type 1
DM2	Myotonic Dystrophy Type 2
DSP	Digital Spatial Profiling
DPR	Dipeptide Repeat
fALS	Familial Amyotrophic Lateral Sclerosis
FBL	Fibrillarin
FTLD-MND	Frontotemporal Lobar Dementia – Motor Neuron Disease
FTD	Frontotemporal Dementia
FUS	Fused in Sarcoma
G_4C_2	GGGGCC
GEF	GDP-GTP Exchange Factor
HRE	Hexanucleotide Repeat Expansion
iPSC	Induced Pluripotent Stem Cell
iPS-MN	Induced Pluripotent Stem Cell derived Motor Neuron
LCL	Lymphablastoid Cell Line

MCP	Bacteriophage MS2 Coat Protein
MND	Motor Neuron Disease
MS2	Bacteriophage MS2 Stem Loop
NCI	Neuronal Cytoplasmic Inclusion
NII	Neuronal Intranuclear Inclusion
NOP56	Nucleolar protein 56
NOP58	Nucleolar protein 58
pTDP-43	Phosphorylated TDP-43
RBP	RNA-binding protein
RP-PCR	Repeat-primed polymerase chain reaction
RAN Translation	Repeat-associated non-AUG translation
sALS	Sporadic Amyotrophic Lateral Sclerosis
SBT	Somatic Brain Transgenesis
SCA	Spinocerebellar Ataxia
SCA3	Spinocerebellar Ataxia Type 3
SCA36	Spinocerebellar Ataxia Type 36
snoRNA	Small Nucleolar RNA
snoRNP	Small Nucleolar Ribonucleoprotein
SNP	Single Nucleotide Polymorphism
SOD1	Cu/Zn superoxide dismutase 1
TARDBP	Transactive response DNA binding protein (gene)
TDP-43	Transactive response DNA binding protein 43 kDa (protein)
TG_3C_2	TGGGCC

SUMMARY

In 2011, two highly related hexanucleotide repeat expansions (HRE) were discovered as the causes of two different diseases – an intronic G_4C_2 HRE in the C9ORF72 gene locus represents the most prevalent genetic cause of Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD), referred to as c9FTD/ALS, and a TG_3C_2 HRE in the first intron of NOP56 gene was independently identified as the genetic cause for a clinically disparate disease, Spinocerebellar Ataxia type 36 (SCA36). The overall aim of this proposal is to elucidate the molecular mechanisms by which hexanucleotide repeat expansions result in neurodegenerative disease, and to identify which mechanisms are responsible for the disease-specific pattern of neuronal degeneration and loss in these disorders. Aim (1) will determine whether these intronic hexanucleotide repeats are unconventionally translated into dipeptide repeats, Aim (2) will identify and compare disease specific RNA-protein interactions, and lastly we will use next generation sequencing to elucidate shared and distinct dysregulated pathways in these two disorders.

We hypothesize that a comprehensive and comparative analysis using c9FTD/ALS and SCA36 patient samples will provide us with a unique opportunity to gain a thorough understanding of the common and disease-specific pathomechanisms of these devastating disorders, for which there are presently no available cures or effective treatments.

CHAPTER 1. INTRODUCTION

"The very large number of members in the families of repeated sequences remains a most surprising feature for which an explanation must be sought. It may be reasonably predicted the large-scale new patterns of relationship among proteins await discovery"

R.J. Britten & D.E. Kohne, *Science*, 1968

All self-replicating, cellular organisms contain deoxyribonucleic acid (DNA). From archaea to humans, the storage and transfer of genetic information relies on various combinations of the four same DNA nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T). In the genomes of all organisms, specific combinations of nucleotides are used as a template to produce the workhorses of the cell – proteins. However, a large portion of the genome, particularly in humans, is made up of stretches of nucleotides that are highly repetitive and do not encode for proteins. Once thought to be "junk DNA", these sequences are now known to play an important role in human health and disease. This thesis is focused on understanding how neurological disorders arise when the integrity of these sequences is compromised.

1.1 Repeat Expansion Disorders

Pathological expansions of short, genomic repeat sequences, referred to as microsatellite expansions, have been implicated in a number of neurodevelopmental, neurodegenerative, or neuromuscular disorders (Pearson et al., 2005). These pathological expansions can be found in both translated (exons) or untranslated (5' UTR, 3' UTR, introns) regions of a gene (DeJesus-Hernandez et al., 2011; Ross and Tabrizi, 2011; Udd and Krahe, 2012) **(Fig 1.1)**. The first repeat expansion, an expanded CGG repeat in the

FMR1 gene, was discovered in 1991 by researchers at Emory University (Verkerk et al., 1991). Since, repeat expansions have been implicated in neurological disorders spanning the clinical spectrum including myotonic dystrophy, ALS/FTD, Huntington's disease and the hereditary ataxias(Paulson, 2018) **(Table 1.1)**. The molecular mechanisms underlying disease in microsatellite expansions are varied and can result from either loss-of-function or gain-of-function mechanisms. These expanded repeats are typically unstable and can grow in repeat size across generations which in turn can modify the toxicity or the underlying mechanism associated with a particular repeat expansion. The focus of this thesis is to provide mechanistic insight into two repeat expansion disorders, c9ALS/FTD and SCA36.

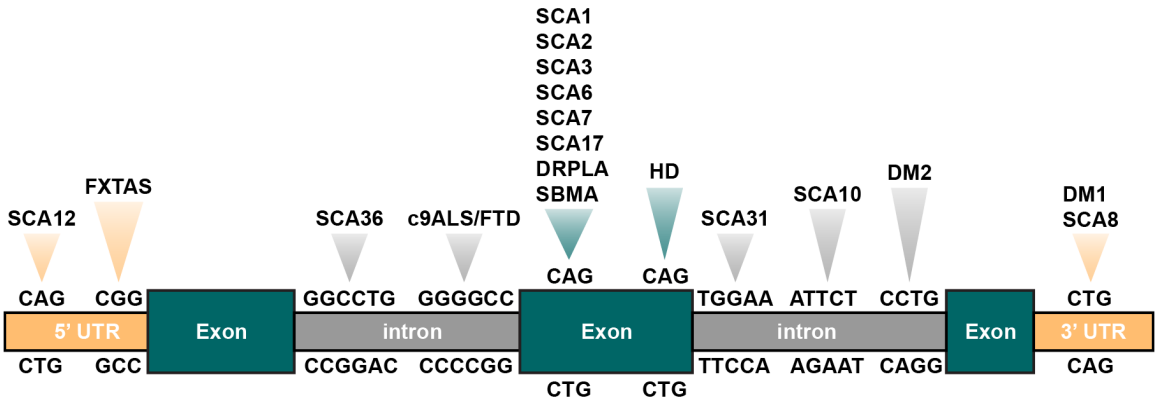


Figure 1.1: Microsatellite expansion mutations. Schematic illustrating the repeat motif and location of microsatellite repeat expansions. Adapted from (Zu et al., 2018).

Table 1.1: Neurological disorders caused by repeat expansions

Disease	Repeat	Gene	Location	Repeat Size†	Reference
Fragile X Syndrome (FXS)	CGG	*FMR1*	5' UTR	> 200	(Verkerk et al., 1991)
Spinal Bulbar Muscular Atrophy (SBMA)	CAG	*AR*	Exon	38 – 62	(Spada et al., 1991)
Spinocerebellar Ataxia Type 1 (SCA1)	CAG	*ATXN1*	Exon	40 – 81	(Orr et al., 1993)
Spinocerebellar Ataxia Type 2 (SCA2)	CAG	*ATXN2*	Exon	32 – 79	(Pulst et al., 1996)
Spinocerebellar Ataxia Type 3 (SCA3)	CAG	*ATXN3*	Exon	52 – 86	(Kawaguchi et al., 1994)
Spinocerebellar Ataxia Type 6 (SCA6)	CAG	*CCNA1A*	Exon	19 – 33	(Zhuchenko et al., 1997)
Spinocerebellar Ataxia Type 7 (SCA7)	CAG	*ATXN7*	Exon	38 – 130	(David et al., 1997)
Spinocerebellar Ataxia Type 8 (SCA8)	CAG	*ATXN8*	3' UTR	71 – 1,300	(Koob et al., 1999)
Spinocerebellar Ataxia Type 10 (SCA10)	ATTCT	*ATXN10*	Intron	400 – 4,500	(Matsuura et al., 2000)
Spinocerebellar Ataxia Type 12 (SCA12)	CAG	*PPP2R2B*	5' UTR	51 – 78	(Fujigasaki et al., 2001)
Spinocerebellar Ataxia Type 17 (SCA17)	CAG	*TBP*	Exon	44 – 67	(Koide et al., 1999)
Fragile X Tremor/Ataxia Syndrome (FXTAS)	CGG	*FMR1*	5' UTR	55 – 200	(Hagerman et al., 2001)
Huntington's Disease (HD)	CAG	*HTT*	Exon	> 37	(MacDonald et al., 1993)
Myotonic Dystrophy Type 1 (DM1)	CUG	*DMPK*	3' UTR	50 – 1,000	(Brook et al., 1992)
Myotonic Dystrophy Type 2 (DM2)	CCTG	*CNBP*	3' UTR	75 – 11,000	(Liquori et al., 2001)
Spinocerebellar Ataxia Type 31 (SCA31)	TGGAA	*BEAN*	Exon	>2.5kb insertion length*	(Sato et al., 2009)
Spinocerebellar Ataxia Type 36 (SCA36)	GGCCTG	*NOP56*	Intron	650 – 2,500	(Kobayashi et al., 2011)
C9orf72-Amyotrophic Lateral Sclerosis / Frontotemporal Dementia (c9ALS/FTD)	GGGGCC	*C9orf72*	Intron	700 – 4,400	(DeJesus-Hernandez et al., 2011; Renton et al., 2011)
Neuronal Intranuclear Inclusion Disease (NIID)	GGC	*NOTCH2NLC*	5' UTR	66 – 517	(Sone et al., 2019; Tian et al.)

†Repeat size refers to number of repeats found in affected patients

*The SCA31 associated expansion is typically a complex repeat. Instead of number of repeats, usually the insertion size is referred to.

1.2 Amyotrophic Lateral Sclerosis

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurological disorder with a fatal prognosis. Characterized by the loss of neurons that are involved in voluntary muscle control, ALS renders those afflicted paralyzed and ultimately unable to breathe (Brown and Al-Chalabi, 2017). While tremendous efforts have been made to combat this devastating disease, biological insights into the disorder have only further added to its mystery and therapies to extend or improve quality of life are limited.

1.2.1 Historical Insights

In the early 19th century, Sir Charles Bell, who is most often associated with the eponymous "Bell's Palsy" disorder, published a seminal text describing the human nervous system (Bell, 1824). A key concept advocated by Bell was the anatomical and functional distinction between sensory and motor nerves. Bell is credited with making the first recorded report of amyotrophic lateral sclerosis (ALS) in which he refers to a patient, 'Mrs. G'., whose case (No. XLVIIl; Case 47) is "*very interesting, and might have been used as illustrative of my* [Bell's] *views of the nerves*" (Bell, 1830). At the time, the patient was in the sixth decade of her life and bed-ridden. She was described as having a progressive weakening of her extremities with "*twitching of the muscles*." Additionally, her "*speech became more and more indistinct"* (dysarthria) and her "*difficulty swallowing gradually increased*" (dysphagia). Notably, however, throughout the course of her disease, she never experienced any loss of sensation and the *"slightest pressure of the legs, toes, fingers, or arms, was immediately perceived*." At autopsy her physician and a former pupil of Bell's, Thomas Ingel, made a keen observation that the anterior part of the spinal cord was in a

"*semifluid state, approaching nearly to the consistence of cream"*, however the posterior part of the spinal cord remained firm. Following Bell's initial description of a progressive neurological disorder affecting only the motor system, a number of prominent neurologists at the time began to present similar case reports including Guillaume-Benjamin Duchenne (Duchenne, 1847), François-Amilcar Aran (Aran, 1850), Jean Cruveilhier (Cruveilhier, 1853), and Jacob Lockhart Clarke (Radcliffe and Clarke, 1862). However, it was not until keen observations made by Jean-Martin Charcot that amyotrophic lateral sclerosis as a distinct clinical disorder was fully realized. Charcot's definition of ALS developed over a number of years and resulted from a number of cases that allowed him to correlate clinical symptoms seen in ALS patients and neuropathological findings. In 1865, Charcot presented a case of a women who had presented with considerable weakness in her extremities in addition to what was termed at the time "contractures" (spasticity). Charcot observed that, at autopsy, there were "sclerotic changes" and that the "lateral columns have in their most superficial and posterior regions, a gray, semitransparent appearance" (Charcot, 1865; Goetz, 2000). A few years later, Charcot had observed that individuals who had pediatric forms of progressive muscular atrophy had considerable degeneration of the anterior horn cells (Charcot, 1869). These two studies allowed Charcot to relate the sclerotic changes in the lateral columns to the clinical finding of "contractures" and degeneration of the anterior horn cells to muscular atrophy. Subsequent encounters with patients presenting with both contractures and muscular atrophy formed the basis for his first diagnosis of ALS. Charcot's naming of this disorder, amyotrophic lateral sclerosis perfectly captured the clinico-anatomical findings: *a* (without) + *myo* (muscle) + *trophy* (support) referring to the degeneration of the anterior horn and *lateral sclerosis* capturing the loss of the lateral columns (i.e. axonal tracts of the upper motor neurons). While

Charcot was able to systematically and wholly relate clinical symptoms with neuropathological findings, it should be noted that both Cruveilhier and Lockhart-Clark had noted in their seminal texts that anterior horn cells of the spinal cord were lost in patients.

1.2.2 The Genetics of ALS

During the 19th century, morbid pathology (i.e. clinico-pathological correlations) provided unprecedented insight into diseases of the nervous system. However, it wasn't until the mid to late 20th century that our understanding of the genetic code, conceptualization of mathematical principles to model population genetics, and the development of tools to analyze genetic sequences allowed for a new scientific revolution to emerge in neurology.

The prevailing thought at the time was that motor neuron disease was an isolated and non-hereditary disorder, with even Charcot claiming that it was never familial (Charcot, 1869). However, the notion that the same disease could afflict other members of a family was noted as early as 1850 when Aran commented that his patient suffering from progressive muscular atrophy had three relatives – a sister and two maternal uncles – who experienced similar symptoms before succumbing to the disease (Aran, 1850; Siddique and Ajroud-Driss, 2011). In the first edition of his text on diseases of the nervous system, the famed US surgeon general William Alexander Hammond comments "*hereditary influence is a well-recognized predisposing cause,*" and offers as an example two cases sent to him that were brothers as well as nine other patients under his care who had relatives afflicted by the same disease (Hammond, 1871). In a later edition he provides additional evidence for a hereditary link to progressive muscular atrophy (Hammond, 1881). He shares an

interesting case that was made known to him by Dr. R.F. Andrews of Gardner, Massachusetts in 1874, concerning a patient first seen at Massachusetts General Hospital in 1872. In addition to comments about the case from Dr. Andrews, there was a pamphlet entitled "Muscular Atrophy" written by the patient himself, E. Wetherbee, describing the remarkable history of what progressive muscular atrophy in his family for several generations. According to E. Wetherbee, this hereditary disorder his family suffered from was known as the "Weatherbee Ail" by their family and friends. Other accounts of familial forms of the disease were also communicated by a number of neurologists. Another remarkable example of familial progressive muscular atrophy is the Bessel family seen by Dr. Bernhard Naunyn in Germany; over six generations of family members had suffered from this disease (Naunyn, 1873). In 1880, William Osler of McGill University in Canada published an article in which he presented a "genealogical chart" of the Farr family and stated it "*illustrates well the hereditary nature of progressive muscular atrophy*" (Osler, 1880). A member of the Farr family had recently been seen by Dr. Osler and diagnosed with progressive muscular atrophy. Intriguingly, the patient indicated that 13 family members across two generations "*have been affected by this disease*." Unbeknownst at the time, the Farr family's role in elucidating the etiology of familial progressive muscular atrophy was much more far-reaching than simply an early account of the hereditary nature of progressive muscular atrophy.

Although these early reports of familial progressive muscular atrophy/ALS provided reasonable evidence of its hereditary nature, it remained widely believed that ALS was not an inherited disorder. However, an article published in 1955 argued that "18 extensive pedigrees compatible with dominant inheritance;" had been described in the literature furthermore, the authors presented six new pedigrees that demonstrated ALS can

be hereditary (Kurland and Mulder, 1955). Subsequent studies and reviews of reported familial cases suggested that 5-10% of ALS cases are familial (Mulder et al., 1986). As our understanding and ability to analyze genomic sequences advanced, investigators began to search for genetic aberrations that may give rise to familial forms of ALS. Using genetic linkage analysis, Siddique and colleagues were able to identify a genetic locus on chromosome 21 that was strongly associated with a number of pedigrees of familial ALS (fALS) (Siddique et al., 1991; Siddique et al., 1989). In 1993, mutations in the Cu/Zn superoxide dismutase 1 gene (SOD1) within the linked region of chromosome 21 were identified as the first genetic cause of fALS (Rosen et al., 1993). This discovery prompted others to search for candidate ALS genes in both familial and sporadic ALS. In the early 2000's genetic based efforts as well as insights from neuropathological findings began to pay off as a number of genetic mutations were implicated in ALS (Kabashi et al., 2008; Kwiatkowski et al., 2009; Neumann et al., 2006; Sreedharan et al., 2008). The discovery that TDP-43 was the main component of ubiquitinated inclusions found in the majority of ALS as well as FTLD-U cases supported a long suspected idea that ALS and FTD are part of a disease continuum (Ferrari et al., 2011; Lillo and Hodges, 2009; Neumann et al., 2006). Concomitantly, genetic linkage studies in a number of large families with individuals suffering from ALS, and FTD identified a region on chromosome 9 as a new ALS-FTD locus (Morita et al., 2006; Valdmanis et al., 2007; Vance et al., 2006). These findings set off a hunt to discover the chromosome 9 linked mutation(Gijselinck et al., 2010; Le Ber et al., 2009; Luty et al., 2008; Pearson et al., 2011). In 2011, these efforts culminated in the discovery by two independent groups that the causal mutation was a massively expanded G_4C_2 repeat in the intron of *C9orf72* (DeJesus-Hernandez et al., 2011; Renton et al., 2011). The discovery of the *C9orf72*-linked mutation was unique because it was the first repeat

expansion implicated in ALS and FTD. Furthermore, the frequency of the *C9orf72*-linked mutation was remarkable as follow up studies estimated this genetic mutation accounted for ~40% of familial ALS cases and ~10% of sporadic ALS cases worldwide (Majounie et al., 2012). Since the initial discovery of the SOD1 mutation, a staggering number of genetic variants have been linked to ALS, however only a subset of these are believed to be truly causal (Brown and Al-Chalabi, 2017; Taylor et al., 2016). Advances in next generation sequencing technologies and concerted worldwide efforts to discover new ALS genes will undoubtedly expand our understanding of ALS. It bears mentioning that an example of these efforts is the multinational collaboration, Project MinE (www.projectmine.com) which aims to whole genome sequence 15,000 ALS patients. While only a fraction of genomes have been sequenced to date, a number of new ALS variants have been identified including variants in *NEK1*, *KIF5A*, and *C21orf2* (Kenna et al., 2016; Nicolas et al., 2018).

1.3 *C9orf72*-linked ALS/FTD

The discovery that a G_4C_2 repeat expansion in the first intron of *C9orf72* is the most common genetic cause of ALS and FTD was a major breakthrough in our understanding of these neurological disorders.

1.3.1 The C9orf72 Gene and Protein

Prior to the discovery of the G_4C_2 repeat expansion, the *C9orf72* gene was only a predicted gene structure in the human genome. Once it was reported to be causal for ALS/FTD, a number of studies began to elucidate its function. Somewhat serendipitously, an early study using bioinformatic methods to discover novel proteins containing the tripartite DENN module, identified the C9orf72 protein as a candidate DENN type GDP-

GTP exchange factor (GEF) (Levine et al., 2013; Zhang et al., 2012). DENN (differentially expressed in normal and neoplastic cells) type proteins are evolutionary conserved proteins that interact with Rab GTPases and are thought to function as Rab GEFs (Marat et al., 2011). This initial report identifying the C9orf72 protein as a Rab-GEF suggested that it plays a role in Rab-dependent vesicular trafficking. Follow up biochemical studies demonstrated that the C9orf72 protein co-localizes with Rab 1, Rab 5, Rab 7, and Rab 11 and provided evidence that the C9orf72 was involved in the regulation of endocytosis and autophagy (Farg et al., 2014). C9orf72 protein was shown to be necessary for the formation of the autophagosome and reduction in C9orf72 resulted in impaired autophagy leading to the accumulation of cytoplasmic p62 aggregates (Sellier et al., 2016; Webster et al., 2016). Recent studies have suggested that the C9orf72 protein forms a tripartite complex with WDR41 and SMRC8 (Amick et al., 2016; Sellier et al., 2016; Sullivan et al., 2016; Yang et al., 2016); interestingly, the SMRC8 protein was also predicted to be a DENN type protein (Zhang et al., 2012) and has been implicated previously in autophagy. Collectively these studies have demonstrated that the C9orf72 protein is likely involved in the autolysosomal pathway.

1.3.2 Disease Mechanism

A number of mechanisms have been proposed for ways in which an expanded microsatellite repeat can elicit toxicity. Similar to other repeat expansions, a number of putative disease mechanisms have been proposed for c9ALS/FTD including haploinsufficiency of the *C9orf72*, RNA mediated toxicity resulting from G_4C_2 repeat containing RNA, and translation of the G_4C_2 repeat to produce aberrant dipeptide repeats **(Fig 1.2)**.

Intranuclear inclusions of repeat containing RNA ("RNA foci") is a hallmark of c9ALS/FTD. A leading hypothesis is that these foci sequester a number of RNA binding proteins (RBPs), thereby reducing the functional pool of these RBPs. Since their discovery in c9ALS/FTD tissue, a number of studies have sought to determine which RNA binding proteins are sequestered by these RNA foci. A number of RBPs have been identified including Pur-alpha , ADARB2 (Donnelly et al., 2013b), ALYREF (Cooper-Knock et al., 2014), various HNRNPs (Conlon et al., 2016; Cooper-Knock et al., 2014; Mori et al., 2013a), and nucleolin (Haeusler et al., 2014a). Furthermore, bidirectional transcription of the *C9orf72* locus results in the formation of antisense transcripts that form RNA foci as well and sequester RNA binding proteins (Cooper-Knock et al., 2015). The downstream molecular consequences of RBP sequestration have only begun to be examined, but recent studies have demonstrated that wide spread changes in alternative splicing events result from RBP sequestration (Conlon et al., 2016; Prudencio et al., 2015).

Shortly after the discovery of the *C9orf72* mutation, it was found that *C9orf72* repeat containing RNA could undergo a non-canonical form of translation, termed repeat-associated non-AUG (RAN) translation resulting in the production of repeating dipeptide species (Mori et al., 2013b). Intriguingly, RAN translation can occur in both the sense and antisense direction to produce five distinct DPR species. Since their initial report, a number of studies have investigated the toxicities of individual DPRs. The arginine-rich DPRs, poly(GR) and poly(PR) have consistently been demonstrated to be toxic *in vitro* and *in vivo* and a number of different mechanisms by which they elicit toxicity have been proposed including impairment of protein translation and ribosomal biogenesis (Choi et al., 2019; Kwon et al., 2014; Lopez-Gonzalez et al., 2016; Yang et al., 2015; Zhang et al., 2018).

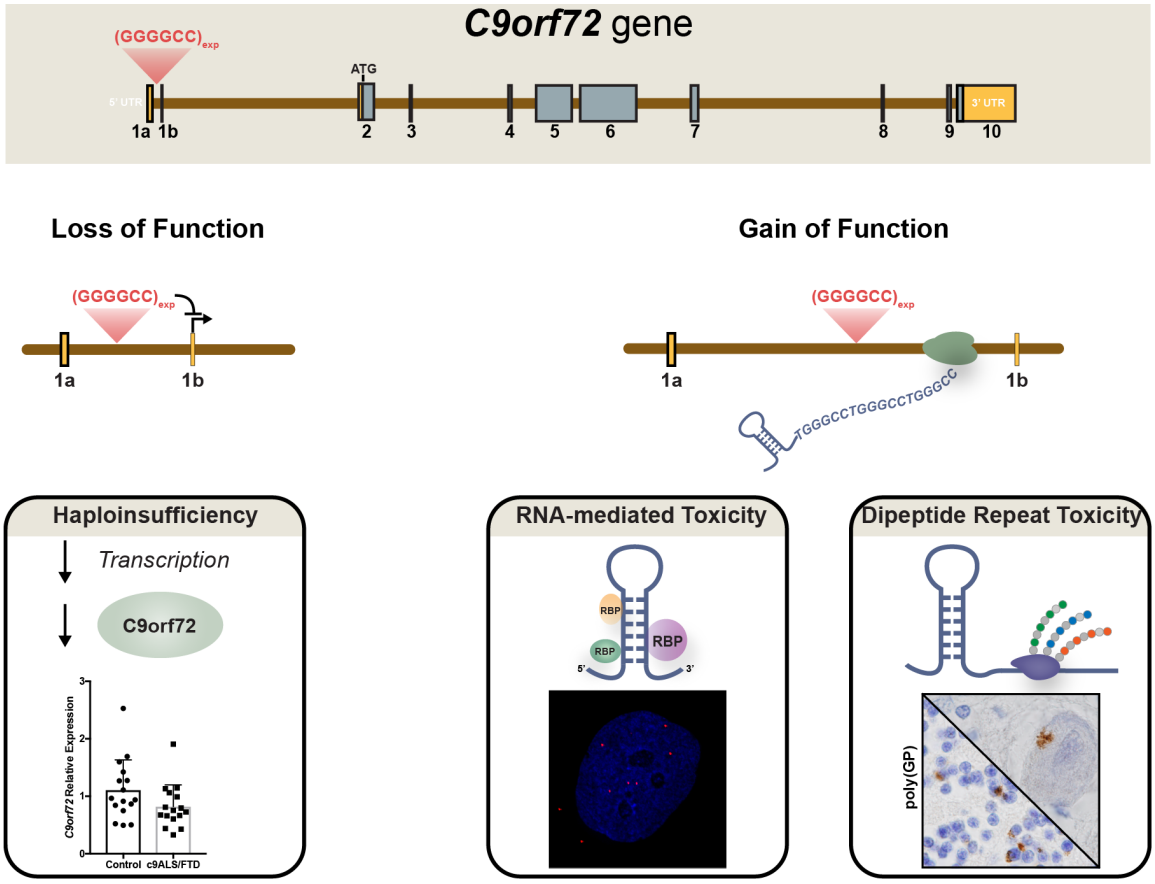


Figure 1.2: Putative Disease Mechanisms in c9ALS/FTD.
Putative disease mechanisms in c9ALS/FTD include 1) haploinsuffiency of the *NOP56* gene, 2) RNA-mediated toxicity via sequestration of RNA binding proteins, and 3) production of aberrant dipeptide repeats. Adapted from (Gitler and Tsuiji, 2016).

1.3.3 Neuropathology of c9ALS/FTD

A number of neuropathological findings are shared between c9ALS/FTD and other forms of ALS and FTD. TDP-43 pathology, i.e. the nuclear clearing, cleavage and phosphorylation of TDP-43, found in most cases of ALS and common subtypes of FTD, is a neuropathological hallmark of c9ALS/FTD as well. Early studies found that a unique neuropathological finding specific to c9ALS/FTD was the presence of neuronal cytoplasmic inclusions (NCI) and neuronal intranuclear inclusions (NII) that are negative for TDP-43, but immunoreactive for ubiquitin and p62, both markers of the ubiquitin

proteasome system (Al-Sarraj et al., 2011; Mackenzie and Neumann, 2016). Subsequent investigations demonstrated that these TDP-43 negative, p62/ubiquitin positive inclusions were in fact aberrant dipeptide repeats (DPR) produced by unconventional translation of the G_4C_2 repeat (Ash et al., 2013; Gendron et al., 2013; Mori et al., 2013b; Zu et al., 2013). Interestingly, the G_4C_2 repeat in *C9orf72* can be translated in all six reading frames, both sense and antisense, to produce five different dipeptide repeat species: Glycine-Alanine (GA), Glycine-Proline (GP), Glycine-Arginine (GR) in the sense direction and Glycine-Proline (GP), Proline-Alanine (PA), Proline-Arginine (PR) in the antisense direction. The sense DPRs are more prevalent in post-mortem samples than the antisense DPRs, with poly(GA) being the most abundant followed by poly(GP) and then poly(GR). Regionally, DPRs are most abundant in neurons of the frontal cortex, hippocampus, and granule layer of the cerebellum; interestingly, they are rarely found in the spinal cord (Gomez-Deza et al., 2015; Vatsavayai et al., 2019). In addition to cells containing single DPR inclusions, there is evidence that the sense DPRs, poly(GP) and poly(GR) co-localize with poly(GA) inclusions. In fact, poly(GP) inclusions were almost exclusively co-localized with poly(GA) in the temporal lobe of three cases (Lee et al., 2017); furthermore, a recent report suggests that poly(GA) can sequester poly(GR) into cytoplasmic inclusions (Yang et al., 2015). A number of studies have reported other proteinaceous inclusions that colocalize with poly(GA) including Unc119 (May et al., 2014), Drosha (Porta et al., 2015), and HR23B (Zhang et al., 2016) while poly(GR) has been shown to co-localize with a number of ribosomal proteins including S6, L21, and eIF3η in the frontal cortex of three patients (Zhang et al., 2018).

1.3.4 Mouse models of c9ALS/FTD

. The G_4C_2 repeat expansion can potentially elicit toxicity through either a loss-of-function or gain-of function mechanism, therefore a number of mouse models have been generated to study c9ALS/FTD that investigate either *C9orf72* haploinsufficiency, RNA mediated toxicity or dipeptide repeat toxicity **(Table 1.2)**. In 2015, the first mouse model of c9ALS/FTD was generated by somatic brains transgenesis (SBT) of an adeno-associated virus (AAV) encoding 66 G_4C_2 repeats (Chew et al., 2015). The AAV-66R mouse model recapitulated many cellular phenotypes characteristic of human c9ALS/FTD including the formation of RNA foci, the production of DPRs, and the cytoplasmic mislocalization and phosphorylation of TDP-43; however, the behavioural and motor phenotypes were mild. The authors noted that at 6 months of age, there was observed upper but not lower motor neuron loss as well increased anxiety and sociability deficits. Shortly thereafter, a number of groups published C9orf72 BAC transgenic models (Jiang et al., 2016; Liu et al., 2016; O'Rourke et al., 2015; Peters et al., 2015). A BAC transgenic model is hypothetically a more disease relevant model since the BAC would contain a large number of G_4C_2 repeats as well as *cis*-regulatory elements associated with endogenous human *C9orf72*. Surprisingly, although all of the BAC models exhibited characteristic RNA foci and dipeptide repeats, behavioural and motor phenotypes were completely absent or mild except for the BAC-C9-500 model developed by Liu et al. This model displayed robust, though inconsistent, phenotypes including upper and lower motor neuron loss, neuronal cell loss throughout the CNS, hindlimb paralysis, seizures, and even early death. Interestingly the BAC-C9-500 model also exhibited robust cytoplasmic TDP-43 inclusions. The phenotypic discrepancy between this model and the other c9BAC models has, understandably, caused consternation in the field. It should be noted that the BAC-C9-500 model developed by Liu et al is on a different mouse background compared to the

other models. The c9BAC models in theory allow for analysis of both RNA gain-of-function toxicity and DPR toxicity; however, it is hard to tease to out which phenotypes may or may not be attributed to RNA or DPR toxicity. Therefore, a number of AAV based SBT models have been developed to study the toxicity associated with individual DPRs in the absence of toxic RNA. Similar to the approach employed in the AAV-66R model, Zhang et al. used an AAV encoding for poly(GA), poly(GR), or poly(PR) dipeptide repeats to assess whether each of these DPRs individually contribute to neurodegeneration (Zhang et al., 2018; Zhang et al., 2016; Zhang et al., 2019). An interesting result from these studies suggests that the individual DPRs (i.e. -GA, -GR, or -PR) studied were not sufficient to promote TDP-43 pathology.

Table 1.2: Models of C9orf72-associated ALS/FTD

Repeat Expansion Models[†]:						
Model	**# of G_4C_2 Repeats**	**Phenotypes[#]**	**Pathology**		**Strain**	**Reference**
			DPRs	**pTDP**		
AAV-66R	66	Upper motor neuron loss, Increased anxiety, reduced motor function	GA, GP, GR	Yes	C57BL/6J	(Chew et al., 2015)
AAV-149	149	Motor deficits, astrogliosis, cognitive impairment	GA, GP, GR, PR, PA	Yes	C57BL/6J	(Chew et al., 2019)
BAC-C9-500/300	300/500	None	GP	No	SJL/6J	(Peters et al., 2015)
BAC-C9-(100-100)	100-1000	None	GP	No	C57BL/C3H	(O'Rourke et al., 2015)
BAC-C9-450	450	Increased anxiety, spatial learning deficits	GA, GP, GR	No[*]	C57BL/6J	(Jiang et al., 2016)
BAC-C9-500	500	Motor neuron loss, weight loss, motor function deficits, seizures, clasping, increased anxiety, hindlimb paralysis, early death	GA, GP	Yes	FVB/NNJ	(Liu et al., 2016)
DPR Models:						
Model	**# of DPR Repeats**	**Phenotypes[#]**	**Pathology**		**Strain**	**Reference**
			DPRs	**pTDP**		

AAV-GA	50	Slight motor deficits as assessed by tail-suspension and rotarod, astrogliosis, cortical neuron loss	GA	No	C57BL/6J	(Zhang et al., 2016)
AAV-GR	100	Progressive motor deficits, astro- & microgliosis in the cortex and hippocampus, cortical neuron loss.	GR	No	C57BL/6J	(Zhang et al., 2018)
AAV-PR	50	Motor Deficits, neuron loss, neuroinflammation, early death.	PR	No	C57BL/6J	(Zhang et al., 2019)
***C9orf72* KO Models:**						
Model	**KO Strategy**	**Phenotypes‡**	**Pathology**		**Strain**	**Reference**
			DPRs	**pTDP**		
	Exon 2-6 KO	Motor deficits (Rotarod), Social interaction deficits, reduced body weight, splenomegaly, lymphadenopathy	N/A	No	C57BL/6J	(Jiang et al., 2016)
	Exon 2-6 KO	Splenomegaly, lymphadenopathy	N/A	No	C57BL/6J	(O'Rourke et al., 2016)
	Exon 2-6 KO CRISPR-KO	Reduced body weight, splenomegaly, lymphadenopathy	N/A	No	C57BL/6J	(Burberry et al., 2016)
	CRISPR-KO	Reduced body weight, splenomegaly, lymphadenopathy, increased in autophagy/lysosome markers	N/A	No	C57BL/6J	(Sullivan et al., 2016)
	Full KO	Reduced body weight, splenomegaly, lymphadenopathy	N/A	No	C57BL/6J	(Sudria-Lopez et al., 2016)

†See (Batra and Lee, 2017) for detailed review of HRE models
*Although no pTDP43 inclusions were detected, the authors report increased pTDP in the insoluble fraction as determined by immunoassay
#Additional phenotype information can be obtained from https://www.alzforum.org/research-models
‡See (Balendra and Isaacs, 2018) for detailed review of KO models

1.4 Spinocerebellar ataxia (SCA)

Derived from Greek, the word *Ataxia* means "without order"(Barboi, 2000). Ataxia is a clinical term that refers to the loss of coordination of voluntary muscle control. Although similar symptoms are experienced by patients suffering from ataxia, the etiology can differ considerably. Ataxia can result from genetic or non-genetic origins (Klockgether et al., 2019). Hereditary forms of ataxia can have either an autosomal dominant or recessive pattern of inheritance; autosomal dominant ataxias are referred to as the

Spinocerebellar Ataxias (SCA). To date there have been over 49 types of spinocerebellar ataxias reported (Pilotto and Saxena, 2018). The spinocerebellar ataxias are a group of progressive degenerative diseases affecting the cerebellum. Cerebellar ataxia (either truncal or appendicular) is a unifying clinical phenotype across all SCAs; however, there exists specific clinical symptoms characteristic of a particular SCA type (Shakkottai and Fogel, 2013). Genetic mutations associated with the SCAs can be broadly grouped into four groups: the polyglutamine ataxias (e.g. SCA1, SCA2, SCA3), mutations in ion channels (e.g. SCA5, SCA6), mutations in signal transduction pathways (e.g. SCA11, SCA12), and mutations associated with RNA toxicity (e.g. SCA31, SCA36) (Shakkottai and Fogel, 2013); interestingly mutations associated with SCAs in each of these groups can result from expansion of microsatellites **(Table 1.1)**. While, in general the prevalence of ataxia is 3 per 100,000 there appears to be a regional and ethnic component to the prevalence of a particular SCA type (Pilotto and Saxena, 2018). The discovery of expanded CAG repeats as the genetic cause of SCA1 prompted considerable efforts focused on elucidating molecular mechanisms result from CAG expansions. However, research efforts focused on other SCA types have been limited, especially for spinocerebellar ataxia type 36.

1.5 Spinocerebellar ataxia type 36 (SCA36)

Spinocerebellar ataxia type 36 (SCA36) is a late-onset, progressive neurodegenerative disease characterized chiefly by truncal ataxia (Arias et al., 2017). First described clinically by two independent groups, SCA36 was initially referred to as *Costa da Morte* ataxia in the Galician region of Spain and *Asidan* in the Chugoku region of Japan (Arias et al., 2008; Ohta et al., 2007). Subsequently, it was discovered that both of these

ataxic syndromes resulted from a massively expanded GGCCTG repeat in the first intron of the gene Nucleolar Protein 56 (*NOP56*) (García-Murias et al., 2012; Kobayashi et al., 2011). SCA36 has since been reported in individuals from Italy, France, China, Taiwan, Poland, and the U.S.A. (Lee et al., 2016b; Obayashi et al., 2015; Sarto et al., 2013; Sulek et al., 2013; Valera et al., 2017; Zeng et al., 2016). In addition to truncal ataxia, patients present other neurological symptoms to varying extents, including appendicular ataxia, sensorineural impairment, and tongue fasciculations and atrophy (Ikeda et al., 2012; Ikeda et al., 2013).

1.5.1 The *NOP56* Gene and Protein

NOP56 encodes for nucleolar protein 56 (NOP56). NOP56 is a evolutionarily conserved 594 amino acid protein involved in ribosomal biogenesis. More specifically, it is a core component of the box C/D small nucleolar ribonucleoprotein (snoRNP) particles. NOP56, in conjunction with NOP58, bind the C'/D' and C/D boxes, respectively of the box C/D snoRNP particle, and serves as a scaffold for the 2'-O-methyltransferase, Fibrillarin (FBL) (Bizarro et al., 2014). NOP56 is important for the final maturation of activation of box C/D snoRNPs (Lykke-Andersen et al., 2018).

1.5.2 Disease Mechanisms

While numerous studies have focused on c9ALS/FTD, efforts targets at elucidating disease mechanisms in SCA36 have been limited. Similar to what has been proposed for c9ALS/FTD, putative disease mechanisms in SCA36 include haploinsufficiency of the *NOP56* gene, RNA mediated toxicity resulting from TG_3C_2 repeat containing RNA, and

translation of the TG_3C_2 repeat to produce aberrant DPRs **(Fig 1.3)** (Arias et al., 2017). In their initial report describing the SCA36 repeat expansion, Kobayashi et al. demonstrated that *NOP56* mRNA levels are not affected by the repeat expansion but that miR1292 which is located in intron of *NOP56* is significantly reduced (Kobayashi et al., 2011). Furthermore, using a FISH probe targeting the TG_3C_2 containing RNA they presented evidence that characteristic RNA foci form in a patient derived lymphoblastoid cell line (LCL). Gel shift assays and immunofluorescence analysis suggests that the RNA binding protein SRSF2 interacts with TG_3C_2 RNA and is sequestered into RNA foci. In a follow up study, Liu et al. demonstrated that these foci are present throughout the CNS of SCA36 patients (Liu et al., 2014). Interestingly, they noted that the size of the foci actually differed across regions and they could group the foci by size (small, medium, or large); large foci (i.e. >10 μm) were predominately in disease relevant cell types including Purkinje neurons, spinal motor neurons, and neurons of the olivary nucleus. A more recent study generated patient derived stem cell models of SCA36 and demonstrated that RNA foci are present in both patient iPS and iPS derived neurons (Matsuzono et al., 2017). Antisense oligonucleotides (ASO) developed to target the repeat expansion reduced RNA foci burden in both iPS and iPS derived neurons, thus providing proof-of-principle that ASOs could be a therapeutic strategy to mitigate RNA toxicity in SCA36.

1.5.3 Neuropathology of SCA36

Neuropathological findings are limited in SCA36. To date, only two publications have reported post-mortem analysis of SCA36 patients (Liu et al., 2014; Obayashi et al., 2015). In their initial report of SCA36 post-mortem analysis, Liu et al., demonstrated that RNA foci were present in a number of CNS regions including Purkinje neurons, spinal

motor neurons, and neurons of the olivary nucleus (discussed in detail above). Immunohistochemistry for TDP-43 was negative throughout the CNS; however, the authors report the presence of large p62 and ubiquitin aggregates in the inferior olive. In a second study by Obayashi et al., neuropathological examination of an asymptomatic patient harboring the SCA36 associated repeat expansion revealed diffuse cerebellar atrophy and mild Purkinje cell degeneration characterized by distorted dendrites and Purkinje cell "torpedos" (Obayashi et al., 2015). Although immunohistochemical staining for ubiquitin, TDP43, FUS and p62 were unremarkable, the authors note that there were numerous amyloid plaques consistent with Braak Stage IV; it should be noted the patient was 90 years old at the time of autopsy.

Presented in more detail in Chapter 2, work done in this thesis greatly expands our knowledge of the neuropathology associated with SCA36. During this project, six SCA36 post-mortem cases were obtained from collaborators around the world (see **Table 2.1** in Chapter 2 for SCA36 cases used in post-mortem analysis). To our knowledge these are the only SCA36 post-mortem cases known to date.

1.5.4 Mouse Models of SCA36

To date a mouse model of SCA36 has not been published; however, as a result of collaborations initiated during this thesis a mouse model is currently being developed by our colleagues.

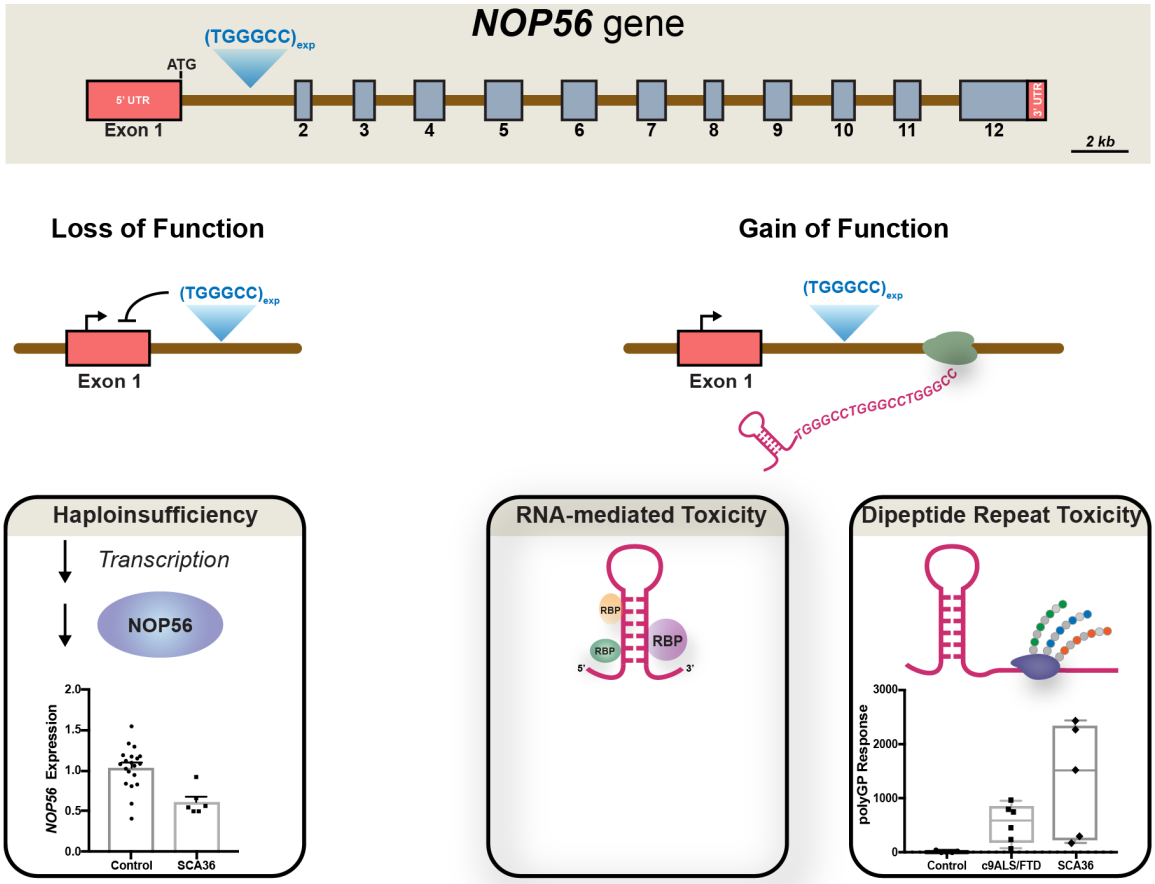


Figure 1.3: Putative Disease Mechanisms in SCA36.
Putative disease mechanisms in SCA36 include 1) haploinsuffiency of the *NOP56* gene, 2) RNA-mediated toxicity via sequestration of RNA binding proteins, and 3) production of aberrant dipeptide repeats.

1.6 Discussion, Thoughts, and Thesis Objectives

The discovery of the c9ALS/FTD repeat expansion is undeniably a breakthrough in our understanding of ALS and FTD. The prevalence of c9ALS/FTD in both familial and sporadic forms of ALS and FTD has placed significant importance on understanding how this mutation results in disease as well as on the development of therapeutic strategies to combat it.

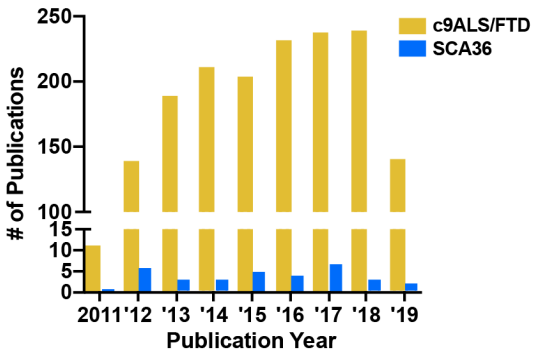


Figure 1.4: Number of publications related to c9ALS/FTD and SCA36 since their discovery in 2011.

Moreover, the type of mutation implicated in c9ALS/FTD, i.e. a repeat expansion, has no doubt promoted heightened interest and research in the field **(Fig 1.4)**. Investigators who previously studied other repeat expansion disorders or researchers with a focus on the development of tools and assays to study repetitive sequence elements have been invaluable to our understanding of c9ALS/FTD pathogenesis. Despite numerous efforts, however, our understanding of the fundamental biology of how the G_4C_2 repeat expansion results in ALS and/or FTD remains elusive. In my opinion, increased efforts indeed provide insight into disease pathogenesis, but perhaps paradoxically, also make it more difficult to separate wheat from chaff. I believe our ability to converge upon a disease mechanism in c9ALS/FTD is confounded by

1) The number of putative disease mechanisms plausible in a repeat expansion disorder. The heterogeneity of a clinical condition caused by a repeat expansion lends itself to the idea that multiple disease mechanisms may be in play. Particularly in the case of c9ALS/FTD, where two clinically disparate disorders can occur, it is highly plausible that while one pathomechanism contributes to the development of ALS, a completely separate mechanism gives rise to the FTD phenotypes.

2) Inconsistencies of c9ALS/FTD animal models. For example, in some animal models, RNA foci are toxic, in others the dipeptide repeats are toxic; Furthermore, multiple c9BAC mouse models lack any robust behavioral or motor phenotypes yet exhibit "toxic" cellular phenotypes such as RNA foci and dipeptide repeats. This is particularly confusing since the c9BAC mouse models are arguably more physiologically similar to the human mutation, and thus should, in theory, more faithfully represent the human condition.
3) Lack of proper controls to identify disease relevant changes in the underlying biology. Leading hypotheses of c9ALS/FTD pathogenesis are based on gain of function toxic mechanisms. This begs the question, what are the proper controls to tease out disease relevant and specific mechanisms?

This last point further raises an interesting question...**what is it about the G_4C_2 repeat expansion specifically that results in cellular death and subsequently ALS and/or FTD?** Would the G_4C_2 repeat expansion result in c9ALS/FTD if it were placed in a different context? More specifically, would the G_4C_2 repeat expansion cause disease if it was located in another intron of *C9orf72?* Or if it was located in the first intron of a different gene? While easy to formulate, these questions are considerably more difficult to answer. The technical challenges associated with working with massively expanded sequences, let alone a sequence that is 100% GC rich limits our ability to address these questions in a disease relevant context. These questions and challenges prompted me to ask whether there are hexanucleotide repeat expansions similar to the G_4C_2 repeat expansion in c9ALS/FTD that could be used to tease out disease specific mechanisms. Surprisingly, the highly related TG_3C_2 repeat expansions had recently been implicated in SCA36. Differing by one base pair in the repeating unit, the SCA36 and c9ALS/FTD

repeat expansion are strikingly similar; they are both GC rich, intronic hexanucleotide repeats and they are both massively expanded in patients (hundreds to thousands of repeats on average). While considerably understudied, investigations into SCA36 will undoubtedly aid in our understanding of c9ALS/FTD.

Therefore, the overarching objective of this thesis was to provide a comprehensive comparison of these genetically similar, yet clinically disparate disorders and provide mechanistic insight into the disease pathogenesis of c9ALS/FTD and SCA36. More specifically, I sought to provide insight into the role and toxicity of dipeptide repeats in these repeat expansion disorders. Furthermore, I aimed to compare RNA-protein interactions between the c9ALS/FTD and SCA36 associated repeats in order to tease out disease specific interactions. Lastly, given the strikingly different clinical phenotype between c9ALS/FTD and SCA36, I sought to elucidate gene expression signatures specific to each disorder in disease relevant regions.

CHAPTER 2. DIVERGENT DIPEPTIDE REPEAT PATHOLOGY IN C9ALS/FTD AND SCA36

2.1 Divergent dipeptide repeat pathology in c9ALS/FTD and SCA36

Several lines of evidence suggest that the translation of toxic homopolymeric peptides may contribute to disease phenotypes in a number of microsatellite repeat expansion disorders (Banez-Coronel et al., 2015; Mori et al., 2013b; Todd et al., 2013; Zu et al., 2017; Zu et al., 2011). This includes the breakthrough discovery that GGGGCC hexanucleotide repeat expansions (HRE) in *C9orf72*, which cause amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), lead to the production of aggregating dipeptide repeat (DPR) proteins via repeat associated non-AUG (RAN) translation (Kwon et al., 2014; Mizielinska et al., 2014; Zhang et al., 2018; Zhang et al., 2016). Here we show that the related intronic GGCCTG HRE (alternatively denoted as TGGGCC) that causes spinocerebellar ataxia type 36 (SCA36) is also translated into DPRs, including poly(GP) and poly(PR), which are present in *C9orf72*-associated ALS/FTD (c9ALS/FTD). We found that although poly(GP) is more abundant in SCA36 patient tissue compared to tissues from c9ALS/FTD patients, it exists as a soluble protein species and does not form insoluble aggregates as in c9ALS/FTD. However, the frequency of the antisense RAN translation product poly(PR) is comparable between SCA36 and c9ALS/FTD patient postmortem samples. TDP-43 pathology is yet another hallmark feature of c9ALS/FTD, and the mechanisms by which the GGGGCC HREs cause TDP-43 abnormalities remain an important question to be answered. SCA36 patients do not exhibit TDP-43 pathology, suggesting that DPRs alone are insufficient to induce this disease-relevant correlate of c9ALS/FTD. These findings reveal key differences in translation, solubility and protein

aggregation of DPRs between SCA36 and c9ALS/FTD and implicates alternate mechanisms for how highly similar repeat expansions can cause markedly distinct forms of neurodegeneration. Ultimately, these results expand our understanding of DPR pathology and their role in neurodegeneration.

2.1.1 The SCA36 associated TG_3C_2 expansion is translated into dipeptide repeats

Translation across the TG_3C_2 repeat results in six putative DPR protein species: poly(GP), poly(WA), and poly(GL) in the sense direction and poly(GP), poly(PR), and poly(AQ) in the antisense direction **(Fig 2.1A)**. Two of these putative DPRs, poly(GP) and poly(PR), are also translated via a RAN mechanism in c9ALS/FTD(Gendron et al., 2013; Mori et al., 2013b). To determine whether the TG_3C_2 repeat could undergo RAN translation *in vitro*, we generated *NOP56* minigenes containing increasing number of TG_3C_2 repeats **(Fig 2.7A, B)**. Using a quantitative poly(GP) immunoassay (Gendron et al., 2017; Gendron et al., 2015) we found that translation across the TG_3C_2 repeat is length-dependent and must meet a certain repeat-length threshold before it can occur **(Fig 2.7C)**. We could detect the presence of all sense reading frame DPRs through immunoblotting or immunocytochemistry against frame-specific c-terminal tags **(Fig 2.7D, E)**. These findings demonstrate that the intronic TG_3C_2 repeat can undergo RAN translation to produce DPRs in SCA36. To test this hypothesis in a disease-relevant context, we developed multiple patient-derived cellular models of SCA36. Quantitative immunoassays revealed that poly(GP) is detectable in SCA36 patient cells. Strikingly, compared to c9ALS/FTD patient fibroblasts, poly(GP) was significantly increased in SCA36 patient fibroblasts **(Fig 2.1B)**. A similar relative increase in poly(GP) abundance was also seen in SCA36 induced pluripotent stem (iPS) cells, lymphoblastoid cell lines (LCLs), and 3D forebrain organoids (day 105). Of interest, poly(GP) abundance in iPS-derived motor

neurons (iPS-MNs) was comparable between c9ALS/FTD and SCA36 samples **(Fig 2.1C-F)**.

2.1.2 Divergent DPR pathology in SCA36 and c9ALS/FTD

Given the significant increase in detectable poly(GP) DPRs in cellular models of SCA36, we assessed whether poly(GP) is present in SCA36 postmortem tissue. Poly(GP) was present in all evaluated regions of the central nervous system (CNS) **(Fig. 2.2A)**, and poly(GP) abundance was comparable in the majority of neuroanatomical regions tested. However, there was a striking increase of poly(GP) in the dentate nucleus of the cerebellum, a region primarily affected in SCA36(Arias et al., 2017; Liu et al., 2014). We next compared the poly(GP) burden in disease relevant CNS regions from c9ALS/FTD and SCA36 cases **(Fig 2.2B-E)**. Although poly(GP) abundance in the cerebellar cortex of SCA36 cases were comparable to c9ALS/FTD cases, there was a marked increase of poly(GP) in the frontal cortex of SCA36 patients, a result supported by the increased poly(GP) detected in iPS-derived 3D forebrain organoid models **(Fig 2.1E)**. Additionally, poly(GP) abundance in the spinal cord and the cerebellar dentate nucleus of SCA36 patients was greater than in c9ALS/FTD. These findings prompted us to assess the relative mRNA expression of *NOP56* and *C9orf72* in neural tissues obtained from healthy control samples. Relative to *C9orf72* mRNA levels, *NOP56* mRNA is expressed at similar levels in the cerebellar cortex, a neuroanatomical region with comparable poly(GP) abundance in patients with SCA36 or c9ALS/FTD. In contrast, *NOP56* expression is significantly higher than *C9orf72* expression in the frontal cortex, cerebellar dentate nucleus, and spinal cord, three regions in which poly(GP) is likewise higher in SCA36 compared to c9ALS/FTD.

A neuropathological hallmark of c9ALS/FTD is the aggregation of both sense and antisense DPRs in patient postmortem tissue (Ash et al., 2013; Gendron et al., 2013; Mori et al., 2013b; Zu et al., 2013). Since poly(GP) was detectable in lysates from postmortem tissues from SCA36 patients we investigated whether poly(GP) pathology was also present. Surprisingly, immunohistochemical analysis of poly(GP) in disease relevant regions of SCA36 did not reveal the characteristic poly(GP) aggregates seen in c9ALS/FTD **(Fig 2.2F, Fig 2.8)**. The lack of poly(GP) aggregation was supported by the absence of pathology for ubiquitin-binding protein p62/SQSTM1, a common component of protein aggregates (Bjorkoy et al., 2006). Although detectable by immunoassay, poly(GP) did not form insoluble inclusions in other CNS regions including the ventral horn of the spinal cord, hypoglossal nucleus of the medulla, inferior olive of the medulla, temporal cortex, occipital cortex, or parietal cortex **(Fig 2.9, 2.10)**. To further investigate the lack of poly(GP) aggregation in SCA36 we sequentially fractionated cerebellar and frontal cortex tissue samples using reagents of increasing solubilizing power. Dotblot analysis for poly(GP) in the urea soluble fraction further supports the finding that poly(GP) is insoluble in c9ALS/FTD but exists in a soluble state in SCA36 **(Fig. 2.2G)**. Quantitative analysis of poly(GP) in each fraction revealed that it was predominately present in detergent soluble fractions in SCA36, and predominately present in the detergent insoluble, urea soluble fraction in c9ALS/FTD **(Fig 2.2H, Fig 2.11)**. Highly multiplexed, spatially resolved protein analysis (Digital Spatial Profiling; DSP) of the cerebellar granule, Purkinje, and molecular layers reveal layer specific divergent pathologies between c9ALS/FTD and SCA36 **(Fig 2.2I, J ; Fig 2.12)**. The presence of poly(GA) and p62 only in c9ALS/FTD cases confirms divergent DPR pathology. Poly(GP) is also present in SCA36 as detected by immunoassay, however we failed to detect with DSP; this is consistent however, with

the lack of apparent poly(GP) via immunohistochemistry. Interestingly, the RNA modification, N^6-methyladenosine, as detected by an antibody to m6A, was significantly increased in the granule specifically of SCA36 but not in c9ALS/FTD.

The chemical properties and lack of immunogenicity of the other putative sense DPRs, poly(GL) and poly(WA), precluded us from generating antibodies targeting these DPRs. However, using previously established polyclonal antibodies to poly(PR), an antisense DPR in c9ALS/FTD, we determined that poly(PR) formed aggregates in the granule cell layer of the cerebellar cortex in SCA36 **(Fig 2.3A)** and similar to c9ALS/FTD, poly(PR) aggregates were relatively rare. In contrast to c9ALS/FTD, immunohistochemistry and immunoblots for phosphorylated TDP-43 in frontal cortex sections or urea fractions, respectively, demonstrates that SCA36 patients do not exhibit TDP-43 pathology characteristic of c9ALS/FTD **(Fig 2.3B, C)**.

2.1.3 Expanded TG3C2 repeats in the first intron of NOP56 are retained in SCA36

Given the intronic nature of the SCA36 repeat expansion, we hypothesized that the massively expanded TG_3C_2 repeat impairs splicing of *NOP56* and results in retention of the repeat containing intron 1. To determine whether expanded TG_3C_2 repeats could impair splicing we adapted a split luciferase reporter that was previously developed to screen small-molecule modulators of splicing (Younis et al., 2010). The full intron 1 sequence of *NOP56* was modified to contain 2, 37, or 82 TG_3C_2 repeats and was inserted into the firefly luciferase ORF at an optimized splice site; defective splicing of the split luciferase reporter results in reduced luciferase activity **(Fig 2.4A)**. Relative to a *NOP56* intron harboring 2 TG_3C_2 repeats we saw a ~15% and ~25% reduction in luciferase activity in *NOP56* introns harboring 37 and 82 TG_3C_2 repeats, respectively **(Fig 2.4B)**. This repeat-length dependent

reduction in luciferase activity suggests that expanded TG_3C_2 repeats in the context of the *NOP56* intron 1 can impair splicing. We next investigated whether intron 1 of *NOP56* is retained in mRNA from SCA36 patient samples. We designed custom Taqman qPCR probes spanning the exon 1 – exon 2 (ex1-2) splice junction **(Fig 2.13A)**. We reasoned that retention of intron 1 would prevent hybridization of the probe spanning the junction, thus resulting in a reduced qPCR signal. We found a significant reduction in ex1-2 *NOP56* transcript levels (~40%) in the cerebellum of SCA36 patients **(Fig 2.4C)**. This reduction was consistent in patient-derived cellular models of SCA36 including fibroblast, iPS, and iPS-motor neurons **(Fig 2.13B)**. Interestingly, when we assessed *NOP56* mRNA levels using probes spanning the downstream ex3-4 and ex6-7 splice junctions we found a marked increase in *NOP56* transcripts in SCA36 compared to controls **(Fig. 2.4C, 2.13B)**. Although unexpected, this finding is consistent with a recent study that demonstrated that the NOP56 protein is autoregulated by modulating the available pool of protein-coding *NOP56* mRNA (Lykke-Andersen et al., 2018). These findings provide evidence that massively expanded TG_3C_2 repeats impair splicing and result in retention of the repeat containing intron 1 of *NOP56* in SCA36 patient.

2.1.4 ATG-mediated translation preferentially drives production of poly(GP) DPRs in SCA36

Retention of *NOP56* intron 1 in SCA36 results in a new mRNA variant. Unlike the *C9orf72* intron 1 that is flanked by untranslated exons, the *NOP56* intron 1 is flanked by coding exons. Intron 1 of *NOP56* is a phase 0 intron positioned immediately after the ATG start codon in exon 1. This type of exon-intron structure is referred to as a "start-codon intron" and is overrepresented in genes encoding highly conserved proteins such as the

ribosomal proteins (Nielsen and Wernersson, 2006; Yoshihama et al., 2002). Given the ATG start site in the first exon, we hypothesized that retention of intron 1 would allow for AUG mediated translation of the expanded TG_3C_2 repeats in SCA36. Interestingly, analysis of the *NOP56* Refseq gene (NM_006392) reveals that AUG mediated translation of a retained intron transcript variant would produce a polypeptide containing repeating Gly-Pro units, i.e. poly(GP) frame **(Fig 2.4D)**. It should also be noted that the retained intron transcript variant maintains a perfect Kozak sequence (GCCATG|g, where | denotes exon-intron boundary). Therefore, we sought to determine whether the ORF of a retained intron transcript variant specifically on the allele harboring TG_3C_2 expansions ("expanded allele") are in frame with poly(GP) in SCA36 patients. We performed repeat-primed PCR (RP-PCR) on gDNA isolated from SCA36 patient cerebellum. RP-PCR amplicons were separated via gel electrophoresis and high-molecular weight amplicons were gel extracted, TOPO-cloned, and Sanger sequenced **(Fig. 2.14A)**. It is possible that RP-PCR would produce amplicons from the WT (non-expanded) allele as well; thus, to ensure that only amplicons specifically from the expanded allele were analyzed, we considered TOPO-clone sequences that 1) contained a greater number of TG_3C_2 repeat units than the corresponding WT allele, as determined by sequencing of the WT specific allele and 2) contain the rs28970277 (-/CGGGCG) polymorphism in the intron region specific to the expanded allele haplotype (**(Fig. 2.14A**). Similar to the *NOP56* Refseq gene sequence, putative AUG mediated translation of the expanded allele would produce a poly(GP) containing peptide species **(Fig 2.4D)**. In light of this finding, it stands to reason that canonical AUG translation could drive production of the poly(GP) DPR in SCA36. Therefore, we next investigated how canonical AUG mediated translation in the GP frame (+0 frame) would affect translation of the other putative SCA36 specific DPR frames,

poly(GL) and poly(WA). To this end we generated a number of translation reporter constructs that contained 23bp of *NOP56* exon 1 and intron 1 containing 82 TG_3C_2 repeats followed by NanoLuc in either the +0, +1, +2 frame **(Fig 2.4E)**. As expected, translation in the +0 (GP) frame was most efficient and resulted in robust luciferase signal; however, luciferase signal in the +1 and +2 frames was significantly reduced. Given that expanded TG_3C_2 repeats can undergo RAN translation **(Fig 2.1A-D)**, we wanted to compare the efficiency of translation of the +1 and +2 frame associated DPRs in both canonical AUG mediated and non-canonical RAN translation mediated contexts. The translational reporter constructs were modified by replacing the 23bp exon 1 of *NOP56* with a 23bp sequence containing six stop codons (two stop codons per frame). Notably, translation of the +1 and +2 frame is significantly reduced in a canonical AUG mediated context compared to the non-canonical RAN translation context **(Fig. 2.4F)**. Collectively, this data suggests that the retention of the repeat containing intron 1 of *NOP56* in SCA36 results in a novel mRNA variant that promotes canonical AUG driven translation of poly(GP) DPRs in SCA36. Moreover, canonical translation preferentially produces the poly(GP) DPR relative to poly(GL) or poly(WA).

2.1.5 Chimeric DPR species underlie divergent DPR pathology between SCA36 and c9ALS/FTD

Next, we sought to determine the underlying mechanism that drives divergent dipeptide repeat pathology in pathology in c9ALS/FTD and SCA36. Transient expression of GP_{80x}-GFP in HEK293T cells suggests that poly(GP) is a soluble peptide species **(Fig 2.15A, B)**; however we wanted to assess whether poly(GP) remains soluble after long term expression in disease relevant cell types both *in vitro* and *in vivo*. To this end, we generated an AAV

expressing GFP alone or GP_{80x}-GFP. Transduction of iPS-derived motor neurons with AAV-GP_{80x}-GFP for 21 days revealed that poly(GP) remained diffuse throughout the cell including the cytoplasm, neurites, and nucleus **(Fig 2.15C)**. Since poly(GP) forms perinuclear inclusions in c9ALS/FTD, it is reasonable to posit that an *in vivo* component is required for the aggregation of poly(GP). To address this, we expressed GFP or GP_{80x}-GFP in the mouse CNS via intracerebrventricular delivery. Immunohistochemical analysis for poly(GP) in GP_{80x}-GFP revealed diffuse poly(GP) staining throughout the CNS. Poly(GP) was diffuse in the cell bodies and dense dendritic neurites of Purkinje cells in the cerebellum as well as neurons of the frontal cortex **(Fig 2.5A, Fig 2.15D)** . Notably, poly(GP) signal was robust throughout the neuropil which is consistent with poly(GP) being diffuse in the neurites in of iPS-derived motor neurons. The c9ALS/FTD specific DPR, poly(GA) is highly aggregate prone, and unlike poly(GP), transient expression of GA encoding constructs in HEK293T cells form perinuclear inclusions. Moreover, poly(GA) inclusions are highly abundant in c9ALS/FTD post mortem tissue. We hypothesized that the aggregate prone propensity of poly(GA) may promote aggregation of poly(GP). We performed dual-immunofluorescence for poly(GA) and poly(GP) in human c9ALS/FTD post-mortem tissue. Similar to a previous report (Lee et al., 2017) we found that near 100% of poly(GP) aggregates co-localized with a poly(GA) inclusion, however only a fraction of poly(GA) aggregates co-localized with poly(GP) **(Fig 2.5B)**. Co-localization of poly(GP) inclusions with poly(GA) further supports the hypothesis that poly(GA) mediates aggregation of poly(GP) in human post-mortem samples. In light of this finding, we reasoned that poly(GA)-mediated aggregation of poly(GP) could result through two different mechanisms: 1) poly(GA) DPRs cross-seed aggregation of poly(GP) or 2) aberrations in the translation of the c9ALS/FTD associated G_4C_2 repeat results in the

production of a chimeric dipeptide repeat ("cDPRs") species consisting of GA and GP DPRs **(Fig 2.5C)**. To test this hypothesis, we generated a number of tagged constructs expressing DPRs alone, GA_{80x}-GFP ("GA"), GA_{80x}-HA ("GA-HA"), GP_{80x}-GFP ("GP") or a chimeric DPR construct expressing $GA_{50x}GP_{50x}$-GFP ("GAGP"). Immunofluorescence analysis for GFP reveals that chimeric $GA_{50x}GP_{50x}$-GFP peptides form perinuclear inclusions in HEK293T cells similar to GA_{80x}-GFP; this is in contrast to the diffuse pattern seen with poly(GP) **(Fig 2.5D)**. While we see co-localization of GA_{80x}-HA and GP_{80x}-GFP when co-transfected in HEK293T cells, it is unclear whether this co-localization is in fact due to co-aggregation of these DPRs or rather a weak hydrophpbic interaction. To address this question, we investigated whether GA_{80x}-GFP ("GP"), GP_{80x}-GFP ("GA"), co-transfection of GA_{80x}-HA and GP_{80x}-GFP ("GA + GP"), or chimeric $GA_{50x}GP_{50x}$-GFP ("GAGP") peptides reside in the soluble (1% triton-x) or insoluble (2% SDS) fraction. As expected, immunoblots for GFP reveal poly(GP) alone resides predominately in the soluble fraction, whereas poly(GA) resided in the insoluble fraction **(Fig 2.5E)**. Interestingly, co-transfection of GP_{80x}-GFP and GA_{80x}-HA demonstrates that poly(GP) in the presence of a separate poly(GA) peptide remains soluble. In contrast, chimeric $GA_{50x}GP_{50x}$-GFP peptides are found in the insoluble fraction. This suggests that co-localization of GA and GP as assessed by immunofluorescence may result from weak protein-protein interactions. Furthermore, these data suggest that chimeric DPRs, which we termed "cDPRs" but not a combination of individual GA-GP interactions can promote the formation of insoluble poly(GP) inclusions as seen in human post-mortem tissue **(Fig 2.2F,H and Fig 2.5B)**. We next investigated whether cDPRs occur in human post-mortem tissue samples. In order to test our hypothesis, we developed a meso-scale discovery (MSD) immunoassay to detect chimeric DPR species **(Fig 2.5F)**. The sensitivity of this

immunoassay to specifically detect chimeric DPR species was tested by assaying HEK293T lysates expressing GP alone, GA alone, chimeric GAGP, or co-transfected GA and GP constructs. We found that the chimeric MSD immunoassay specifically and robustly detects chimeric GAGP DPRs **(Fig 2.5G)**. Importantly, we found that chimeric GAGP DPRs could be detected in both detergent soluble and detergent-insoluble, urea-soluble fractions in the cortex of c9ALS/FTD patients (n=4) (**Fig 2.5H,I)**. The absence of GAGP cDPRs in controls and SCA36 support the specificity of this assay to detect cDPRs. Collectively, these data provide evidence that the production of chimeric dipeptide repeats underlie the divergent DPR pathology between c9ALS/FTD and SCA36.

2.1.6 Antisense oligonucleotides reduce poly(GP) in SCA36 cellular models

Finally, we evaluated the therapeutic potential of antisense oligonucleotides (ASOs) to mitigate DPR burden in SCA36. RNAse H-dependent ASOs were designed to target various regions of the *NOP56* locus **(Fig. 2.6A)**. An initial screen in SCA36 fibroblasts transfected with 100nM ASOs for 48 hours revealed that ASOs specifically targeting the HRE-harboring intron 1 of the *NOP56* locus significantly reduced poly(GP) abundance by ~50% **(Fig 2.6B)**. Similarly, expression of *NOP56* mRNA was reduced by intron 1 targeting ASOs **(Fig 2.6C)**. Given the limitations of lipid-based delivery methods as a viable *in vivo* therapeutic option, we tested whether gymnotic delivery of ASOs could reduce poly(GP) in patient fibroblast and LCLs **(Fig 2.6D)**. We found that ASOs targeting the expanded repeat robustly reduced poly(GP). Lastly, we demonstrate that poly(GP) is detectable in the cerebrospinal fluid (CSF) of SCA36 patients as observed in c9ALS/FTD patients(Gendron et al., 2017) **(Fig 2.6E)**. These results suggest that ASOs could be a potential therapeutic option for reducing repeat-RNA and DPRs in SCA36 and that

detection of poly(GP) in the CSF may prove to be a useful biomarker for ASO target engagement.

2.1.7 Discussion

Here we show that the intronic SCA36-associated TG_3C_2 repeat expansion can undergo RAN translation in both the sense and antisense direction to produce DPR proteins. We found that poly(GP) levels were higher in SCA36 than in c9ALS/FTD, but that inclusions of poly(PR), reported to be highly toxic in several model systems (Kwon et al., 2014; Mizielinska et al., 2014), were rare in both SCA36 and c9ALS/FTD. With regards to poly(GP), there is an AUG initiation site in the first exon of *NOP56*; it is thus possible that the increase in poly(GP) abundance in SCA36 results from a combination of non-canonical RAN translation and canonical AUG-mediated translation. Of interest, despite poly(GP) abundance being higher in SCA36 than in c9ALS/FTD, poly(GP) is predominately soluble in SCA36 but predominately insoluble in c9ALS/FTD. These data suggests that a secondary mechanism initiates aggregation of poly(GP) in c9ALS/FTD. Although poly(GP) is not overtly toxic in some cellular and animal models of c9ALS/FTD, these models are limited in their potential to study the long-term effect of poly(GP) burden. Moreover, it is possible that CNS regions with relatively high soluble poly(GP) burden in SCA36 such as the cerebellar dentate nucleus are less efficient in clearing DPRs and may in fact lead to their selective vulnerability in SCA36.

A hallmark of c9ALS/FTD as well as sporadic and other genetic forms of ALS and FTD is the nuclear clearing of TDP-43 and the accumulation of phosphorylated and cleaved TDP-43 into cytoplasmic aggregates. Previous studies have demonstrated that TDP-43 pathology correlates with neurodegeneration and clinical phenotypes(Davidson et al.,

2016). Although a recent study demonstrated that c9ALS/FTD-associated DPRs can cause TDP-43 mislocalization in *Drosophila* (Solomon et al., 2018), previous work in mice provided evidence that poly(GA) (Zhang et al., 2016) and poly(GR) (Zhang et al., 2018) alone cannot promote the formation of TDP-43 inclusions. Nevertheless, expression of the c9ALS/FTD-associated expanded G_4C_2 repeat in mice does cause TDP-43 pathology (Chew et al., 2015), suggesting that TDP-43 dysfunction in c9ALS/FTD results from a combination of various DPRs, from RNA-mediated toxicity or from some combination of these mechanisms. Consistent with these data, we show that the presence of poly(GP) and poly(PR) are insufficient to cause TDP-43 pathology in SCA36.

Overall, this study reveals key similarities and differences in the solubility and protein aggregation of DPRs between SCA36 and c9ALS/FTD and implicates alternate mechanisms for how highly similar repeat expansions can cause markedly distinct forms of neurodegeneration. Comparing these related yet different diseases provides us with a unique opportunity to address key questions about HRE-dependent neurodegeneration, and to expand our understanding of DPR properties and their role in neurological disorders.

2.1.8 Figures

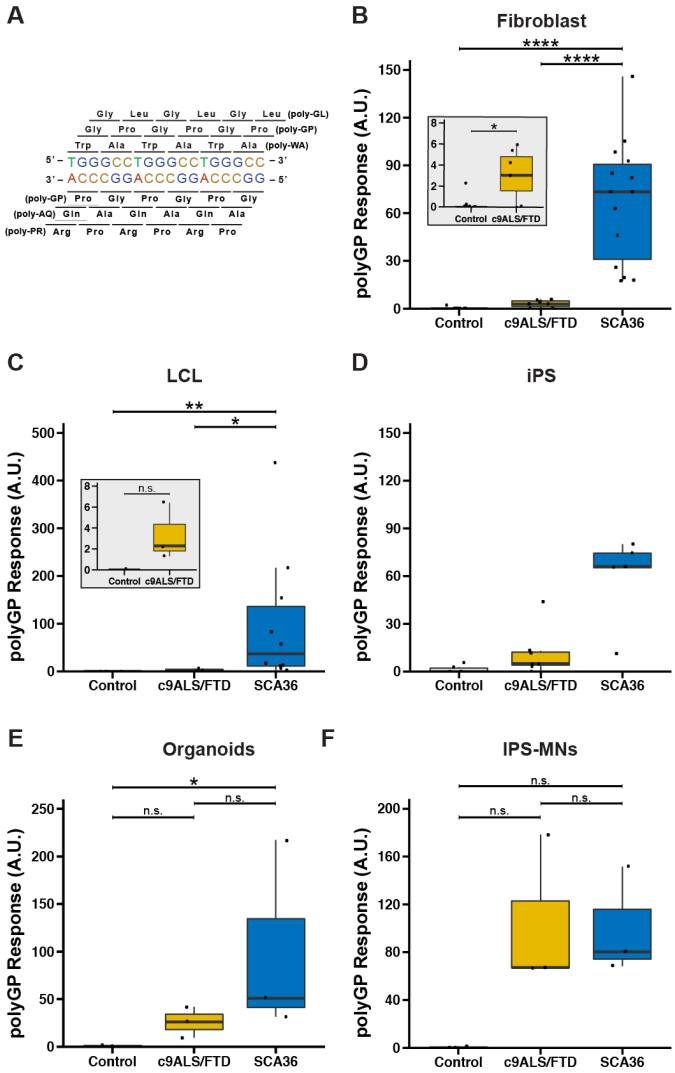


Figure 2.1: Poly(GP) is detectable in patient derived cellular models of SCA36.

(A) Schematic of putative dipeptide repeats in SCA36. **(B to F)** Immunoassay for poly(GP) in SCA36 fibroblast, lymphoblastoid cell lines (LCL), induced pluripotent stem cell (iPS), 3D forebrain organoid (Day 105), and iPS-derived motor neurons (iPS-MNs). Data are presented as mean ± SEM; *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001, **(B-F)** Kruskal-Wallis one way analysis of variance followed by pairwise Mann-Whitney test.

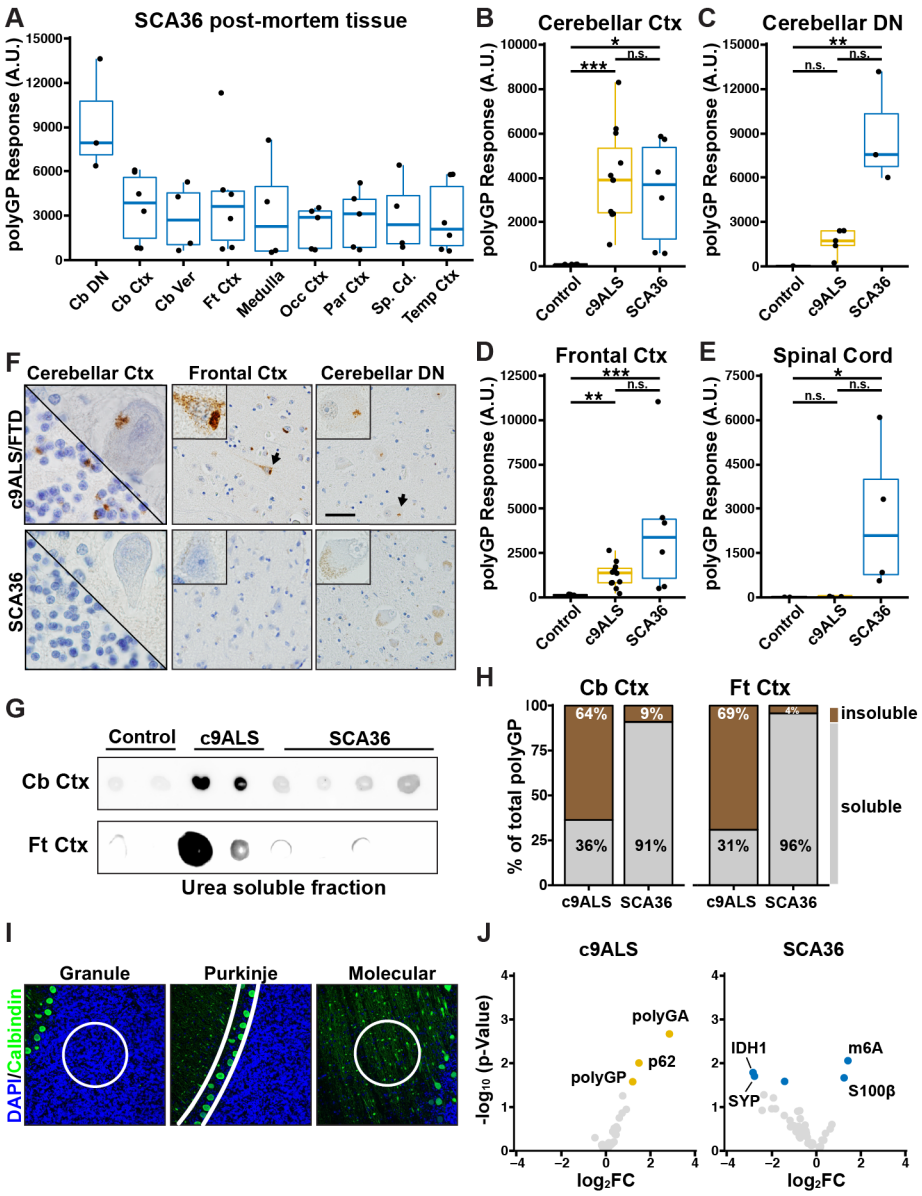


Figure 2.2: Poly(GP) is abundant but does not form insoluble inclusions in SCA36 postmortem tissue.

(A) polyGP is detectable throughout the CNS of SCA36 patients. **(B-E)** polyGP is expressed at similar or increased abundance in SCA36 samples compared to c9ALS/FTD samples in disease relevant CNS regions. **(F)** poly(GP) aggregates in c9ALS/FTD but not SCA36 patient postmortem tissue. Scale Bar: 100 µm. **(G)** poly(GP) is detectable in the

urea soluble fraction of c9ALS/FTD but not SCA36 samples as assessed by dotblot. **(H)** Quantitative immunoassay of sequentially fractionated tissue samples reveals poly(GP) predominately exist as a soluble protein species in SCA36 whereas in c9ALS/FTD a large fraction of poly(GP) exists in the insoluble fraction. **(I)** Representative images of granule, Purkinje, and molecular cell layer ROI selection (white ouline) for digital spatial profiling of cerebellar samples. Calbinding and DAPI were used to delineate cerebellar architecture. **(J)** Volcano plots illustrating differentially abundant proteins as determined by digital spatial profiling in the cerebellar granule layer of c9ALS/FTD and SCA36 compared to controls (n=7 c9ALS/FTD, n=5 SCA36, n=7 control cases). Significant proteins ($p<0.05$) are colored. (B-E) *$p<0.05$, **$p<0.005$, ***$p<0.0005$, Kruskal-Wallis one way analysis of variance followed by Dunn's multiple comparison test.

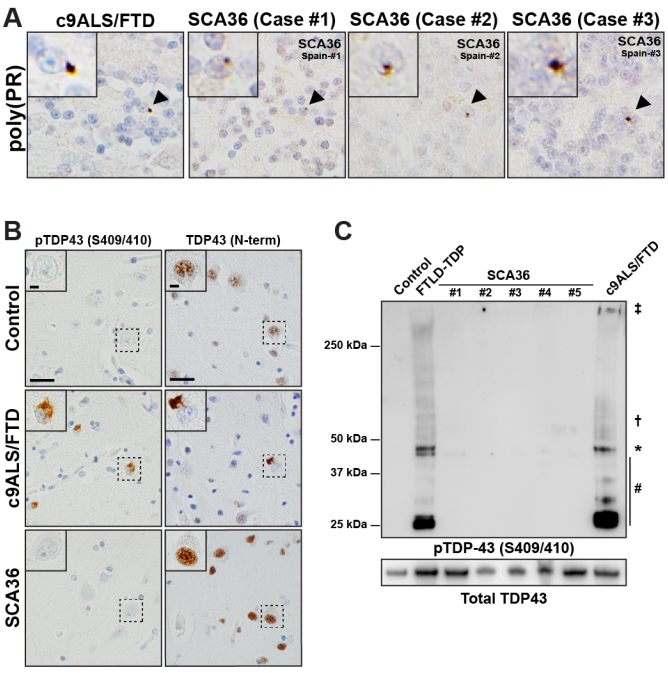


Figure 2.3: SCA36 contains polyPR aggregates but lacks TDP43 pathology.

(A) The antisense DPR, poly(PR) aggregates in both c9ALS/FTD and SCA36 cerebellar tissue. (B) SCA36 frontal cortex tissue lacks TDP43 pathology characteristic of c9ALS/FTD. (C) pTDP43 is not found in the detergent insoluble, urea soluble fraction of SCA36 patient frontal cortex. #cleavage products, *full-length TDP43, †oligomeric species, ‡high molecular weight species.

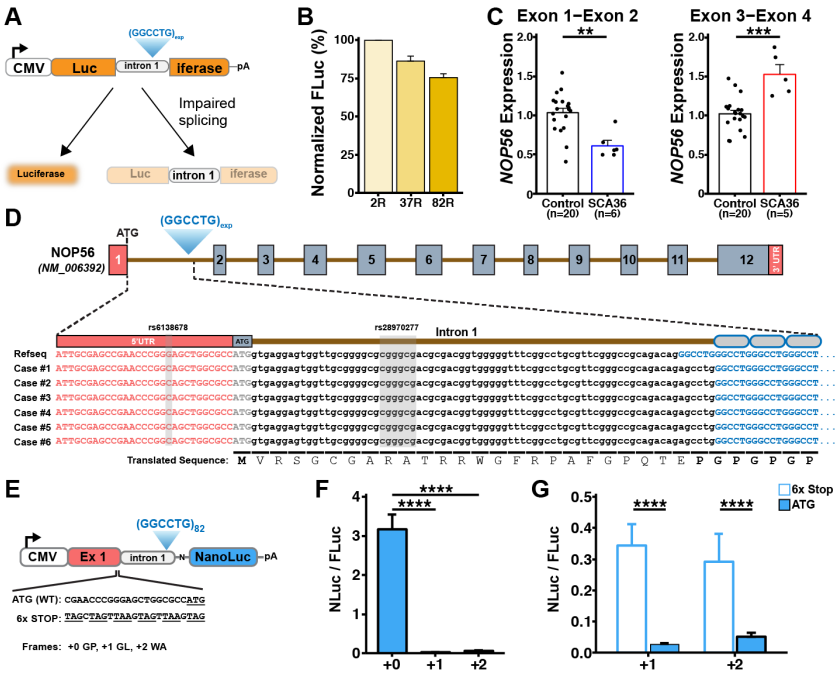


Figure 2.4: Expanded TG3C2 impair splicing and result in intron 1 retention of NOP56 in SCA36.

(A) Schematic of split luciferase expression construct. Intron 1 refers to the intron 1 sequence of *NOP56* containing 2, 37, or 82 TG_3C_2 repeats. (B) Normalized expression (to 2R) of split luciferase reporter constructs (n=2 biological replicates). C) Relative expression of *NOP56* using TaqMan probes spanning exon-exon junctions in control and SCA36 cerebellar cortex samples (n=20 control, n=5 SCA36). (D) Alignment of the 5' flanking sequence of intron 1 from the expanded allele of SCA36 cases. (E) Schematic figure of $(GGCCTG)_{82}$ NanoLuc (NLuc) reporter constructs. (F) Expression of ATG $(GGCCTG)_{82}$ NanoLuc reporter constructs (n=3 biological replicates). Values normalized to co-transected firefly luciferase (FLuc). (G) Expression of 6x Stop and ATG - $(GGCCTG)_{82}$ NanoLuc reporter constructs for the +1 (GL) and +2 (WA) frame (n=3 biological replicates).

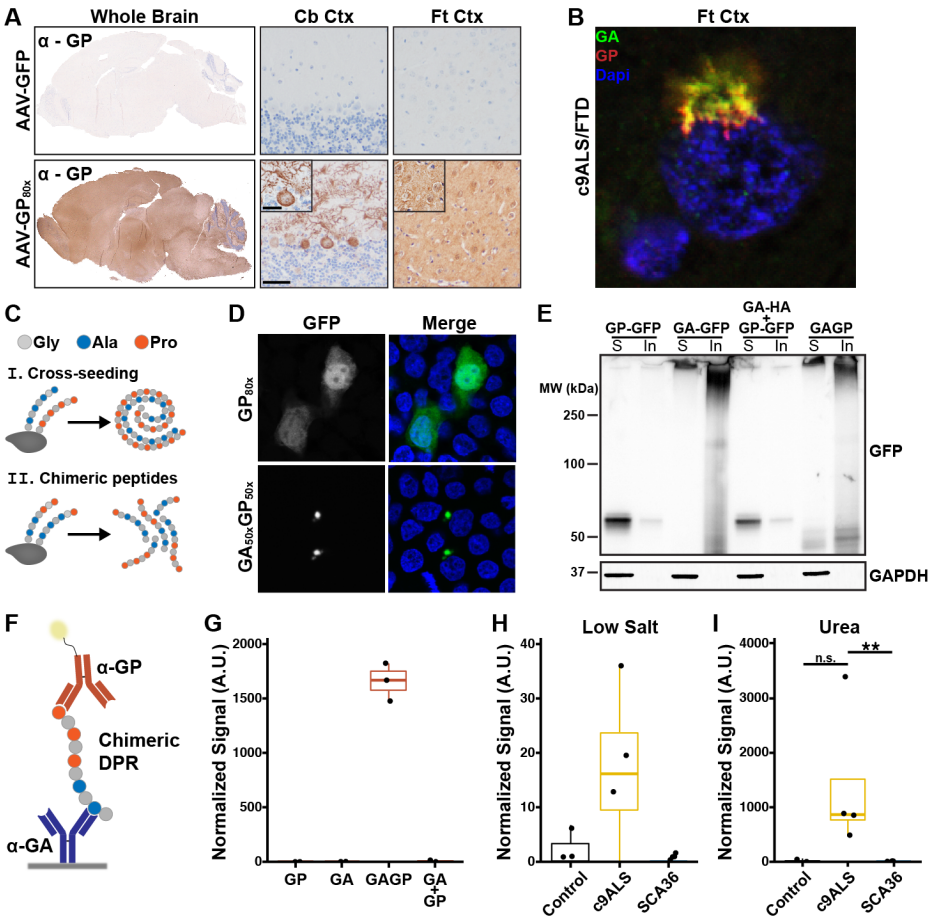


Figure 2.5: Chimeric peptides underlie divergent DPR pathology in c9ALS/FTD and SCA36.

(A) polyGP immunohistochemistry in AAV-GFP and AAV-GP_{80x}-GFP. polyGP remains soluble after 2 months *in vivo*. **(B)** poly(GA) and poly(GP) colocalize in perinuclear aggregates in the frontal cortex of a c9ALS/FTD patient. **(C)** Schematic illustrating putative mechanisms of poly(GA)-dependent poly(GP) aggregation. **(D)** Immunofluorescence for GFP 48 hours after transfection in HEK293T cells expressing GP_{80x}-GFP or chimeric $GA_{50x}GP_{50x}$-GFP to assess the aggregation propensity of a chimeric GAGP peptide. **(E)** Immunoblot for the indicated proteins to determine whether GP_{80x}-GFP, GA_{80x}-GFP, chimeric $GA_{50x}GP_{50}$-GFP, or 1:1 co-transfection of GP_{80x}-GFP and GA_{80x}-HA is present in the soluble (S) or insoluble (In) fraction. **(F)** Schematic

illustrating the meso-scale discovery (MSD) immunoassay developed to detect chimeric DPRs. **(G)** Chimeric GA/GP immunoassay was used to detect chimeric GAGP DPRs in HEK293T cells expressing either GP_{80x}-GFP, GA_{80x}-GFP, chimeric $GA_{50x}GP_{50}$-GFP, or 1:1 co-transfection of GP_{80x}-GFP and GA_{80x}-HA. (n=3 separate transfections) **(H, I)** Chimeric GA/GP immunoassay was used to detect chimeric DPRs in soluble (low salt buffer) or insoluble (urea) fractions from control (n=3), c9ALS/FTD (n=4), or SCA36 (n=5) patient frontal cortex. *$p<0.05$, **$p<0.005$, Kruskal-Wallis one way analysis of variance followed by Dunn's multiple comparison test.

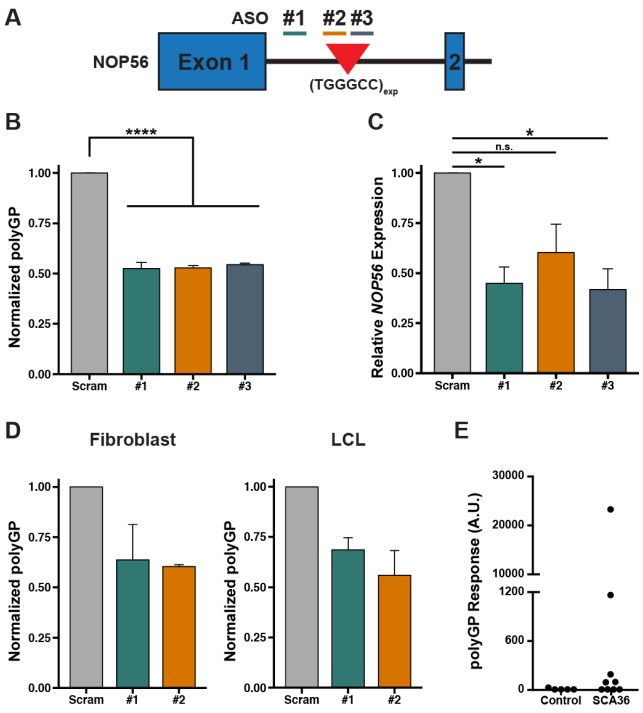


Figure 2.6: Antisense oligonucleotides reduce poly(GP) in patient derived cellular models of SCA36.
(A) Schematic of ASOs targeting the SCA36 associated repeat expansion harboring intron 1 of NOP56. **(B, C)** SCA36 fibroblast were treated with 100nM non-targeting control ASO or an ASO targeting intron 1of NOP56 delivered via cationic-lipid mediated transfection. Protein lysates or RNA were collected 48hrs after treatment. n=3 replicates of one SCA36 patient cell line. **(B)** Relative poly(GP) levels from cell lysates after treatment. **(C)** NOP56 transcript levels after treatment. **(D)** SCA36 fibroblast (n=2 patient cell lines) and LCLs (n=2 patient cell lines) were treated with 5µM non-targeting control ASO or an ASO targeting intron 1of NOP56 delivered via gymnosis. Protein lysates were collected 10 days after treatment and poly(GP) levels measured. Data are presented as mean ± SEM; (B) to (E), *P<0.05, **P<0.01, ***P<0.001 one way analysis of variance (ANOVA) followed by Dunnett's multiple comparisons test. **(E)** poly(GP) in cerebrospinal fluid (CSF) from healthy controls (n=5) and SCA36 patients (n=9).

2.1.9 *Supplemental Figures*

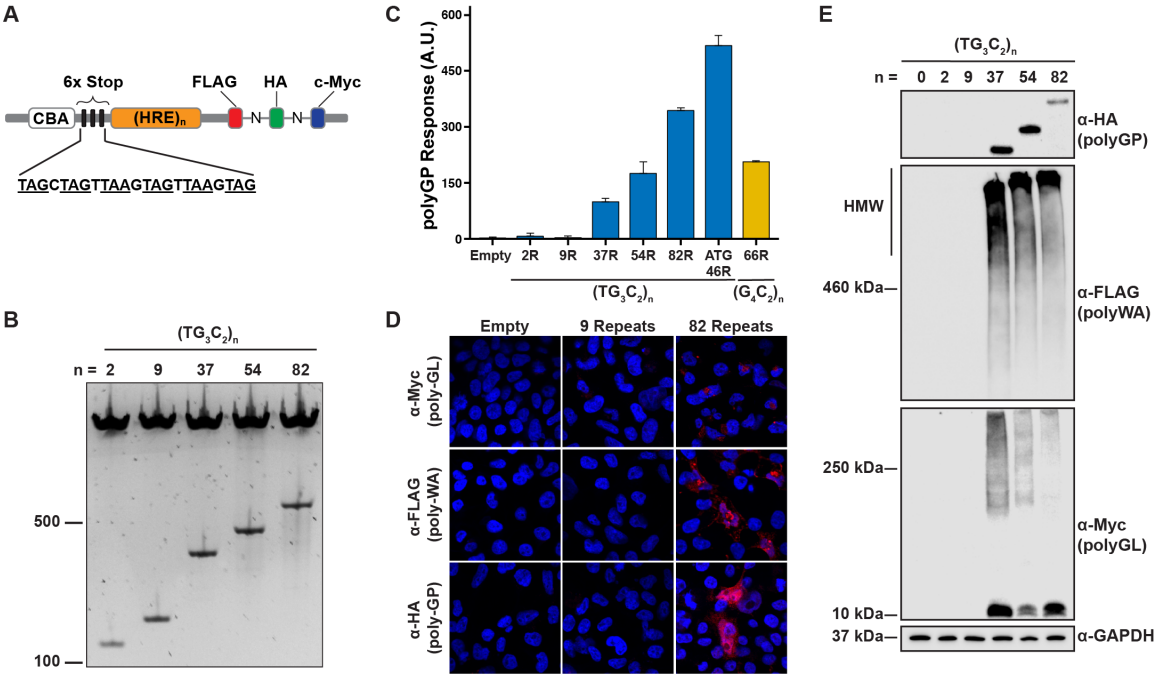


Figure 2.7: (Supplemental) Expanded TG3C2 repeats can undergo RAN translation in vitro.

(A) Schematic figure of $(TG_3C_2)_n$-3T constructs used for *in vitro* RAN translation studies. **(B)** Restriction digest and gel electrophoresis of $(TG_3C_2)_n$-3T constructs containing various TG_3C_2 repeats. **(C)** The expression of poly(GP) DPR protein in cells expressing expanded $(TG_3C_2)_n$ repeats is detectable by poly(GP) immunoassay. **(D)** Immunofluorescence analysis of $(TG_3C_2)_n$-3T constructs in HEK293T cells. **(E)** Sense SCA36 DPRs can be translated as detected by western blot analysis of HEK293T lysates transfected with $(TG_3C_2)_n$-3T constructs.

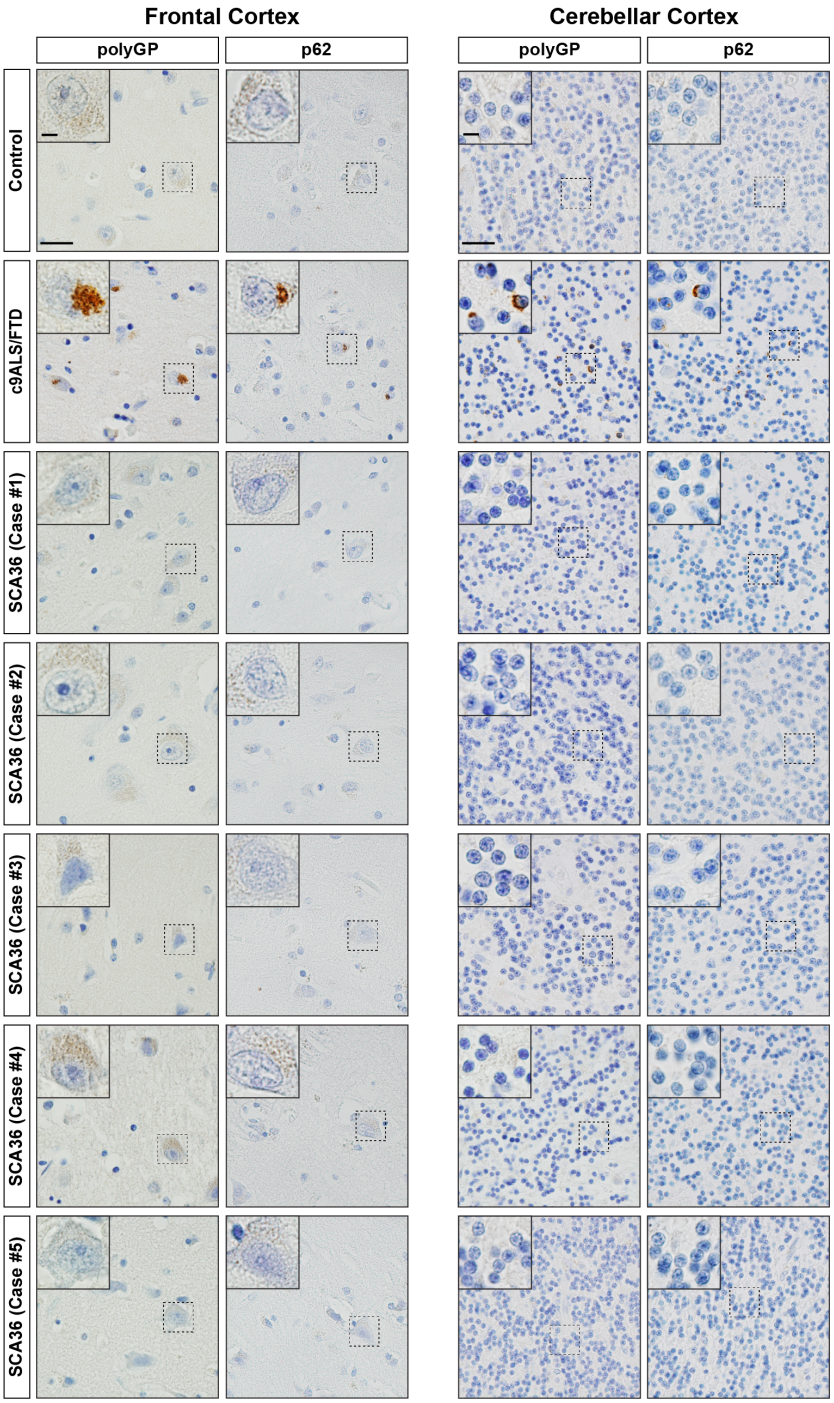


Figure 2.8: (Supplemental) Lack of poly(GP) and p62-positive aggregates in SCA36 patient postmortem tissue.

Immunohistochemical staining of poly(GP) and p62/SQSTM1 in the cerebellar cortex and frontal cortex of control, c9ALS/FTD, and SCA36 patients (n=5 SCA36 cases). Scale bars: 50µm, *inset* 10µm

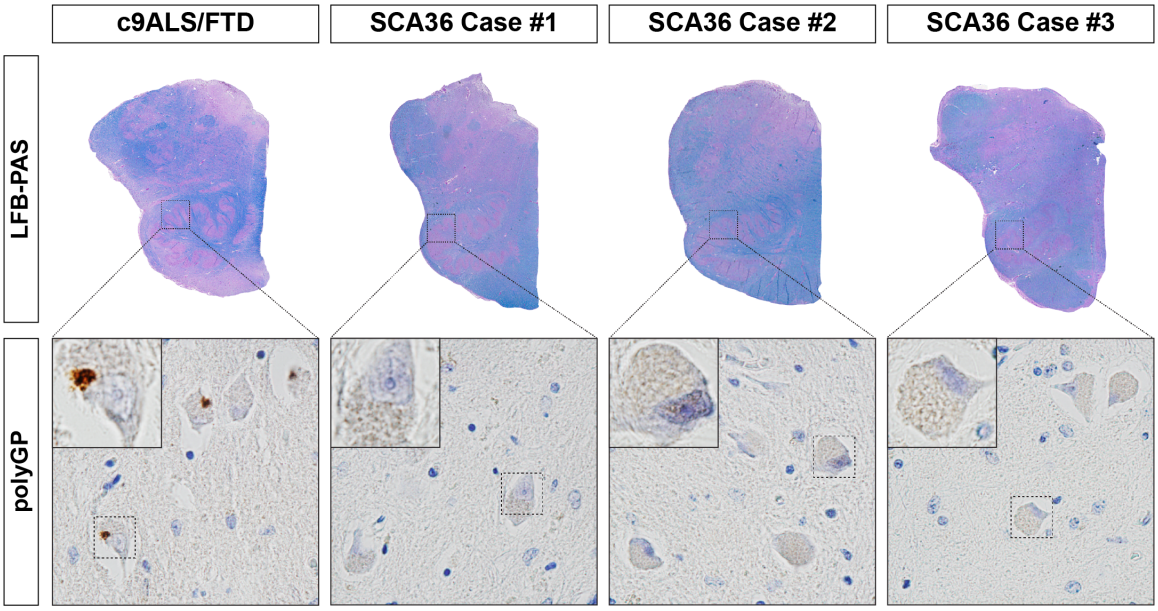


Figure 2.9: (Supplemental) Lack of poly(GP) aggregate in the inferior olivary nucleus of SCA36 patients.

Poly(GP) aggregates are absent in the inferior olive in SCA36 patients as determined by immunohistochemistry (n=3 SCA36 cases).

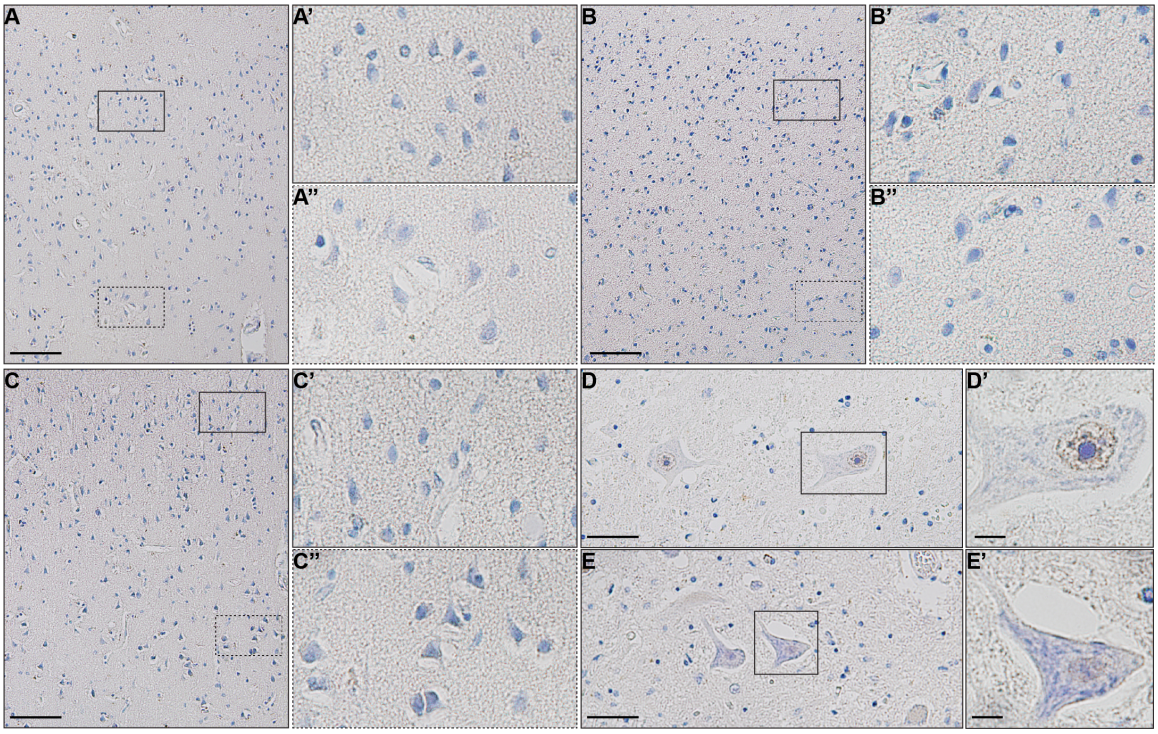


Figure 2.10: (Supplemental) Lack of poly(GP) inclusions in SCA36 throughout the CNS.

Lack of poly(GP) aggregates in the parietal cortex **(A-A'')**, occipital cortex **(B-B'')**, temporal cortex **(C-C'')**, spinal cord **(D, D')**, and hypoglossal nucleus **(E, E')**.

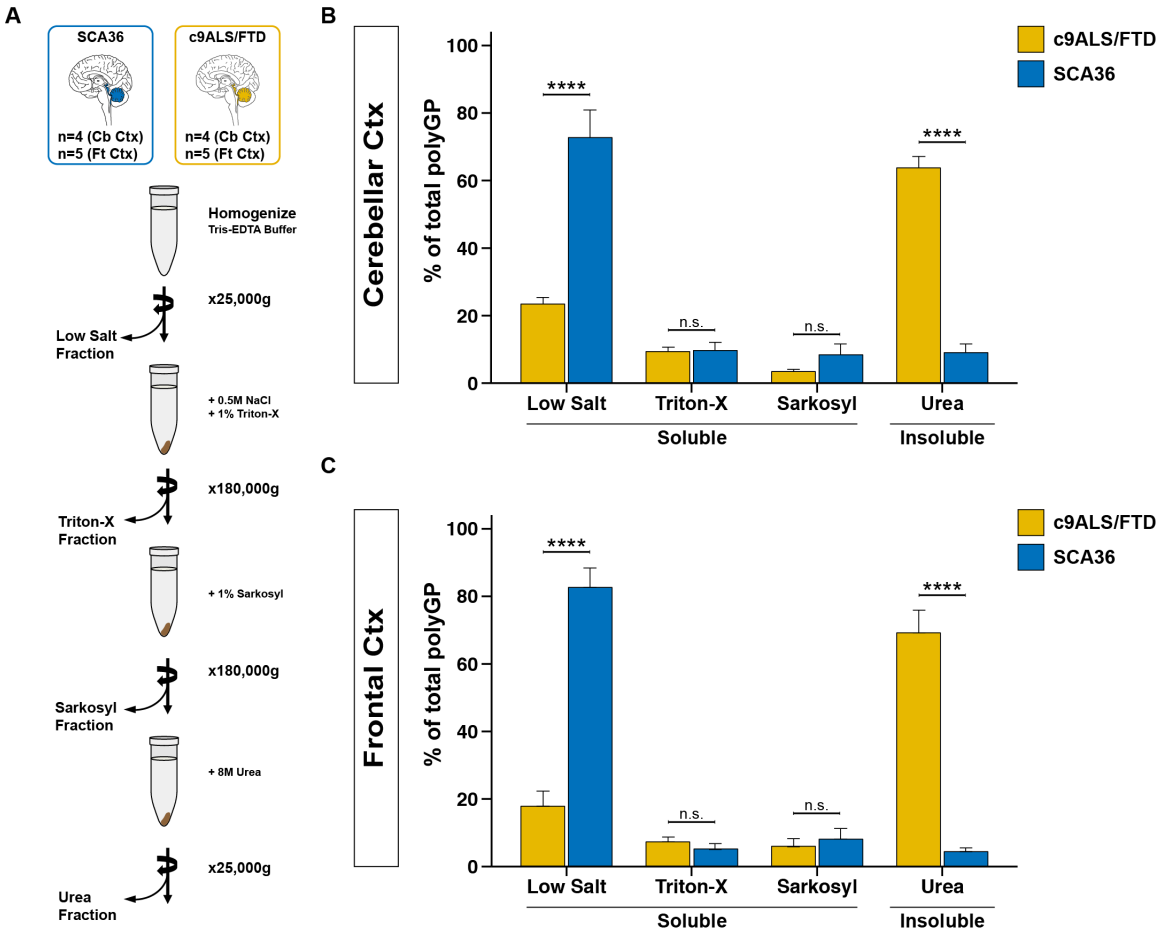


Figure 2.11: (Supplemental) Poly(GP) immunoassay of sequential fractions with increasing solubilizing power.

(A) Schematic of sequential fractionation protocol. **(B, C)** Quantitative immunoassays for poly(GP) in each fraction. Poly(GP) is most abundant in the low salt fraction in SCA36. Cb Ctx: cerebellar cortex. Ft Ctx: frontal cortex. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 two-way analysis of variance (ANOVA) followed by Sidak's multiple comparisons test.

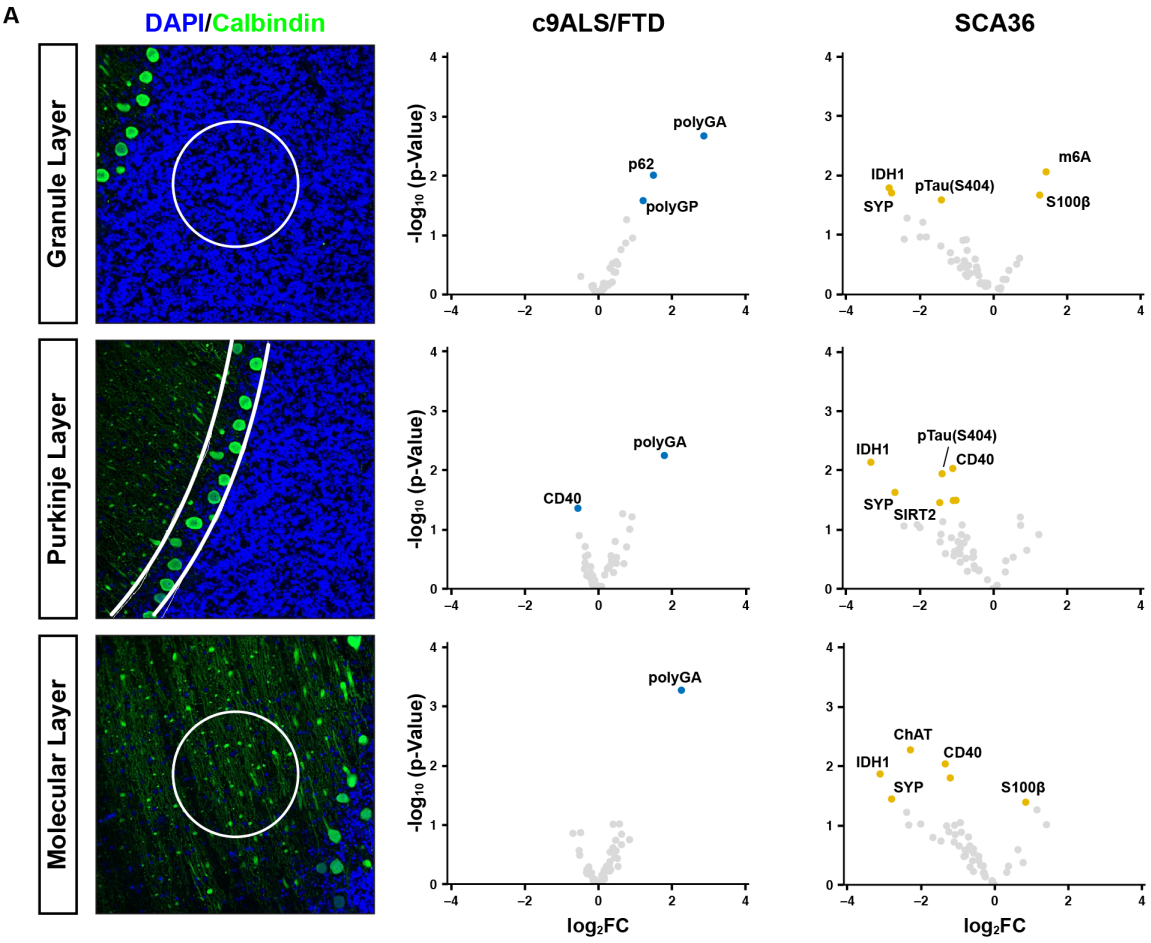


Figure 2.12: (Supplemental) Digital spatial profiling in the cerebellar cortex of c9ALS and SCA36.

(A) Representative images of granule, Purkinje, and molecular cell layer ROI selection (white ouline) for digital spatial profiling of cerebellar samples. Calbinding and DAPI were used to delineate cerebellar architecture. **(B)** Volcano plots illustrating differentially abundant proteins as determined by digital spatial profiling in the cerebellar granule, Purkinje, and molecular layer of c9ALS/FTD and SCA36 compared to controls (n=7 c9ALS/FTD, n=5 SCA36, n=7 control cases). Significant proteins ($p<0.05$) are colored.

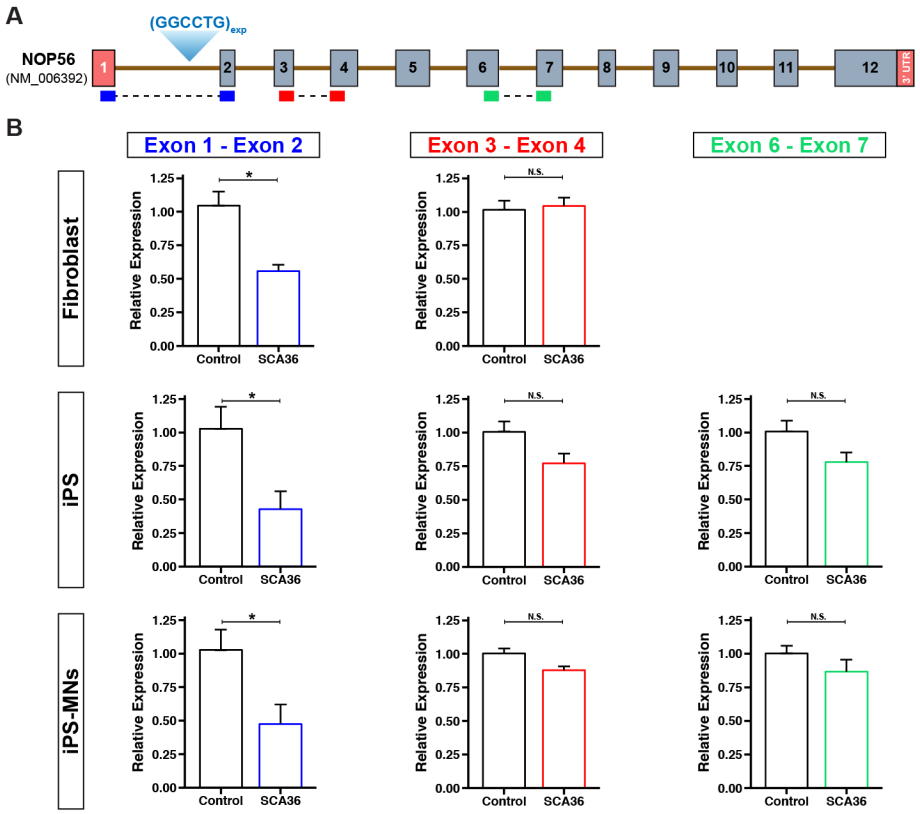
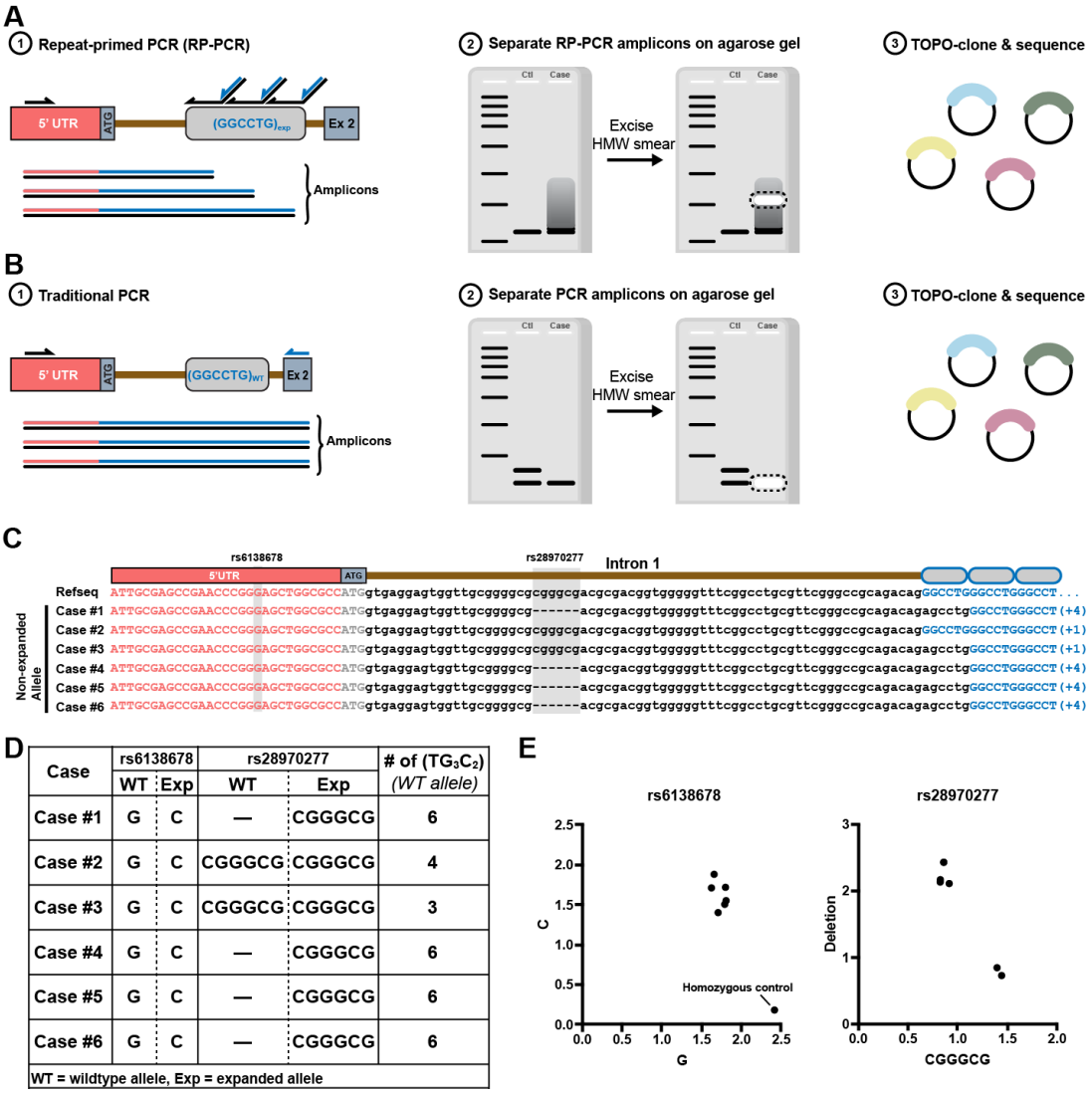


Figure 2.13: (Supplemental) Intron of *NOP56* is retained in SCA36.

(A) Schematic of *NOP56* gene with relative locations of SCA36 associated repeat expansion and custom Taqman probes to assess splicing. **(B)** qPCR for indicated Taqman probes in control and SCA36 patient derived cellular models.



Case	rs6138678		rs28970277		# of (TG_3C_2) *(WT allele)*
	WT	Exp	WT	Exp	
Case #1	G	C	—	CGGGCG	6
Case #2	G	C	CGGGCG	CGGGCG	4
Case #3	G	C	CGGGCG	CGGGCG	3
Case #4	G	C	—	CGGGCG	6
Case #5	G	C	—	CGGGCG	6
Case #6	G	C	—	CGGGCG	6

WT = wildtype allele, Exp = expanded allele

Figure 2.14: (Supplemental) Outline of WT and expanded allele sequencing.

(A) (1) Repeat-primed PCR (RP-PCR) was performed using unmodified oligonucleotides, (2) RP-PCR reaction products were separated on a 2% agarose gel. In order to preferentially select amplicons for the exon 1 – intron 1 region of *NOP56* from the expanded allele, high-molecular weight amplicons were gel extracted, (3) high-molecular weight amplicons were TOPO-cloned and Sanger sequenced. **(B)** (1) PCR was performed using unmodified oligonucleotides to preferentially amplify the WT allele for the exon 1 – intron 1 region of *NOP56*, (2) PCR reaction products were separated on a 2% agarose gel

and single bands for the SCA36 cases were gel extracted, (3) amplicons were TOPO-cloned and Sanger sequenced. **(C)** Alignment of the 5' flanking sequence of intron 1 from the WT allele of SCA36 cases. Numbers in parenthesis on the right side represent number of additional GGCCTG repeats from the WT allele. **(D)** Table summarizing genotype data for the WT and expanded alleles in each SCA36 cases. **(E)** Allelic discrimination plots for the SNPs, rs6138678 and rs28970277.

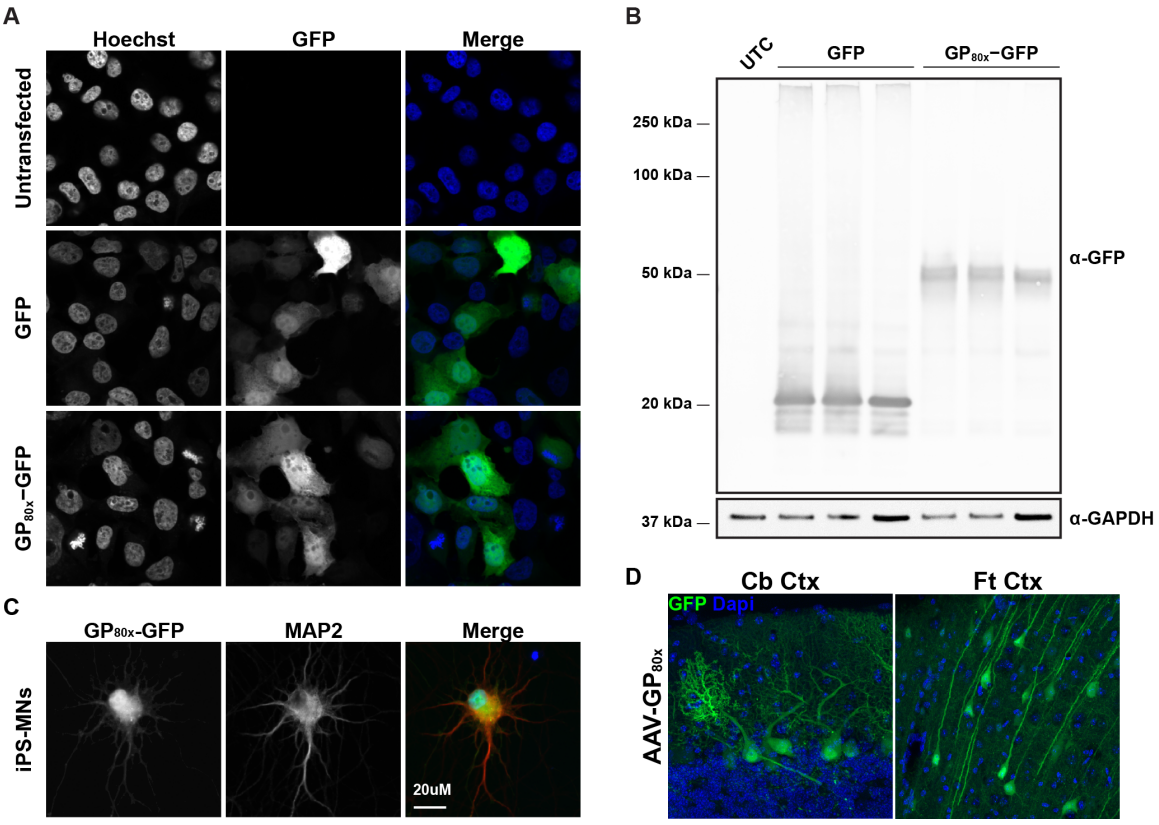


Figure 2.15: (Supplemental) poly(GP) is soluble in *in vitro* and *in vivo* models.

(A) Immunofluorescence of GP_{80x}-GFP transiently transfected in HEK293T cells for 48 hours, **(B)** Immunoblot for indicated proteins of GP_{80x}-GFP transiently transfected in HEK293T cells for 48 hours. **(C)** AAV-mediated transduction of GP_{80x}-GFP in iPS-derived motor neurons for 24 days. **(D)** Diffuse poly(GP) as detected in the cerebellar cortex and frontal cortex in GP_{80x}-GFP mice at 2 months of age

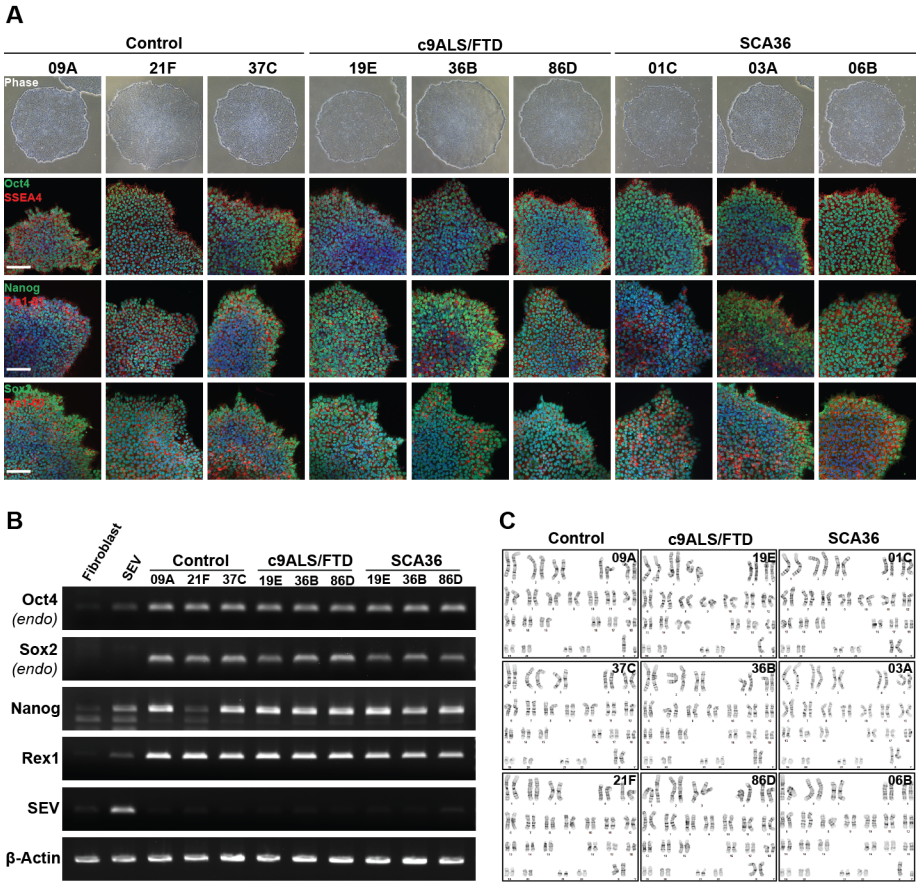


Figure 2.16: (Supplemental) Characterization of Control, c9ALS/FTD, and SCA36 iPS lines.

(A) Phase image of iPS colony morphology and immunofluorescence staining of pluripotent stem cell markers. **(B)** RT-PCR analysis for mRNA expression of stem cell-specific markers; early passage (<5) control iPS colony was used for the SEV positive control. **(C)** Karyotype analysis of iPS cells.

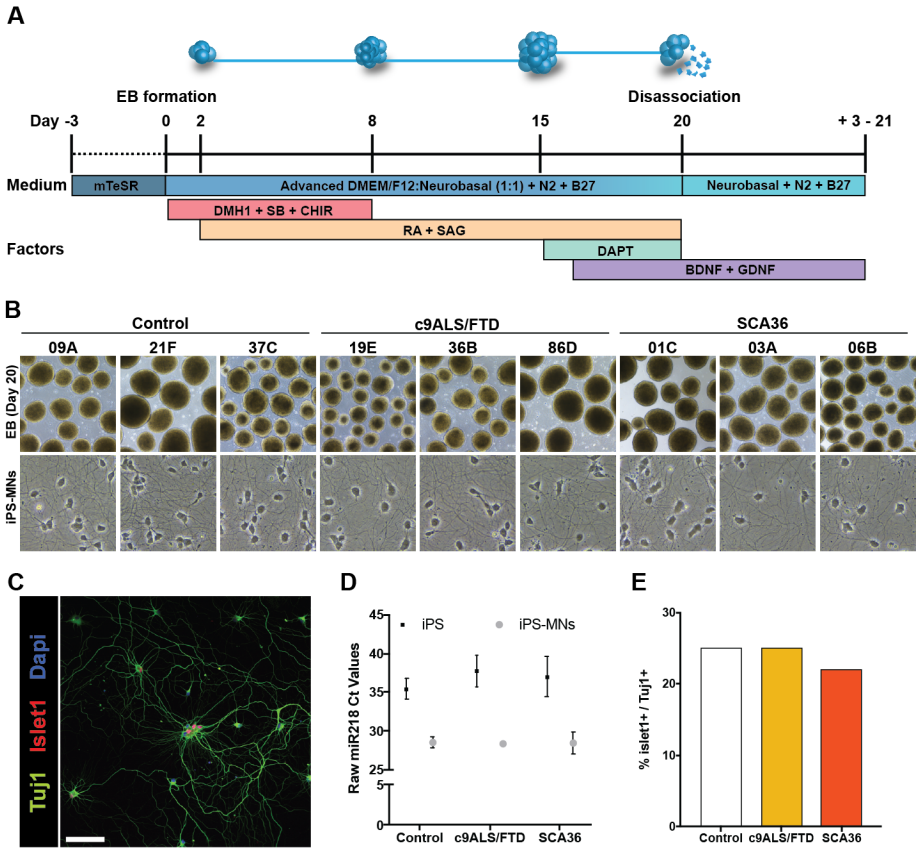


Figure 2.17: (Supplemental) Schematic of the differentiation protocol and characterization of iPS-derived motor neurons (iPS-MNs).

(A) Schematic of iPS-MN differentiation protocol. **(B)** Phase images of embryoid bodies during differentiation and iPS-MNs at day 3 post dissociation. **(C)** Immunofluorescence staining of motor neuron markers in iPS-MN. **(D)** iPS-MNs express mir-218, a motor neuron specific microRNA. **(E)** Quantification of islet1-positive iPS-MNs.

Table 2.1: Characteristics of patients with post mortem analysis

ID	Gender	Age of onset	Age at autopsy	Survival (years)	Diagnosis	Genetic Status	Country
1	M	60	91	31	SCA	SCA36	Spain
2	M	60	95	35	SCA	SCA36	Spain
3	F	63	77	14	SCA	SCA36	Spain
4	F	83	90	7	SCA	SCA36	France
5	M	58	79	21	SCA	SCA36	Japan
6	M	57	86	29	SCA	SCA36	Japan
7	M	58	71	13	ALS/FTLD-TDP	C9orf72	U.S.A.
8	M	62	66	4	ALS/FTLD-TDP	C9orf72	U.S.A.
9	M	57	66	9	FTLD-TDP	C9orf72	U.S.A.
10	M	65	67	1.5	ALS/FTLD-TDP	C9orf72	U.S.A.
11	M	58	67	9	ALS/FTLD-TDP	C9orf72	U.S.A.
12	F	54	57	2.5	ALS/FTLD-TDP	C9orf72	U.S.A.
13	F	50	57	7	ALS/FTLD-TDP	C9orf72	U.S.A.
14	F	60	67	7	ALS/FTLD-TDP	C9orf72	France
15	M	-	57	-	Control	-	U.S.A.
16	M	-	70	-	Control	-	U.S.A.
17	M	-	94	-	Control	-	U.S.A.
18	F	-	57	-	Control	-	U.S.A.
19	F	-	61	-	Control	-	U.S.A.
20	F	-	91	-	Control	-	U.S.A.
21	F	-	92	-	Control	-	France

ALS/FTLD-TDP=amyotrophic lateral sclerosis and frontotemporal lobar degeneration with TDP43 inclusions, SCA36=spinocerebellar ataxia type 36

Table 2.2: Sequence, target, and design of antisense oligonucleotides (ASOs)

ASO Name	Sequence	*NOP56* Target	ASO Design
Scrambled	GCGACTATACGCGCAAUAUG	Nontargeting	• 5-10-5 Gapmer: • phosphorothioate backbone; • 2'-O-Me modifications on 5' and 3' wing;
ASO-#1	GCGCCCCGCAACCACUCCUC	Intron 1	
ASO-#2	CAGGCCCAGGCCCAGGCCCA	(TG_3C_2) repeat	
ASO-#3	GCCCAGGCCCAGGCCCAGGC	(TG_3C_2) repeat	

CHAPTER 3. RNA-MEDIATED TOXICITY IN C9ALS/FTD AND SCA36

3.1 RNA-mediated toxicity in repeat expansion disorders

Repeat containing RNA have been shown to interact and sequester RNA-binding proteins (RBPs), thus effectively reducing the available pool of functional RNA-binding proteins. This putative disease mechanism has been proposed for a number of repeat expansion disorders (Zhang and Ashizawa, 2017); however, the most extensively studied and well-established example of RNA-mediated toxicity occurs in myotonic dystrophy type 1 (DM1). In DM1, expanded CUG repeats in the gene DMPK1 form intranuclear RNA foci (Davis et al., 1997) and sequester the splicing associated RNA binding proteins Muscleblind 1 – 3 (MBNL1-3) (Miller et al., 2000). Depletion of the MBNLs results in mRNA and miRNA misprocessing (Ho et al., 2004; Rau et al., 2011; Wang et al., 2012). Mouse models of DM1 mimic disease phenotypes seen in humans. A homozygous transgenic mouse model expressing expanded CUG repeats in the context of the human *DMPK* gene developed cellular phenotypes such as intranuclear CUG foci as well as myotonia and progressive muscle weakness (Gomes-Pereira et al., 2007; Seznec et al., 2000; Vignaud et al., 2010). The MBNL KO model replicates many of the phenotypes seen in the transgenic CUG model, including clinically relevant features such as myotonia (Kanadia et al., 2003). This further supports the hypothesis that RNA-mediated depletion of MBNL may be the underlying mode of toxicity in DM1. These findings have been the basis for numerous studies investigating this mode of RNA-mediated toxicity in other repeat expansion disorders.

3.2 RNA-Protein Interactions in c9ALS/FTD and SCA36

3.2.1 The c9ALS/FTD and SCA36 associated repeat expansions form atypical nucleic acid structures

The intronic hexanucleotide repeat, G_4C_2, implicated in ALS/FTD has been shown to form both DNA and RNA G-quadruplex structures (Fratta et al., 2012; Haeusler et al., 2014b; Reddy et al., 2013; Su et al., 2014). Since leading hypotheses in c9ALS/FTD rely on the idea that the atypical nucleic acid structures formed by expanded G_4C_2 repeats elicit cellular toxicity and initiate cascades of disease, it is critical to determine whether the SCA36 HRE forms similar atypical nucleic acid structures. It is known that G-quadruplex containing nucleic acids are capable of interacting with and sequestering various RNA binding proteins, thereby preventing the proteins from functioning normally (Lee et al., 2013b). This model of RNA toxicity has been hypothesized to be a possible disease mechanism in c9ALS/FTD and recent studies have demonstrated, using a biotinylated RNA pulldown assay, that various RNA binding proteins indeed interact with G_4C_2 repeats (Donnelly et al., 2013b; Lee et al., 2013b). This, mechanism has not been explored in SCA36. Knowledge of the structure and stability of the SCA36 HRE compared to the c9ALS/FTD HRE will allow us to better understand why shared or unique proteins are able to interact and potentially be sequestered by one or both of the HREs. This understanding, coupled with validation of either similar or different toxicities of these repeats in *in vitro* models, will allow us to test more targeted hypotheses and elucidate why these two similar HREs lead to clinically different phenotypes.

In collaboration with the Hud group at GaTech, we used CD spectroscopy, to demonstrate that $(TG_3C_2)_8$ RNA forms a parallel G-quadruplex structure as evidenced by a

characteristic peak at 263nm and trough at 240nm; similar to the c9ALS/FTD associated $(G_4C_2)_8$ RNA **(Fig 3.1)**.

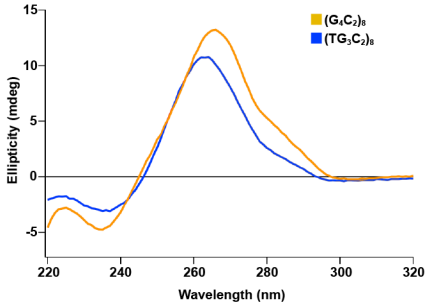


Figure 3.1: CD spectra of G_4C_2 and TG_3C_2 RNA oligonucleotides
CD spectra shows K+ - dependent RNA G-quadruplex formation with $(G_4C_2)_8$ and $(TG_3C_2)_8$ repeats.

To further investigate whether $(TG_3C_2)_8$ oligonucleotides can form non-canonical structures we employed an enzymatic assay, referred to as a G-quadruplex-hemin DNA/RNAzyme assay. Numerous studies have shown that stacked G-tetrads structures can bind and activate hemin, thus promoting the hemin-mediated peroxidation reaction (Cheng et al., 2009; Li et al., 2016). This technique has been used in recent studies to assess the ability of the c9ALS/FTD associated G_4C_2 repeat to form G-quadruplex structures (Grigg et al., 2014; Haeusler et al., 2014a). Surprisingly, we found that unlike $(G_4C_2)_8$ repeats, the $(TG_3C_2)_8$ repeat is unable to bind hemin and facilitate its peroxidase activity **(Fig 3.2A)**. In support of this finding, the $(G_4C_2)_8$ repeats but not the $(TG_3C_2)_8$ were able to bind a G-quadruplex triggered fluorogenic dye (ISCH-oa1) (Chen et al., 2016) **(Fig 3.2B)**. The discrepancy between the spectroscopic and enzymatic/fluorogenic G-quadruplex results for the SCA36-associated TG_3C_2 repeat is intriguing. The G-quadruplex-hemin DNA/RNAzyme assay is mediated by binding of hemin on the planar surface of the stacked G-tetrads (Alizadeh et al., 2017), therefore these data suggest that despite forming a G-quadruplex, the structure of the TG_3C_2 prevents it from interacting

with hemin. Extending this observation to a cellular system, it stands to reason that structures formed from G_4C_2 and TG_3C_2 repeats would potentially interact with unique proteins, in addition to several overlapping proteins.

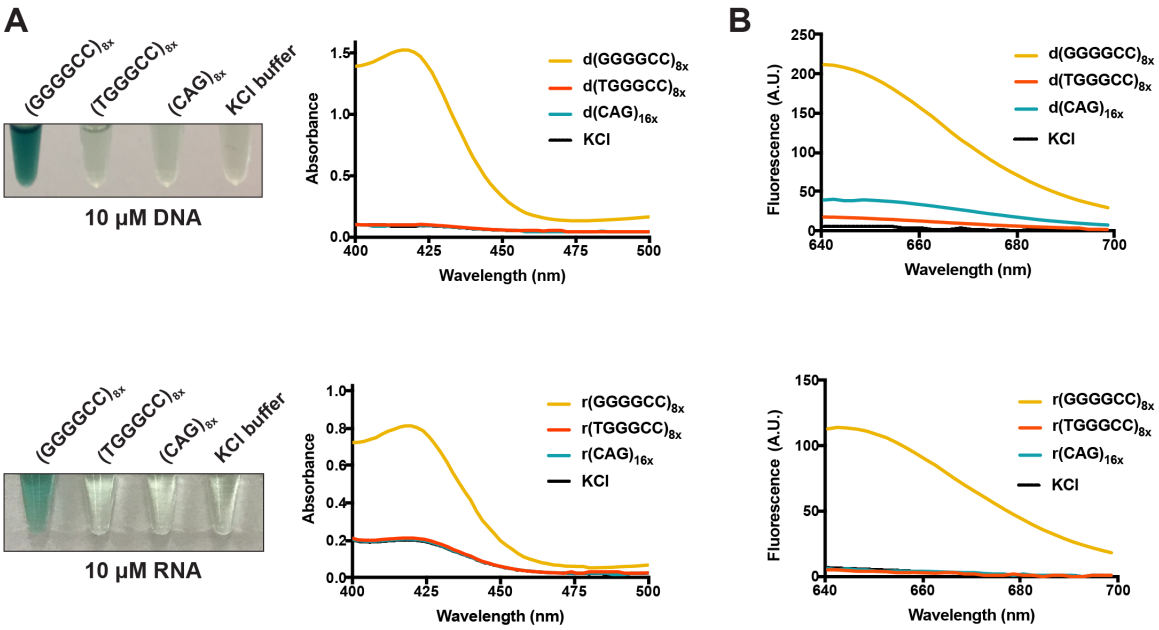


Figure 3.2: Enzymatic and fluorogenic G-quadruplex assays reveal divergent structural properties between the c9ALS/FTD and SCA36 associated repeats. **(A)** G-quadruplex-hemin DNA/RNAzyme assay. $(G_4C_2)_8$, but not $(TG_3C_2)_8$ can bind and activate hemin. *Left*: images of colorimetric changes. *Right*: UV-Vis spectroscopy of 10µM $(G_4C_2)_8$, $(TG_3C_2)_8$ and $(CAG)_{16}$ repeats. **(B)** Fluorescence spectroscopy of 2µM $(G_4C_2)_8$, $(TG_3C_2)_8$ and $(CAG)_{16}$ incubated with 1 µM ISCH dye.

3.2.2 The c9ALS/FTD and SCA36 associated repeat expansions interact with a shared and unique subset of RNA binding proteins.

Previous studies have demonstrated that the c9ALS/FTD-associated G_4C_2 HRE can form both DNA and RNA G-quadruplexes (Haeusler et al., 2014a; Reddy et al., 2013; Zamiri et al., 2014) and that G_4C_2 RNA oligonucleotides can interact with a number of RNA-binding

proteins (RBPs) (Cooper-Knock et al., 2014; Donnelly et al., 2013a; Haeusler et al., 2014a; Mori et al., 2013a); however, these studies together have identified over >900 different potential RNA-interacting proteins (Haeusler et al., 2016), an intractable number to test each interactor independently. Noting the sequence similarity and comparable HRE size between c9ALS/FTD and SCA36 patients, **we hypothesized that comparison of proteins sequestered uniquely by c9ALS/FTD and not SCA36-associated repeats would allow us to narrow down sequestered proteins relevant to c9ALS/FTD pathogenesis**. Using a protein microarray containing over 19,000 human proteins, we performed an unbiased RNA-protein interaction screen (**Fig. 3.3A, B**). Unique to our study is the experimental design of our RNA-protein microarray screen: 1) The SCA36 associated repeat, TG_3C_2, was used as the control/normalizing probe instead of a scrambled control, 2) Both the c9ALS/FTD and SCA36 associated probes were analyzed on the same array, 3) We performed each screen both in the absence and in the presence of physiological levels of KCl (a cation required for G-quadruplex stability), 4) Each screen was performed in duplicate and the repeat associated fluorophore was switched for each replicate. Our results showed that **G_4C_2 and TG_3C_2 RNAs share common interacting proteins, but they also bind to a unique set of RNA-binding proteins** (**Fig. 3.3C**). Interestingly, in the presence of KCl, the number of total G_4C_2 repeat protein interactions increased considerably (**Fig. 3.3D**), suggesting that a subset of these proteins interact in a structure dependent manner.

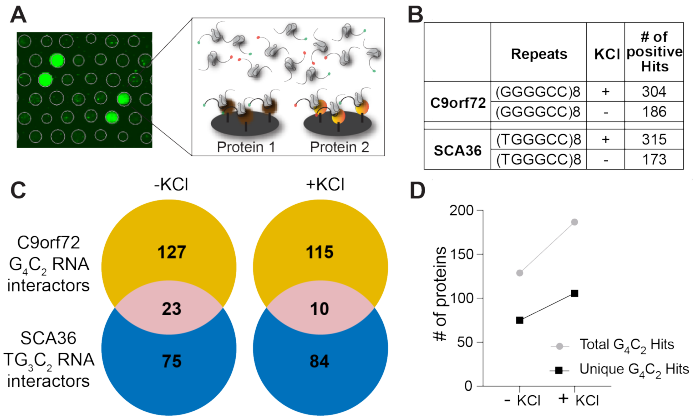


Figure 3.3: Both c9ALS/FTD and SCA36 HRE form G-quadruplexes and interact with similar, but also unique RNA binding proteins.
(A) Schematic showing the experimental design of the RNA-protein interaction microarray. **(B)** Total number of hits identified. **(C)** Venn diagram of overlapping and unique RNA binding interactors between the G_4C_2 and fluorescence intensities for $(G_4C_2)_8$ and $(TG_3C_2)_8$ RNA. **(D)** Number of proteins identified for the $(G_4C_2)_8$ RNA in the absence and presence of KCl.

Similar to previous studies, an intractable number of "positive" hits were identified using this RNA-protein microarray approach. Moreover, the proteins on the microarray may have different abundances in different tissues and thus some RNA-protein interactions may be irrelevant when considering a particular cell/tissue type. Therefore, we reasoned that cross-referencing putative RNA-protein interactors with RNAseq expression data from iPS-derived motor neurons (iPS-MNs) would allow us to narrow down disease relevant interactors. To this end we performed RNA-seq on day 14 iPS-MNs generated from non-expansion control patients and cross-referenced genes expressed in iPS-MNs to proteins we identified as specific interactors of G_4C_2 repeats. Approximately 50% of unique G_4C_2 interacting proteins identified in our protein microarray were indeed expressed in motor neurons thus significantly reducing the number of proteins to consider **(Fig 3.4A-B)**. Furthermore, ranking of RNA-protein interactors by relative gene expression and repeat specificity resulted in a more concise list of top candidates to pursue for further analysis **(Fig 3.4D)**. Interestingly, a number of motor neuron expressed, G_4C_2 specific interactors

are involved in alternative splicing (*e.g.* HNRNPD) and oxidative stress (*e.g.* PIP4K2B), pathways implicated in ALS pathogenesis (Barber and Shaw, 2010; Geuens et al., 2016).

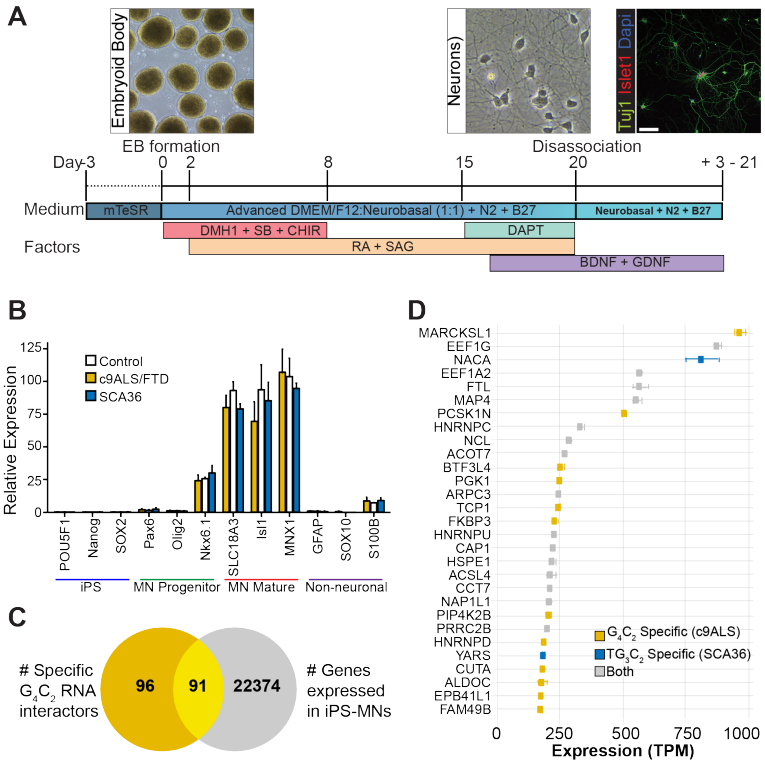


Figure 3.4: RNAseq analysis of iPS-derived MNs reveal motor neuron specific G_4C_2 interactors.
(A) Schematic of iPS motor neuron differentiation. **(B)** RNAseq analysis of iPS-derived MNs reveal a MN specific transcriptome profile. **(C)** Venn diagram of unique G_4C_2 interactors (+ KCl) and motor neuron expressed genes. **(D)** Ranking of repeat specific interactors by relative gene expression (TPMs) in iPS-derived motor neurons (only genes with TPM >150 shown). For ranking, only expression profiles of control (N=3) iPS-motor neurons were used.

3.2.3 Development of an MS2-MCP in vitro model to identify RNA-protein interactions

Although these data support this approach to compare RNA-protein interactions in the c9ALS/FTD and SCA36, to date, studies have not been conducted with G_4C_2 repeat sizes believed to be disease relevant, i.e. ≥30 repeats in a cellular system (Donnelly et al., 2013b; Lee et al., 2013b); instead, these interactions have only been explored using synthesized

short repeat RNAs after incubation with protein extracts from cells/tissues or protein arrays. We believe that using (G_4C_2) constructs of pathologic size to study RNA-protein interactions in disease-relevant model organisms will allow us to better understand the molecular cascades that initiate neurodegeneration in c9ALS/FTD.

In order to capture TG_3C_2 and G_4C_2 specific RNA-protein interactions we proposed to use the MS2-MCP tagging system. The MS2 aptamer is 23nt long RNA stem-loop sequence that binds with high affinity to the *E. Coli* bacteriophage MS2 coat protein (MCP). When the aptamer is added to an RNA sequence of interest, the RNA can be visualized or isolated via the MS2 coat protein (Bardwell and Wickens, 1990). The MS2 system has been used in a number of studies specifically to capture RNA binding proteins (Slobodin and Gerst, 2011), and in a recent study it was used to visualize phase transitions of RNA foci in myotonic dystrophy type 1 and c9ALS/FTD (Jain and Vale, 2017). In addition to the MS2 aptamer system, a number of techniques to capture specific RNA-protein interactions have been developed including biotin tagging of *in vitro* transcribed RNA (Grabowski and Sharp, 1986), PP7 aptamer tagging (Carey et al., 1983), S1/D8 aptamer tagging (Srisawat and Engelke, 2001) and more recently the CRISPR/Csy4 system (Lee et al., 2013a).

Our repeat tagged MS2-MCP system was composed of two parts: 1) Plasmids expressing either G_4C_2 or TG_3C_2 repeats with 3' 24x MS2 stem loop repeats and 2) a stable HEK293T cell line expressing a tandem MCP-HaloTag fusion. A number of considerations were taken into account when developing this system. Typical MS2 stem-loop sequences are highly repetitive and technically challenging to manipulate. Moreover, the GC rich nature of G_4C_2 or TG_3C_2 repeats make them equally difficult to manipulate and they are notoriously prone to recombination during plasmid propagation. Therefore to circumvent the potential complications associated with tagging repetitive sequences with additional

repetitive stem loop sequences, we took advantage of a recently developed synonymous MS2-stem loop, MBSV5 (Wu et al., 2015). These stem loop sequences were designed via a SELEX based approach to maintain high affinity binding with MCP but minimize repetitive loop structures. We generated $(G_4C_2)_2$ control, $(G_4C_2)_{66}$ pathogenic, $(TG_3C_2)_2$ control, and $(TG_3C_2)_{82}$ pathogenic constructs containing 24x MBSV5 stem loops at the 3' end **(3.5 A,B)**. Next, we developed a stable HEK293T cell line expressing a tandem MCP fused to HaloTag (tdMCP-Halo). The MS2-MCP system is most commonly used to visualize single molecule RNAs in the cytoplasm thus most MCP proteins are engineered to contain a nuclear localization sequence (NLS) in order to promote cytoplasmic clearance of unbound MCP proteins and reduce background signal in the cytoplasm **(Fig 3.6A)**. However, repeat containing RNA form intranuclear foci which are thought to be potential sites of RNA toxicity. We found that inclusion of an NLS on MCP-HaloTag resulted in excess background signal in the nucleus and we could not visualize distinct RNA foci (data not shown). Therefore, in order to promote both nuclear and cytoplasmic distribution of the MCP-HaloTag while minimizing nuclear background signal, we added a strong nuclear export signal and a weak nuclear localization signal. Fluorescence microscopy revealed diffuse nuclear and cytoplasmic distribution of tdMCP-HaloTag **(Fig 3.6D).** Transient expression of MS2-tagged G_4C_2 repeats in the stable MCP-HaloTag HEK293T cell line and subsequent visualization by addition of the Halo binding dye, TMR, demonstrated that only G_4C_2 repeats of pathogenic length formed distinct intranuclear foci, similar to what is detected by fluorescent *in situ* hybridization of G_4C_2 foci **(Fig 3.5 C, D)**

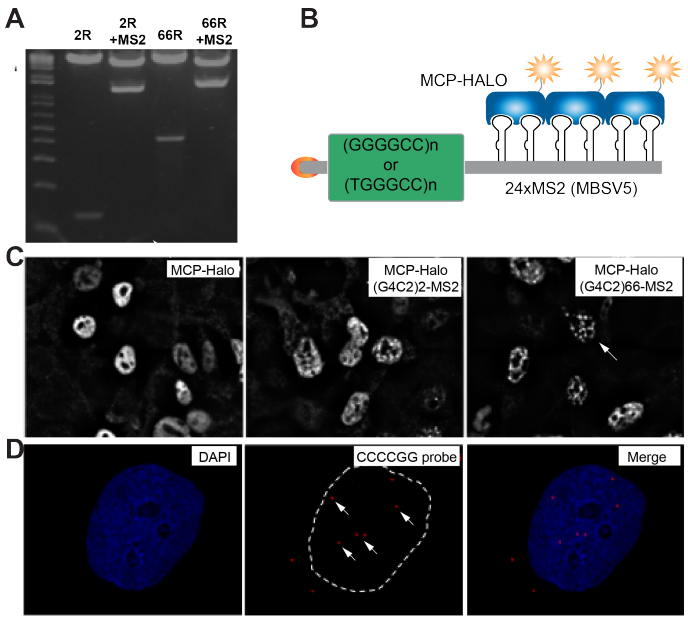


Figure 3.5: MS2-tagged RNA affinity pull-down (MS2-TRAP) assay to identify RNA binding proteins interacting with G4C2 or TG3C2 repeat RNAs.
(A) Restriction digest of (G_4C_2)-MS2 constructs. **(B)** Schematic of (G_4C_2)-MS2/MCP-HaloTag system. **(C)** MCP-HaloTag was transiently transfected into HEK293T cells alone or with (G_4C_2)-MS2 constructs. Live-cell visualization of HaloTag using the TMRdirect ligand show distinct formation of RNA foci (arrow) in the 66 repeat constructs but not the 2R constructs. **(D)** RNA fluorescent *in situ* hybridization using an LNA probe targeting the (G_4C_2) transcript in c9ALS/FTD fibroblast identify RNA foci (arrows).

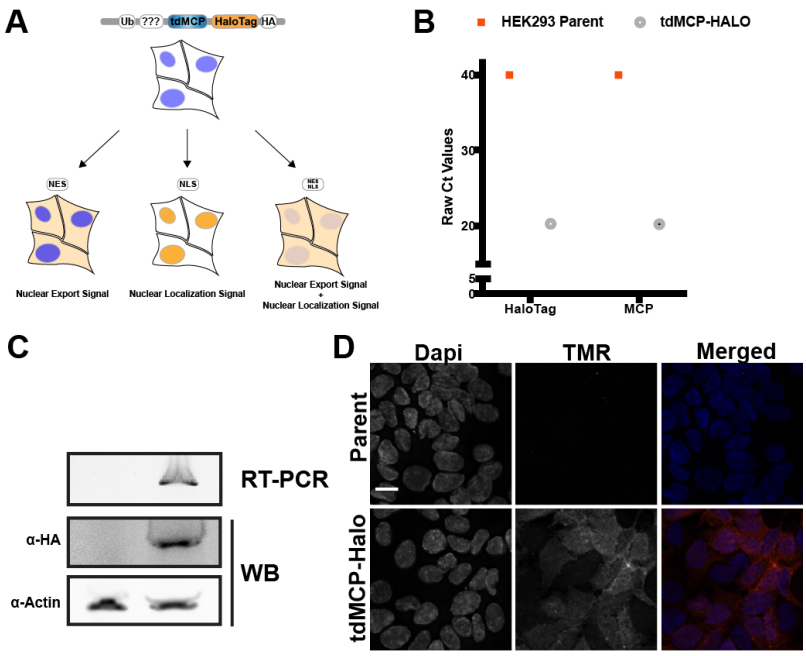


Figure 3.6: Development and characterization of a stable MCP HEK293T cell line. **(A)** Schematic illustrating localization signal considerations for MCP construct. **(B)** qPCR using Taqmans targeting either the HaloTag sequence or MCP sequence in the parental HEK293T and stable TdMCP-Halo HEK293T cell line. No amplification in parent HEK293; Ct values set at 40 for graphical purposes. **(C)** RT-PCR of HaloTag sequence and immunoblots for indicated proteins reveals expression of TdMCP-HaloTag. **(D)** Fluorescence microscopy for the HaloTag binding dye, TMR demonstrates that the TdMCP-HaloTag protein localizes to both the nucleus and cytoplasm.

CHAPTER 4. GENERAL DISCUSSION

4.1 Summary

In the preceding chapters I sought to provide mechanistic insight into the hexanucleotide repeat expansions disorders, c9ALS/FTD and SCA36. More specifically, I hoped to gain a better understanding of how proposed gain-of-function toxic mechanisms may contribute to neurodegeneration in these two disorders. The genetic mutations underlying these two clinically distinct disorders are highly similar in sequence and location. Therefore, I hypothesized that a comprehensive and comparative analysis of molecular phenotypes in c9ALS/FTD and SCA36 would provide a unique opportunity to gain a thorough understanding of common and disease-specific pathomechanisms in these two disorders. In **Chapter 2** I demonstrated for the first time that the SCA36 associated TG_3C_2 repeat expansion can be translated into both sense and antisense dipeptide repeats. Furthermore, I provided evidence from patient brain tissue that chimeric, aggregate-prone DPRs are produced specifically in c9ALS/FTD, which may explain the divergent pathology between SCA36 and c9ALS/FTD In **Chapter 3** I demonstrated that the c9ALS/FTD and SCA36 repeat expansions interact with shared and unique subsets of RNA-binding proteins, and I developed an *in vitro* model to further study RNA-protein interactions in a disease relevant context. Additionally, I used targeted and whole transcriptome sequencing to understand which molecular pathways are disrupted in the CNS of patients with repeat expansion disorders. Future studies I intend to pursue will further two key findings from this thesis and be focused on: 1) defining the role of chimeric dipeptide repeats in the pathogenesis of c9ALS/FTD and related repeat expansions and 2) defining the role of RNA-mediated toxicity in the development of TDP-43 pathology in c9ALS/FTD. I believe this comparative approach to study repeat expansions that are similar in sequence but result in vastly different clinical presentations will aid in our understanding of other microsatellite expansion disorders. Lastly, I hope the knowledge

generated from this thesis will prove useful in the development of therapeutic strategies to combat these devastating disorders.

4.2 The role of DPRs in the pathogenesis of c9ALS/FTD and SCA36

The discovery that microsatellite repeat expansions in non-coding regions of a gene can undergo non-canonical translation to produce homopolymeric peptides (Zu et al., 2011) prompted significant interest in whether these peptides could contribute to neurodegeneration. Indeed, subsequent studies have provided evidence that homopolymeric peptides are produced and elicit toxicity in several microsatellite expansion disorders (Banez-Coronel et al., 2015; Mizielinska et al., 2014; Sellier et al., 2017; Zu et al., 2017). In c9ALS/FTD, a number of mechanisms by which dipeptide repeats induce neurotoxicity have been proposed including inhibition of the ubiquitin-proteasome system (May et al., 2014), disruption of nucleocytoplasmic transport (Jovicic et al., 2015; Shi et al., 2017), impairment of stress granule dynamics (Boeynaems et al., 2017; Lee et al., 2016a; Zhang et al., 2018) and more recently mitochondrial dysfunction (Choi et al., 2019). Furthermore, long-term CNS expression of poly(GA), -(GR), or -(PR) in mice result in behavioral and motor impairments. Nonetheless, the overt toxicity associated with DPRs in model systems contrast to observations made in human c9ALS/FTD patients. While it is possible (and perhaps even likely) that DPRs contribute in some manner to the vulnerability of certain neuronal and/or non-neuronal populations, their exact role in the pathogenesis of ALS and/or FTD remain elusive. Here, I discuss a number of points that I believe are important in the consideration of dipeptide repeat toxicity in c9ALS/FTD and in totality suggest that DPRs do not directly promote neurodegeneration in c9ALS/FTD.

DPR burden and neurodegeneration: Dipeptide repeats, especially poly(GA) and poly(GP), have been detected throughout the CNS in c9ALS/FTD including in the frontal cortex, cerebellum, motor cortex, and hippocampus (Al-Sarraj et al., 2011; Cooper-Knock et al., 2012; Troakes et al., 2012; Vatsavayai et al., 2019). Although they have been detected, DPR inclusions are rare in the spinal cord (Gomez-Deza et al., 2015); likewise in this study we found that poly(GP) levels in the spinal cord of c9ALS/FTD cases were comparable to the background signal detected in controls (see **Chapter 2**, Fig 2.2E). It is possible that the lack of DPRs in the spinal cord of patients could be attributed to the significant loss of motor neurons at autopsy. In contrast to the spinal cord, in CNS regions unaffected in c9ALS/FTD (e.g. the granule layer of the cerebellum) the abundance of DPRs, particularly poly(GA) and poly(GP), is striking. It is not unreasonable, therefore, to speculate that the cerebellum and cerebellar associated function would be significantly affected if DPR inclusions alone were sufficient for neurodegeneration.

Relevance of arginine-rich DPRs: The arginine rich DPRs, poly(GR) and poly(PR) are undoubtedly toxic in model systems (Kwon et al., 2014; Mizielinska et al., 2017; Zhang et al., 2018; Zhang et al., 2019). However, both DPRs, particularly the antisense poly(PR), are rare relative to poly(GA) and poly(GP). Moreover, in this thesis (see **Chapter 3**, Fig 3.1A) we provide evidence that the antisense DPR, poly(PR), is also present in SCA36, a clinically distinct disorder from c9ALS/FTD. Furthermore, previous studies have not been able to correlate poly(GR) or poly(PR) to neurodegeneration (Davidson et al., 2016; Mackenzie et al., 2015; Schludi et al., 2015), although a recent report demonstrated that poly(GR) correlates with neurodegeneration in the specific clinicopathological subtype, FTLD-MND; however, this correlation was only seen in the hippocampus (Sakae et al., 2018). Despite being toxic in model systems, the relatively low abundance of arginine rich

DPRs in patient brains, the fact that poly(PR) is also found in SCA36, and the lack of a direct correlation of arginine rich DPRs with neurodegeneration call into question the contribution of these DPRs in disease pathogenesis.

DPRs and TDP-43 pathology: While the exact role of TDP-43 dysfunction in ALS and FTD pathogenesis remains elusive, TDP-43 pathology is a hallmark of the majority of ALS cases and common FTD subtypes, including c9ALS/FTD. Previous studies have demonstrated that TDP-43 pathology correlates with neurodegeneration and clinical phenotypes (Davidson et al., 2016). Although one study demonstrated that c9ALS/FTD-associated DPRs can cause TDP-43 mislocalization in *Drosophila* (Solomon et al., 2018), previous work in mice suggest that poly(GA) (Zhang et al., 2016), poly(GR) (Zhang et al., 2018) , and poly(PR) (Zhang et al., 2019) alone cannot promote the formation of TDP-43 inclusions. Additionally in this thesis, I demonstrate that the presence of either poly(GP) or poly(PR) alone is insufficient to cause TDP-43 pathology in SCA36. Together these finding suggest that DPRs do not cause TDP-43 pathology.

Insight from atypical C9orf72 cases: A number of unique cases associated with the *C9orf72* repeat expansion have been reported that provide unique insight into putative pathomechanisms. One such case describes a patient who had abundant DPR pathology and RNA foci but no apparent TDP-43 pathology in a part of temporal lobe that had been surgically resected five year prior to disease onset (Vatsavayai et al., 2016). Interestingly, at the time of autopsy, the patient had developed TDP-43 pathology. This case demonstrated that DPR pathology was present in the CNS before onset of disease and preceded TDP-43 pathology. Furthermore, it seems plausible that DPR pathology is present in early life. Other neurodegenerative disease pathologies such as TDP-43, Tau,

or Aß pathology are dependent on some upstream dysfunction occurring prior to aggregation. The aggregate prone nature of DPRs, specifically, poly(GA) suggests that production of these DPRs from *C9orf72* RNA would be the only requirement for them to aggregate, thus it is likely that DPR pathology occurs in early life. A second intriguing case reported a healthy individual who harbored only 70 G_4C_2 repeats but had robust DPR pathology and RNA foci throughout the CNS(McGoldrick et al., 2018). The patient's daughter presented with ALS and harbored ~1750 G_4C_2 repeats. The authors argue that these findings do not negate the toxicity of DPRs but suggest that DPRs alone are insufficient to cause neurodegeneration.

Individual vs mixed DPR populations: To date, the majority of studies have focused on investigating mechanisms of toxicity associated with one DPR species at a time. However, it is well established that more than one DPR inclusion can exist within the same cell. Therefore, it will be necessary to see how interactions among DPRs could modulate their toxicity. **A key finding from my thesis work is that dipeptide repeats are not homogenous throughout the peptide and may in fact form chimeric proteins consisting of more than one dipeptide repeat, e.g. GAGP**. In **Chapter 2** I demonstrate that chimeric GAGP DPRs shift the solubility of poly(GP) from soluble to insoluble, and result in poly(GP) inclusions, similar to observed poly(GP) neuropathology in c9ALS/FTD. Investigations into the toxicity of poly(GP) have only considered soluble poly(GP), however my findings suggest that chimeric species may modulate the toxicity of both poly(GA) or poly(GP). The implications for chimeric DPRs are potentially more important in the case of GAGR chimeric DPRs. Our finding of chimeric GAGP DPRs, supports the idea that other chimeric DPR species may exist, though this has not yet been experimentally verified. Given the robust toxicity associated with poly(GR) it stands to

reason that changes in poly(GR) localization or solubility could modify GR associated toxicity.

4.3 The argument for RNA-mediated toxicity in c9ALS/FTD

Toxicity resulting from repeat containing RNA is a likely culprit in many microsatellite expansion disorders. The archetypical example is that of myotonic dystrophy type 1. Characteristic of many microsatellite expansion disorders and inherent to arguments in favor of repeat toxicity is the formation of intranuclear RNA inclusions (RNA "foci") (La Spada and Taylor, 2010). In the initial report describing the *C9orf72* mutation, RNA foci were reported (DeJesus-Hernandez et al., 2011) and a number of follow up studies have thus focused on understanding how toxic G_4C_2 containing RNA results in neurodegeneration. The fundamental hypothesis is that repeat containing RNA sequester RNA-binding proteins, effectively resulting in their loss of function; however, to date, findings from various studies have been contradictory (Haeusler et al., 2016). This may be due to the fact that while RNA foci are found in both affected and non-affected regions of the CNS (DeJesus-Hernandez et al., 2017), the toxicity of repeat containing RNAs has not been considered in the context of disease relevant cell types. In **Chapter 3**, I provided preliminary data to show that including only those RNA-protein interactions that are plausible in a given cell type based on transcriptome-wide expression profiles, narrows down potential hits. While I only cross-referenced interactors with iPS-MN RNA expression, a better approach would be to compare to the proteome of a given a cell type. Nonetheless, whether considering the transcriptome or the proteome to cross-reference to, identification of RNA-protein interactions is limited to the assay (e.g. protein array vs. MS2) and/or cell-type (e.g. HEK293T) used. A better approach would be to identify RNA-

protein interactions directly in iPS-derived motor neurons through the use of RNA tagging (e.g. MS2) or RNA targeting systems (e.g. Cas13a). Additionally, the prevailing thought is that RNA toxicity is mediated through formation of RNA foci, however soluble repeat containing RNA species have not been considered. I believe moving forward, studies should consider, for example, sequestered RNA binding proteins in nuclear (RNA foci) and cytoplasmic fractions (presumably soluble RNA species).

Converging evidence suggests that dysfunction in RNA metabolism and processing plays a key role in ALS pathogenesis. Mutations in a number of RNA-binding proteins have been linked to ALS including variants in *TARDBP*, *FUS*, and *HNRNPA2B1*. Furthermore, I believe the role of RNA-mediated toxicity is further supported by data generated in this thesis as well as from previous reports in *C9orf72* mouse models. An intriguing finding from this study is that TDP-43 pathology, a hallmark of c9ALS/FTD, is absent in SCA36 patients. The highly similar hexanucleotide repeat sequences in these two disorders reside on different genes, and thus expression differences could contribute to the selective vulnerability of certain neuronal populations. However, we have found that *NOP56* expression is higher than *C9orf72* in all regions surveyed. Additionally, *NOP56* is essential for ribosomal biogenesis and thus it is likely expressed in the majority of neuronal and non-neuronal populations. In support of an RNA over DPR mode of toxicity are findings associated with AAV mouse models. Expression of the c9ALS/FTD-associated expanded G_4C_2 repeat in mice results in TDP-43 pathology (Chew et al., 2019; Chew et al., 2015) while expression of DPRs alone are insufficient (Zhang et al., 2018; Zhang et al., 2016; Zhang et al., 2019), suggesting that TDP-43 dysfunction in c9ALS/FTD results from RNA-mediated toxicity, from a combination of various DPRs, or from some combination of these mechanisms. While the exact role of TDP-43 pathology in the pathogenesis of ALS

remains to be determined, the majority of cases exhibit this pathological hallmark and TDP-43 pathology is correlated with neurodegeneration. Moving forward, an important question to shed light on in the field is why TDP-43 pathology occurs in ALS and FTD.

4.4 Concluding Remarks

In the conclusion of Jean Cruveilhier's 1853 report of progressive muscular atrophy/ALS he poses the question "*What is the cause of the atrophy of the anterior spinal nerve roots*" (Cruveilhier, 1853; Drouin et al., 2016)? Despite more than 100 years of efforts since to answer this question, we are only incrementally closer to understanding what molecular mechanism underlies degeneration of motor neurons. In this thesis, I hoped to shed insight into the pathogenesis of ALS, and by extension FTD and SCA36, by comparing molecular mechanisms associated with genetic forms of these disorders. I demonstrated that neuropathological hallmarks of the most common genetic form of ALS and FTD are also present in a seemingly disparate disorder, SCA36. Intriguingly these neuropathological hallmarks, while composed of the same proteins, differ in their solubility and aggregation propensity in the human brain. This comparative approach allowed me to demonstrate that the divergence in the observed pathology between these two disorders results from the production of chimeric dipeptide repeats in c9ALS/FTD. The implications of chimeric DPRs in disease pathogenesis are far-reaching, as our current understanding associated with DPR pathology and toxicity is based on studies using single DPR species. It remains to be seen if these chimeric DPRs modify (by either mitigating or exacerbating) the toxicity associated with DPRs. Similarly, comparison of c9ALS/FTD and SCA36 repeat RNA-protein interactions yielded insight into shared and unique repeat RNA-protein

interactions. Ultimately, I hope that this thesis proves the utility of a comprehensive and comparative approach to elucidating relevant mechanisms in neurological disorders.

CHAPTER 5. METHODS AND MATERIALS

Lysate preparation

Adherent cells (HEK293T, fibroblast, iPS, iPS-motor neurons) were directly lysed in ice-cold co-IP buffer (50mM Tris–HCl, pH 7.4, 300mM NaCl, 1% Triton X-100, 5mM EDTA, 2% SDS) with 1x HALT protease and phosphatase inhibitors (Pierce). Lysates were sonicated at 25% amplitude for 3 cycles of 5 seconds on/5 seconds off. Cells in suspension (lymphoblastoid cell lines, 3D forebrain organoids) were first pelleted by centrifugation for 5 minutes at 300xg and subsequently lysed as was done for adherent cells.

Immunoblotting

For western blots, protein lysates were prepared in 4x loading buffer, heat-denatured at 70^{o}C for 15 minutes. Samples were normalized to equal amounts and resolved on 10% Bis-Tris gels (Thermo Fisher), transferred to 0.44 μm PVDF membranes, and blocked in Licor Odyssey blocking buffer (Licor) for 1 hour. The membrane was incubated overnight at 4^{o}C with primary antibodies diluted in blocking buffer. Secondary antibodies were diluted in blocking buffer and applied to membrane for 1 hour at R.T. Primary antibodies used in immunoblotting experiments included anti-pTDP43 (1:3000, Cosmo Bio), anti-TDP (1:1000, Protein Tech), anti-FLAG (1:1000, Sigma), anti-HA (1:1000, Cell Signaling), and anti-cMyc (1:1000, Roche) For dot blots, 1.5 μL of urea soluble protein fractions were applied to a pre-wetted PVDF membrane and allowed to air dry for ~1.5 hours at R.T. The membrane was blocked in 5% non-fat dry milk for 1 hour and subsequently probed with purified anti-poly(GP) (Rb9258) overnight at 4^{o}C.

Sequential biochemical fractionation

Tissue lysates were fractionated according to previous published protocols(Neumann et al., 2006). In brief, ~200mg tissue was homogenized in low salt buffer (10mM Tris, 5mM EDTA, 1mM DTT, 10% sucrose, 1x HALT protease/phosphatase inhibitor). Lysates were pelleted at 25,000g for 30 min. Supernatants were collected as the "low salt" fraction. The resulting pellet was solubilized in Triton-X buffer (low salt buffer + 1% triton X-100 + 0.5M NaCl). Lysates were subsequently pelleted at 180,000g for 30 min. Supernatants were collected as the "Triton-X" fraction. The resulting pellet was solubilized in Triton-X buffer + 30% sucrose. Lysates were subsequently pelleted at 180,000g for 30 min. Supernatants from this fraction were not used for analysis. The resulting pellet was solubilized in Sarkosyl buffer (low salt buffer + 1% sarkosyl + 0.5M NaCl). Lysates were subsequently pelleted at 180,000g for 30 min. Supernatants were collected as the "Sark" fraction. The resulting insoluble pellet was resolubilized in 8M urea.

Immunofluorescence

Cells were fixed with 4% paraformaldehyde for 10 minutes at R.T. Cells were washed 3x with 1x PBS and permeabilized for 10 min in 1x PBST (0.1% Triton-X). Cells were then blocked in 5% normal donkey serum (Jackson Labs) for 1 hour at R.T. Cells were incubated with primary antibodies (Table S3) in 5% normal donkey serum overnight at 4^{o}C. The next day cells were washed 3x in PBS and incubated with secondary antibodies (Alexa Fluor, 1:1500) for 1 hour at R.T.

Meso-Scale Discovery (MSD) immunoassay

The expression of poly(GP) in brain, organoid or cell lysates, as well as CSF, was measured blinded using a previously described sandwich immunoassay that utilizes MSD electrochemiluminescence detection technology, and an affinity purified rabbit polyclonal

poly(GP) antibody (Rb9259) for both the capture and the detection antibody(Gendron et al., 2017; Gendron et al., 2015). In brief, lysates were diluted in Tris-buffered saline (TBS) such that all samples of a given type were made up to the same concentration and an equal amount of protein for samples was tested in duplicate wells. CSF samples were tested in duplicate wells using 90 µl per well. Response values corresponding to the intensity of emitted light upon electrochemical stimulation of the assay plate using the MSD QUICKPLEX SQ120 were acquired. These response values were background corrected by subtracting response values from corresponding negative controls (e.g., lysates from tissues or cells lacking a repeat expansion).

Animals

Wildtype pregnant mice (E13) maintained on a C57BL/6 background and C57BL/6 wild type (WT) mice were obtained from *Charles River.* The animal protocol was approved by the Institutional Animal Care and Use Committees of Emory University and complied with the Guide for the Care and Use of Laboratory Animals.

Intracerebroventricular injection (ICV) of AAV in newborn pups:

Neonates at postnatal day 0 after birth are collected from the cage and cryoanesthesized at 0°C for ca.5 min until movement stops before injection. A solution of recombinant AAV virus diluted in sterile saline containing the opaque tracer Evans blue (0.1%; to visualize the borders of the lateral ventricle after injection) is injected stereotactically into the ventricles using a 10 µL Nanofil syringe with a 33 gauge needle. The injection site is located 1 mm lateral to the sagittal suture, halfway between lambda and bregma. The needle was held perpendicular to the skull surface during insertion into the lateral ventricle. Once the needle was in place, 1-2 µL of viral solution were injected slowly (33 nl/sec) with a

MicroSyringe Pump Controller into the lateral ventricle. One minute after the injection is complete, the needle was withdrawn slowly and the injection procedure was repeated for the other ventricle. Post injection the pups were placed on another polystyrene weight boat with cage bedding on a warming pad (Deltaphase Isothermal Pad) until they regained normal color and resumed movement. Post recovery, the whole litter was returned to the mother together.

Tissue Processing

Mice were transcardially perfused with ice cold PBS followed by cold 4% Paraformaldehyde (Sigma). The brains were post fixed overnight in PFA followed by cryoprotection in 30% sucrose for ~ 48 hours (brain sank to the bottom)

Immunohistochemistry

Paraffin embedded sections (5µm or 8µm) were deparaffinized in Histo-clear (National Diagnostics) and rehydrated in ethanol. Antigen retrieval was performed in ddH_2O by steam heat for 30 minutes. Endogenous peroxidase activity was quenched using hydrogen peroxide and washed 3x in TBS-T. Tissue sections were blocked using serum-free protein block (Dako) for 1 hour. Primary antibodies were applied for 45 minutes at R.T. and washed 3x in TBS-T. HRP-conjugated secondaries (Dako) were applied for 30 minutes at R.T. Peroxidase labeling was visualized with 3,3'-diaminobenzidine (DAB). Sections were subsequently counterstained with Gill's hematoxylin and blued in Scott's tap water substitute.

Digital Spatial Profiling

An automated setup capable of imaging and sample collection was developed by modifying a standard microscope. For protein detection, a multiplexed cocktail of primary antibodies, each with a unique, UV photocleavable indexing oligo, and/or 1-3 fluorescent markers (antibodies and/or DNA dyes) was applied to a slide-mounted FFPE tissue section.

The tissue slide was placed on the stage of an inverted microscope. A custom gasket was then clamped onto the slide, allowing the tissue to be submerged in 1.5 mL of buffer solution.

The microcapillary tip is connected to a syringe pump primed with buffer solution, allowing for accurate aspiration of small volumes (<2.5 µL). Under the microscope, wide field fluorescence imaging was performed with epi-illumination from visible LED light engine. The tissue area of interest was then located using fluorescence imaging. 20x image corresponds to 650µm x 650µm of tissue area with a CMOS camera. The 20x images were stitched together to yield a high-resolution image of the tissue area of interest. The regions of interest (ROIs) were then selected based on the fluorescence information and sequentially processed by the microscope automation.

The steps performed for each ROI by the microscope automation were as follows: First, the microcapillary tip was washed by dispensing clean buffer out the capillary and into a wash station. Next, the tissue slide was bulk washed by exchanging the buffer solution on the slide via the inlet and outlet wash ports on the gasket clamp. The microcapillary tip was then moved into position directly over the ROI with a distance of 50 µm from the tissue. The local region of tissue around the ROI was washed by dispensing 100 µL of buffer solution from the microcapillary. Then, the area of tissue within the ROI was selectively illuminated with UV light to release the indexing oligos. UV LED light was collimated to

be reflected from the DMD surface into the microscope objective and focused at the sample tissue. Each micro mirror unit in the DMD corresponds to ~1μm^2 area of sample and reflects the UV light in controlled pattern based on the ROI selection in the image. Following each UV illumination cycle, the eluent was collected from the local region via microcapillary aspiration and transferred to an individual well of a microtiter plate or strip tubes. Once all ROIs were processed, indexing oligos were hybridized to NanoString optical barcodes for *ex situ* digital counting and subsequently analyzed with an nCounter® Analysis System.

Genomic DNA (gDNA) extraction and repeat-primed PCR (RP-PCR)

Genomic DNA was extracted from patient cerebellum using the Puragene Kit (Qiagen) as per the manufacturer's protocol. RP-PCR was performed as previously described(DeJesus-Hernandez et al., 2011) with slight modifications. PCR reactions were performed in 50uL total reaction volumes containing 100ng sample gDNA, 1x SuperFi Buffer (Invitrogen), 0.2mM dNTPs (Invitrogen), 0.5μM primer mix, 1x SuperFi GC enhancer, and 0.02 U/μL Platinum SuperFi DNA polymerase (Invitrogen) PCR cycling conditions were as follows: 95ºC for 2 minutes, 34 cycles of 98ºC for 10 seconds, 64.7ºC for 10 seconds, 72ºC for 40 seconds, followed by a final elongation step at 72ºC for 5 minutes.

RNA extraction and quantitative reverse transcription with PCR (RT-PCR)

RNA was extracted using the Quick RNA kit (Zymo Research) with a combined on-column DNase I digestion step. Adherent cells were directly lysed in the culture dish and RNA extraction proceeded per the manufacture's protocol. For RNA extraction from post-mortem tissue, ~30mg of tissue was homogenized in lysis buffer using a bullet blender

tissue homogenizer (Next Advance). RNA lysates were cleared by spinning samples at 10,000g for 1 minute. Cleared lysates were used for RNA extraction as per the manufacture's protocol. cDNA was obtained via RT-PCR using the High Capacity cDNA Reverse Transcription Kit (Thermo).

To quantify relative mRNA expression for *C9orf72* and *NOP56*, qPCR was peformed for each sample using Taqman gene expression assays (Thermo) on a Quantstudio 6 Flex system (Applied Biosystems). For qPCR analysis of tissue samples, the following FAM labelled Taqman assays were used Hs00376619 (*C9orf72*) and Hs0197340 (*NOP56*). Relative tissue RNA expression was normalized to the geometric mean Ct values of *UBE2D2* and *Cyc1*, endogenous reference transcripts previously demonstrated to be stable across human brain regions in neurodegenerative diseases(Rydbirk et al., 2016). For qPCR analysis of ASO treated fibroblast a custom FAM labelled Taqman assay spanning the exon 1-exon 2 splice junction of *NOP56* was used (AI89MBZ). Relative RNA expression was normalized to *RPLP0*.

RNA sequencing

For RNA sequencing of iPS-derived neurons, strand specific poly-A selected libraries were constructed (Kapa) and sequenced on an Illumina NextSeq500 sequencer with 30 million reads per replicate (2 x 150bp). For RNA sequencing of cerebellar cortex samples, poly-A selected libraries were constructed (NEB) and sequenced on an Illumina Novaseq sequencer with 80 million pair-end reads per replicate (2 x 150bp). Transcript abundances were quantified as transcripts per million (TPMs) and then collapsed to gene abundances using Salmon. Alternative splicing events were identified based on RNA-seq coverage using rMATS and PSI values of differential splicing were obtained. Stringent cutoffs for

differential splicing events were set to FDR < 0.05, Inclusion level difference > -0.1, and min counts ≥ 20. Splicing scatter plots were generated using R-studio and Sashimi plots were generated for the visualization of splicing events. Pathway enrichment of genes in RNAseq were performed on gene ontology (GO) terms using Gorilla.

Generation of induced pluripotent stem cells (iPS)

iPSCs were generated as previously described(Holler et al., 2016) using the Cytotune 2.0 Kit (Life Technologies) per the manufacturer's protocol. In brief, early passage fibroblast (<P10) were grown to \ in fibroblast medium consisting of 10% ES-qualified FBS (Life Technologies), 0.1 mM NEAA, 55μM β-mercaptoethanol, high glucose DMEM with Glutamax (Life Technologies). On Day 0, fibroblasts were transduced with Sendai virus encoding KOS, hc-Myc, hKlf4, each at a MOI of 5. Cells were fed with fibroblast medium every other day for one week. On day 7, cells were passaged onto vitronectin (Life Technologies) coated dishes at a density of 250,000 to 500,000 cells/well. Beginning on day 8, cells were fed every day in Essential 8 medium (Life Technologies). iPS colonies were manually picked and transferred to a dish coated with either vitronectin or Matrigel (BD). iPSCs were maintained on Matrigel coated dishes and mTesR1 medium (Stem Cell Technologies). iPSCs were passaged every 5-7 days using ReLeSR.

Generation of 3D forebrain organoid

iPSCs colonies were detached from Matrigel coated plates with collagenase (1mg/ml; Invitrogen) treatment for 1hr and suspended in EB medium, comprising of FGF-2-free hESC medium supplemented with Dorsomorphin (2μM; Tocris) and A-83 (2μM; Tocris), in non-treated polystyrene plates for 4 days with a daily medium change. On days 5-6, half of the medium was replaced with induction medium consisting of DMEM:F12, N2

Supplement (1X; Invitrogen), 10µg/ml Heparin (Sigma), NEAA (1X), Glutamax (1X), WNT-3A (4 ng/ml; R&D Systems), CHIR99021 (1µM; Cellagentech), and SB-431542 (1µM; Cellagentech). On day 7, organoids were embedded in Matrigel (Corning) and continued to grow in induction medium for 6 more days. On day 14, embedded organoids were mechanically dissociated from Matrigel by pipetting up and down onto the plate with a 5ml pipette tip. Typically, 10 – 20 organoids were transferred to each well of a 12-well spinning bioreactor (SpinΩ)(Qian et al., 2016) containing differentiation medium, consisting of DMEM:F12, N2 (1X), B27 (1X; Invitrogen), 2-Mercaptoenthanol (100µM), NEAA (1X), Insulin (2.5µg/ml; Sigma). At day 71, differentiation medium was exchanged with maturation medium, consisting of Neurobasal (Invitrogen), B27 (1X), 2-Mercaptoenthanol (100 µM), 0.2mM Ascorbic Acid (Sigma), 20 ng/ml BDNF (Peprotech), 20ng/ml GDNF (Peprotech), 1 ng/ml TFGβ (Peprotech), and 0.5 mM cAMP (Sigma). Media was changed every other day until day 105 when organoids were collected for protein lysates.

iPS-motor neuron differentiation

Prior to differentiation, iPS colonies were treated with 10µM ROCK inhibitor, Y-27632 (Stem Cell Technologies), for ~1 hour. iPSCs were treated with Accutase (Stem Cell Technologies) for ~8 min to obtain a single cell suspension. Cells were spun out of Accutase and resuspended in N2B27 differentiation medium (1:1 Advanced DMEM-F12/Neurobasal, 1x N2, 1x B27, 0.2% Penstrep, 1x Glutamax, 110µM β-mercaptoethanol) and seeded in 10cm Ultra-Low Attachment dishes (Corning) in order to form embryoid bodies. Cells were maintained as embryoid bodies throughout the differentiation and were fed every 2 days. For the first 2 days, the differentiation medium contained 3µM

CHIR99021 (Stem Cell Technologies), 10μM SB431542 (Stemgent), 5μM DMH1 (Tocris), and 10μM Y-27632 (Tocris). Starting on day 2, 1μM Retinoic Acid (Sigma) and 500nM Smoothened Agonist (Millipore) were added to the medium and Y-27632 was removed. On Day 8, CHIR99021 was removed from the medium. On Day 8, SB and DMH1 were removed from the medium. On day 15, 10μM DAPT (Tocris) was added. 10ng/mL BDNF (Peprotech) and 10ng/mL GDNF (Peprotech) were added to the media on day 16. On day 20, embryoid bodies were disassociated to single cells using papain/DNase (Worhtington Bio) and plated on polyornithine/laminin (3.3 μg/mL) coated glass coverslips or cell culture plates. Disassociated neurons were maintained in N2B27 maintainence medium (Neurobasal, 1x N2, 1x B27 plus, 0.2% Penstrep, 1x Glutamax, 110μM β-mercaptoethanol) supplemented with 10ng/mL BDNF and 10ng/mL GDNF. Neurons were refed every 2-3 days.

Statistics

All data is presented as mean ± standard error of the mean (SEM). Data was analyzed using unpaired, two-tailed, Student's *t*-test, Kruskal-Wallis analysis of variance, one-way ANOVA followed by Dunn's *post hoc* or Dunnett's *post hoc* analysis, or two-way ANOVA followed by Sidak's *post hoc* analysis (GraphPad Prism 7). Data obtained from poly(GP) immunoassays were considered to be non-normally distributed. Data from qPCR was assumed to be normally distributed. Statistical analysis of qPCR data was performed on delta Ct values.

Antibodies

Table 5.1: Antibodies used in this study

Antibody	Dilution	Experiment	Form	Company	Catalog #	Notes
poly(GP)	1:5000	IHC	Sera	Millipore	ABN455	
poly(GP)	1:1000	Dotblot	AP	Petrucelli Lab	Rb9258	
poly(PR)	1:50	IHC	AP	Petrucelli Lab	Rb8736	See Ref [31]
pTDP-43	1:3000	IB/IHC	Sera	Cosmo Bio	TIP-PTD-P02	1:1 glycerol
TDP-43	1:1000	IB/IHC	AP	Protein Tech	10782-2	
p62	1:250	IHC	AP	BD Biosciences	610832	
FLAG	1:1000	IB/IF	AP	Sigma	F1804	1:1 glycerol
cMyc	1:1000	IB/IF	AP	Roche	11667149001	1:1 glycerol
HA	1:1000	IB/IF	AP	CST	3724	
GFP	1:3000	IB/IF	AP	Clonetech	632381	1:1 glycerol
GAPDH	1:3000	IB	AP	CST	5174	
Islet1	1:250	IF	Sera	DSHB	40.2D6	
Tuj1	1:1500	IF	AP	Covance	MRB-435P	
Oct4	1:400	IF	AP	CST	2840	
SSEA3/4	1:500	IF	AP	CST	4755	
Nanog	1:200	IF	AP	R&D Systems	AF1997	1:1 glycerol
Sox2	1:200	IF	AP	Thermo	PA1-16968	
Tra1-60	1:150	IF	AP	Millipore	MAB4360	
Tra1-81	1:150	IF	AP	Millipore	MAB4381	

AP=affinity purified, IB=immunoblot, IHC=immunohistochemistry, IF=immunofluorescence

CHAPTER 6. REFERENCES

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

Alizadeh, N., Salimi, A., and Hallaj, R. (2017). Hemin/G-Quadruplex Horseradish Peroxidase-Mimicking DNAzyme: Principle and Biosensing Application. Advances in biochemical engineering/biotechnology.

Amick, J., Roczniak-Ferguson, A., and Ferguson, S.M. (2016). C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. Mol Biol Cell *27*, 3040-3051.

Aran, F.A. (1850). Recherches sur une maladie non encore décrite du systéme musculaire (Atrophie musculaire progressive). Arch Gén Méd *24*, 172-214.

Arias, M., Arias-Rivas, S., Blanco-Arias, P., Dapena, D., Vazquez, F., and Rossi, M. (2008). SCA from the Costa da Morte: A new SCA. Description of the phenotype. Neurologia *23*, 628-629.

Arias, M., Garcia-Murias, M., and Sobrido, M.J. (2017). Spinocerebellar ataxia 36 (SCA36): <<Costa da Morte ataxia>>. Neurologia *32*, 386-393.

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

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

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

Barber, S.C., and Shaw, P.J. (2010). Oxidative stress in ALS: key role in motor neuron injury and therapeutic target. Free Radic Biol Med *48*, 629-641.

Barboi, A.C. (2000). Cerebellar Ataxia. JAMA Neurology *57*, 1525-1527.

Bardwell, V.J., and Wickens, M. (1990). Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. Nucleic acids research *18*, 6587-6594.

Batra, R., and Lee, C.W. (2017). Mouse Models of C9orf72 Hexanucleotide Repeat Expansion in Amyotrophic Lateral Sclerosis/ Frontotemporal Dementia. Frontiers in cellular neuroscience *11*, 196-196.

Bell, C. (1824). An Exposition of the Natural Nervous System of the Nerves of the Human Body (London: Spottiswoode, A.R.).

Bell, C. (1830). The Nervous System of the Human Body (London: Rees, Orme, Brown, and Green).

Bizarro, J., Charron, C., Boulon, S., Westman, B., Pradet-Balade, B., Vandermoere, F., Chagot, M.-E., Hallais, M., Ahmad, Y., Leonhardt, H.*, et al.* (2014). Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. *207*, 463-480.

Bjorkoy, G., Lamark, T., and Johansen, T. (2006). p62/SQSTM1: a missing link between protein aggregates and the autophagy machinery. Autophagy *2*, 138-139.

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

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.-P., Hudson, T.*, et al.* (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell *68*, 799-808.

Brown, R.H., and Al-Chalabi, A. (2017). Amyotrophic Lateral Sclerosis. *377*, 162-172.

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

Carey, J., Cameron, V., de Haseth, P.L., and Uhlenbeck, O.C. (1983). Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. Biochemistry *22*, 2601-2610.

Charcot, J.M. (1865). Sclérose des cordons latéraux de la moelle épiniére chez une femme hystérique atteinte de contracture permanente des quatre membres. Bull Soc Med, 24-35.

Charcot, J.M., Joffroy, A. (1869). Deux cas d'atrophie musculaire pro-gressive avec lésions de la substance grise et de faisceaux antérolatéraux de la moelle épiniére. Arch Physiol Norm Pathol, 1: 354-367; 352:628-649; 353:744-757.

Chen, S.-B., Hu, M.-H., Liu, G.-C., Wang, J., Ou, T.-M., Gu, L.-Q., Huang, Z.-S., and Tan, J.-H. (2016). Visualization of NRAS RNA G-Quadruplex Structures in Cells with an

Engineered Fluorogenic Hybridization Probe. Journal of the American Chemical Society *138*, 10382-10385.

Cheng, X., Liu, X., Bing, T., Cao, Z., and Shangguan, D. (2009). General Peroxidase Activity of G-Quadruplex−Hemin Complexes and Its Application in Ligand Screening. Biochemistry *48*, 7817-7823.

Chew, J., Cook, C., Gendron, T.F., Jansen-West, K., Del Rosso, G., Daughrity, L.M., Castanedes-Casey, M., Kurti, A., Stankowski, J.N., Disney, M.D.*, et al.* (2019). Aberrant deposition of stress granule-resident proteins linked to C9orf72-associated TDP-43 proteinopathy. Mol Neurodegener *14*, 9.

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

Choi, S.Y., Lopez-Gonzalez, R., Krishnan, G., Phillips, H.L., Li, A.N., Seeley, W.W., Yao, W.D., Almeida, S., and Gao, F.B. (2019). C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. Nat Neurosci *22*, 851-862.

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

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

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

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

Cruveilhier, J. (1853). Sur la paralysie musculaire progressive atrophique. Archives Générales de Médecine *1*, 561-603.

David, G., Abbas, N., Stevanin, G., Dürr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E.*, et al.* (1997). Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nature Genetics *17*, 65-70.

Davidson, Y., Robinson, A.C., Liu, X., Wu, D., Troakes, C., Rollinson, S., Masuda-Suzukake, M., Suzuki, G., Nonaka, T., Shi, J.*, et al.* (2016). Neurodegeneration in

frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. Neuropathol Appl Neurobiol *42*, 242-254.

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

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

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

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

Donnelly, C.J., Zhang, P.W., Pham, J.T., Heusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M.*, et al.* (2013b). RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. Neuron *80*, 415-428.

Drouin, E., Drouin, A.-S., and Péréon, Y. (2016). Cruveilhier versus Charcot. The Lancet Neurology *15*, 362.

Duchenne, G. (1847). Recherche faites à l'aide du galvanisme sur l'état de contractilité et de la sensibilité électromusculaire dans les paralysies des membres supérieurs. Compt Rennd Acad Sci *29*, 667-670.

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

Ferrari, R., Kapogiannis, D., Huey, E.D., and Momeni, P. (2011). FTD and ALS: a tale of two diseases. Curr Alzheimer Res *8*, 273-294.

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

Fujigasaki, H., Verma, I.C., Camuzat, A., Margolis, R.L., Zander, C., Lebre, A.-S., Jamot, L., Saxena, R., Anand, I., Holmes, S.E.*, et al.* (2001). SCA12 is a rare locus for autosomal dominant cerebellar ataxia: A study of an Indian family. *49*, 117-121.

García-Murias, M., Quintáns, B., Arias, M., Seixas, A.I., Cacheiro, P., Tarrío, R., Pardo, J., Millán, M.J., Arias-Rivas, S., Blanco-Arias, P.*, et al.* (2012). 'Costa da Morte' ataxia is spinocerebellar ataxia 36: clinical and genetic characterization. Brain *135*, 1423-1435.

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

Gendron, T.F., Chew, J., Stankowski, J.N., Hayes, L.R., Zhang, Y.J., Prudencio, M., Carlomagno, Y., Daughrity, L.M., Jansen-West, K., Perkerson, E.A.*, et al.* (2017). Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. Sci Transl Med *9*.

Gendron, T.F., van Blitterswijk, M., Bieniek, K.F., Daughrity, L.M., Jiang, J., Rush, B.K., Pedraza, O., Lucas, J.A., Murray, M.E., Desaro, P.*, et al.* (2015). Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. Acta Neuropathol *130*, 559-573.

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. Hum Genet *135*, 851-867.

Gijselinck, I., Engelborghs, S., Maes, G., Cuijt, I., Peeters, K., Mattheijssens, M., Joris, G., Cras, P., Martin, J.J., De Deyn, P.P.*, et al.* (2010). Identification of 2 Loci at chromosomes 9 and 14 in a multiplex family with frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Arch Neurol *67*, 606-616.

Gitler, A.D., and Tsuiji, H. (2016). There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS. Brain Res *1647*, 19-29.

Goetz, C.G. (2000). Amyotrophic lateral sclerosis: Early contributions of Jean-Martin Charcot. Muscle & Nerve *23*, 336-343.

Gomes-Pereira, M., Foiry, L., Nicole, A., Huguet, A., Junien, C., Munnich, A., and Gourdon, G. (2007). CTG trinucleotide repeat "big jumps": large expansions, small mice. PLoS Genet *3*, e52.

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

Grabowski, P., and Sharp, P. (1986). Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. *233*, 1294-1299.

Grigg, J.C., Shumayrikh, N., and Sen, D. (2014). G-Quadruplex Structures Formed by Expanded Hexanucleotide Repeat RNA and DNA from the Neurodegenerative Disease-Linked C9orf72 Gene Efficiently Sequester and Activate Heme. PLOS ONE *9*, e106449.

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

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.-S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R.*, et al.* (2014b). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *advance online publication*.

Haeusler, A.R., Donnelly, C.J., and Rothstein, J.D. (2016). The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. Nat Rev Neurosci *17*, 383-395.

Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P.J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *57*, 127-130.

Hammond, W. (1871). A Treatise on the Diseases of the Nervous System.

Hammond, W. (1881). A Treatise on the Diseases of the Nervous System, Seventh edn.

Ho, T.H., Charlet-B, N., Poulos, M.G., Singh, G., Swanson, M.S., and Cooper, T.A. (2004). Muscleblind proteins regulate alternative splicing. The EMBO journal *23*, 3103-3112.

Holler, C.J., Taylor, G., McEachin, Z.T., Deng, Q., Watkins, W.J., Hudson, K., Easley, C.A., Hu, W.T., Hales, C.M., Rossoll, W.*, et al.* (2016). Trehalose upregulates progranulin expression in human and mouse models of GRN haploinsufficiency: a novel therapeutic lead to treat frontotemporal dementia. Mol Neurodegener *11*, 46.

Ikeda, Y., Ohta, Y., Kobayashi, H., Okamoto, M., Takamatsu, K., Ota, T., Manabe, Y., Okamoto, K., Koizumi, A., and Abe, K. (2012). Clinical features of SCA36: a novel spinocerebellar ataxia with motor neuron involvement (Asidan). Neurology *79*, 333-341.

Ikeda, Y., Ohta, Y., Kurata, T., Shiro, Y., Takao, Y., and Abe, K. (2013). Acoustic impairment is a distinguishable clinical feature of Asidan/SCA36. J Neurol Sci *324*, 109-112.

Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. Nature *546*, 243-247.

Jiang, J., Zhu, Q., Gendron, T.F., Saberi, S., McAlonis-Downes, M., Seelman, A., Stauffer, J.E., Jafar-Nejad, P., Drenner, K., Schulte, D.*, et al.* (2016). Gain of Toxicity from

ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. Neuron *90*, 535-550.

Jovicic, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S.B., Paul, J.W., 3rd, Sun, S., Herdy, J.R., Bieri, G.*, et al.* (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat Neurosci *18*, 1226-1229.

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F.*, et al.* (2008). TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet *40*, 572-574.

Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W., and Swanson, M.S. (2003). A muscleblind knockout model for myotonic dystrophy. Science *302*, 1978-1980.

Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I.*, et al.* (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nature Genetics *8*, 221-228.

Kenna, K.P., van Doormaal, P.T.C., Dekker, A.M., Ticozzi, N., Kenna, B.J., Diekstra, F.P., van Rheenen, W., van Eijk, K.R., Jones, A.R., Keagle, P.*, et al.* (2016). NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nature Genetics *48*, 1037.

Klockgether, T., Mariotti, C., and Paulson, H.L. (2019). Spinocerebellar ataxia. Nature Reviews Disease Primers *5*, 24.

Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., Habu, T., Liu, W., Okuda, H., and Koizumi, A. (2011). Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet *89*, 121-130.

Koide, R., Kobayashi, S., Shimohata, T., Ikeuchi, T., Maruyama, M., Saito, M., Yamada, M., Takahashi, H., and Tsuji, S. (1999). A Neurological Disease Caused By an Expanded CAG Trinucleotide Repeat in The TATA-Binding Protein Gene: A New Polyglutamine Disease? Human Molecular Genetics *8*, 2047-2053.

Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W., and Ranum, L.P.W. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nature Genetics *21*, 379-384.

Kurland, L.T., and Mulder, D.W. (1955). Epidemiologic Investigations of Amyotrophic Lateral Sclerosis. 2 Familial Aggregations Indicative of Dominant Inheritance Part II *5*, 249-249.

Kwiatkowski, T.J., Bosco, D.A., LeClerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T.*, et al.* (2009). Mutations in the

FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. *323*, 1205-1208.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science *345*, 1139-1145.

La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. Nature reviews Genetics *11*, 247-258.

Le Ber, I., Camuzat, A., Berger, E., Hannequin, D., Laquerriere, A., Golfier, V., Seilhean, D., Viennet, G., Couratier, P., Verpillat, P.*, et al.* (2009). Chromosome 9p-linked families with frontotemporal dementia associated with motor neuron disease. Neurology *72*, 1669-1676.

Lee, H.Y., Haurwitz, R.E., Apffel, A., Zhou, K., Smart, B., Wenger, C.D., Laderman, S., Bruhn, L., and Doudna, J.A. (2013a). RNA-protein analysis using a conditional CRISPR nuclease. Proc Natl Acad Sci U S A *110*, 5416-5421.

Lee, K.H., Zhang, P., Kim, H.J., Mitrea, D.M., Sarkar, M., Freibaum, B.D., Cika, J., Coughlin, M., Messing, J., Molliex, A.*, et al.* (2016a). C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell *167*, 774-788 e717.

Lee, Y.-B., Baskaran, P., Gomez-Deza, J., Chen, H.-J., Nishimura, A.L., Smith, B.N., Troakes, C., Adachi, Y., Stepto, A., Petrucelli, L.*, et al.* (2017). C9orf72 poly GA RAN-translated protein plays a key role in amyotrophic lateral sclerosis via aggregation and toxicity. Human Molecular Genetics *26*, 4765-4777.

Lee, Y.B., Chen, H.J., Peres, J.N., Gomez-Deza, J., Attig, J., Stalekar, M., Troakes, C., Nishimura, A.L., Scotter, E.L., Vance, C.*, et al.* (2013b). Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic. Cell Rep.

Lee, Y.C., Tsai, P.C., Guo, Y.C., Hsiao, C.T., Liu, G.T., Liao, Y.C., and Soong, B.W. (2016b). Spinocerebellar ataxia type 36 in the Han Chinese. Neurol Genet *2*, e68.

Levine, T.P., Daniels, R.D., Gatta, A.T., Wong, L.H., and Hayes, M.J. (2013). The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics (Oxford, England) *29*, 499-503.

Li, W., Li, Y., Liu, Z., Lin, B., Yi, H., Xu, F., Nie, Z., and Yao, S. (2016). Insight into G-quadruplex-hemin DNAzyme/RNAzyme: adjacent adenine as the intramolecular species for remarkable enhancement of enzymatic activity. Nucleic Acids Research *44*, 7373-7384.

Lillo, P., and Hodges, J.R. (2009). Frontotemporal dementia and motor neurone disease: Overlapping clinic-pathological disorders. Journal of Clinical Neuroscience *16*, 1131-1135.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P.W. (2001). Myotonic Dystrophy Type 2 Caused by a CCTG Expansion in Intron 1 of ZNF9. *293*, 864-867.

Liu, W., Ikeda, Y., Hishikawa, N., Yamashita, T., Deguchi, K., and Abe, K. (2014). Characteristic RNA foci of the abnormal hexanucleotide GGCCUG repeat expansion in spinocerebellar ataxia type 36 (Asidan). Eur J Neurol *21*, 1377-1386.

Liu, Y., Pattamatta, A., Zu, T., Reid, T., Bardhi, O., Borchelt, D.R., Yachnis, A.T., and Ranum, L.P. (2016). C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. Neuron *90*, 521-534.

Lopez-Gonzalez, R., Lu, Y., Gendron, T.F., Karydas, A., Tran, H., Yang, D., Petrucelli, L., Miller, B.L., Almeida, S., and Gao, F.B. (2016). Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron *92*, 383-391.

Luty, A.A., Kwok, J.B.J., Thompson, E.M., Blumbergs, P., Brooks, W.S., Loy, C.T., Dobson-Stone, C., Panegyres, P.K., Hecker, J., Nicholson, G.A.*, et al.* (2008). Pedigree with frontotemporal lobar degeneration--motor neuron disease and Tar DNA binding protein-43 positive neuropathology: genetic linkage to chromosome 9. BMC neurology *8*, 32-32.

Lykke-Andersen, S., Ardal, B.K., Hollensen, A.K., Damgaard, C.K., and Jensen, T.H. (2018). Box C/D snoRNP Autoregulation by a cis-Acting snoRNA in the NOP56 Pre-mRNA. Mol Cell *72*, 99-111 e115.

MacDonald, M.E., Ambrose, C.M., Duyao, M.P., Myers, R.H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S.A., James, M., Groot, N.*, et al.* (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell *72*, 971-983.

Mackenzie, I.R., Frick, P., Grasser, F.A., Gendron, T.F., Petrucelli, L., Cashman, N.R., Edbauer, D., Kremmer, E., Prudlo, J., Troost, D.*, et al.* (2015). Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathol *130*, 845-861.

Mackenzie, I.R., and Neumann, M. (2016). Molecular neuropathology of frontotemporal dementia: insights into disease mechanisms from postmortem studies. J Neurochem *138 Suppl 1*, 54-70.

Majounie, E., Renton, A.E., Mok, K., Dopper, E.G., Waite, A., Rollinson, S., Chio, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J.*, et al.* (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol *11*, 323-330.

Marat, A.L., Dokainish, H., and McPherson, P.S. (2011). DENN domain proteins: regulators of Rab GTPases. The Journal of biological chemistry *286*, 13791-13800.

Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L.*, et al.* (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nature Genetics *26*, 191-194.

Matsuzono, K., Imamura, K., Murakami, N., Tsukita, K., Yamamoto, T., Izumi, Y., Kaji, R., Ohta, Y., Yamashita, T., Abe, K.*, et al.* (2017). Antisense Oligonucleotides Reduce RNA Foci in Spinocerebellar Ataxia 36 Patient iPSCs. Mol Ther Nucleic Acids *8*, 211-219.

May, S., Hornburg, D., Schludi, M.H., Arzberger, T., Rentzsch, K., Schwenk, B.M., Grasser, F.A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J.*, et al.* (2014). C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol *128*, 485-503.

McGoldrick, P., Zhang, M., van Blitterswijk, M., Sato, C., Moreno, D., Xiao, S., Zhang, A.B., McKeever, P.M., Weichert, A., Schneider, R.*, et al.* (2018). Unaffected mosaic C9orf72 case: RNA foci, dipeptide proteins, but upregulated C9orf72 expression. Neurology *90*, e323-e331.

Miller, J.W., Urbinati, C.R., Teng-umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., and Swanson, M.S. (2000). Recruitment of human muscleblind proteins to (CUG) expansions associated with myotonic dystrophy. The EMBO Journal *19*, 4439.

Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C.E., Clayton, E.L., Devoy, A., Moens, T., Norona, F.E., Woollacott, I.O.C., Pietrzyk, J.*, et al.* (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science *345*, 1192-1194.

Mizielinska, S., Ridler, C.E., Balendra, R., Thoeng, A., Woodling, N.S., Grasser, F.A., Plagnol, V., Lashley, T., Partridge, L., and Isaacs, A.M. (2017). Bidirectional nucleolar dysfunction in C9orf72 frontotemporal lobar degeneration. Acta Neuropathol Commun *5*, 29.

Mori, K., Lammich, S., Mackenzie, I.R., Forne, I., Zilow, S., Kretzschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M.*, et al.* (2013a). hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol *125*, 413-423.

Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C.*, et al.* (2013b). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science *339*, 1335-1338.

Morita, M., Al-Chalabi, A., Andersen, P.M., Hosler, B., Sapp, P., Englund, E., Mitchell, J.E., Habgood, J.J., de Belleroche, J., Xi, J.*, et al.* (2006). A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. *66*, 839-844.

Mulder, D.W., Kurland, L.T., Offord, K.P., and Beard, C.M. (1986). Familial adult motor neuron disease: amyotrophic lateral sclerosis. Neurology *36*, 511-517.

Naunyn (1873). Ueber Heredität der progressiven Muskelatrophie. Berliner klinische Wochenschrift, 42-43.

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

Nicolas, A., Kenna, K.P., Renton, A.E., Ticozzi, N., Faghri, F., Chia, R., Dominov, J.A., Kenna, B.J., Nalls, M.A., Keagle, P.*, et al.* (2018). Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron *97*, 1268-1283.e1266.

Nielsen, H., and Wernersson, R. (2006). An overabundance of phase 0 introns immediately after the start codon in eukaryotic genes. BMC Genomics *7*, 256-256.

O'Rourke, J.G., Bogdanik, L., Muhammad, A.K., Gendron, T.F., Kim, K.J., Austin, A., Cady, J., Liu, E.Y., Zarrow, J., Grant, S.*, et al.* (2015). C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. Neuron *88*, 892-901.

O'Rourke, J.G., Bogdanik, L., Yanez, A., Lall, D., Wolf, A.J., Muhammad, A.K., Ho, R., Carmona, S., Vit, J.P., Zarrow, J.*, et al.* (2016). C9orf72 is required for proper macrophage and microglial function in mice. Science *351*, 1324-1329.

Obayashi, M., Stevanin, G., Synofzik, M., Monin, M.L., Duyckaerts, C., Sato, N., Streichenberger, N., Vighetto, A., Desestret, V., Tesson, C.*, et al.* (2015). Spinocerebellar ataxia type 36 exists in diverse populations and can be caused by a short hexanucleotide GGCCTG repeat expansion. J Neurol Neurosurg Psychiatry *86*, 986-995.

Ohta, Y., Hayashi, T., Nagai, M., Okamoto, M., Nagotani, S., Nagano, I., Ohmori, N., Takehisa, Y., Murakami, T., Shoji, M.*, et al.* (2007). Two cases of spinocerebellar ataxia accompanied by involvement of the skeletal motor neuron system and bulbar palsy. Internal medicine (Tokyo, Japan) *46*, 751-755.

Orr, H.T., Chung, M.-y., Banfi, S., Kwiatkowski, T.J., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P.W., and Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genetics *4*, 221-226.

Osler, W. (1880). On heredity in progressive muscular atrophy as illustrated in Farr family of Vermont. Arch Med, 316-320.

Paulson, H. (2018). Repeat expansion diseases. Handb Clin Neurol *147*, 105-123.

Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet *6*, 729-742.

Pearson, J.P., Williams, N.M., Majounie, E., Waite, A., Stott, J., Newsway, V., Murray, A., Hernandez, D., Guerreiro, R., Singleton, A.B.*, et al.* (2011). Familial frontotemporal dementia with amyotrophic lateral sclerosis and a shared haplotype on chromosome 9p. J Neurol *258*, 647-655.

Peters, O.M., Cabrera, G.T., Tran, H., Gendron, T.F., McKeon, J.E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J.*, et al.* (2015). Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. Neuron *88*, 902-909.

Pilotto, F., and Saxena, S. (2018). Epidemiology of inherited cerebellar ataxias and challenges in clinical research. Clinical and Translational Neuroscience *2*, 2514183X18785258.

Porta, S., Kwong, L.K., Trojanowski, J.Q., and Lee, V.M. (2015). Drosha inclusions are new components of dipeptide-repeat protein aggregates in FTLD-TDP and ALS C9orf72 expansion cases. J Neuropathol Exp Neurol *74*, 380-387.

Prudencio, M., Belzil, V.V., Batra, R., Ross, C.A., Gendron, T.F., Pregent, L.J., Murray, M.E., Overstreet, K.K., Piazza-Johnston, A.E., Desaro, P.*, et al.* (2015). Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat Neurosci *18*, 1175-1182.

Pulst, S.-M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.-N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A.*, et al.* (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nature Genetics *14*, 269-276.

Qian, X., Nguyen, H.N., Song, M.M., Hadiono, C., Ogden, S.C., Hammack, C., Yao, B., Hamersky, G.R., Jacob, F., Zhong, C.*, et al.* (2016). Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. Cell *165*, 1238-1254.

Radcliffe, C.B., and Clarke, J.L. (1862). An Important Case of Paralysis and Muscular Atrophy, with Disease of the Nervous Centres. Br Foreign Med Chir Rev *30*, 215-225.

Rau, F., Freyermuth, F., Fugier, C., Villemin, J.P., Fischer, M.C., Jost, B., Dembele, D., Gourdon, G., Nicole, A., Duboc, D.*, et al.* (2011). Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. Nat Struct Mol Biol *18*, 840-845.

Reddy, K., Zamiri, B., Stanley, S.Y., Macgregor, R.B., Jr., and Pearson, C.E. (2013). The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J Biol Chem *288*, 9860-9866.

Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L.*, et al.* (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72*, 257-268.

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X.*, et al.* (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature *362*, 59-62.

Ross, C.A., and Tabrizi, S.J. (2011). Huntington's disease: from molecular pathogenesis to clinical treatment. Lancet Neurol *10*, 83-98.

Rydbirk, R., Folke, J., Winge, K., Aznar, S., Pakkenberg, B., and Brudek, T. (2016). Assessment of brain reference genes for RT-qPCR studies in neurodegenerative diseases. Sci Rep *6*, 37116.

Sakae, N., Bieniek, K.F., Zhang, Y.-J., Ross, K., Gendron, T.F., Murray, M.E., Rademakers, R., Petrucelli, L., and Dickson, D.W. (2018). Poly-GR dipeptide repeat polymers correlate with neurodegeneration and Clinicopathological subtypes in C9ORF72-related brain disease. Acta Neuropathologica Communications *6*, 63.

Sarto, E., Magri, S., Mariotti, C., Nanetti, L., Gellera, C., Di Bella, D., and Taroni, F. (2013). SCA36 molecular analysis in patients with spinocerebellar ataxia. Paper presented at: International Conference on Spinocerebellar Degenerations (Paris, France).

Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., Takahashi, M., Matsuura, T., Flanigan, K.M., Iwasaki, S.*, et al.* (2009). Spinocerebellar Ataxia Type 31 Is Associated with “Inserted” Penta-Nucleotide Repeats Containing (TGGAA)_n. The American Journal of Human Genetics *85*, 544-557.

Schludi, M.H., May, S., Grasser, F.A., Rentzsch, K., Kremmer, E., Kupper, C., Klopstock, T., Arzberger, T., and Edbauer, D. (2015). Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol *130*, 537-555.

Sellier, C., Buijsen, R.A., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziane, H., Vincent, A.*, et al.* (2017). Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron *93*, 331-347.

Sellier, C., Campanari, M.L., Julie Corbier, C., Gaucherot, A., Kolb-Cheynel, I., Oulad-Abdelghani, M., Ruffenach, F., Page, A., Ciura, S., Kabashi, E.*, et al.* (2016). Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. Embo j *35*, 1276-1297.

Seznec, H., Lia-Baldini, A.S., Duros, C., Fouquet, C., Lacroix, C., Hofmann-Radvanyi, H., Junien, C., and Gourdon, G. (2000). Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. Hum Mol Genet *9*, 1185-1194.

Shakkottai, V.G., and Fogel, B.L. (2013). Clinical neurogenetics: autosomal dominant spinocerebellar ataxia. Neurol Clin *31*, 987-1007.

Shi, K.Y., Mori, E., Nizami, Z.F., Lin, Y., Kato, M., Xiang, S., Wu, L.C., Ding, M., Yu, Y., Gall, J.G., *et al.* (2017). Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. Proc Natl Acad Sci U S A *114*, E1111-E1117.

Siddique, T., and Ajroud-Driss, S. (2011). Familial amyotrophic lateral sclerosis, a historical perspective. . *Acta Myol*, 117-120.

Siddique, T., Figlewigz, D.A., Pericak-Vance, M.A., Haines, J.L., Rouleau, G., Jeffers, A.J., Sapp, P., Hung, W.-Y., Bebout, J., McKenna-Yasek, D., *et al.* (1991). Linkage of a Gene Causing Familial Amyotrophic Lateral Sclerosis to Chromosome 21 and Evidence of Genetic-Locus Heterogeneity. *324*, 1381-1384.

Siddique, T., Pericak-Vance, M.A., Brooks, B.R., Roos, R.P., Hung, W.Y., Antel, J.P., Munsat, T.L., Phillips, K., Warner, K., Speer, M., *et al.* (1989). Linkage analysis in familial amyotrophic lateral sclerosis. Neurology *39*, 919-925.

Slobodin, B., and Gerst, J.E. (2011). RaPID: an aptamer-based mRNA affinity purification technique for the identification of RNA and protein factors present in ribonucleoprotein complexes. Methods Mol Biol *714*, 387-406.

Solomon, D.A., Stepto, A., Au, W.H., Adachi, Y., Diaper, D.C., Hall, R., Rekhi, A., Boudi, A., Tziortzouda, P., Lee, Y.-B., *et al.* (2018). A feedback loop between dipeptide-repeat protein, TDP-43 and karyopherin-α mediates C9orf72-related neurodegeneration. Brain *141*, 2908-2924.

Sone, J., Mitsuhashi, S., Fujita, A., Mizuguchi, T., Mori, K., Koike, H., Hashiguchi, A., Takashima, H., Sugiyama, H., Kohno, Y., *et al.* (2019). Long-read sequencing identifies GGC repeat expansion in human-specific NOTCH2NLC associated with neuronal intranuclear inclusion disease. 515635.

Spada, A.R.L., Wilson, E.M., Lubahn, D.B., Harding, A.E., and Fischbeck, K.H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature *352*, 77-79.

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., *et al.* (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science *319*, 1668-1672.

Srisawat, C., and Engelke, D.R. (2001). Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. RNA (New York, NY) *7*, 632-641.

Su, Z., Zhang, Y., Gendron, T.F., Bauer, P.O., Chew, J., Yang, W.Y., Fostvedt, E., Jansen-West, K., Belzil, V.V., Desaro, P., *et al.* (2014). Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. Neuron *83*, 1043-1050.

Sudria-Lopez, E., Koppers, M., de Wit, M., van der Meer, C., Westeneng, H.J., Zundel, C.A., Youssef, S.A., Harkema, L., de Bruin, A., Veldink, J.H., *et al.* (2016). Full ablation

of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. Acta Neuropathol *132*, 145-147.

Sulek, A., Krysa, W., Elert-Dobkowska, E., Rajkiewicz, M., STepniak, I., Rakowicz, M., Rola, R., Dusza-Rowinska, M., and Zaremba, J. (2013). Spinocerebellar ataxias (SCAs) and hereditary spastic paraplegias (HSP) - rare movement disorders prevalence in Poland. Paper presented at: International Conference on Spinocerebellar Degenerations (Paris, France).

Sullivan, P.M., Zhou, X., Robins, A.M., Paushter, D.H., Kim, D., Smolka, M.B., and Hu, F. (2016). The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. Acta Neuropathol Commun *4*, 51.

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

Tian, Y., Wang, J.-L., Huang, W., Zeng, S., Jiao, B., Liu, Z., Chen, Z., Li, Y., Wang, Y., Min, H.-X.*, et al.* Expansion of Human-Specific GGC Repeat in Neuronal Intranuclear Inclusion Disease-Related Disorders. The American Journal of Human Genetics.

Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K.C., Scaglione, K.M., Basrur, V.*, et al.* (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron *78*, 440-455.

Troakes, C., Maekawa, S., Wijesekera, L., Rogelj, B., Siklos, L., Bell, C., Smith, B., Newhouse, S., Vance, C., Johnson, L.*, et al.* (2012). An MND/ALS phenotype associated with C9orf72 repeat expansion: abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline. Neuropathology *32*, 505-514.

Udd, B., and Krahe, R. (2012). The myotonic dystrophies: molecular, clinical, and therapeutic challenges. Lancet Neurol *11*, 891-905.

Valdmanis, P.N., Dupre, N., Bouchard, J.-P., Camu, W., Salachas, F., Meininger, V., Strong, M., and Rouleau, G.A. (2007). Three Families With Amyotrophic Lateral Sclerosis and Frontotemporal Dementia With Evidence of Linkage to Chromosome 9p. JAMA Neurology *64*, 240-245.

Valera, J.M., Diaz, T., Petty, L.E., Quintans, B., Yanez, Z., Boerwinkle, E., Muzny, D., Akhmedov, D., Berdeaux, R., Sobrido, M.J.*, et al.* (2017). Prevalence of spinocerebellar ataxia 36 in a US population. Neurol Genet *3*, e174.

Vance, C., Al-Chalabi, A., Ruddy, D., Smith, B.N., Hu, X., Sreedharan, J., Siddique, T., Schelhaas, H.J., Kusters, B., Troost, D.*, et al.* (2006). Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2,Äì21.3. Brain *129*, 868-876.

Vatsavayai, S.C., Nana, A.L., Yokoyama, J.S., and Seeley, W.W. (2019). C9orf72-FTD/ALS pathogenesis: evidence from human neuropathological studies. Acta Neuropathologica *137*, 1-26.

Vatsavayai, S.C., Yoon, S.J., Gardner, R.C., Gendron, T.F., Vargas, J.N.S., Trujillo, A., Pribadi, M., Phillips, J.J., Gaus, S.E., Hixson, J.D.*, et al.* (2016). Timing and significance of pathological features in C9orf72 expansion-associated frontotemporal dementia. Brain : a journal of neurology *139*, 3202-3216.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P.*, et al.* (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905-914.

Vignaud, A., Ferry, A., Huguet, A., Baraibar, M., Trollet, C., Hyzewicz, J., Butler-Browne, G., Puymirat, J., Gourdon, G., and Furling, D. (2010). Progressive skeletal muscle weakness in transgenic mice expressing CTG expansions is associated with the activation of the ubiquitin-proteasome pathway. Neuromuscular disorders : NMD *20*, 319-325.

Wang, E.T., Cody, N.A.L., Jog, S., Biancolella, M., Wang, T.T., Treacy, D.J., Luo, S., Schroth, G.P., Housman, D.E., Reddy, S.*, et al.* (2012). Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell *150*, 710-724.

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

Wu, B., Miskolci, V., Sato, H., Tutucci, E., Kenworthy, C.A., Donnelly, S.K., Yoon, Y.J., Cox, D., Singer, R.H., and Hodgson, L. (2015). Synonymous modification results in high-fidelity gene expression of repetitive protein and nucleotide sequences. Genes & development *29*, 876-886.

Yang, D., Abdallah, A., Li, Z., Lu, Y., Almeida, S., and Gao, F.-B. (2015). FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. Acta Neuropathologica *130*, 525-535.

Yang, M., Liang, C., Swaminathan, K., Herrlinger, S., Lai, F., Shiekhattar, R., and Chen, J.F. (2016). A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. Sci Adv *2*, e1601167.

Yoshihama, M., Uechi, T., Asakawa, S., Kawasaki, K., Kato, S., Higa, S., Maeda, N., Minoshima, S., Tanaka, T., Shimizu, N.*, et al.* (2002). The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. Genome research *12*, 379-390.

Younis, I., Berg, M., Kaida, D., Dittmar, K., Wang, C., and Dreyfuss, G. (2010). Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. Mol Cell Biol *30*, 1718-1728.

Zamiri, B., Reddy, K., Macgregor, R.B., Jr., and Pearson, C.E. (2014). TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. J Biol Chem *289*, 4653-4659.

Zeng, S., Zeng, J., He, M., Zeng, X., Zhou, Y., Liu, Z., Xia, K., Pan, Q., Jiang, H., Shen, L.*, et al.* (2016). Genetic and clinical analysis of spinocerebellar ataxia type 36 in Mainland China. Clin Genet *90*, 141-148.

Zhang, D., Iyer, L.M., He, F., and Aravind, L. (2012). Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. Frontiers in genetics *3*, 283.

Zhang, N., and Ashizawa, T. (2017). RNA toxicity and foci formation in microsatellite expansion diseases. Current opinion in genetics & development *44*, 17-29.

Zhang, Y.J., Gendron, T.F., Ebbert, M.T.W., O'Raw, A.D., Yue, M., Jansen-West, K., Zhang, X., Prudencio, M., Chew, J., Cook, C.N.*, et al.* (2018). Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. Nat Med *24*, 1136-1142.

Zhang, Y.J., Gendron, T.F., Grima, J.C., Sasaguri, H., Jansen-West, K., Xu, Y.F., Katzman, R.B., Gass, J., Murray, M.E., Shinohara, M.*, et al.* (2016). C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. Nat Neurosci *19*, 668-677.

Zhang, Y.J., Guo, L., Gonzales, P.K., Gendron, T.F., Wu, Y., Jansen-West, K., O'Raw, A.D., Pickles, S.R., Prudencio, M., Carlomagno, Y.*, et al.* (2019). Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. Science *363*.

Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D.W., Amos, C., Dobyns, W.B., Subramony, S.H., Zoghbi, H.Y., and Lee, C.C. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the α1A-voltage-dependent calcium channel. Nature Genetics *15*, 62-69.

Zu, T., Cleary, J.D., Liu, Y., Banez-Coronel, M., Bubenik, J.L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H.B., Yachnis, A.T.*, et al.* (2017). RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. Neuron *95*, 1292-1305 e1295.

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.*, et al.* (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A *108*, 260-265.

Zu, T., Liu, Y., Banez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T.M., Harms, M.B., Falchook, A.E., Subramony, S.H.*, et al.* (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci U S A *110*, E4968-4977.

Zu, T., Pattamatta, A., and Ranum, L.P.W. (2018). Repeat-Associated Non-ATG Translation in Neurological Diseases. Cold Spring Harb Perspect Biol *10*.