Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Emma Davis March 24, 2021

Repeat size dependent-toxicity of dipeptide repeat proteins in C9orf72 ALS/FTD

by

Emma Davis

Dr. Jie Jiang
Adviser

Biology Department

Dr. Jie Jiang

Adviser

Dr. Gary Bassell

Committee Member

Dr. Kate O'Toole

Committee Member

2021

Repeat size dependent-toxicity of dipeptide repeat proteins in C9orf72 ALS/FTD

By

Emma Davis

Dr. Jie Jiang

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Biology

2021

Abstract

Repeat size dependent-toxicity of dipeptide repeat proteins in C9orf72 ALS/FTD
By Emma Davis

Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Degeneration (FTD) are two devastating neurodegenerative diseases with clinical, pathological and genetic overlaps. These two diseases both can be caused by the same most common genetic mutation: a hexanucleotide GGGGCC (G4C2) repeat expansion in the C9orf72 intron region. Toxicity through the translation of certain dipeptide repeat proteins (DPRs), GR and PR, have been shown to likely lead to disease; however, literature has focused on short repeat lengths rather than the long repeat lengths that are seen in patients. Furthermore, the translational mechanism of the repeat translation is relatively unknown, particularly in the antisense direction. In order to gain a better understanding of the translational mechanism of DPRs in C9orf72 ALS/FTD, we first looked at what role the intron plays in the translation of antisense DPRs. We found that when the intron sequence is removed, there is a significant reduction in DPR production, indicating that the intron region is likely involved in antisense DPR translation. Preliminary data suggests that this translation within the intron region is driven by start codons in frame with various DPRs. Due to this start codon driven translation, longer repeats within patients are likely to be completely translated. Comparing the shorter repeats used in literature with longer repeats, we found a change in the localization of repeats and a change in stress granule dynamics, a downstream pathway. Longer GR repeats localized within the cytoplasm, as do patient samples, while shorter GR repeats localized within the nucleolus. Furthermore, shorter PR repeats lead to stress granule formation, while longer PR repeats do not. These discrepancies, along with the insight into the translational mechanism of disease, indicate that researchers may need to reevaluate the conclusions that have been made using short DPRs in their applicability to the actual disease.

Repeat size dependent-toxicity of dipeptide repeat proteins in C9orf72 ALS/FTD

By

Emma Davis

Dr. Jie Jiang

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology Department

2021

Acknowledgements

I would like to thank Dr. Jie Jiang who has provided me with unconditional support, mentorship and advice throughout the three years of working in his lab. I would also like to thank Dr. Janani Parameswaran for her support and guidance throughout my research.

Table of Contents

INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Degeneration (FTD) are two devastating neurodegenerative diseases with clinical, pathological and genetic overlaps (Ling et al., 2013). ALS, also known as Lou Gehrig's disease after the famous New York Yankees baseball player who died from the disease in 1941, causes accelerating upper and lower motor neuron loss, leading to eventual paralysis (Rowland and Shneider, 2001). FTD, caused by frontal and temporal lobe neuron degeneration, can be characterized by an array of symptoms such as personality and behavioral alterations, language dysfunction, and dementia (Olney et al., 2017). Though different in their clinical presentations, these two diseases both can be caused by the same most common genetic mutation: a hexanucleotide GGGGCC (G4C2) repeat expansion in the C9orf72 intron region (DeJesus-Hernandez et al., 2011).

Healthy humans have up to about 30 G4C2 repeats, however, ALS/FTD patients with the C9orf72 mutation (referred to as C9ALS/FTD thereafter) can harbor hundreds to thousands of repeats (**Fig. 1A**). There are three proposed disease mechanisms by which the G4C2 repeat expansions can cause diseases. The first is through the loss of C9orf72 function, whose function is still largely unknown. The expanded repeats prevent the transcription of the C9orf72 gene, leading to C9orf72 protein haploinsufficiency (**Fig. 1B**). Supporting this mechanism, the C9orf72 mRNA is reduced by half in patients carrying the G4C2 repeat expansions. The second

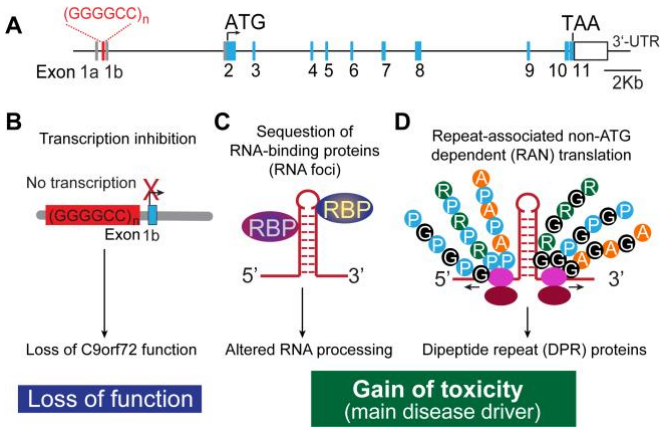

Figure 1. Gain of toxicity from C9orf72 GGGGCC repeat expansions plays a central role in driving ALS/FTD. (A) Schematic representation of the human C9orf72 gene with expanded GGGGCC hexanucleotide repeats. Proposed disease mechanisms include: 1) loss of C9orf72 protein function **(B)**, and/or 2) a gain of toxicity from the sense or antisense repeat-containing RNAs, mediated either by sequestration of RNA binding proteins into RNA foci **(C)** or by production of at least five different aberrant dipeptide repeat (DPR) proteins [GA, GP, GR, PR and PA] through a novel repeat-associated non-AUG-dependent (RAN) translation mechanism **(D)**. Several studies, including my own work, support gain of toxicity as a central component of disease mechanism. Figure adopted from Jie Jiang.

potential disease mechanism is a gain of toxicity from the bi-directionally transcribed repeat-containing RNAs via sequestration of key RNA binding proteins (RBPs) into RNA foci, a structure thought to be formed by the repeat-containing RNAs that form complicated secondary structures and interact with RBPs. The third potential disease mechanism is a gain of toxicity through the production of dipeptide repeat proteins (DPRs) from the intronic repeat sequences via a non-canonical translation mechanism called repeat-associated non-AUG dependent (RAN) translation. At least five different DPRs proteins can be made. More specifically, GA (glycine-alanine) and GR (glycine-arginine) DPRs are translated by the sense G4C2 repeat RNAs, whereas PR (proline-arginine) and PA (proline-alanine) DPRs are translated by the antisense CCCCGG (C4G2) RNAs. GP (glycine-proline) DPRs can be translated by the either strand. Supporting the gain of toxicity mechanism, both sense and antisense RNA foci, as well as the accumulation of five DPR proteins, have been observed in C9ALS/FTD patients.

Of these three potential disease mechanisms, studies have shown that the gain of toxicity from the repeat-containing RNAs is likely to be the main driver of C9ALS/FTD. Expressing either G4C2 or C4G2 repeats in cultured neuronal cells, C. elegans, Drosophila, and mice can produce RNA foci and DPR proteins, and lead to toxicity (Jiang and Ravits, 2019). The gain of toxicity mechanism was further supported by studies expressing individual DPR proteins (discussed below). In contrast, reduced or even complete loss of the C9orf72 gene has been insufficient to cause any ALS/FTD-like behavioral symptoms in mouse models (Koppers et. al, 2015; Burberry et al., 2016; Ugolino et al., 2016). Furthermore, the loss of C9orf72 function is likely not the sole or main disease driver because there have been no mutations identified in ALS/FTD patients that cause a loss of function (Harms et al., 2013).

Whether RNA foci or DPR proteins are the main toxic species is currently debated. Isolating RNA foci-mediated toxicity from the DPR-mediated toxicity is also highly difficult due to their inherent biological makeup of being RNA, and thus are translated into DPRs. Many repeat expansion

disorders are characterized by the presence of RNA foci. One primary hypothesis for how RNA foci cause toxicity is that they sequester key RBPs, thus decreasing their ability to properly function. The RNA-foci mediated toxicity is most heavily researched in Myotonic dystrophy type 1 (DM1), a multisystem disorder that mainly affects skeletal and smooth muscles that is caused by a CUG repeat expansions in the DMPK1 gene (Zhang and Ashizawa, 2017). In this disease, RNA foci are formed from the CUG repeat expanded RNAs in the nucleus (Davis et al., 1997), and sequester the RBP Muscleblind 1-3 (MBNL1-3) (Miller et al., 2000). Both mouse models with expanded CUG repeats and patients with DM1 have shown colocalization of MBNL1 with RNA foci. Further supporting loss of MBNL1 function as disease mechanism, MBNL1 knockout mice recapitulated many DM1 patient phenotypes. Finally, transgenic mice carrying with CUG repeats also exhibited myotonia and myopathy, similarly to patients, (Kanadia et al., 2003), and overexpression of MBNL1 reduces some disease phenotypes (Kanadia et al., 2006). These studies support that MBNL sequestration by RNA foci containing CUG repeat expanded RNAs is a major cause of DM1, and set a foundation for demonstrating RNA foci-mediated toxicity in other repeat expansion disorders.

For C9ALS/FTD, several RBPs have been proposed to be sequestered by either sense or antisense RNA foci. McEachin et al., 2020 cites many RBPs, including ADARB2, Pur-alpha, ALYREF, SRSF1, SRSF2, various hnRNPs, nucleolin, Zfp106, and several paraspeckle proteins such as SFPQ, NONO, RBM14 and FUS (Donnelly et al., 2013; Xu et al., 2013; Cooper-Knock et al., 2014; Conlon et al., 2016; Celona et al., 2017; Mori et al., 2013; Haeusler et al., 2014; Celona et al., 2017; Bajc Česnik et al., 2019). However, there has been insufficient evidence to show that loss of the proposed RBPs can lead to ALS/FTD, nor do replenishing any RBPs rescue toxicity from the C9orf72 repeat-containing RNAs. Thus, whether and how RNA foci observed in C9ALS/FTD patients contribute to disease pathogenesis is not clear.

To determine the toxicity of DPRs, many studies have been carried out by overexpressing different DPRs in varying models, including mice, zebrafish, yeast, mammalian cells and Drosophila

(Jovičić et al., 2015; Wen et al., 2014; Tao et al., 2015; Lee et al., 2016; Zu et al., 2011; Kanekura et al., 2016; Mizielinska et al., 2014; Freibaum et al., 2015; Yang et al., 2015; Swinnen et al., 2018; Ohki et al., 2017; Zhang et al., 2016; Schludi et al., 2017; Zhang et al., 2018; Choi et al., 2019). Codon-optimized DNA sequences that disrupt (G4C2) or (C4G2) repeats were used to ensure that only DPRs will be made, but not RNA foci. Table 1 is a compilation of many of these studies, which provides compelling evidence that certain DPR proteins are toxic. As illustrated in the table, polymeric GR and PR are the most toxic. There are numerous examples in which GR and PR cause cell death and toxicity in cellular models including U2OS cells, cultured human astrocytes, Hek293T cells and iPSC-derived neurons (Kwon et al., 2014; Wen et al., 2014; Tao et al., 2015; Lee et al., 2016; Jiang and Ravits, 2019). In mice, GR and PR led to behavioral phenotypes as those found in patients with ALS or FTD (Zhang et al., 2018; Choi et al., 2019; Hao et al., 2019). While GA also shows toxicity in some cases, PA and GP are unlikely to cause diseases due to their low toxicity in model systems.

How do these repeat-containing RNAs, especially the arginine-containing RAN-translated DPRs, cause neuronal toxicity? Several cellular dysfunctions have been proposed (Jiang and Ravits, 2019). One example of these is the altered stress granule dynamics. Under stress, cytoplasmic ribonucleoprotein (RNP) granules form, sequestering many messenger RNAs (mRNAs), translation initiation factors, and other RNA-binding proteins (Jiang and Ravits, 2019). In this way, only essential proteins that are key for cell survival are translated. Once stress is alleviated, stress granules will dissipate or be cleared by the autophagy pathways. Stress granules are thus vital for cell health and functionality, and any disruption in a cell's ability to make or dissipate stress granules in specific situations can harm the cell. Alterations in stress granule dynamics have been suggested to contribute to ALS/FTD diseases (Li et al., 2013). There has been extensive research on how the arginine-containing DPRs affect stress granule dynamics. GR and PR have been shown to interact with proteins that lead to the disruption of stress granule dynamics (Boeynaems et al., 2017).

			Size of DPRs					
Study	**Model**	**constructs**	**GR**	**PR**	**GA**	**PA**	**GP**	**Measure of toxicity**
Kwon 2014	U2OS/Human astrocyte	peptide	20	20				Survival (ATP)
Zu, 2013	HEK293T	ATG repeats		80			80	Survival (LDH, MTT)
May 2014	**HEK293T**	**codon optimized**	149	175	175	175	80	Survival (LDH)
Lee 2017	HEK293	codon optimized	125	125	125	125	125	Survival (PARP cleavage)
Tao 2015	HEK293/NSC34	codon optimized	30,60	30	30;60	30	30	Survival (PI)
Tao 2015	PHEK293/NSC34	ATG repeats	30,60	30	30;60	30	30	Survival (PI)
Zhang 2014	HEK293T	codon optimized			50			Survival (caspase 3, LDH)
Yamakawa 2014	Neuro-2a	codon optimized	100	100	100	100	100	survival (count)
Lee 2016	Neuro-2a	codon optimized	50	50	50	50	47	Survival (LDH)
Chang 2016	Neuroblastoma	peptide			15			Survival (LDH)
Callister 2016	SH-SY5Y	codon optimized	1136	1100	1020	1024		Electrophysiology
Kanekura 2016	NSC34	peptide	20	20	20			Survival (LDH)
Zhang 2014	Primary neurons	codon optimized			50			Survival (caspase 3, LDH)
Wen 2014	Primary neurons	codon optimized	25-400	25-200	25-400	25-200	25-50	Survival (count)
Gupta 2017	Primary neurons	peptide	20	20				Survival (count)
May 2014	Primary neurons	codon optimized			150			Survival (TUNEL), dendritic morphology
Wen 2014	human iPSNs	codon optimized		50	50			Survival (count)
Lopez-Gonzalez 2016	Human iPSNs	codon optimized	80		80			Survival (TUNEL)
Shi 2018	Human iPSNs	codon optimized	50	50				Survival (count)
Jovicic 2015	yeast	codon optimized	50;100	50	50	50		growth
Wen 2014	Drosophila- Eye (GMR driver); Motor neuron	codon optimized		50	50	50		eye degeneration survival
Mizielinska 2014	Drosophila- Eye (GMR driver); Pan neuron	codon optimized	36;100	36;100	36;100	36;100		eye degeneration survival
Freibaum 2015	Drosophila- Eye (GMR driver);	codon optimized	50		50		47	eye degeneration, survival
Yang 2015	Drosophila- Eye (GMR driver); Pan neuron, Motor neuron	codon optimized	80	80	80			eye degeneration survival
Boeynaems 2016	Drosophila- Eye (GMR driver); Pan neuron, Motor neuron	codon optimized	50	25;50	25;50	25;50		eye degeneration survival
Lee 2016	Drosophila- Eye (GMR driver); Motor neuron	codon optimized	50	50	50	50	47	eye degeneration survival
Baldwin 2016	Drosophila- Motor neuron	codon optimized		36				Mitochondrial transport
Lee 2017	Chick	codon optimized	125	125	125	125	125	TUNEL
Rudich 2017	C elegan	codon optimized	50	50	50	50		motor phenotype
Ohki 2017	Zebrafish	ATG repeats			80			pericardial edema
Swinnen 2018	Zebrafish	codon optimized	50	50	50	50	50	motor axonopathy
Swaminathan 2018	Zebrafish	codon optimized	>100	>1000	>200	>1000		motor phenotype
Zhang 2016	Mouse (AAV)	codon optimized			50			Clinical/pathological
Schludi 2017	Mouse (transgenic)	codon optimized			149			Clinical/pathological
		Consensus	GR	PR	GA	PA	GP	
		Legend	Toxic	mildly toxic	not toxic			

Table 1: A compilation of literature studies on the toxicity of different DPRs of varying lengths. As indicated in the key, red is toxic to the model studied, orange is mildly toxic and green is not toxic. The length of DPR is indicated by the number in each box. As indicated by the coloring, GR and PR are more toxic in model systems than GA, PA or GP. Adapted and edited from (Swinnen et al., 2020).

Furthermore, overexpression of GR and PR increased stress granule formation, which sequester multiple nuclear pore proteins, leading to disrupted nucleocytoplasmic transport (Wen et al., 2014, Zhang, 2018).

Although current research supports that the arginine-containing DPRs, GR and PR, are toxic in model systems by affecting multiple molecular pathways such as alteration in stress granule dynamics, most DPRs that have been studied are "short", and tend to be about 50 to 100 in length. However, as stated previously, C9ALS/FTD patients can have hundreds to thousands of repeats, thus likely producing DPRs with hundreds to thousands of di-peptides. Indeed, while aggregations of GR and PR DPRs in patients are generally cytoplasmic and perinuclear, GR and PR with sizes of 50-100 accumulate in cell nucleolus. The molecular mechanisms of DPR production from the intronic repeat-containing RNAs are recently emerging. Instead of a random process that is independent of the canonical AUG start codon, Green et al. showed of DPRs translated by the sense G4C2 repeat expanded RNAs are strongly facilitated by the intronic sequence in the 5' region of the repeats (Green et al., 2017). More specially, when a near cognate start codon "CUG" was mutated in the intron region, it would lead to a significant reduction of GA-DPRs that are in frame. This suggests that not only is the intron sequence in front of the sense G4C2 repeats involved in DPR translation, but that it is likely there are direct start codons or near cognate start codon beginning such translation. This insight into the potential mechanism of the disease is significant because if this is occurring in patients, and a start codon is initiating the production of DPRs, then it is likely that the entire hundred to thousands of repeats in patients are being translated, rather than the translation beginning at a random point within the sequence. This would produce much longer DPR proteins than those that have been used in existing studies. It is thus important to determine if long DPRs have the same pathogenesis and molecular mechanisms as short 50-100 DPRs.

In my thesis project, I will 1) determine whether intronic sequence in front of the antisense C4G2 repeats play a role in antisense DPR translations and 2) compare how different sizes of DPRs affected cells in order to better understand if the longer DPRs likely produced in patients will have the same effects as short-length DPRs commonly used in existing research.

RESULTS

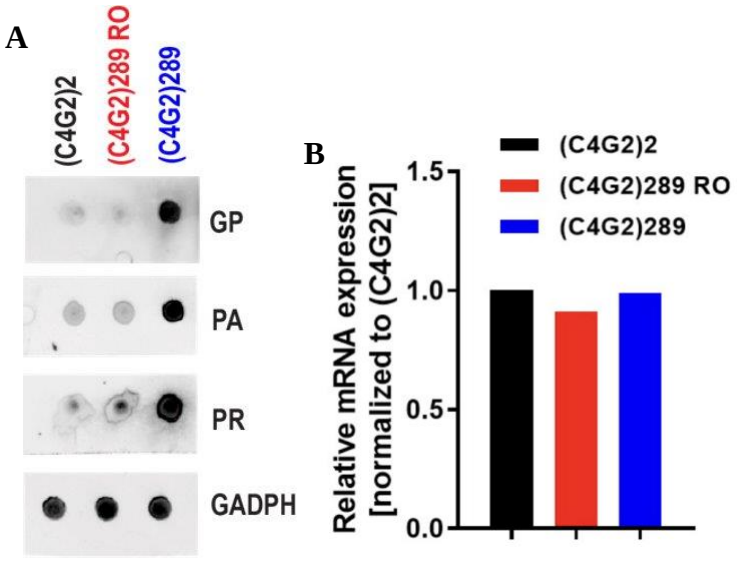

Figure 2. 5' intronic region of C9orf72 antisense drives DPR protein translation. A) A dot plot performed after transfection into Hek293T cells. Darker dots indicate more DPR present, and different antibodies specific to DPRs were used. A negative control with only 2 repeats, (C4G2)2, was also used to compare (C4G2)289 with (C4G2)289 RO. (C4G2)289RO was comparable to the control, while (C4G2)289 had a significantly larger amount of DPRs translated. B) qRT-PCR was used to determine the relative mRNA levels in (C4G2)2, (C4G2)289 and (C4G2)289 RO, showing all are relatively the same. Similar mRNA levels indicates that transfection efficiency is similar throughout, and differences in in DPR levels in Fig. 2A are due to differences in translation of DPRs.

Intron sequence implicated in antisense DPR translation.

As described, the intron sequence plays a role in the translation of DPRs in the sense direction. However, it is currently unknown whether adjacent nucleotide sequences near the antisense C4G2 repeats influence the translation of antisense DPRs. To address this question, it is important to know where and how the antisense transcription starts. Using 5'-RACE, a technique to determine the transcription start sites, recent literature has found that transcription of the C9orf72 antisense RNAs begins at 450 base pairs ahead of the repeats (Zu et al., 2013). In order to understand if it plays a similar role in the translation of antisense DPRs, two constructs were created. One containing the 450 base pairs of the intron sequence ahead of 289 C4G2 repeats, called (C4G2)289, and a second with the intron sequence removed, called (C4G2)289 RO (repeats only). As a control, construct containing 2 C4G2 repeats were made. I transfected Hek293T cells with these constructs, collected proteins and performed a dot plot analysis using antibodies against respective DPRs. Cells expressing (C4G2)289 accumulate significant levels of PR, PA and GP DPRs. However, little to no accumulation of PR, PA and GP DPRs are observed in cells those expressing (C4G2)289 RO(**Fig. 2A**), similar to those expressing (C4G2)2.

The mRNA expression of each constructs was relatively similar (**Fig. 2B**), indicating that this difference in DPR levels observed by dot plot is not due to any discrepancies in the expression at the mRNA level.

Start codons in antisense intron sequence likely initiating translation.

We next wanted to test what specifically in the intron sequence was driving translation. We hypothesized that start codons in frame with DPRs in the intron sequence drive translation, and thus attempted to create constructs with these start codons mutated to compare DPR translation.

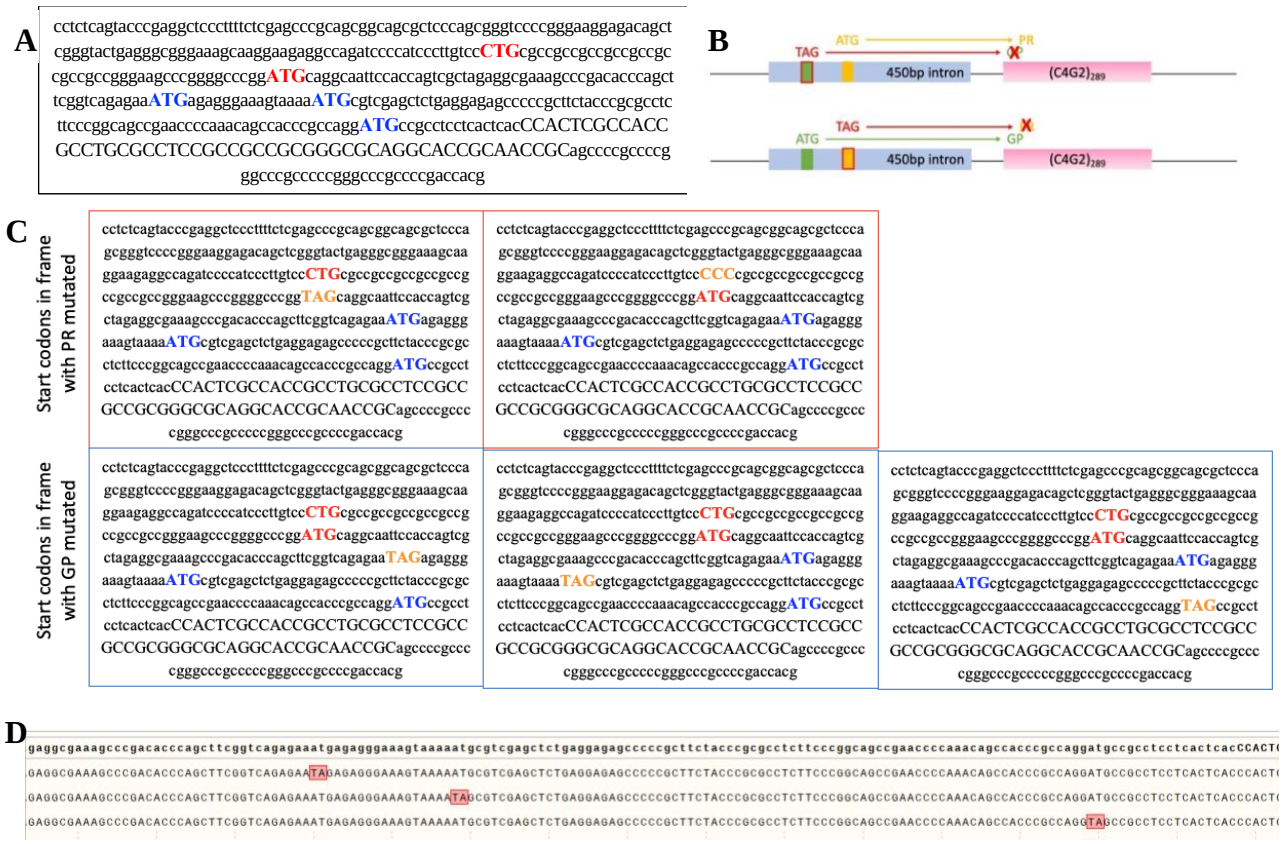

Figure 3. Start codons in C9orf72 intron region identified. A) 450 base pairs of the C9orf72 intron region were sequenced and start codons ATG and CTG were identified. Red indicates start codons in frame with PR frame, blue indicates start codons in frame with GP frame. B) Shows summary of different constructs made with start codons mutated with 73 repeats inserted after intron region. C) TAG and CCC mutations were introduced via site directed mutagenesis in different DPR frames. Orange indicates start codon mutated. D) Sequencing results of 3 vectors in which ATG in frame with GP successfully mutated to TAG after site-directed mutagenesis.

Multiple start codons were identified within the intron region in frame with either the GP or PR DPR reading frame (**Fig. 3A**), and these were mutated (**Fig. 3B, 3C**) to be stop codons**.** We hypothesized that these start codons were likely driving translation of the DPRs, and thus that the constructs with mutated start codons would theoretically have less DPR production. For the start codons in frame with PR, analysis of the level of DPRs comparing constructs with unmutated intron ((C4G2)289 ATG-PR) were compared to mutated ATG to TAG stop codon in frame with PR ((C4G2)289 TAG-PR) (**Fig. 4**). DPR levels were consistent for GP and PA, however, there was a decrease in PR levels of (C4G2)289 TAG-PR compared to (C4G2)289 ATG-PR, indicating that ATG in frame with PR is likely driving some translation of PR-DPRs. Due to timing restrictions resulting from COVID-19, the constructs for the GP start codons are still in progress. However, vectors containing ATG to TAG mutations in frame with GP were successfully created (**Fig. 3D**). More details on the creation of the constructs can be found in the materials and methods section.

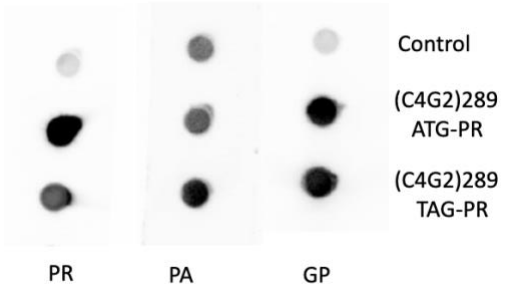

Figure 4. Mutated ATG in frame with PR shows decreased PR-DPR formation. A dot plot performed after transfection into Hek293T cells. Darker dots indicate more DPR present, and different antibodies specific to DPRs were used. A negative control with only 2 repeats, (C4G2)2 was used. Constructs with 289 repeats and 450bp of intron sequence ((C4G2)289 ATG-PR) were compared to mutated ATG in frame with PR ((C4G2)289 TAG-PR). In only PR staining, (C4G2)289 TAG-PR) showed some decreased levels of PR DPR compared to (C4G2)289 ATG-PR).

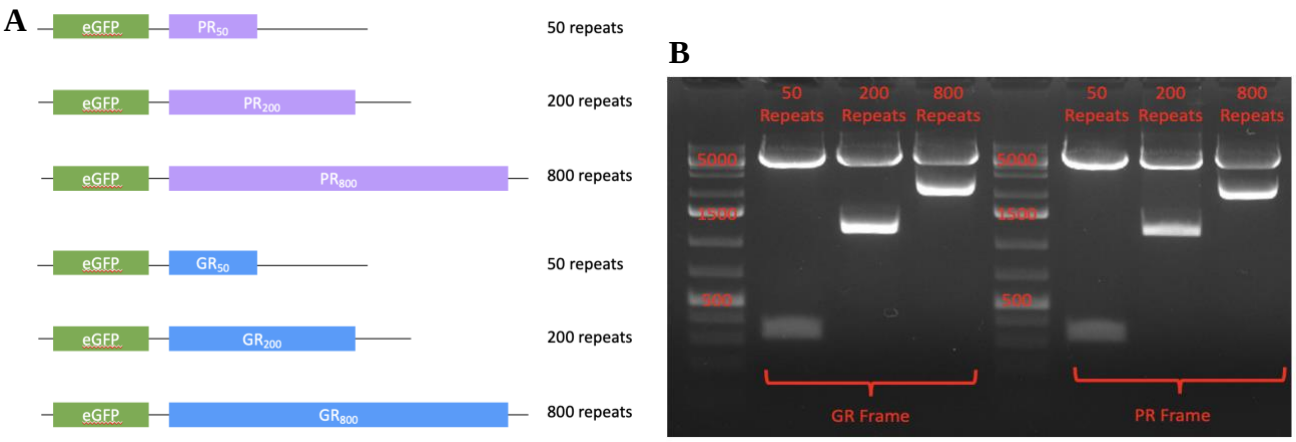

Figure 5. Long repeat constructs confirmed in PR and GR reading frames. A) Graphical representation of constructs built with 50, 200 and 800 repeats in frame with eGFP in both PR and GR reading frames. B) A gel electrophoresis of six constructs made, each increasing in the number of repeats.

Long GR DPRs form cytoplasmic aggregates as in observed in patients.

As outlined in the introduction, C9ALS/FTD patients have hundreds to thousands of repeats and my data and others support that translation for DPRs starts in the intronic region in front of repeats, thus DPRs in patients might also be hundreds to thousands in size. However, existing studies generally use DPRs with 50-100 dipeptides in size, and thus analysis of longer DPRs will likely more closely mimic what is found in patients. This is vital to better understanding the underlying disease mechanisms of C9ALS/FTD. To address this question, codon optimized constructs were created of varying lengths in both the GR and PR reading frames to ensure no RNA foci will be formed. Constructs with 50 ("short"), and 200 and 800 ("long") repeats were successfully created for both GR and PR (**Fig. 5B**). Once constructs were confirmed, they were transfected into Hek293T cells and immunostaining was used to visualize the location of the DPRs. **Fig. 6** shows the comparison of 50, 200 and 800 repeats in the GR frame. Cells transfected with (GR)50 showed a localization of proteins within the nucleolus, which is consistent to what has been observed in literature expressing GR-DPRs with 50-100 repeats in size. However patient samples show aggregation of DPRs within the

cytoplasm, not within the nucleolus (Balendra and Isaacs, 2019). Interestingly, cells transfected with (GR)200 and (GR)800, showed cytoplasmic aggregates, similarly to what is seen in patient samples.

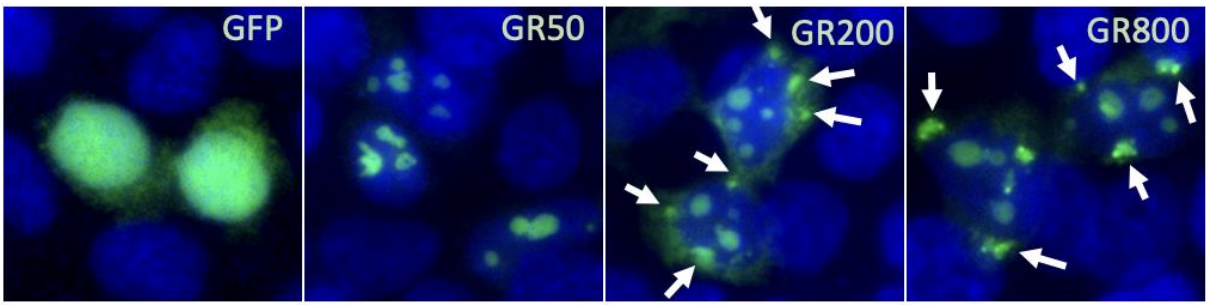

Figure 6. Long GR repeats localize in cytoplasm. Immunostaining images of Hek293T cells transfected with varying GR repeat length constructs. The green indicates eGFP, which is translated at the N-terminus of each construct, indicating translation of the DPRs. A negative control was used with just eGFP. The GR50 shows aggregation of DPRs in the nucleolus, while the GR200 and GR800 both show aggregates in the cytoplasm as well as the nucleolus.

Long PR DPRs do not trigger stress granule formation.

After determining localization of repeats, we next wanted to analyze stress granule dynamics of cells with longer repeats compared to shorter repeats. Immunostaining of transfected Hek293T cells was again performed, and BG3P was used as a marker for stress granules. Similarly to literature, PR(50) had a large number of stress granules in transfected cells. However, (PR)200 and (PR)800 had a significant reduction in the number of stress granules, comparable to the negative control (**Fig. 7**). These findings suggest that shorter PR repeats lead to stress granule formation, while longer PR repeats do not.

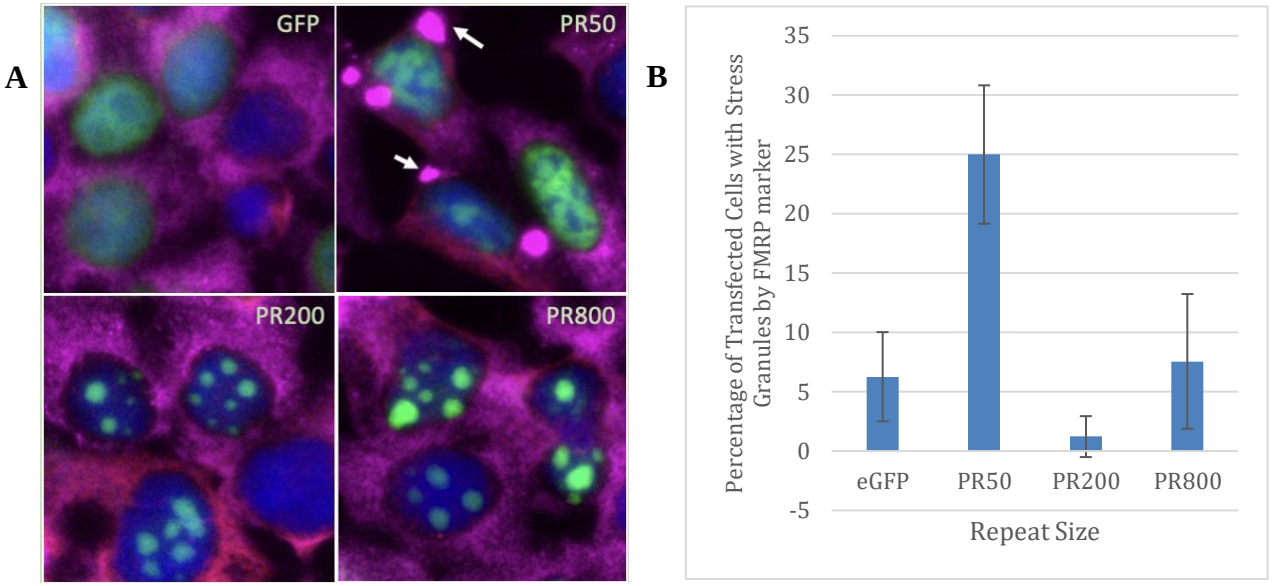

Figure 7. Long PR DPRs do not trigger stress granule formation. A) Immunostaining images of Hek293T cells transfected with varying repeat length constructs. The green indicates EGFP, which is translated at the N-terminus of each construct, indicating translation of the DPRs. The purple stained for B3GP, a maker for stress granules. All repeat containing constructs had DPR localization in the nucleus. PR50 had a large number of stress granules, while PR200 and PR800 both saw negligible stress granules, similar to the negative control. B) Quantification of transfected cells with stress granules. PR50 had a significantly larger number of stress granules than PR200 and PR800, which are comparable to the negative control.

DISCUSSION

In order to gain a better understanding of the translational mechanism of DPRs in C9orf72 ALS/FTD, we first looked at what role the intron plays in the translation of antisense DPRs. We found that when 450 base pairs of the intron sequence are present, there is a significantly larger level of DPRs that are produced when compared to when that intron sequence is removed. The removal of the intron sequence leads to a level of DPRs that is comparable to the control. These results are indicative that the intron sequence is involved in the translation of the antisense DPRs. Previously, the translation of DPRs has been considered to be a somewhat random process known as repeat-associated non-ATG (RAN) translation. However, these insights that the intron sequence is directly involved in translation indicate that DPR translation is likely to not be random, as previously thought. Furthermore, there is some evidence that specific start codons and near-start codons are involved in sense translation of DPRs. Green found that mutation of specific start and near-start codons in the

intron sequence significantly reduced DPR translation in the sense direction (Green et al., 2017). It is likely that the DPR translation is thus being driven by start codons within the intron sequence, and we hypothesize that this is also occurring in the antisense direction. My data suggests that there is similarly an ATG start codon that drives the translation of PR-DPRs. If this is the case, then the entire insertion is likely to be translated until it reaches a stop codon. In patients, this means that the full 100 to 1000 repeats are likely being translated. Though patients have such long repeats, studying these long DPRs is quite difficult due to the nature of only two nucleotides being repeated for such long sequences. Literature has thus focused primarily on shorter repeats, looking at the effects of mostly only 50 to 100 repeats. However, if patients have their full long repeats translated, an important question is will there be differences in the downstream disease mechanisms between the short and long DPRs?

To understand if there is a repeat-length associated DPR toxicity in C9orf72, we successfully created constructs with GR or PR repeated from 50 to 800 in size. EGFP, a green fluorescent protein, was placed at the N-terminus to monitor DPR expressions. Once transfected in Hek293T cells, we found that the longer (GR)200 and (GR)800 form cytoplasmic aggregates, while the shorter (GR)50 were localized in the nucleolus. Not only does this point to a discrepancy between what is likely occurring in patients and what literature is focusing on, but it directly mimics patient samples. Patient samples have been shown to aggregate in the cytoplasm, a phenotype that has not been shown with shorter repeats in vitro (Balendra and Isaacs, 2019). If research focuses primarily on the study of shorter repeats, then the phenotypes that are exhibited by these shorter DPRs are not going to be as applicable to what is occurring in patients, and might lead to misunderstandings about disease mechanisms. We also found a distinct difference in stress granule dynamics between short and long PR-DPRs. As shown in previous literature, transfection of the shorter (50 repeats) PR-DPR led to a significant number of cells with stress granules. However, the longer (PR)200 and (PR)800 showed almost no stress granule formation. As discussed in the introduction, stress granules are

important for cells to protect themselves from different forms of stress. If those dynamics change, whether by preventing the granules from dissipating or by preventing them from forming at all, it can harm the cell and lead to further downstream complications. There is a clear difference in how short and long DPRs affect stress granule dynamics, and it seems that longer repeats prevent stress granules from forming. Again, this discrepancy between short and long DPRs is important in our understanding of the disease mechanism. It is likely that the longer DPRs are more closely mimicking what is occurring in patients, and therefore it is possible that researchers will need to reevaluate the conclusions that have been made using short DPRs in their applicability to the actual disease.

MATERIALS AND METHODS

Molecular Cloning Mutated Start Codons

To study what role the intron played in antisense DPR translation, two plasmids were created using molecular cloning: one with the c9orf72 intron and 289 repeats, (C4G2)289 and one with the intronic sequence removed, (C4G2)289 RO, via restriction enzymes. Stable3 competent cells were used for transformation, and bacteria were grown at 32 degrees Celsius on kanamycin plates. DNA gel electrophoresis was used to confirm plasmid sizes. Hek293T cells were transfected in 12 well plates, and proteins were extracted 48 hours after transfection. 2ul of protein was used for each dot in the dot plot, and lab made antibodies for the PR, GR and PA tags were used to stain for DPRs.

For study of specific start codon and near-start codons' roles in antisense DPR translation, the intronic region of C9orf72 was sequenced via Sanger Sequencing by GeneWiz. Site directed mutagenesis was performed to alter ATG start codons to TAG, and CUG near-start codons to CCC. Repeats were inserted via restriction enzymes by molecular cloning.. Stable3 competent cells were used for transformation, and bacteria were grown at 32 degrees Celsius on kanamycin plates. Hek293T cells were transfected in 12 well plates, proteins were extracted 48 hours after

transfection. 2ul of protein was used for each dot in the dot plot, and lab made antibodies for the PR, GR and PA tags were used to stain for DPRs.

Molecular Cloning Long Repeats

For creation of longer repeats starting from 50 repeats, repeats were cut via restriction enzymes and ligated seamlessly to ensure continuation of proper DPR reading frames. To create constructs with long repeats, molecular cloning was used to insert varying repeat length inserts into eGFP containing vectors. Stable3 competent cells were used for transformation, and bacteria were grown at 32 degrees Celsius on ampicillin plates

Immunostaining

Hek293T cells were transfected with (GR)50, (GR)200, (GR)800, (PR)50, and (PR)800 in 12 well plates to compare differences between different lengths of DPRs. Immunostaining was performed 48 hours after transfection. Stress granules were used to measure one of the possible phenotypic effects of long DPRs, and the marker BG3P, a common marker of stress granules, was used in immunostaining. EGFP was used as a marker for DPR production within the cells.

REFERENCES

Bajc Česnik A, Darovic S, Prpar Mihevc S, et al. Nuclear RNA foci from *C9ORF72* expansion mutation form paraspeckle-like bodies. *J Cell Sci*. 2019;132(5):jcs224303. Published 2019 Mar 7. doi:10.1242/jcs.224303

Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol*. 2018;14(9):544-558. doi:10.1038/s41582-018-0047-2

Boeynaems S, Bogaert E, Kovacs D, et al. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol Cell*. 2017;65(6):1044-1055.e5. doi:10.1016/j.molcel.2017.02.013

Burberry A, Suzuki N, Wang JY, et al. Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci Transl Med*. 2016;8(347):347ra93. doi:10.1126/scitranslmed.aaf6038

Celona B, Dollen JV, Vatsavayai SC, et al. Suppression of *C9orf72* RNA repeat-induced neurotoxicity by the ALS-associated RNA-binding protein Zfp106. *Elife*. 2017;6:e19032. Published 2017 Jan 10. doi:10.7554/eLife.19032

Choi SY, Lopez-Gonzalez R, Krishnan G, et al. C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. *Nat Neurosci*. 2019;22(6):851-862. doi:10.1038/s41593-019-0397-0

Conlon EG, Lu L, Sharma A, et al. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife*. 2016;5:e17820. Published 2016 Sep 13. doi:10.7554/eLife.17820

Cooper-Knock J, Walsh MJ, Higginbottom A, et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*. 2014;137(Pt 7):2040-2051. doi:10.1093/brain/awu120

Davis BM, McCurrach ME, Taneja KL, Singer RH, Housman DE. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A*. 1997;94(14):7388-7393. doi:10.1073/pnas.94.14.7388

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72(2):245-256. doi:10.1016/j.neuron.2011.09.011

Donnelly CJ, Zhang PW, Pham JT, et al. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention [published correction appears in Neuron. 2013 Nov 20;80(4):1102. Heusler, Aaron R [corrected to Haeusler, Aaron R]]. *Neuron*. 2013;80(2):415-428. doi:10.1016/j.neuron.2013.10.015

Freibaum BD, Lu Y, Lopez-Gonzalez R, et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature*. 2015;525(7567):129-133. doi:10.1038/nature14974

Green KM, Glineburg MR, Kearse MG, et al. RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat Commun*. 2017;8(1):2005. Published 2017 Dec 8. doi:10.1038/s41467-017-02200-0

Haeusler AR, Donnelly CJ, Periz G, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*. 2014;507(7491):195-200. doi:10.1038/nature13124

Hao Z, Liu L, Tao Z, et al. Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR. *Nat Commun*. 2019;10(1):2906. Published 2019 Jul 2. doi:10.1038/s41467-019-10956-w

Harms MB, Cady J, Zaidman C, et al. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. *Neurobiol Aging*. 2013;34(9):2234.e13-2234.e2.234E19. doi:10.1016/j.neurobiolaging.2013.03.006

Jiang J, Ravits J. Pathogenic Mechanisms and Therapy Development for C9orf72 Amyotrophic Lateral Sclerosis/Frontotemporal Dementia. *Neurotherapeutics*. 2019;16(4):1115-1132. doi:10.1007/s13311-019-00797-2

Jovičić A, Mertens J, Boeynaems S, et al. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci*. 2015;18(9):1226-1229. doi:10.1038/nn.4085

Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, Timmers AM, Hauswirth WW, Swanson MS. A muscleblind knockout model for myotonic dystrophy. Science. 2003 Dec 12;302(5652):1978-80. doi: 10.1126/science.1088583. PMID: 14671308.

Kanadia RN, Shin J, Yuan Y, et al. Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc Natl Acad Sci U S A*. 2006;103(31):11748-11753. doi:10.1073/pnas.0604970103

Kanekura K, Yagi T, Cammack AJ, et al. Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum Mol Genet*. 2016;25(9):1803-1813. doi:10.1093/hmg/ddw052

Koppers M, Blokhuis AM, Westeneng HJ, et al. C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Ann Neurol*. 2015;78(3):426-438. doi:10.1002/ana.24453

Kwon I, Xiang S, Kato M, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science*. 2014;345(6201):1139-1145. doi:10.1126/science.1254917

Lee KH, Zhang P, Kim HJ, et al. C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell*. 2016;167(3):774-788.e17. doi:10.1016/j.cell.2016.10.002

Li YR, King OD, Shorter J, Gitler AD. Stress granules as crucibles of ALS pathogenesis. *J Cell Biol*. 2013;201(3):361-372. doi:10.1083/jcb.201302044

Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron*. 2013;79(3):416-438. doi:10.1016/j.neuron.2013.07.033

McEachin ZT, Parameswaran J, Raj N, Bassell GJ, Jiang J. RNA-mediated toxicity in C9orf72 ALS and FTD. *Neurobiol Dis*. 2020;145:105055. doi:10.1016/j.nbd.2020.105055

Miller JW, Urbinati CR, Teng-Umnuay P, et al. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J*. 2000;19(17):4439-4448. doi:10.1093/emboj/19.17.4439

Mizielinska S, Ridler CE, Balendra R, et al. Bidirectional nucleolar dysfunction in C9orf72 frontotemporal lobar degeneration. *Acta Neuropathol Commun*. 2017;5(1):29. Published 2017 Apr 18. doi:10.1186/s40478-017-0432-x

Mori K, Lammich S, Mackenzie IR, et al. 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*. 2013;125(3):413-423. doi:10.1007/s00401-013-1088-7

Olney NT, Spina S, Miller BL. Frontotemporal Dementia. *Neurol Clin*. 2017;35(2):339-374. doi:10.1016/j.ncl.2017.01.008

Rowland LP, Shneider NA. Amyotrophic lateral sclerosis. *N Engl J Med*. 2001;344(22):1688-1700. doi:10.1056/NEJM200105313442207

Schludi MH, Becker L, Garrett L, et al. Spinal poly-GA inclusions in a C9orf72 mouse model trigger motor deficits and inflammation without neuron loss. *Acta Neuropathol*. 2017;134(2):241-254. doi:10.1007/s00401-017-1711-0

Swinnen B, Bento-Abreu A, Gendron TF, et al. A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta Neuropathol*. 2018;135(3):427-443. doi:10.1007/s00401-017-1796-5

Swinnen B, Robberecht W, Van Den Bosch L. RNA toxicity in non-coding repeat expansion disorders. *EMBO J*. 2020;39(1):e101112. doi:10.15252/embj.2018101112

Tao Z, Wang H, Xia Q, et al. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet*. 2015;24(9):2426-2441. doi:10.1093/hmg/ddv005

Ugolino J, Ji YJ, Conchina K, et al. Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling. *PLoS Genet*. 2016;12(11):e1006443. Published 2016 Nov 22. doi:10.1371/journal.pgen.1006443

Wen X, Tan W, Westergard T, et al. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron*. 2014;84(6):1213-1225. doi:10.1016/j.neuron.2014.12.010

Xu Z, Poidevin M, Li X, et al. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc Natl Acad Sci U S A*. 2013;110(19):7778-7783. doi:10.1073/pnas.1219643110

Yang D, Abdallah A, Li Z, Lu Y, Almeida S, Gao FB. FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathol*. 2015;130(4):525-535. doi:10.1007/s00401-015-1448-6

Zhang N, Ashizawa T. RNA toxicity and foci formation in microsatellite expansion diseases. *Curr Opin Genet Dev*. 2017;44:17-29. doi:10.1016/j.gde.2017.01.005

Zhang YJ, Gendron TF, Ebbert MTW, et al. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat Med*. 2018;24(8):1136-1142. doi:10.1038/s41591-018-0071-1

Zhang YJ, Gendron TF, Grima JC, et al. C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci*. 2016;19(5):668-677. doi:10.1038/nn.4272

Zu T, Gibbens B, Doty NS, et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A*. 2011;108(1):260-265. doi:10.1073/pnas.1013343108

Zu T, Liu Y, Bañez-Coronel M, et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A*. 2013;110(51):E4968-E4977. doi:10.1073/pnas.1315438110