S.I. : Genetic pathways to Neurodegeneration

Models and mechanisms of repeat expansion disorders: a worm's eye view

Paige Rudich[1] and Todd Lamitina[1,2,3]

[1]Graduate Program in Cell Biology and Molecular Physiology, University of Pittsburgh Medical Center

[2]Division of Child Neurology, Department of Pediatrics, Children's Hospital of Pittsburgh, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

[3]Department of Cell Biology, University of Pittsburgh School of Medicine

Address correspondence to Todd Lamitina, Ph.D., Children's Hospital of Pittsburgh of the University of Pittsburgh Medical Center, Departments of Pediatrics and Cell Biology, 4401 Penn Avenue, Rangos 7122, Pittsburgh, PA 15224, stl52@pitt.edu

Abstract

The inappropriate genetic expansion of various repetitive DNA sequences underlies over 20 distinct inherited diseases. The genetic context of these repeats in exons, introns, and untranslated regions has played a major role in thinking about the mechanisms by which various repeat expansions might cause disease. Repeat expansions in exons are thought to give rise to expanded toxic protein repeats (i.e. polyQ). Repeat expansions in introns and UTRs (i.e. FXTAS) are thought to produce aberrant repeat bearing RNAs that interact with and sequester a wide variety of essential proteins, resulting in cellular toxicity. However, a new phenomenon termed 'Repeat-associated non-AUG dependent (RAN) translation' paints a new and unifying picture of how distinct repeat expansion bearing RNAs might act as substrates for this non-canonical form of translation, leading to the production of a wide range of repeat sequence specific-encoded toxic proteins. Here, we review how the model system *C. elegans* has been utilized to model many repeat disorders and discuss how RAN translation could be a previously unappreciated contributor to the toxicity associated with these different models.

Introduction

Repeat expansions and neurodegenerative diseases

The presence of simple microsatellite repeats is a common feature of most genomes. These repeat sequences are usually replicated during mitosis and meiosis with high fidelity. However, they can undergo unregulated expansions in both somatic and germline contexts and produce offspring with increased repeat numbers as compared to the parents. Unregulated repeat expansions are the underlying genetic cause of at least 20 different disorders, most of which involve some type of central or peripheral nervous system neurodegeneration. Many of these disorders share common clinical features, including an autosomal dominant inheritance pattern, an inverse relationship between disease onset and severity and repeat length, and major phenotypic effects in specific classes of neurons, although non-neuronal cells are sometimes also affected.

In 1991, the first disease-associated repeat expansions were discovered in two genes – the fragile X mental retardation syndrome 1 gene FMR1 and the spinal bulbar muscular atrophy gene SBMA (La Spada, Wilson et al. 1991, Verkerk, Pieretti et al. 1991). Since then, repeat expansions have been associated with a wide variety of diseases, including many neurodegenerative diseases, such as Huntington's Disease (HD), multiple spinocerebellar ataxias (SCAs), Fragile X syndrome (FXS), myotonic dystrophy (DM1 and DM2) and C9orf72 associated Amyotrophic Lateral Sclerosis / Frontotemporal Degeneration (ALS/FTD). While the molecular nature of each repeat expansion varies, almost all repeats are GC rich. Disease-causing GC rich repeat

expansions occur in many genetic contexts, including exons, introns, 3'UTR's and 5'UTR's. Repeats located in non-coding regions are hypothesized to exert their toxicity through the presence of repeat-bearing RNAs, which sequester various RNA binding proteins and result in their toxic functional depletion. Repeats located in coding regions are hypothesized to cause toxicity via their potential to encode amino acid repeats (i.e. polyQ), which form protein aggregates and overwhelm protein quality control mechanisms, leading to the wide-spread accumulation of damaged and misfolded proteins. Recently, another mechanism to explain the toxicity of repeat expansions, termed repeat-associated non-AUG dependent (RAN) translation, has emerged (Cleary and Ranum 2014). In RAN translation, repeat expansions in coding and non-coding contexts can be translated in all three reading frames to produce multiple repeat-encoded peptides. This remarkable discovery was first made in 2011 while investigating mechanisms of toxicity in the CAG repeat expansion within the coding sequence of the SCA8 gene (Zu, Gibbens et al. 2011). Since then, RAN translated products have been found in patient samples representing several repeat expansion disorders (Cleary and Ranum 2014). The molecular mechanisms of RAN translation are just beginning to be studied and there are many unanswered questions (Green, Linsalata et al. 2016). For example, it is unclear whether the RNA substrate for translation is the unspliced pre-mRNA, mature spliced mRNA, or spliced out intron RNA (in cases where the repeat is intronic). Recent work suggests that RAN translation does require some canonical translational components, such as the 7'methylguaninie cap at the 5'

end of the RNA (Kearse, Green et al. 2016, Green, Glineburg et al. 2017). Nevertheless, the data strongly suggest that RAN translation plays a significant but previously unappreciated role in the pathogenesis of repeat expansion disorders. In this review, we will discuss various diseases caused by repeat expansions and highlight how the unique advantages of the model system *C. elegans* have led to new and unexpected discoveries that have informed our understanding of these disease mechanisms. We will conclude by discussing how these models might be re-interpreted and further explored in the context of RAN translation.

Developing *C. elegans* as a Model to Study Repeat Expansion Disease: advantages and disadvantages

Understanding the molecular mechanisms underlying repeat expansions disorders is experimentally challenging. Such mechanistic studies are not well suited to human subjects due to the limited availability of patient tissue, heterogeneous genetic backgrounds, and difficult ethical and technical considerations associated with gene manipulation *in vivo*. To circumvent these problems, several cellular or animal models recapitulating various aspects of repeat expansion mutations in various contexts have been developed, including yeast, worms, fruit flies, zebrafish, rodents, and human induced pluripotent stem cells (iPSCs). It is critical to recognize that each of these models has their own unique mix of strengths and weaknesses and that no single model fully recapitulates disease pathogenesis. However, each model, through the

uniqueness of their biology and experimental resources, illuminates distinct and conserved aspects of disease pathology. Utilization of each of these models is necessary if we hope to gain a detailed understanding of the molecular mechanisms underlying repeat expansion pathogenesis.

C. elegans are small (1 millimeter long), transparent nematodes that have been studied as a genetic model system for over 50 years. Worms offer several advantages not present in other model systems for the study of repeat associated genetic disorders. First, transgenic methods in *C. elegans* are highly efficient. Following the creation of a suitable DNA expression construct, transgenic worms expressing a repeat sequence of interest are obtained in ~ 1 week. Second, *C. elegans* are optically transparent across their entire lifespan, facilitating observation of neuron and tissue morphology in live animals, as well as subcellular localization of GFP fusion proteins. Third, *C. elegans* have a highly conserved genome with humans, with 60-80% of the ~20,000 worm genes having a human homolog. Fourth, the function of these genes can be rapidly inhibited, either using conventional mutants or RNA interference. Finally, ageing, which is a major risk factor for all repeat expansion diseases, has been extensively studied in *C. elegans*. Mutant animals that alter the rate of ageing allow us to ask how disease-associated toxicity is influenced by specific ageing pathways over the relatively short 2-week lifespan of the animal.

Despite these experimental advantages, worms are not humans and cannot model all potential aspects of disease pathophysiology. *C. elegans* motor neurons lack astrocytes and glia, which play a significant functional role in

several neurodegenerative diseases (Ilieva, Polymenidou et al. 2009). While worms possess a conserved primitive innate immune system (Ewbank and Pujol 2016), they lack adaptive immunity, which contributes to many pathological aspects of repeat expansion disorders (Cappellano, Carecchio et al. 2013). Similar limitations impact other model systems that are used to investigate repeat associated neurodegenerative diseases (Donnelly, Zhang et al. 2013, Haeusler, Donnelly et al. 2014, Jovicic, Mertens et al. 2015, Zhang, Donnelly et al. 2015, Boeynaems, Bogaert et al. 2016, Kramer, Carlomagno et al. 2016). Despite these limitations, *C. elegans* are a uniquely valuable model whose experimental advantages complement many of the limitations present in other systems for the study of repeat expansion disorders.

Worm models of repeat expansions

C. elegans HD and polyQ models

Huntington's disease (HD) is a classic genetic disease caused by a single mutation, a CAG repeat expansion in the first exon of the huntingtin (HTT) gene (1993). Most people have ~20 CAG repeats in HTT, but patients with HD have 36 or more CAG repeats. People with 36-39 CAG repeats in HTT are at risk for developing HD, and people with 40 or more CAG repeats in HTT will develop HD at some point in their lifespan. There is a strong inverse correlation between the length of the CAG repeat and time of onset of HD, with the average age of onset ~40 (Wexler, Lorimer et al. 2004). These patients have degeneration of the striatum and cerebral cortex leading to mental, psychological and physical

declines (Warby, Graham et al. 1993). There are several proposed pathological mechanisms thought to contribute to HD (Figure 1). One major hypothesis is that the long stretch of glutamines (polyQ) encoded by the CAG repeat present in the mutant HTT (mHTT) predisposes mHTT to aggregate, precipitating large-scale disruption of the cellular protein quality control system (Satyal, Schmidt et al. 2000). mHTT also disrupts HTT's normal role in autophagy and is believed to have a toxic gain of function interaction with the mitochondria (Ashkenazi, Bento et al. 2017). However, mHTT might not be the only cause of toxicity. Both HTT and mHTT can also be transcribed in the antisense direction to produce a CTG repeat containing RNA (Chung, Rudnicki et al. 2011). Recent data show that both the CAG sense and CTG antisense repeat transcripts can serve as substrates for RAN translation, resulting in the production of five RAN peptides that are all detectable in HD patient neurons (poly-Glutamine, poly-Serine, poly-Alanine, poly-Leucine, and poly-Cysteine) (Banez-Coronel, Ayhan et al. 2015). Remarkably, many of these newly discovered RAN products are present in regions that undergo significant neurodegeneration even in the absence of detectable polyQ protein aggregates, suggesting that they may play a significant pathological role in disease. While the properties of the polyQ proteins are well known, the properties of these new HD RAN peptides have not yet been described.

Several models of HD and other CAG repeat expansion disorders have been developed in *C. elegans* (Table 1). Most of the models include surrounding HTT sequence, as genetic context can both enhance and suppress the toxicity of

repeats (Van Assche, Borghgraef et al. 2017). In one model, pure CAG repeats were set in the genetic context of the sequence encoding the first 171 amino acids of HTT (171HTT(CAG)) (Faber, Alter et al. 1999). The Htt-polyQ was expressed in ASH neurons, which are polymodal sensory neurons exposed to the environment (Hart, Kass et al. 1999). Exposure of worms to a fluorescent dye in their environment robustly labels the ASH neurons, as well as other environmentally exposed neurons. Worms expressing 171mHTT(CAG)$_{150}$ had an age-dependent loss of ASH dye filling due to cellular dysfunction, but did not exhibit cell death. Ageing enhanced the aggregation of 171mHTT(CAG)$_{150}$, although no aggregation was observed with shorter repeats. This model has facilitated both biased and unbiased genetic screens, as well as small molecule drug screens, for suppressor of CAG repeat expansion toxicity. For example, Bates et al. discovered that the specific HDAC homolog HDA-1 suppresses 171mHTT(CAG)$_{150}$ toxicity, while HDA-3 enhances 171mHTT(CAG)$_{150}$ toxicity. HDACs are now known to play significant roles in the pathogenesis of polyQ toxicity across a wide range of cellular and animal model systems (Pandey, Nie et al. 2007, Jia, Pallos et al. 2012). These studies, as well as many others that utilized this model (Jia, Hart et al. 2007, Voisine, Varma et al. 2007, Jeong, Then et al. 2009), have provided major insights into our understanding of polyQ toxicity.

Another model for HD in *C. elegans* utilized expanded pure CAG repeats within the coding sequence for the first 57 amino acids of HTT (57mHTT(CAG)) (Parker, Connolly et al. 2001) (Table 1). In contrast to the 171mHTT(CAG)

model that was expressed only in ASH sensory neurons, this 57mHTT(CAG) model was expressed in a different class of sensory neurons called touch neurons. Inhibition of touch neuron function rendered animals unable to respond to light touch by altering their direction of movement (Tavernarakis and Driscoll 2001). 57mHTT$(CAG)_{128}$ caused morphological defects in the sensory neurons and a defective touch response (Parker, Connolly et al. 2001) but no cell death until late in life. Because functional defects precede cell death, this model is particularly useful for modeling the 'early-onset' aspects of CAG/polyQ pathology. For example, Huntingtin interacting protein 1 (HIP1) is involved in clathrin mediated endocytosis and protein trafficking within the cell (Metzler, Legendre-Guillemin et al. 2001, Legendre-Guillemin, Metzler et al. 2005). HIP1 normally interacts with wild-type Htt, but it has weaker interaction with mHtt. Knocking down the *C. elegans* homologue of HIP1 specifically increased expanded polyQ toxicity, as did mutations in other synaptic endocytic proteins, suggesting that expanded polyQ proteins can disrupt synaptic endocytosis (Parker, Metzler et al. 2007). In another example, Lejeune et al. utilized RNAi screening in the 57mHTT$(CAG)_{128}$ line to identify suppressors and enhancers of expanded CAG toxicity (Lejeune, Mesrob et al. 2012). Many of the 662 genes identified in this screen were previously known to be involved in HD or other neurodegenerative diseases, including 49 of which that are dysregulated in the striatum of HD mouse models. Rather than identifying 'worm-specific' pathways, these data demonstrate that for CAG repeat expansion toxicity, RNAi screens performed in

C. elegans lead to the identification of conserved pathological mechanisms that play similar biological roles in higher organisms, including mammals.

A recent *C. elegans* model of HD mimics a cleaved fragment of mHTT seen in patients, which is believed to be toxic (Lee, Ung et al. 2017) (Table 1). The model expresses the first 513 amino acids of HTT with the polyglutamine repeats (Lee, Ung et al. 2017). Caspase-3 cleaves at amino acid 513 of mHTT in patients and the resulting fragment is believed to be a driving cause of toxicity (Wellington, Ellerby et al. 2002). $(CAG)_{15}$ or $(CAG)_{128}$ were expressed as part of the 513 HTT fragment in the *C. elegans* body wall muscle cells. PolyQ aggregation was still length dependent. 513mHTT$(CAG)_{128}$ had a detectable motility defect, but this phenotype was not age-dependent. 513mHTT$(CAG)_{128}$ have a shorter lifespan compared to animals with the shorter repeat length, which has not been reported for the other *C. elegans* CAG models. The differences between this model and the previous *C. elegans* CAG models emphasize how specific aspects of genetic context can modify toxicity in different cellular setting.

The previous models strived to examine CAG repeats in the presence of the HTT protein. However, there are ten age-onset neurodegenerative diseases caused by CAG repeat expansions, suggesting the CAG repeat itself, in many different genetic contexts, is sufficient to cause age-onset toxicity. To explore CAG toxicity independent from these different contexts, a pure CAG model was generated (Table 1). In this model, pure CAG repeats are fused with YFP in the polyQ reading frame and expressed in muscle tissue (Morley, Brignull et al.

2002). Although HD is not a muscle disease, expression of CAG in the muscle tissue of *C. elegans* provides several experimental advantages (large cell size, simple feeding-based RNAi gene knockdown, sensitive age-dependent phenotypic outputs, etc.) that are not available when expressed in neurons. Degeneration of muscle cells leads to easily observable movement disruptions and paralysis that can be monitored across the lifespan of the animals. In this model, aggregation of polyQ is repeat length dependent, with Q_{82} causing complete polyQ aggregation in young animals, and Q_{33} being diffuse in young animals. Q_{40} is initially localized in a diffuse and non-aggregated manner in young animals, but the protein transitions from a soluble to an aggregated state over the short lifespan of the animal (Morley, Brignull et al. 2002). Aggregate formation coincides with the onset of motility defects, suggesting a link between aggregation and toxicity. However, recent work implies that polyQ aggregation and toxicity are genetically separable events, suggesting that mechanisms other than polyQ aggregation may contribute to toxicity in this model (Gidalevitz, Wang et al. 2013). The toxicity and aggregation of these pure polyQ proteins are strongly influenced by the ageing process, since mutants that modify ageing (i.e. in insulin/IGF signaling pathway) also modify polyQ aggregation and CAG toxicity (Hsu, Murphy et al. 2003, Morley and Morimoto 2004, Teixeira-Castro, Ailion et al. 2011, Moronetti Mazzeo, Dersh et al. 2012).

More than any other repeat expansion model, the context-independent pure CAG repeat model has played a profound role in advancing our understanding of the roles of ageing, protein aggregation, and repeat expansion

associated toxicity. For example, expression of expanded CAG repeats in neurons, the major cell type affected in HD and other CAG repeat expansion disorders, shows heterogenous aggregation and toxicity depending on the neuron type (Brignull, Moore et al. 2006). While demonstrating that the behavior of CAG repeat expansion proteins depends on cellular context, these observations significantly complicate efforts to identify genes that might modify the aggregation and/or toxicity of polyQ proteins (or other CAG-derived translation products) in neurons. In this respect, muscle cells, while not directly translationally relevant, have provided a much more amenable and homogenous cellular context for *in vivo* genetic screens that have ultimately been proven to inform our understanding of CAG toxicity in neurons. For example, in one of the first examples of genome-wide RNAi screening in *C. elegans*, Nollen et al. screened ~17,000 gene knock-downs for enhancers of muscle polyQ aggregation, i.e. their normal function is to oppose polyQ aggregation. These findings revealed that the breadth of the protein homeostasis machinery that regulates protein misfolding extends far beyond chaperones and the protein degradation machinery and involves a wide variety of biological processes included RNA synthesis and processing, protein biosynthesis, and protein trafficking (Nollen, Garcia et al. 2004) (Figure 1). Many of the genes discovered in this screen have subsequently been shown to play roles in mediating polyQ aggregation and/or toxicity in mammalian cells, thus validating that screens utilizing *C. elegans* muscle models of repeat expansion diseases have high

translational relevance (Kitamura, Kubota et al. 2006, Teuling, Bourgonje et al. 2011).

Spinocerebellar ataxia

Ataxias are a group of inherited neurodegenerative disorders with age-dependent progressive ataxia (loss of voluntary coordinated movement) as a common and defining clinical feature (Paulson, Shakkottai et al. 2017). Hereditary ataxias can present with either autosomal dominant, autosomal recessive, X-linked or mitochondrial modes of inheritance. SCAs are autosomal dominant inherited ataxias, of which there are currently >40 distinct genetic subtypes. Twelve SCAs, including the most common SCA mutation (SCA3, also known as Machado-Joseph disease), are caused by repeat expansions in exons, introns or promoters. Eight of these repeats (SCA1,2,3,6,7,12,17, and DRPLA) are CAG expansions, with the normal repeat length between 20-40 and the disease causing repeat length over 30-50, depending on the type of SCA (Sandford and Burmeister 2014). These mutations frequently lead to degradation of the Purkinje cells of the cerebellar cortex, although there is significant pathological variability within each type of CAG-associated SCA. Patients start developing symptoms of SCA during middle age (average age onset of 37), displaying disrupted gaits, slurred speech, and poor hand-eye coordination. SCA3 is the most common type of SCA, with 10-~80% of families with autosomal dominant ataxias having a CAG-expanded SCA3 allele (Silveira, Lopes-Cendes et al. 1996, Lopes-Cendes, Teive et al. 1997), depending on the specific

population. SCA3 is caused by a CAG repeat expansion in the exon of ATXN3 (Kawaguchi, Okamoto et al. 1994). Because it is the most common SCA subtype, many cell and animal models have been developed to study SCA3 mutations (Ingram, Orr et al. 2012). The major mechanism of toxicity for all the SCAs with CAG repeat expansions in coding regions is thought to be related to the translation of an in-frame stretch of polyQ within the respective SCA protein. Recent data also suggests that for some SCA repeat expansions, RAN translation of the CAG repeat and production of RAN peptides may also play a significant pathogenic role (Zu, Gibbens et al. 2011, Scoles, Ho et al. 2015).

There are several models for SCA3-related atxia in *C. elegans*. One model focuses on both the truncated and full length ATXN3 with CAG expansions expressed either pan-neuronally (Teixeira-Castro, Ailion et al. 2011) or in muscle (Christie, Lee et al. 2014). Another SCA3 model expresses the full length ATXN3 solely in GABAergic motor neurons (Fardghassemi, Tauffenberger et al. 2017) (Table 1). As was observed for the HD polyQ models in worms, expanded CAG repeats within SCA3 are strongly dependent on both genetic and cellular contexts. Worms expressing CAG repeats (14, 75, or 130) in either the context of the full length ATXN3 or a truncated ATXN3 containing the last 257 amino acids were expressed across all *C. elegans* neurons (Teixeira-Castro, Ailion et al. 2011). The full length ATXN3 with 130 CAG repeats, but not 75 repeats, showed polyQ aggregates in some, but not all neurons, which led to an 'uncoordinated' (Unc) phenotype. In contrast, truncated ATXN3 showed significant aggregation even at 75 repeats. Interestingly, ageing appears to play

a complex role in this model, with some neurons exhibiting age-dependent aggregation and others showing no aggregation. However, ageing did not appear to play a role in ATXN3 toxicity when expressed in muscle (Christie, Lee et al. 2014), suggesting that ageing can have disparate effects on aggregation and toxicity depending on the genetic and cellular context of the CAG repeat.

The broad expression pattern and phenotypic variability amongst neuron subtypes severely limits the utility of these pan-neuronal ATXN3 models for genetic studies aimed at understanding pathophysiological mechanisms underlying CAG expanded ATXN3. To address these limitations, Fardghassemi et al. expressed full length ATXN3 with 10 or 89 CAG repeats in the 26 GABAergic motor neurons (Fardghassemi, Tauffenberger et al. 2017) (Table 1). GABA neurons are required for coordinating the contraction and relaxation of body wall muscle to drive sinusoidal crawling movement across the growth substrate (McIntire, Jorgensen et al. 1993). Toxic proteins that disrupt the function or integrity of GABA neurons produce animals with strong motility defects. GABA neurons are not needed for survival or reproduction in *C. elegans*, since animals can be maintained as hermaphrodites that undergo internal fertilization and do not require any movement-based mating behaviors to reproduce. Therefore, animals expressing toxic proteins in these cells are still viable and fertile (McIntire, Jorgensen et al. 1993). Individual GABA neurons are easily visualized in live animals with GFP reporters and signs of neurodegeneration, such as aberrant varicosities and process breakage, are easily scored with single cell resolution. In this setting, expression of ATXN3 with

$(CAG)_{89}$ inhibited GABA signaling and caused age-dependent neurodegeneration (Fardghassemi, Tauffenberger et al. 2017). These cellular phenotypes led to behavioral consequences, since animals with enhanced neurodegeneration also exhibited reduced movement and eventual paralysis. Based on these phenotypes, three small molecules which had previously suppressed toxicity in an ALS *C. elegans* model, also suppressed ATXN3 $(CAG)_{89}$ toxicity and reduced their levels of ER stress (Fardghassemi, Tauffenberger et al. 2017). This example shows how restricting the expression of a repeat expansion protein to a specific and homogenous cell type in *C. elegans* can enable highly sensitive *in vivo* genetic and pharmacological screens aimed at identifying conserved regulators of repeat expansion toxicity.

Myotonic dystrophy

There are two types of myotonic dystrophy, myotonic dystrophy type 1 (DM1) and myotonic dystrophy type 2 (DM2). DM1 is the most common and is caused by a CTG expansion (usually over 50 repeats) in the 3'UTR of the DM1/DMPK gene (Brook, McCurrach et al. 1992, Fu, Pizzuti et al. 1992, Mahadevan, Tsilfidis et al. 1992). DM2 is caused by a CCTG expansion (usually hundreds to thousands of repeats) in the intron of the DM2/CNBP/ZNF9 gene (Liquori, Ricker et al. 2001). Patients with DM1 and DM2 exhibit muscular dystrophy, with prolonged muscle contraction after stimulation (myotonia) as a common and distinguishing clinical feature. Patients with DM1 have an average age of onset of 26, although childhood symptoms are sometimes observed (Bird

1993). Patients with DM2 repeat expansions have an average age of onset of 34 (Dalton, Ranum et al. 1993). Due to the presence of the DM1 and DM2 repeats in non-coding regions, disease pathogenesis is thought to occur via RNA toxicity mechanisms (Ranum and Cooper 2006). However, work on both the DM1 CAG expansion and the DM2 CCTG expansion shows that they both undergo RAN translation and produce RAN peptides that are present in patient cells or mouse models of disease (Zu, Gibbens et al. 2011, Zu, Cleary et al. 2017), suggesting that RAN peptides may play a previously unappreciated role in the pathogenesis of myotonic dystrophies.

To date, only DM1 has been modeled in *C. elegans* (Table 1). In this model, worms express a CUG repeat in the 3'UTR of the GFP coding sequence from a muscle-specific promoter (Chen, Pan et al. 2007). Animals expressing over 200 CUG repeats died during embryogenesis, while animals with 125 repeats developed normally but exhibited post-developmental defects in motility. Surprisingly similar phenotypic defects were observed when the CUG DM1 repeat was replaced with a CAG repeat, such as that found in the 5'UTR of SCA12 (Holmes, O'Hearn et al. 1999). The CAG expansion model showed embryonic lethality at over 200 CAG repeats, while animals with 124 CAG repeats developed normally but exhibited post-developmental motility defects (Wang, Chen et al. 2011). Both the CAG/CUG RNA formed foci in the muscle nuclei and co-localized with RNA binding protein Muscleblind, as has also been demonstrated to occur in humans (Mankodi, Urbinati et al. 2001). Moreover, overexpression of Muscleblind was able to partially rescue the phenotypes

associated with 3' UTR $(CAG)_{125}$ or $(CUG)_{125}$. The striking similarity between the different models strongly suggest CAG repeat expansions and CUG repeat expansions in UTR's are toxic through similar, if not the same, pathways.

Fragile X disorders

Expansion of a CGG repeat sequence in the 5'UTR of the fragile X mental retardation 1 gene FMR1 gives rise to a spectrum of neurological disorders (Saul and Tarleton 1993). Fragile X syndrome is found in patients with >200 CGG repeats and is the most common genetic cause of intellectual disability and autism. Shorter repeat expansion lengths from 55-200 cause the neurodegenerative disorder fragile X-associated tremor/ataxia syndrome (FXTAS), as well as the reproductive disorder fragile X-associated primary ovarian insufficiency. FXTAS is a progressive movement disorder usually observed in the sixth decade of life. Due to the location of the CGG repeat expansion in the non-coding 5'UTR of the Fragile X gene, the prevailing hypothesis is that the mechanism of toxicity involves the repeat-bearing RNA interacting with and/or sequestering multiple CGG RNA binding proteins, leading to their functional depletion (Galloway and Nelson 2009). Recent data also suggests that, despite their presence in the 5'UTR, CGG repeats in the Fragile X RNA can be translated, either via RAN translation or the utilization of non-canonical start codons to produce expanded polypeptides, such as poly-Glycine, which cause cellular toxicity (Todd, Oh et al. 2013, Kearse, Green et al. 2016, Krans, Kearse et al. 2016).

The *C. elegans* genome does not contain an FMR1 homolog. Therefore, worms have received limited attention for modeling Fragile X syndromes. Recently, a transgenic worm model of Fragile X was generated by expressing the human 5'UTR with 0 or 99 CGG upstream of the GFP coding sequence (Juang, Ludwig et al. 2014) (Table 1). Expression was driven in the sensory AWC neuron, which mediates both primary and adaptive olfactory responses to attractive volatile chemical stimulants (Colbert and Bargmann 1995, L'Etoile, Coburn et al. 2002). Neither 0 nor 99 CGG repeats led to degeneration of the AWC neuron. However, expression of the expanded 99 CGG repeat RNA, but not the 0 repeat RNA, disrupted the ability of the AWC neurons to adapt their response to a stimulus. This disruption in AWC neuronal plasticity required the *alg-2* gene, which plays a role in the processing of microRNAs. These findings led to the hypothesis that microRNAs are critical for the CGG repeat bearing RNA to exert toxicity. While CGG repeats in mammalian models are known to undergo RAN or non-canonical initiation codon-induced translation, it is not known if the CGG repeats expressed in *C. elegans* undergo similar translation and if so, whether these CGG-derived peptides are involved in the toxicity of this model.

C9 ALS/FTD

A repeat expansion in the C9orf72 gene is the most common inherited cause of ALS/FTD (DeJesus-Hernandez, Mackenzie et al. 2011, Renton, Majounie et al. 2011). Unlike many other tri-nucleotide repeat expansion

mutations, the C9orf72 repeat involves the expansion of a GGGGCC (G_4C_2) hexanucleotide repeat sequence in the first intron of C9orf72. Unaffected individuals harbor 10-20 repeats while ALS/FTD patients can have hundreds to thousands of repeats. The C9orf72 repeat expansion is responsible for as much as 50% of all inherited cases of ALS and 25% of FTD cases. The high frequency of C9orf72 repeat expansions in ALS/FTD, as well as its association with several other less common neurodegenerative conditions (Cooper-Knock, Shaw et al. 2014), makes C9orf72 repeat expansions the most common known neurodegenerative disease mutation.

The mechanism(s) by which C9orf72 repeat expansions lead to neurotoxicity and disease are unknown. Patients with loss-of-function mutations in C9orf72 have not been identified, arguing against a loss-of-function mechanism. Clinical presentation of C9orf72 repeat expansion carriers are consistent with an autosomal dominant genetic pattern which could manifest pathologically through at least two non-exclusive molecular mechanisms (Ling, Polymenidou et al. 2013). First, given that the repeat expansion is found in an intron, one hypothesis is that toxicity is mediated via a repeat containing RNA. In patients, pathologically expanded RNAs are known to cluster to foci within the nucleus, while non-expanded RNAs do not form nuclear foci (Lee, Chen et al. 2013, Zu, Liu et al. 2013). Although the significance of these RNA foci remains unknown, G_4C_2 RNA is known to bind several RNA binding proteins, including some that regulate the nuclear import/export cycle and nuclear pore complex function (Lee, Chen et al. 2013, Freibaum, Lu et al. 2015, Zhang, Donnelly et al.

2015). Disruption of neuronal nuclear transport via G_4C_2 RNA may therefore play a pathological role in disease onset and/or severity. A second potential mechanism involves RAN translation of the intronic repeat RNA in all three reading frames. As is the case for other repeat expansions, the C9orf72 expanded repeat additionally produces an antisense transcript, which also undergoes RAN translation (Zu, Liu et al. 2013). Altogether, G_4C_2 and C_4G_2 repeat expanded RNAs can be RAN translated in six different reading frames to produce five distinct dipeptide repeat proteins (DPRs) (Cleary and Ranum 2014). Currently, it is unclear whether the RAN DPRs are translated from the spliced out G_4C_2-containing intron or from the intact pre-mRNA. However, factors that promote the cytoplasmic localization of the repeat bearing RNA appear to facilitate RAN translation, suggesting that the observed RNA foci may serve a protective role (Mori, Nihei et al. 2016, Hautbergue, Castelli et al. 2017). The sense-derived DPRs Glycine-Arginine (GR_n) and antisense derived Proline-Arginine (PR_n) are toxic across a wide range of model systems (Mizielinska, Gronke et al. 2014, Wen, Tan et al. 2014, Freibaum, Lu et al. 2015, Jovicic, Mertens et al. 2015, Tran, Almeida et al. 2015, Boeynaems, Bogaert et al. 2016, Lee, Zhang et al. 2016, Rudich, Snoznik et al. 2017). While RAN DPRs are thought to play a significant role in the toxicity of the C9orf72 repeat expansion, there are some notable inconsistencies that mitigate this hypothesis. For example, in post-mortem brain samples from C9orf72 patients, anti-DPR antibody staining patterns are not well correlated with tissue neurodegeneration (Mackenzie, Arzberger et al. 2013, Davidson, Barker et al. 2014, Schludi, May et

al. 2015). One possible explanation for this discrepancy is that in post-mortem samples, cells exhibiting robust RAN translation may undergo cell death, which results in low levels of observable DPRs. Alternatively, DPRs may be localized to sites of cell death but the antibodies used for their detection fail to recognize them due to post-translational modification or unusual structural configurations of the peptide.

C. elegans have recently been utilized to investigate the mechanisms of C9orf72 repeat expansion toxicity (Table 1). To date, three distinct models have been generated. One model expresses 9 or 29 G_4C_2 repeats upstream from the GFP coding sequence, where both RNA toxicity and RAN product toxicity could occur (Burgess, Snutch et al. 2016, Kramer, Carlomagno et al. 2016). The repeats were placed in the 5'UTR between the transcriptional start site and the translational start site for GFP and contains additional C9orf72 intronic sequence surrounding the G_4C_2 repeat. Global expression of $(G_4C_2)_{29}$ repeats, but not $(G_4C_2)_9$, caused motility defects and increased lethality. Because the repeat was expressed globally, it is unclear whether the observed phenotypic defects result from G_4C_2 toxicity in any one specific tissue. Nevertheless, nuclear transport was impaired by $(G_4C_2)_{29}$ expression, which is consistent with other research showing that G_4C_2 RAN DPRs disrupt nucleocytoplasmic transport (Freibaum, Lu et al. 2015, Zhang, Donnelly et al. 2015, Boeynaems, Bogaert et al. 2016). A mutagenesis screen for suppressors of the motility defect caused by $(G_4C_2)_{29}$ repeats identified two genes, F57A10.2 and *acp-2*, as loss-of-function suppressors. F57A10.2 shares homology of a human sperm protein (HSP)

domain with human VAMPS, including VAPB. Mutations in VAPB in the HSP domain have been found in ALS8 patients. *acp-2* is predicted to be a lysosomal acid phosphatase. Studies of G_4C_2 repeat toxicity in other model systems have not identified roles for F57A10.2 or *acp-2* homologs in C9orf72 toxicity, so the relevance of these findings is unknown.

In a second G_4C_2 *C. elegans* model, worms express 66 repeats of G_4C_2 across all neurons (Kramer, Carlomagno et al. 2016). Stop codons were placed in all three reading frames before the G_4C_2 repeats to prevent non-canonical translation initiation. Whether or not the G_4C_2 repeat was placed within the context of surrounding C9orf72 intronic sequence was not specified. In this model, G_4C_2 sense RNA formed RNA foci and produced at least one RAN DPR (Gly-Pro). $(G_4C_2)_{66}$ animals exhibited shortened lifespan which was further shortened by overexpression of the conserved transcription elongation factor Spt4. Knockdown of the worm homolog, *spt-4*, had the opposite phenotype and extended the lifespan of $(G_4C_2)_{66}$ animals, phenotypes that were conserved in yeast and mammalian cells (Kramer, Carlomagno et al. 2016). While both (G_4C_2) models have the advantage of looking at the disease-causing repeat, they do not differentiate between RNA toxicity and dipeptide toxicity due to RAN translation products derived from the RNA.

To address these limitations, a newly reported *C. elegans* model specifically examined individual RAN translation products in the absence of repeat containing RNA (Rudich, Snoznik et al. 2017). In this model, codon variation was utilized to encode a specific amino acid sequence but eliminate

(G_4C_2) sequence and any potential RNA secondary structure and therefore RNA based toxicity. The reading frame was defined using a canonical start codon and Kozak sequence. Four C9orf72-derived dipeptide sequences (Gly-Ala, Gly-Arg, Pro-Ala, Pro-Arg; all 50 repeats and lacking additional C9orf72 context) were tagged with GFP at the C-terminus and FLAG at the N-terminus and expressed individually in muscle cells or in GABA motor neurons. The arginine rich dipeptides, PR and GR, were toxic in both cellular contexts. When expressed embryonically in muscle, $(GR)_{50}$ and $(PR)_{50}$ caused developmental arrest. Developmental expression could be suppressed using feeding-based RNAi directed towards the GFP fusion protein. Removal of animals from *gfp(RNAi)* and subsequent re-expression of the DPR revealed an age associated paralysis phenotype. Motor neuron expression of $(GR)_{50}$ and $(PR)_{50}$ caused signs of neurodegeneration, such as commissure blebbing and breakage. The loss of motor neuron integrity resulted in striking motility defects in $(GR)_{50}$ and $(PR)_{50}$ animals but not in $(GA)_{50}$ or $(PA)_{50}$ animals. Interestingly, both toxic DPRs failed to form protein aggregates, as is commonly seen for other repeat containing peptides. Instead, $(GR)_{50}$ and $(PR)_{50}$ were primarily localized to the nucleolus and nuclear localization was necessary and sufficient for their toxicity, suggesting that DPR toxicity occurs through a nuclear mechanism(s). The highly penetrant phenotypes of $(GR)_{50}$ and $(PR)_{50}$ animals make this an ideal *in vivo* system for future genetic screens aimed at identifying modifiers of DPR toxicity. Such modifiers could illuminate potential new disease mechanisms, identify novel biomarkers, or suggest new potential therapeutic targets.

Could RAN translation and RAN peptides contribute to the toxicity in previous worm models of repeat expansions disorders?

RAN translation was first discovered in 2011. Since its discovery, RAN-derived peptides have been detected in seven different repeat expansion disorders (Todd, Oh et al. 2013, Zu, Liu et al. 2013, Banez-Coronel, Ayhan et al. 2015, Zu, Cleary et al. 2017). With this discovery, RAN peptides add a new and previously unappreciated dimension to the pathological mechanisms underlying these distinct disorders. This discovery necessitates a re-evaluation of the several existing *C. elegans* repeat expansion models to consider the possibility that RAN translation may be a contributing factor to the observed toxicity. Such an interpretation was never considered when these models were first developed. Re-examining these existing models may further enhance their utility and provide additional insights into their suitability, or lack thereof, for investigating human disease mechanisms.

In this respect, it is most important to ask whether *C. elegans* models of repeat expansion mutations can exhibit RAN translation. In the several existing CAG *C. elegans* models, there is currently no evidence for or against the existence of RAN translation. RAN translation of the sense strand CAG repeat could give rise to poly-Serine and poly-Alanine proteins, in addition to the poly-Glutamine protein. RAN translation of a putative antisense strand, if one is produced in these models, could give rise to poly-Leucine, poly-Cysteine, and poly-Alanine. Studies to examine whether these additional sense and antisense reading frames could be toxic, such as the use of codon-varied transgenes to

specifically encode each peptide, have not been reported. It is also not known whether the production of poly-Glutamine in these CAG models is dependent on the presence of a start ATG. The ATG-independent translation of poly-Glutamine from the SCA8 CAG repeat was one of the initial observations underlying the discovery of RAN translation (Zu, Gibbens et al. 2011). In the future, it will be critical to determine if CAG repeats can drive RAN translation in *C. elegans* and whether such RAN peptides exhibit any toxicity, either through established polyQ mechanisms or independent pathways (Figure 1).

In contrast to the lack of evidence for RAN translation in *C.* elegans CAG models, there is some evidence that RAN translation occurs in *C. elegans* models for G_4C_2 repeats. In the *C. elegans* model expressing 29 G_4C_2 repeats (Wang, Hao et al. 2016), the presence of RNA foci and RAN proteins was not determined. However, in the *C. elegans* model expressing 66 G_4C_2 repeats (Kramer, Carlomagno et al. 2016), worms exhibited sense strand RNA foci, a feature also observed in human patient cells harboring expanded repeats. Lysates from these worms contain enhanced levels of the RAN product poly-GP (Kramer, Carlomagno et al. 2016), although the cellular and subcellular *in vivo* expression pattern of this RAN product, which is reported to be non-toxic and soluble in other systems (Wen, Tan et al. 2014) was not reported. It is also not clear whether this poly-GP RAN protein was produced from the sense strand or the antisense strand or whether the production was repeat-length dependent. Whether additional RAN products were produced, including sense strand poly-GA and poly-GR or anti-sense strand poly PA and poly-PR was also not

described. Despite these open questions, the available data do support the conclusion that *C. elegans* expressing a pathogenic repeat expansion can execute RAN translation *in vivo*.

Given that RAN translation of the G_4C_2 repeat expansion occurs in yeast(Kramer, Carlomagno et al. 2016), *C. elegans(Kramer, Carlomagno et al. 2016)*, *Drosophila* (Freibaum, Lu et al. 2015), and humans, the general mechanism(s) of RAN translation are likely to be highly conserved. It also seems likely that there will be sequence-specific mechanisms governing the regulation of RAN translation among distinct repeat expansion sequences. For example, in DM2, the CCTG repeat expansion is sequestered in nuclear RNA foci in part by association with the RNA binding protein Muscleblind (Zu, Cleary et al. 2017). This interaction prevents cytoplasmic accumulation of the repeat expansion DM2 RNA and subsequent RAN translation and accumulation of toxic RAN proteins. However, Muscleblind does not regulate nuclear sequestration or RAN translation of G_4C_2 RNA, suggesting that distinct mechanisms regulate the sequestration and/or RAN translation of these repeat expansion sequences. Given its tremendous genetic advantages, as well as its ability to examine these events in neuronal and ageing contexts within a live animal, *C. elegans* could be powerful system for defining the general and sequence-specific determinants of RAN translation across a wide array of repeat expansions.

Conclusion

As these examples show, *C. elegans* is a well-established system for modeling repeat expansion disorders. Conclusions drawn from these models are largely consistent with findings from other models, including human patients, and suggest that discoveries made in worms should not be dismissed as trivial 'worm-specific' mechanisms. Like yeast, flies, cell culture, and human iPSC cells, worms act as a powerful and complementary model to other established systems and can illuminate aspects of disease toxicity that are either difficult or impossible to study using other approaches. Like any system, *C. elegans* disease models have their own unique strengths and weaknesses that must be considered when interpreting their relevance to disease pathology. However, history has shown that worm models have much to contribute to our understanding of repeat expansion disorders and it would be misguided to dismiss these efforts as irrelevant. Revisiting the many models of repeat expansion mutations in the context of new discoveries, such as that of RAN translation, may bring added relevance to these models and enhance their continued utility as genetic models of human disease.

References

(1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72**(6): 971-983.
Ashkenazi A., Bento C.F., Ricketts T., Vicinanza M., Siddiqi F., Pavel M., et al. (2017). Polyglutamine tracts regulate beclin 1-dependent autophagy. *Nature* **545**(7652): 108-111.
Banez-Coronel M., Ayhan F., Tarabochia A.D., Zu T., Perez B.A., Tusi S.K., et al. (2015). RAN Translation in Huntington Disease. *Neuron* **88**(4): 667-677.
Bird T.D. (1993). Myotonic Dystrophy Type 1. *GeneReviews((R))*. M. P. Adam, H. H. Ardinger, R. A. Pagon et al. Seattle (WA).
Boeynaems S., Bogaert E., Michiels E., Gijselinck I., Sieben A., Jovicic A., et al. (2016). Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci Rep* **6**: 20877.
Brignull H.R., Moore F.E., Tang S.J. and Morimoto R.I. (2006). Polyglutamine proteins at the pathogenic threshold display neuron-specific aggregation in a pan-neuronal Caenorhabditis elegans model. *J Neurosci* **26**(29): 7597-7606.
Brook J.D., McCurrach M.E., Harley H.G., Buckler A.J., Church D., Aburatani H., 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* **69**(2): 385.
Burgess R.W., Snutch T.P., Liu P., Wang X., Hao L., Saur T., et al. (2016). Forward Genetic Screen in Caenorhabditis elegans Suggests F57A10.2 and acp-4 As Suppressors of C9ORF72 Related Phenotypes. *Mol. Neurosci* **9**.
Cappellano G., Carecchio M., Fleetwood T., Magistrelli L., Cantello R., Dianzani U., et al. (2013). Immunity and inflammation in neurodegenerative diseases. *Am J Neurodegener Dis* **2**(2): 89-107.
Chen K.-Y., Pan H., Lin M.-J., Li Y.-Y., Wang L.-C., Wu Y.-C., et al. (2007). Length-dependent toxicity of untranslated CUG repeats on Caenorhabditis elegans. *Biochemical and Biophysical Research Communications* **352**: 774-779.
Christie N.T.M., Lee A.L., Fay H.G., Gray A.A. and Kikis E.A. (2014). Novel polyglutamine model uncouples proteotoxicity from aging. *PloS one* **9**: e96835.
Chung D.W., Rudnicki D.D., Yu L. and Margolis R.L. (2011). A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Hum Mol Genet* **20**(17): 3467-3477.
Cleary J.D. and Ranum L.P. (2014). Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Curr Opin Genet Dev* **26**: 6-15.
Colbert H.A. and Bargmann C.I. (1995). Odorant-specific adaptation pathways generate olfactory plasticity in C. elegans. *Neuron* **14**(4): 803-812.
Cooper-Knock J., Shaw P.J. and Kirby J. (2014). The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype. *Acta Neuropathol* **127**(3): 333-345.

Dalton J.C., Ranum L.P.W. and Day J.W. (1993). Myotonic Dystrophy Type 2. *GeneReviews((R))*. M. P. Adam, H. H. Ardinger, R. A. Pagon et al. Seattle (WA).
Davidson Y.S., Barker H., Robinson A.C., Thompson J.C., Harris J., Troakes C., et al. (2014). Brain distribution of dipeptide repeat proteins in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathol Commun* **2**: 70.
DeJesus-Hernandez M., Mackenzie I.R., Boeve B.F., Boxer A.L., Baker M., Rutherford N.J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**(2): 245-256.
Donnelly C.J., Zhang P.W., Pham J.T., Haeusler A.R., Mistry N.A., Vidensky S., et al. (2013). RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**(2): 415-428.
Ewbank J.J. and Pujol N. (2016). Local and long-range activation of innate immunity by infection and damage in C. elegans. *Curr Opin Immunol* **38**: 1-7.
Faber P.W., Alter J.R., MacDonald M.E. and Hart A.C. (1999). Polyglutamine-mediated dysfunction and apoptotic death of a Caenorhabditis elegans sensory neuron. *Proc Natl Acad Sci U S A* **96**(1): 179-184.
Fardghassemi Y., Tauffenberger A., Gosselin S. and Parker J.A. (2017). Rescue of ATXN3 neuronal toxicity in C. elegans by chemical modification of ER stress. *Disease models & mechanisms*: dmm.029736.
Freibaum B.D., Lu Y., Lopez-Gonzalez R., Kim N.C., Almeida S., Lee K.H., et al. (2015). GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**(7567): 129-133.
Fu Y.H., Pizzuti A., Fenwick R.G., Jr., King J., Rajnarayan S., Dunne P.W., et al. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**(5049): 1256-1258.
Galloway J.N. and Nelson D.L. (2009). Evidence for RNA-mediated toxicity in the fragile X-associated tremor/ataxia syndrome. *Future Neurol* **4**(6): 785.
Gidalevitz T., Wang N., Deravaj T., Alexander-Floyd J. and Morimoto R.I. (2013). Natural genetic variation determines susceptibility to aggregation or toxicity in a C. elegans model for polyglutamine disease. *BMC Biol* **11**: 100.
Green K.M., Glineburg M.R., Kearse M.G., Flores B.N., Linsalata A.E., Fedak S.J., et al. (2017). RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat Commun* **8**(1): 2005.
Green K.M., Linsalata A.E. and Todd P.K. (2016). RAN translation-What makes it run? *Brain Res* **1647**: 30-42.
Haeusler A.R., Donnelly C.J., Periz G., Simko E.A., Shaw P.G., Kim M.S., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**(7491): 195-200.
Hart A.C., Kass J., Shapiro J.E. and Kaplan J.M. (1999). Distinct signaling pathways mediate touch and osmosensory responses in a polymodal sensory neuron. *J Neurosci* **19**(6): 1952-1958.
Hautbergue G.M., Castelli L.M., Ferraiuolo L., Sanchez-Martinez A., Cooper-Knock J., Higginbottom A., et al. (2017). SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nat Commun* **8**: 16063.

Holmes S.E., O'Hearn E.E., McInnis M.G., Gorelick-Feldman D.A., Kleiderlein J.J., Callahan C., et al. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat Genet* **23**(4): 391-392.
Hsu A.L., Murphy C.T. and Kenyon C. (2003). Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* **300**(5622): 1142-1145.
Ilieva H., Polymenidou M. and Cleveland D.W. (2009). Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J Cell Biol* **187**(6): 761-772.
Ingram M.A., Orr H.T. and Clark H.B. (2012). Genetically engineered mouse models of the trinucleotide-repeat spinocerebellar ataxias. *Brain Res Bull* **88**(1): 33-42.
Jeong H., Then F., Melia T.J., Jr., Mazzulli J.R., Cui L., Savas J.N., et al. (2009). Acetylation targets mutant huntingtin to autophagosomes for degradation. *Cell* **137**(1): 60-72.
Jia H., Pallos J., Jacques V., Lau A., Tang B., Cooper A., et al. (2012). Histone deacetylase (HDAC) inhibitors targeting HDAC3 and HDAC1 ameliorate polyglutamine-elicited phenotypes in model systems of Huntington's disease. *Neurobiol Dis* **46**(2): 351-361.
Jia K., Hart A.C. and Levine B. (2007). Autophagy genes protect against disease caused by polyglutamine expansion proteins in Caenorhabditis elegans. *Autophagy* **3**(1): 21-25.
Jovicic A., Mertens J., Boeynaems S., Bogaert E., Chai N., Yamada S.B., et al. (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18**(9): 1226-1229.
Juang B.T., Ludwig A.L., Benedetti K.L., Gu C., Collins K., Morales C., et al. (2014). Expression of an expanded CGG-repeat RNA in a single pair of primary sensory neurons impairs olfactory adaptation in Caenorhabditis elegans. *Hum Mol Genet* **23**(18): 4945-4959.
Kawaguchi Y., Okamoto T., Taniwaki M., Aizawa M., Inoue M., Katayama S., et al. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* **8**(3): 221-228.
Kearse M.G., Green K.M., Krans A., Rodriguez C.M., Linsalata A.E., Goldstrohm A.C., et al. (2016). CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**(2): 314-322.
Khan L.A., Bauer P.O., Miyazaki H., Lindenberg K.S., Landwehrmeyer B.G. and Nukina N. (2006). Expanded polyglutamines impair synaptic transmission and ubiquitin-proteasome system in Caenorhabditis elegans. *J Neurochem* **98**(2): 576-587.
Kitamura A., Kubota H., Pack C.G., Matsumoto G., Hirayama S., Takahashi Y., et al. (2006). Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. *Nat Cell Biol* **8**(10): 1163-1170.
Kramer N.J., Carlomagno Y., Zhang Y.J., Almeida S., Cook C.N., Gendron T.F., et al. (2016). Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. *Science* **353**(6300): 708-712.
Krans A., Kearse M.G. and Todd P.K. (2016). Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Ann Neurol* **80**(6): 871-881.

L'Etoile N.D., Coburn C.M., Eastham J., Kistler A., Gallegos G. and Bargmann C.I. (2002). The cyclic GMP-dependent protein kinase EGL-4 regulates olfactory adaptation in C. elegans. *Neuron* **36**(6): 1079-1089.
La Spada A.R., 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**(6330): 77-79.
Lee A.L., Ung H.M., Sands L.P. and Kikis E.A. (2017). A new Caenorhabditis elegans model of human huntingtin 513 aggregation and toxicity in body wall muscles. *PLoS One* **12**(3): e0173644.
Lee K.H., Zhang P., Kim H.J., Mitrea D.M., Sarkar M., Freibaum B.D., et al. (2016). C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **167**(3): 774-788 e717.
Lee Y.B., Chen H.J., Peres J.N., Gomez-Deza J., Attig J., Stalekar M., et al. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep* **5**(5): 1178-1186.
Legendre-Guillemin V., Metzler M., Lemaire J.F., Philie J., Gan L., Hayden M.R., et al. (2005). Huntingtin interacting protein 1 (HIP1) regulates clathrin assembly through direct binding to the regulatory region of the clathrin light chain. *J Biol Chem* **280**(7): 6101-6108.
Lejeune F.X., Mesrob L., Parmentier F., Bicep C., Vazquez-Manrique R.P., Parker J.A., et al. (2012). Large-scale functional RNAi screen in C. elegans identifies genes that regulate the dysfunction of mutant polyglutamine neurons. *BMC Genomics* **13**: 91.
Ling S.C., Polymenidou M. and Cleveland D.W. (2013). Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**(3): 416-438.
Liquori C.L., Ricker K., Moseley M.L., Jacobsen J.F., Kress W., Naylor S.L., et al. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **293**(5531): 864-867.
Lopes-Cendes I., Teive H.G., Calcagnotto M.E., Da Costa J.C., Cardoso F., Viana E., et al. (1997). Frequency of the different mutations causing spinocerebellar ataxia (SCA1, SCA2, MJD/SCA3 and DRPLA) in a large group of Brazilian patients. *Arq Neuropsiquiatr* **55**(3B): 519-529.
Mackenzie I.R., Arzberger T., Kremmer E., Troost D., Lorenzl S., Mori K., et al. (2013). Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* **126**(6): 859-879.
Mahadevan M., Tsilfidis C., Sabourin L., Shutler G., Amemiya C., Jansen G., et al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**(5049): 1253-1255.
Mankodi A., Urbinati C.R., Yuan Q.P., Moxley R.T., Sansone V., Krym M., et al. (2001). Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum Mol Genet* **10**(19): 2165-2170.
McIntire S.L., Jorgensen E. and Horvitz H.R. (1993). Genes required for GABA function in Caenorhabditis elegans. *Nature* **364**(6435): 334-337.
Metzler M., Legendre-Guillemin V., Gan L., Chopra V., Kwok A., McPherson P.S., et al. (2001). HIP1 functions in clathrin-mediated endocytosis through binding to clathrin and adaptor protein 2. *J Biol Chem* **276**(42): 39271-39276.

Mizielinska S., Gronke S., Niccoli T., Ridler C.E., Clayton E.L., Devoy A., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**(6201): 1192-1194.
Mori K., Nihei Y., Arzberger T., Zhou Q., Mackenzie I.R., Hermann A., et al. (2016). Reduced hnRNPA3 increases C9orf72 repeat RNA levels and dipeptide-repeat protein deposition. *EMBO Rep* **17**(9): 1314-1325.
Morley J.F., Brignull H.R., Weyers J.J. and Morimoto R.I. (2002). The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **99**(16): 10417-10422.
Morley J.F. and Morimoto R.I. (2004). Regulation of longevity in Caenorhabditis elegans by heat shock factor and molecular chaperones. *Mol Biol Cell* **15**(2): 657-664.
Moronetti Mazzeo L.E., Dersh D., Boccitto M., Kalb R.G. and Lamitina T. (2012). Stress and aging induce distinct polyQ protein aggregation states. *Proc Natl Acad Sci U S A* **109**(26): 10587-10592.
Nollen E.A., Garcia S.M., van Haaften G., Kim S., Chavez A., Morimoto R.I., et al. (2004). Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation. *Proc Natl Acad Sci U S A* **101**(17): 6403-6408.
Pandey U.B., Nie Z., Batlevi Y., McCray B.A., Ritson G.P., Nedelsky N.B., et al. (2007). HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. *Nature* **447**(7146): 859-863.
Parker J.A., Connolly J.B., Wellington C., Hayden M., Dausset J. and Neri C. (2001). Expanded polyglutamines in Caenorhabditis elegans cause axonal abnormalities and severe dysfunction of PLM mechanosensory neurons without cell death. *Proc Natl Acad Sci U S A* **98**(23): 13318-13323.
Parker J.A., Metzler M., Georgiou J., Mage M., Roder J.C., Rose A.M., et al. (2007). Huntingtin-interacting protein 1 influences worm and mouse presynaptic function and protects Caenorhabditis elegans neurons against mutant polyglutamine toxicity. *J Neurosci* **27**(41): 11056-11064.
Paulson H.L., Shakkottai V.G., Clark H.B. and Orr H.T. (2017). Polyglutamine spinocerebellar ataxias - from genes to potential treatments. *Nat Rev Neurosci* **18**(10): 613-626.
Ranum L.P. and Cooper T.A. (2006). RNA-mediated neuromuscular disorders. *Annu Rev Neurosci* **29**: 259-277.
Renton A.E., Majounie E., Waite A., Simon-Sanchez J., Rollinson S., Gibbs J.R., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**(2): 257-268.
Rudich P., Snoznik C., Watkins S.C., Monaghan J., Pandey U.B. and Lamitina S.T. (2017). Nuclear localized C9orf72-associated arginine-containing dipeptides exhibit age-dependent toxicity in C. elegans. *Hum Mol Genet* **26**(24): 4916-4928.
Sandford E. and Burmeister M. (2014). Genes and genetic testing in hereditary ataxias. *Genes (Basel)* **5**(3): 586-603.

Satyal S.H., Schmidt E., Kitagawa K., Sondheimer N., Lindquist S., Kramer J.M., et al. (2000). Polyglutamine aggregates alter protein folding homeostasis in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **97**(11): 5750-5755.
Saul R.A. and Tarleton J.C. (1993). FMR1-Related Disorders. *GeneReviews((R))*. M. P. Adam, H. H. Ardinger, R. A. Pagon et al. Seattle (WA).
Schludi M.H., May S., Grässer F.A., Rentzsch K., Kremmer E., Küpper C., et al. (2015). Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta neuropathologica* **130**: 537-555.
Scoles D.R., Ho M.H., Dansithong W., Pflieger L.T., Petersen L.W., Thai K.K., et al. (2015). Repeat Associated Non-AUG Translation (RAN Translation) Dependent on Sequence Downstream of the ATXN2 CAG Repeat. *PLoS One* **10**(6): e0128769.
Silveira I., Lopes-Cendes I., Kish S., Maciel P., Gaspar C., Coutinho P., et al. (1996). Frequency of spinocerebellar ataxia type 1, dentatorubropallidoluysian atrophy, and Machado-Joseph disease mutations in a large group of spinocerebellar ataxia patients. *Neurology* **46**(1): 214-218.
Tavernarakis N. and Driscoll M. (2001). Mechanotransduction in Caenorhabditis elegans: the role of DEG/ENaC ion channels. *Cell Biochem Biophys* **35**(1): 1-18.
Teixeira-Castro A., Ailion M., Jalles A., Brignull H.R., Vilaca J.L., Dias N., et al. (2011). Neuron-specific proteotoxicity of mutant ataxin-3 in C. elegans: rescue by the DAF-16 and HSF-1 pathways. *Hum Mol Genet* **20**(15): 2996-3009.
Teuling E., Bourgonje A., Veenje S., Thijssen K., de Boer J., van der Velde J., et al. (2011). Modifiers of mutant huntingtin aggregation: functional conservation of C. elegans-modifiers of polyglutamine aggregation. *PLoS Curr* **3**: RRN1255.
Todd Peter K., Oh Seok Y., Krans A., He F., Sellier C., Frazer M., et al. (2013). CGG Repeat-Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome. *Neuron* **78**: 440-455.
Todd P.K., Oh S.Y., Krans A., He F., Sellier C., Frazer M., et al. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**(3): 440-455.
Tran H., Almeida S., Moore J., Gendron T.F., Chalasani U., Lu Y., et al. (2015). Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* **87**(6): 1207-1214.
Van Assche R., Borghgraef C., Vaneyck J., Dumoulin M., Schoofs L. and Temmerman L. (2017). In vitro aggregating beta-lactamase-polyQ chimeras do not induce toxic effects in an in vivo Caenorhabditis elegans model. *J Negat Results Biomed* **16**(1): 14.
Verkerk A.J., Pieretti M., Sutcliffe J.S., Fu Y.H., Kuhl D.P., Pizzuti A., 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**(5): 905-914.
Voisine C., Varma H., Walker N., Bates E.A., Stockwell B.R. and Hart A.C. (2007). Identification of potential therapeutic drugs for huntington's disease using Caenorhabditis elegans. *PLoS One* **2**(6): e504.
Wang L.-C., Chen K.-Y., Pan H., Wu C.-C., Chen P.-H., Liao Y.-T., et al. (2011). Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in Caenorhabditis elegans. *Cellular and Molecular Life Sciences* **68**: 1255-1267.

Wang X., Hao L., Saur T., Joyal K., Zhao Y., Zhai D., et al. (2016). Forward Genetic Screen in Caenorhabditis elegans Suggests F57A10.2 and acp-4 As Suppressors of C9ORF72 Related Phenotypes. *Front Mol Neurosci* **9**: 113.
Warby S.C., Graham R.K. and Hayden M.R. (1993). Huntington Disease. *GeneReviews((R))*. M. P. Adam, H. H. Ardinger, R. A. Pagon et al. Seattle (WA).
Wellington C.L., Ellerby L.M., Gutekunst C.A., Rogers D., Warby S., Graham R.K., et al. (2002). Caspase cleavage of mutant huntingtin precedes neurodegeneration in Huntington's disease. *J Neurosci* **22**(18): 7862-7872.
Wen X., Tan W., Westergard T., Krishnamurthy K., Markandaiah S.S., Shi Y., et al. (2014). Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron* **84**(6): 1213-1225.
Wexler N.S., Lorimer J., Porter J., Gomez F., Moskowitz C., Shackell E., et al. (2004). Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci U S A* **101**(10): 3498-3503.
Zhang K., Donnelly C.J., Haeusler A.R., Grima J.C., Machamer J.B., Steinwald P., et al. (2015). The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**(7567): 56-61.
Zu T., Cleary J.D., Liu Y., Banez-Coronel M., Bubenik J.L., Ayhan F., et al. (2017). RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. *Neuron* **95**(6): 1292-1305 e1295.
Zu T., Gibbens B., Doty N.S., Gomes-Pereira M., Huguet A., Stone M.D., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* **108**(1): 260-265.
Zu T., Liu Y., Banez-Coronel M., Reid T., Pletnikova O., Lewis J., 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**(51): E4968-4977.
Zu T., Liu Y., Bañez-Coronel M., Reid T., Pletnikova O., Lewis J., et al. (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proceedings of the National Academy of Sciences of the United States of America* **110**: E4968-4977.

Table 1. *C. elegans* models of repeat expansion disorders

Disease gene	Repeat	Additional context [1]	Location[2]	Tissue[3]	Agg.[4]	Phenotype[5]	RAN[6]	Ref.
Htt	CAG	Yes	Coding	ASH sensory neurons	Yes	Dyf	N.D.	(Faber, Alter et al. 1999)
Htt	CAG	Yes	Coding	touch sensory neurons	Yes	Mec	N.D.	(Parker, Connolly et al. 2001)
Htt	CAG	No	Coding	Muscle	Yes	Age-dependent Unc	N.D.	(Morley, Brignull et al. 2002)
Htt	CAG	No	Coding	All neurons	Yes	Unc	N.D.	(Brignull, Moore et al. 2006)
Htt	CAG	Yes	Coding	Muscle	Yes	Age independent Unc, shortened lifespan	N.D.	(Lee, Ung et al. 2017)
SCA3	CAG	Yes	Coding	All neurons	Yes	Age-dependent Unc	N.D.	(Khan, Bauer et al. 2006)
SCA3	CAG	Yes	Coding	All neurons	Yes	Let, Unc, Egl, Exp,	N.D.	(Khan, Bauer et al. 2006)
SCA3	CAG	Yes	Coding	All neurons	Yes	Gro, Unc, Tax	N.D.	(Teixeira-Castro, Ailion et al. 2011)
SCA3	CAG	Yes	Coding	All neurons	Yes	Unc, Tax	N.D.	(Teixeira-Castro, Ailion et al. 2011)
SCA3	CAG	Yes	Coding	Muscle	Yes	Unc	N.D.	(Christie, Lee et al. 2014)
SCA3	CAG	Yes	Coding	GABA neurons	Yes	Unc, Shortened lifespan,	N.D.	(Fardghassemi, Tauffenberger et al. 2017)
SCA12	CAG	No	Non-coding	Muscle	N/A	Let, Unc, Shortened lifespan	N.D.	(Wang, Chen et al. 2011)
FMR1	CGG	Yes	Non-coding	AWC sensory neurons	N/A	Tax	N.D.	(Juang, Ludwig et al. 2014)
DM1	CTG	No	Non-coding	Muscle	N/A	Let, Unc, Shortened lifespan	N.D.	(Chen, Pan et al. 2007)
C9orf72	G_4C_2	Unk[7]	N.D.	All neurons	N/A	Shortened lifespan	Yes	(Kramer, Carlomagno et al. 2016)
C9orf72	G_4C_2	Yes	Non-coding	All neurons	N.D.	Unc, Shortened lifespan	N.D.	(Wang, Hao et al. 2016)

C9orf72	PR, GR, GA, PA	No	Coding	Muscle / GABA neurons	No[8]	Let, Unc	N.R.[9]	(Rudich, Snoznik et al. 2017)

[1] Additional human genetic sequence surrounding the repeat expansion
[2] Location of the repeat sequence in coding (exon) or non-coding (5' UTR, 3'UTR, intron) regions of the transgene
[3] Site of tissue expression
[4] Presence of protein aggregates
[5] Phenotype of model, 'Dyf' – Dye-filling defective neuron; 'Mec' – Mechanosensory defective; 'Unc' – uncoordinated motility defect; 'Let' – embryonic/larval lethal; 'Egl' – egg laying defective; 'Exp' – defective in expulsion phase of defecation cycle; 'Gro' – slow growing; 'Tax' – chemotaxis defective
[6] RAN translation products detected, 'N.D.' – Not determined
[7] The presence or absence of additional sequence context was not described
[8] The toxic dipeptides (GR,PR) did not aggregate, but one nontoxic dipeptide (GA) did aggregate
[9] 'N.R.' – Not relevant. Codon varied transcripts do not contain nucleotide repeat expansions and are not a potential substrate for RAN translation

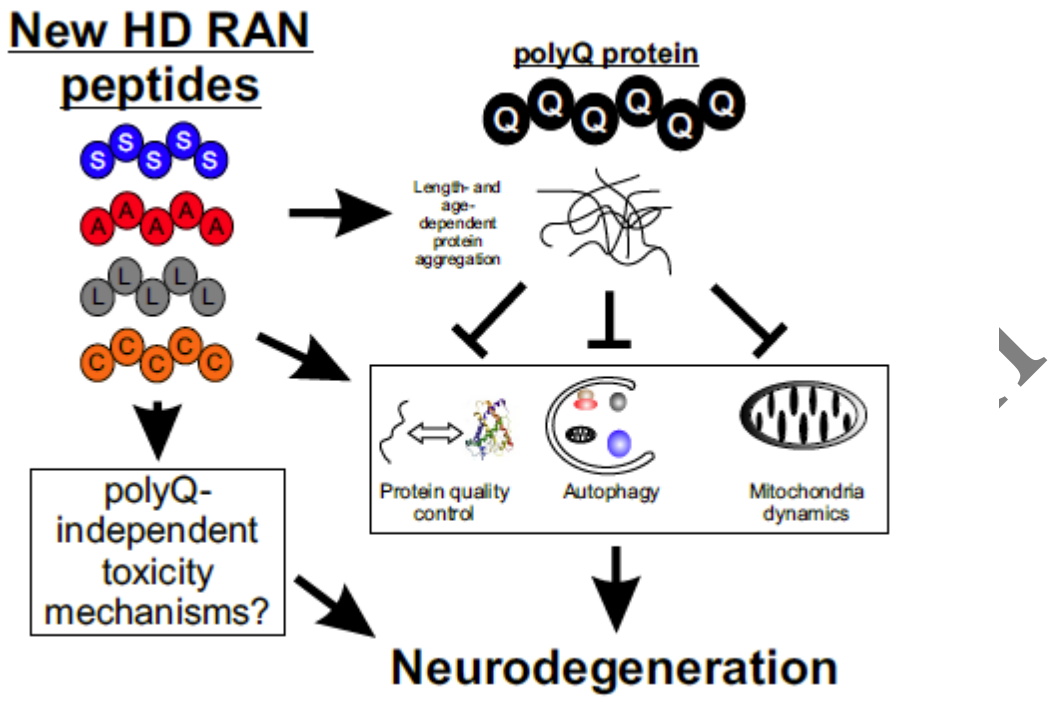

Figure 1. Newly discovered HD RAN peptides may contribute to the pathogenesis of Huntington's disease and other CAG repeat repeat expansion disorders. PolyQ proteins in *C. elegans* and other models exert toxicity through multiple cellular pathways including disruptions in protein quality control, autophagy, and mitochondrial function. RAN translation of the CAG repeat gives rise to four additional peptides which could contribute to HD neurodegeneration, either by enhancing polyQ toxicity, inhibiting known polyQ targets, or acting via novel polyQ-independent pathways.