Defining the Properties and Toxicity Mechanisms of Non-Canonical Translation Products Associated with ALS and Huntington's Disease Using *C. elegans*

by

Paige Davison Rudich

B.S., Carnegie Mellon University, 2014

Submitted to the Graduate Faculty of

the School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented by

by

Paige Davison Rudich

It was defended on

May 15, 2019

and approved by

Jeffrey Brodsky, Professor and Avinoff Chair, Department of Biological Sciences

Michael J. Palladino, Professor and Vice Chair of Academics, Department of Pharmacology and Chemical Biology

Udai Pandey, Associate Professor, Department of Human Genetics

Michael Tsang, Associate Professor, Department of Developmental Biology

Judith Yanowitz, Associate Professor, Department of Obstetrics, Gynecology, and Reproductive Sciences

Dissertation Director: S. Todd Lamitina, Associate Professor, Department of Cell Biology

Defining the Properties and Toxicity Mechanisms of Non-Canonical Translation Products Associated with ALS and Huntington's Disease Using *C. elegans*

Paige Davison Rudich, Ph.D.

University of Pittsburgh, 2019

Expanded guanine/cytosine (G/C)-rich nucleotide repeats are the underlying genetic cause of many age-dependent neurodegenerative diseases. An emerging mechanism underlying disease pathology is an unusual type of protein translation called Repeat-Associated non-AUG (RAN) translation. RAN translation requires extended G/C-rich repeats and occurs independently of a canonical start codon, allowing translation in all three reading frames. Antisense RNA from G/C-rich repeats also gives rise to RAN products, causing up to six distinct protein products from one repeat expansion. The toxicity of RAN products *in vivo* is beginning to be explored. We created codon-varied RAN product models for two different repeats: an intronic GGGGCC repeat expansion that is the most common genetic cause of Amyotrophic Lateral Sclerosis (ALS), and the CAG repeat expansion that causes Huntington's disease (HD).

In our *C. elegans* ALS model, we discovered that the arginine-containing dipeptides, proline-arginine (PR) and glycine-arginine (GR), were the most toxic dipeptides. PR and GR exhibited age-dependent toxicity when expressed in multiple cell types, including motor neurons. Both PR and GR exhibited nuclear localization that was necessary for toxicity. An unbiased whole-genome RNAi screen for suppressors of PR toxicity identified twelve genes. Four of the genes were previously identified in PR modifier screens performed in other systems, suggesting mechanisms of PR toxicity are conserved.

My studies are the first to model codon-varied CAG RAN polypeptides in a multicellular animal. Every polypeptide, except for polyLeucine (polyLeu), formed immobile protein aggregates at ~38 repeats. Surprisingly, polyLeu was the most toxic HD RAN polypeptide in every tissue studied. A forward mutagenesis screen combined with a candidate RNAi screen identified three genes that suppressed polyLeu toxicity. Two of the genes encoded transmembrane proteins, and the third gene encoded a

deubiquitinase enzyme, suggesting that polyLeu toxicity occurs via disrupted folding of transmembrane proteins. Understanding how different RAN products contribute to HD and ALS is vital for developing appropriate treatments for these diseases.

Table of Contents

List of Tables

List of Figures

Preface

This work would not have been possible without the help of my friends, both in lab and out of lab. I am extremely lucky to have found such a great lab and community. Along with the members of my committee, for being supportive and giving practical advice, I would like to thank some key people.

I would like to thank Todd Lamitina who was the perfect mentor for me. Todd gave me the freedom to switch the aim of my thesis halfway through my PhD research. He was supportive as I struggled through characterizing the peptides and had great ideas for experiments to determine what was happening. I learned so much from all the spontaneous hour-long "Todd Talks" about statistics, baseball, and worm history.

I would also like to thank my two lab mates, Sarel and Carley, who helped both physically, with experiments, and emotionally, with commiserating about the craziness of biology. Sarel helped me realize I was not alone and laughed about all the ridiculous parts of this path we have both decided to take.

I would also like to thank Francis, another worm compatriot, who made sure I also took time to enjoy life. He forced me to become friends with him and I value that bond. He was instrumental in causing Worm Club Happy Hours and making sure we all had some level of skill in both darts and tennis.

I would like to thank my family who gave me the drive required to complete a PhD. My mother, Kathy Lucchesi, always supported my quest for knowledge, taking me as a kid to the library for the next thirty books I wanted to devour, and supporting me going to undergrad across the country. My father, Ron Davison, and step-mother, Sue Spivey,

were always there for me and I knew whatever happened, they would have my back. My brothers (Terry Kent, Brett Davison, and Ryan Davison) always pushed my limits and made me realize how different each person can be. While Ryan died in the middle of my graduate career, I will always feel his presence and the effect he has had on my life. I am lucky to have had him as a brother, I just wish it could have been for longer.

Most importantly, I want to thank my partner-in-life, Isaac. He helped me get through the stress of lab and the stresses of life. He listened to me agonize about lab life, fret over whether I belonged in science, and then forget all that nonsense and giddily talk ad nauseum whenever experiments actually worked. I could not have finished graduate school without his constant love and support, including his taking care of the laundry, doing grocery shopping when I was unavailable, and telling me for the thousandth time that yes, I did look good in whatever outfit I had picked out, and yes I would do a fantastic job with my talk/poster/interview etc. Thank you for being my best friend.

1.0 Introduction

Neurodegenerative diseases (NDDs) are affecting an expanding percentage of the world population [1]. This is due to an increase in the percentage of the population who is most at risk for NDDs, the aging population. There are twice as many people over the age of 60 alive today compared to 1980 [2], and that population is predicted to double again by 2050 [2]. This is a growing health crisis as there are no curative treatments for any NDD, and the majority of NDDs are fatal. Despite intensive research, the initiating cellular causes of most NDDs are unknown. To identify druggable targets, the known genetic causes of NDDs have been modeled and studied. However, the mechanism(s) by which these genetic mutations lead to disease is poorly understood.

One type of genetic mutation, nucleotide repeat expansions, is specifically associated with NDDs. Repetitive DNA or DNA derived from repeats [3] comprises up to 60% of the human genome. However, DNA repeats can expand to pathogenic lengths, which predominantly cause NDDs. The link between repeat expansions and NDDs is unclear, but a recent discovery may be the key. The Ranum lab discovered that guanine/cytosine (G/C)-rich nucleotide repeat expansions can undergo a unique type of translation, Repeat-Associated non-AUG (RAN) translation [4]. RAN translation occurs independently of a start codon and can consequently occur in any of the three possible reading frames [4]. Antisense RNA is commonly transcribed from nucleotide repeat expansions [5-7] and can also undergo RAN translation. Thus, six possible RAN products are produced from one repeat expansion [4]. RAN translation has been predominantly observed in neuronal tissue, and it is unclear if it occurs at similar levels in other tissues.

Either RAN translation itself or the RAN products may be why repeat expansions predominantly cause neurodegeneration.

To better define how RAN products cause toxicity, I focused my thesis research on RAN products from two prominent repeat expansions that cause NDDs. The first mutation, a GGGGCC (G_4C_2) repeat expansion in *chromosome 9 open reading frame 72* (*C9orf72)*, is the most common genetic cause of Amyotrophic Lateral Sclerosis (ALS), as well as Frontotemporal Dementia (FTD) [8, 9]. G_4C_2 RAN products can be highly toxic and are associated with neurodegeneration [10]. The second genetic mutation I focused on is a CAG repeat expansion, which causes thirteen different NDDs depending on the genetic location of the repeat expansion [11]. RAN products have been identified in three of these diseases, including the most commonly studied CAG repeat expansion disease, Huntington's disease (HD) [12]. My work creating and characterizing *C. elegans* models of RAN products from these two mutations will aid in identifying molecular pathways that may be relevant to patients.

1.1 Common Features of Repeat Expansions

Over 30 nucleotide repeat expansions cause age-dependent neurodegenerative or neuromuscular diseases (Appendix B-Table 4.1). Interestingly, repeat expansions are commonly the only mutation of the relevant gene that causes a neurodegenerative or neuromuscular disease. For example, the most common genetic cause of ALS, a G_4C_2 repeat expansion in *C9orf72*, is the only mutation in *C9orf72* that has been proven to cause ALS. In contrast, the second most common genetic cause of ALS is mutations in

superoxide dismutase 1 (*sod1*), with over 150 identified point mutations, insertions, deletions and truncations in *sod1* that cause ALS [13]. Therefore, nucleotide repeat expansions having toxic properties beyond disrupting their respective gene and nucleotide repeat expansions are innately toxic. To determine how the repeat expansions are toxic, it is first important to understand the common features of disease-causing repeat expansions.

1.1.1 High G/C Content Causes Repeat Instability

The most noticeable feature of toxic repeat expansions is that most have a high G/C content (Appendix B-Table 4.1). This attribute drives the instability of the repeats. Guanines and cytosines form strong bonds that can cause improper interactions between DNA and RNA during transcription [14] or mismatching between the sense and antisense strand of DNA (slipped-strand DNA) during DNA synthesis [15]. To resolve these improper interactions, DNA repair pathways are activated, such as the double-strand break repair pathway [16]. However, the repetitive nature of the DNA can cause defects in DNA repair and give rise to shrinkage or expansion of the DNA repeats. For example, G/C-rich DNA and its transcribed G/C-rich RNA can form hybrid structures called R loops [14]. While R loops are normally removed by RNA helicases, remaining R loops activate DNA repair mechanisms, which cause expansion or shrinkage of the repeat [14]. Further evidence of the interrelationship between DNA repair pathways and repeat expansions is that a mismatch repair factor, MSH3, is required for CAG repeat expansions within the pathological and pre-pathological range of repeats [17]. Genes in the double-strand break repair pathway have also been identified as modifiers of HD [18]. These repair

pathways can be activated and cause repeat expansions during mitosis, meiosis, or transcription [14, 19]. Interestingly, G/C-rich repeats are only unstable after the repeat passes a certain length. In the HD-associated CAG repeat, CAG repeats up to a length of 35 repeats exhibit high intergenerational stability. However, CAG repeats in the potentially pathogenic range of 36-39 CAG repeats exhibit low intergenerational stability and thus expand [20]. Although the precise mechanism(s) underlying pathogenic repeat expansion is not definitively established, the change in stability is likely due to expanded repeats being more likely to activate DNA repair pathways. Thus, expanded CAG repeats are susceptible to further repeat expansions during meiosis [21].

Most disease-causing repeat expansions are hereditary mutations [22, 23] or occur during germline replication [24]. Therefore, every cell contains the repeat expansion. The effects of germline expansion were documented in HD before the genetic cause of HD was known. Successive generations of HD patients had earlier onsets of the disease, known as genetic anticipation [25]. Genetic anticipation is now understood to be caused by repeat instability and intergenerational expansion of the CAG repeats, as described above. In some, but not all, repeat expansion diseases, repeat length is directly associated with disease onset and/or severity. For example, repeat length is inversely associated with the age of onset in HD [25]. However, repeat length is not a clearly established predictor of age-of-onset or disease severity in another repeat expansion disease, *C9orf72*-associated ALS/FTD [26-29].

Somatic repeat expansions can modify the severity or age of onset in diseases with repeat length associated toxicity, such as HD [30, 31]. A postmortem study of 48 HD patients found that every patient had somatic instability of the CAG repeat within the

mutant *huntingtin* (*mHTT*) in neurons [30]. The somatic repeat expansions ranged from 1 additional repeat to >55 additional repeats, but the majority of somatic repeat expansions were <10 repeats [30]. Larger somatic repeat expansions were correlated with an earlier age of onset. Somatic repeat instability and an earlier age of onset are both associated with mutations in double-strand break repair machinery [17, 18]. Unstable G/C-rich repeats can expand at any stage of development and those expansions can not only cause neurodegeneration but also modify the age of onset of NDD symptoms.

1.1.2 Repeat Expansions Can Occur in Any Genetic Context

Nucleotide repeat expansions happen in every genetic context and can be toxic even when located in introns or untranslated regions (UTRs). The first identified disease-associated repeat expansions were a CGG repeat expansion in a 5'UTR (fragile X syndrome (FXS) [32]) and a CAG repeat expansion in an exon (spinal-bulbar muscular atrophy (SBMA) [33]). Since then, 18 exonic repeat expansions and 16 intronic/UTR localized repeat expansions have been identified (Appendix B-Table 4.1). Exonic repeat expansions are either polyglutamine-encoding CAG repeat expansions ('polyQ' diseases) or polyalanine-encoding GCC repeat expansions. Unlike exonic CAG repeat expansions and most intronic/UTR repeat expansions, polyalanine (GCC) repeat expansions cause developmental defects and do not cause neurodegeneration. For these reasons, polyalanine diseases are likely toxic through distinct mechanisms compared to exonic CAG repeat expansions and untranslated repeat expansions.

In contrast, exonic CAG repeat expansions and untranslated repeat expansions share many features, which will be discussed in-depth in the following sections. They exhibit age-dependent toxicity of the associated NDDs, even though the repeat expansions are present embryonically. The length of the repeat expansions can also modify the age of onset for many of these diseases. Antisense RNA transcription (see below) occurs in both exonic repeat expansions and untranslated repeat expansions. Due to these shared features, it is likely that exonic CAG repeat expansions and untranslated repeat expansions have some shared pathological mechanisms

1.1.3 Bidirectional Transcription of Repeat Expansions

Bidirectional transcription of nucleotide repeat expansions, producing both sense and antisense RNA, was first noted in 2005 with the 5' CTG repeat expansion that causes myotonic dystrophy type 1 (DM1) [7]. Interestingly, the bidirectional transcription was not dependent on the length of the repeat expansion, as similar levels of antisense RNA were transcribed in wild-type cells. Antisense RNA has since been identified in most nucleotide repeat expansions, including HD [5] and *C9orf72* [6]. In some repeat expansion diseases, such as HD, the repeat expansion causes a decrease in antisense RNA [5], while in other diseases, such as *C9orf72*-ALS, the repeat expansion causes an increase in antisense RNA [34]. Therefore, repeat expansions can be modifiers of bidirectional transcription, but the mechanism of modification is not consistent between diseases or well understood.

Bidirectional transcription occurs in 10-40% of protein-coding genes in the human genome [35, 36] and is not unique to pathogenic repeat expansions [35, 36]. Bidirectional transcription is a general feature observed in most, if not all, cells. However, the genes

undergoing bidirectional transcription differ depending on cell type [36]. As native antisense RNA production is widespread, the antisense RNA is likely to serve a biological function. The role of native antisense RNA is varied and can either inhibit or stimulate transcription of the corresponding sense RNA [5, 37] or RNA with high homology to the sense RNA [38]. Antisense RNA is well known to suppress the expression of the sense RNA through RNAi, but the other regulatory methods of antisense RNA are not as well understood. Antisense RNA can stabilize the sense RNA or spliced versions of the sense RNA, leading to increased expression of spliced versions of the gene. Antisense RNA does this through binding to and 'masking' the sense RNA to prevent binding of splicing elements or other regulatory elements [39, 40]. Most commonly, antisense RNA causes silencing of the sense gene through epigenetic changes, such as methylation of promoters or heterochromatin modifications [7, 41, 42]. Depending on the disease, repeat expansions can either increase or decrease the levels of native antisense RNA, causing the epigenetic effects of the antisense RNA to cause a corresponding increase in sense RNA production (HD [5]) or decrease in sense RNA production (FXS [42]). The biological role of native antisense RNA is just beginning to be understood, and the ramifications of alterations to the endogenous level of antisense RNA need to be individually studied with each NDD-associated repeat expansion.

1.2 Shared Disease Characteristics of Repeat-Expansion-Associated NDDs

Like the disease-causing repeat expansions, the diseases caused by repeat expansions also have shared characteristics. Repeat expansion associated NDDs

typically occur during advanced age, even though the causative mutation is present since birth. The length of the repeat expansion can also modify the age when the NDD occurs. Finally, the repeat expansion, which is present in every cell, only causes toxicity in a subset of neurons. Many mechanisms have been identified that may lead to the temporal and tissue specificity of these NDDs. However, it is still unclear which feature or combination of features is the causative mechanism.

1.2.1 Age-Associated Toxicity of Repeat Expansions

While repeat expansions are present from before birth, they do not typically cause neurodegeneration until middle to late age (Appendix B-Table 4.1). The delayed manifestation of the diseases with age could be due to many features of aging neurons. Two possible factors are that aging neurons accumulate somatic mutations [43] and that aging neurons have increased damage from reactive oxygen species (ROS) [44]. Both issues can increase levels of misfolded proteins by altering the protein structure directly (somatic mutations) or by altering the protein folding environment (oxidative stress induced global protein damage due to ROS). These and possibly other factors may contribute to why aging causes an increase in misfolded proteins and protein aggregates [45]. Other effects of aging likely contribute to neurodegeneration beyond inhibiting protein folding. However, the relationship between aging, neurodegeneration, and misfolded proteins is a well-established feature of most NDDs.

Misfolded proteins and protein aggregates are pathological characteristics of many neurodegenerative diseases. For example, polyQ aggregates form in HD [21], TDP-43 aggregates form in ALS [46], and amyloid plaques form in Alzheimer's

Disease [25]. Manipulations that inhibit misfolded protein clearance mechanisms, such as defects in the autophagy pathway [47, 48] or the proteasome pathway [49], cause neurodegeneration in mice. Mutations in protein degradation pathways can also directly cause neurodegenerative diseases (e.g., mutations in genes associated with mitophagy in Parkinson's disease [50] and mutations in genes associated with autophagy in ALS [51]). Conversely, increased proteasome activity [52, 53] or chaperones [54] can suppress age-induced protein aggregation. While it is not clear what the initiating factor is that causes age-induced toxicity of repeat expansions, defects in the degradation of misfolded proteins and protein aggregates are likely a contributing factor.

1.2.2 Length-Dependent Toxicity of Repeat Expansions

Another modifier of the toxicity of repeat expansion mutations is the length of the repeat. Many repeat expansion diseases, such as the CAG/CTG repeat expansion diseases, have a clear inverse correlation between the length of the repeat and the age of onset of the disease [55, 56]. This relationship was first discovered with the CAG repeat expansion in HD, where CAG repeat length explains 50-70% of the variation in disease onset [57]. For example, the average HD patient has 42 CAG repeats and an age onset of ~40 years old, but the youngest documented HD patient had 210-250 CAG repeats and an age onset of 18 months [58]. On the other hand, some repeat expansion diseases, such as the *C9orf72* G_4C_2 repeat expansion, do not have clear length-dependent toxicity. For example, G_4C_2 repeat lengths do not consistently correlate with the age-of-onset of neurodegeneration or predict which regions of the brain will be affected [26, 29]. The lack of length-dependent toxicity from the G_4C_2 repeat expansion

is not due to the untranslated nature of the repeat, as length-dependent toxicity is observed in the 3'UTR localized CTG repeat expansion that causes DM1 [59]. While most repeat expansions have an inverse correlation between repeat length and age of disease onset, G_4C_2 diseases are unique in having an unclear interaction between repeat length and disease characteristics.

1.2.3 Tissue Specificity of Repeat Expansion Diseases

Disease-causing repeat expansions affect a subset of neurons, even though the mutation is normally present in every cell of the body. The likely causes for this selective vulnerability vary based upon the disease and are still not clearly understood. For example, in HD, the huntingtin protein (HTT) is expressed ubiquitously, yet γ-aminobutyric acid (GABA) producing medium spiny neurons in the striatum are the most severely affected cells [60]. This is not due to higher expression of HTT in the striatum, as HTT is expressed at higher levels in other brain regions [61]. Instead, the susceptibility of medium spiny GABAergic neurons may be due to an increased sensitivity to toxicity from overactivation of glutamate receptors (excitotoxic stress). Consistent with this hypothesis, HD neurodegeneration can be mimicked in mice through neural injections of glutamate or glutamate analogs [62, 63]. Additional factors may enhance the vulnerability of these neuronal populations or be the main cause of vulnerability.

In contrast to HD, the vulnerability of specific neurons in *C9orf72*-associated ALS/FTD patients does appear to be affected by expression patterns of the gene containing the repeat expansion. The most sensitive neuronal population also has the highest expression of *C9orf72* [64]. However, expression level alone does not explain

C9orf72-associated neurodegeneration, as different neuronal populations are susceptible in patients who display ALS, FTD, or ALS/FTD [65]. The *C9orf72* repeat expansion may also cause neurodegeneration in other brain regions, as it has been linked with Alzheimer's Disease [66] and Huntington-like diseases [67]. The additional environmental or genetic factors that determine the vulnerable neuronal populations are not known.

In addition to gene expression and vulnerability to excitotoxic stress, there are several other factors that may impact the selective vulnerability of neurons. In contrast to stress from overactivation of excitatory receptors, stress may also be caused by inhibitory receptors. For instance, neuronal expression of inhibitory GABA receptors strongly correlates with neuronal susceptibility to degeneration in ALS patients [68]. Another factor that may contribute to differential vulnerability is that neurons have different rates of misfolded protein clearance, which could affect the susceptibility of neurons. For example, striatal neurons clear polyglutamine more slowly than cortical neurons [69]. This may be one explanation as to why HD and several other polyQ diseases cause striatal degeneration. Differences in the ability of neuronal populations to neutralize oxidative stress are also likely to affect the susceptibility of neurons. Increased oxidative stress is seen in many neurodegenerative diseases [70-72], and differences in susceptibility to oxidative stress are seen in neuronal tissues [73]. Overall, a single overarching cause of selective neuron vulnerability does not appear to exist. Instead, multiple weaknesses likely synergize to sensitize specific neuronal populations to repeat expansion toxicity.

1.3 Repeat Expansions: Mechanisms of Toxicity

Neurodegenerative repeat expansions cause a gain-of-function (GOF) toxicity, as both intronic and exonic repeat expansion mutations are toxic when present in only one allele of a gene. Initially, the cause of gain-of-function toxicity was thought to differ based on the genetic context of the repeat expansions. Untranslated repeats were thought to be primarily toxic through RNA-based GOF toxicity, while exonic repeats were thought to be primarily toxic through protein-based GOF toxicity. RNA toxicity was due to the G/C-rich nature of the RNA. RNA toxicity was first described in DM1, where a CTG repeat expansion occurs in the 3'UTR of *DM1 protein kinase (DMPK)* [59]. The CTG repeat expansion was toxic independent of its genetic context, as the CTG repeat was toxic when cloned into the UTR of an unrelated gene [74]. The repetitive RNA also formed RNA foci in patient tissue and sequestered RNA splicing regulators [59]. RNA foci are a common feature of repeat expansions and are associated with RNA toxicity.

While RNA foci also occur with exonic repeat expansions, the repetitive amino acid sequence produced from exonic repeat expansions are also innately toxic and were thus thought to be the driving cause of protein GOF toxicity. A classic example of these protein GOF diseases is HD. The expanded CAG repeat in the first exon of *mHTT* encodes a polyglutamine repeat that causes aggregation of the mutant huntingtin protein (mHTT). The aggregated mHTT sequesters chaperone resources, leading to a global disruption of protein folding [75-77]. The aggregation of mHTT depends on the age of the animals as well as the length of the CAG repeat, mimicking HD age-dependent toxicity and repeat length dependency [55, 75]. Exonic CAG repeat expansions in other diseases are also thought to be toxic due to the glutamine repeats produced from the repeat.

Until recently, these two modes of toxicity were accepted as the predominant drivers of repeat expansions toxicity. However, in 2011, an unusual type of protein translation called RAN translation was discovered to occur in repetitive G/C-rich RNA [4]. RAN translation produces six peptides from six distinct sense and antisense reading frames and can occur in RNA from either exons or untranslated RNA, making it a unifying feature of repeat expansion diseases. My studies focus on this recently discovered class of proteins produced through RAN translation, many of which are now known to be highly toxic.

1.4 RAN Translation

RAN translation does not require a start codon [4]. Therefore, RAN translation can occur in the 0, +1, and +2 reading frame leading to three potential protein products from one strand of RNA. As mentioned above, repeat expansions are commonly transcribed in the antisense direction as well [5, 7, 78-80]. As the antisense RNA of a G/C-rich repeat is also G/C-rich and repetitive, it also undergoes RAN translation in all three reading frames. Consequently, six genetically distinct protein products are produced from one nucleotide repeat expansion (Figure 1.1). RAN translation has so far been detected in seven repeat expansion diseases: *C9orf72*-ALS/FTD [81, 82], DM1 [4], myotonic dystrophy type 2 (DM2) [83], fragile x-associated tremor/ataxia syndrome (FXTAS) [84] Fuchs Endothelial Corneal Dystrophy (FECD) [85], HD [12], and spinocerebellar ataxia type 8 (SCA8) [4].

Figure 1.1 Example Diagram of RAN Translation Products

The six possible RAN products of a CAG repeat expansion. From the sense RNA strand, polyGlutamine is translated in the 0 reading frame, polySerine is translated in the +1 reading frame, and polyAlanine is translated in the +2 reading frame. From the antisense RNA strand, polyLeucine is translated in the 0 reading frame, polyCys is translated in the +1 reading frame, and polyAlanine is translated in the +2 reading frame.

The molecular biological and biochemical requirements for RAN translation are beginning to be elucidated. Cellular models of RAN translation suggest that RAN translation initiates from alternate start codons upstream of the repetitive sequence [86, 87], and the initiating codon varies between reading frames [87]. As a result, the efficiency of RAN translation between reading frames is highly variable, resulting in dramatic differences in RAN peptide accumulation [86, 87]. The site of termination of RAN products is unknown, as translation could end due to either a stop codon or stalling of translation machinery because of tertiary structures formed by the repetitive RNA, such as G-quadruplexes [88]. In cellular models, RAN translation increases with activation of the integrated stress response system [86, 89, 90]. Therefore, the level of expression of RAN products is not static throughout the progression of a disease and likely increases as cells undergo escalating levels of stress.

To better understand RAN translation, it has been studied in several cellular and animal models expressing repeat expansion sequences [12, 86, 87, 89, 91, 92]. While most of these models were generated after the discovery of RAN translation, some were generated prior to this discovery [12]. In these models, toxicity was previously only considered to be due to repeat bearing RNA or protein. However, RAN translation has now been shown to occur in yeast [91], *C. elegans* [91], *Drosophila* [93], and mice [12]. Therefore, earlier repeat models need to be reexamined in order to determine whether or not they produce RAN peptides from these repeat sequences. For example, a recent reexamination of a CAG repeat mouse model (N171-82Q) revealed that the CAG repeat construct undergoes RAN translation and produced multiple RAN peptides beyond just polyGln, as was previously assumed [12]. Further studies are needed to determine if

other existing cell and animal models expressing repeat expansions are purely RNA toxicity models, purely protein toxicity models, or are RAN models as well.

1.4.1 Modeling RAN Products

To understand the specific toxic properties of RAN products, they need to be modeled independently of RNA toxicity. Creating RAN-only models is challenging as RAN translation depends on the G/C-rich repetitive nature of the RNA [4], which also causes RNA toxicity. Therefore, the RAN peptide amino acid sequence needs to be preserved, but the repetitive and G/C rich nature of the DNA and RNA needs to be eliminated. To accomplish this, the degenerative nature of the genetic code is commonly utilized [12, 94-97], as up to 6 different codons can encode the same amino acid. By varying the underlying genetic code while preserving the overlying encoded amino acid, RAN peptides can be individually expressed in the absence of any repetitive, G/C rich RNA repeats. While powerful, this approach has several caveats. First, codon-varied RAN products are usually overexpressed at non-physiological levels, which may drive aspects of toxicity that are not normally present in patients. Second, RAN products are often, but not always, co-expressed in patient tissue [34]. Individual RAN peptide expression is not sufficient to mimic any possible functional interactions between the RAN peptides that might synergistically enhance or suppress toxicity in patients [96], so, studying each RAN product individually does not precisely model the condition in patients. Finally, most codon-varied RAN peptide models lack upstream and downstream genetic context, which is unique to each peptide. Such sequences might dramatically alter the

toxicity of the peptide. Despite these caveats, characterizing each RAN product individually is an important first step in understanding the toxicity of the RAN products.

While methods of modeling individual RAN products have limitations, they are crucial tools for developing a general understanding of the toxicity of individual RAN products. Findings from these simplistic models will aid in identifying the cellular pathways disrupted by RAN products in each disease. The first question to address with RAN products is if and how they are toxic. As our understanding of RAN translation and RAN products grows, genetic models used to study RAN products will need to be adjusted to better model disease relevant length, genetic context, and peptide-peptide interactions. To address these questions, several cellular or animal models expressing codon-varied RAN products from different repeats have been developed, including in yeast, worms, fruit flies, zebrafish, rodents and human-induced pluripotent stem cells (iPSCs) [10, 95, 98-100].

In my research, I focus on RAN products from two different repeat expansions, the ALS-associated G_4C_2 repeat expansion and the HD-associated CAG repeat expansion. There are twelve other CAG/CTG repeat expansions diseases (Appendix B-Table 4.1), two of which are also confirmed to undergo RAN translation [4]. Multiple CAG/CTG diseases can phenocopy HD (e.g., Huntington disease-like 2 (HDL2), SCA17, DM1, and dentatorubral-pallidoluysian atrophy (DRPLA)), suggesting CAG/CTG diseases may have similar disease mechanisms independent of genetic context [101]. One such common mechanism may be the production of RAN peptides. Despite both CAG and G_4C_2 repeats undergoing repeat expansion and RAN translation, the respective NDDs, repeats, and RAN products are distinct. To understand the cellular characteristics

of the RAN products from each of these diseases and to study their toxicity in a multicellular system, we modeled RAN products from each repeat expansion in *C. elegans*.

1.5 Using *C. elegans* to Model NDDs

C. elegans have been used to model a wide variety of NDDs, as they have unique features that make them advantageous for studying aging-associated diseases with neurodegeneration [102]. *C. elegans* are small (1 mm long), transparent nematodes that have been studied as a genetic model system for over 40 years [103]. Worms offer several advantages not present in other model systems for the study of RAN products and 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 [104]. Second, *C. elegans* are optically transparent across their entire lifespan, facilitating observation of neuron and tissue morphology in living animals, as well as the 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 homologue [105]. Fourth, the function of these genes can be rapidly inhibited, either using chemically or CRISPR induced mutations or RNA interference (RNAi) [106, 107]. Finally, aging, which is a major risk factor for all repeat expansion diseases, has been extensively studied in *C. elegans* as worms have a short lifespan of ~3 weeks [108].

In addition to their short lifespan, several signaling pathways are known to regulate the aging process in worms, as well in other species. One of these pathways, the insulin pathway, was first discovered in nematodes [109, 110]. The insulin signaling pathway is now the most well-characterized pathway that regulates lifespan and youthfulness from worms to humans [111]. Mutations in this pathway are thought to extend lifespan by promoting a youthful cellular physiological state and can be used to separate chronological and physiological states of aging. In *C. elegans*, the *daf-2* gene encodes the homolog of the insulin/IGF receptor. Loss of function mutations in *daf-2* exert their effects via signaling-dependent activation of the FOXO transcription factor *daf-16*, as well as the heat shock transcription factor *hsf-1* [112]. Mutations in *daf-16* and *hsf-1* suppress phenotypes associated with *daf-2* mutations and accelerate the rate of aging [111]. Thus, the role of aging in a neurodegenerative pathway can be studied in *C. elegans* by using *daf-2* mutants to delay cellular aging or *daf-16* or *hsf-1* to accelerate cellular aging.

Despite these experimental advantages, worm models have several caveats for modeling NDDs. *C. elegans* motor neurons lack astrocytes and glia, which play a significant functional role in several NDDs [113]. While worms possess a conserved primitive innate immune system [114], they lack adaptive immunity, which contributes to many pathological aspects of repeat expansion disorders [115]. Similar limitations impact other model systems that are used to investigate repeat-associated NDDs [91, 98, 99, 116-118]. Despite these limitations, *C. elegans* is a uniquely valuable model system whose experimental advantages complement many of the limitations present in other systems for the study of repeat expansion disorders.

1.5.1 Previous Models of *mHTT* in *C. elegans*

Several models of HD and other CAG repeat expansion disorders have been developed in *C. elegans*. Most of the models have the CAG repeat within the surrounding *HTT* sequence, as genetic context can either enhance and suppress the toxicity of repeats [119]. In one model, 150 pure CAG repeats were set in the genetic context of the sequence encoding the first 171 amino acids of HTT [120]. The mHTT was expressed in amphid neurons with single ciliated endings (ASH neurons). Cellular dysfunction could be detected by impaired dye filling of the ASH neurons. This model facilitated both biased and unbiased genetic screens, as well as small molecule drug screens, for suppressors of CAG repeat expansion toxicity. For example, Bates et al. discovered that histone deacetylases (HDACs) can modify mHTT toxicity [121]. HDACs are now known to play significant roles in the pathogenesis of polyQ toxicity across a wide range of cellular and animal model systems [122, 123].

Another model for HD in *C. elegans* modeled 'early-onset' aspects of CAG/polyQ pathology. 128 pure CAG repeats were placed within the coding sequence of the first 57 amino acids of HTT [124]. In contrast to the previous model expressed in ASH sensory neurons, this 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 [125]. Lejeune et al. utilized RNAi screening and sensitized genetic backgrounds in this system to identify suppressors and enhancers of expanded CAG toxicity [126]. Many of the 662 genes identified in this screen were previously known to be involved in HD or other NDDs, including 49 that are dysregulated in the striatum of HD mouse models [126]. This demonstrates 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.

1.5.2 Modeling Repeat Expansions Independent of Genetic Context

The models just described examined CAG repeats in the presence of the *HTT* sequence. However, there are nine other age-dependent NDDs caused by CAG repeat expansions, suggesting the CAG repeat itself, in multiple genetic contexts, is sufficient to cause age-dependent toxicity. To explore CAG toxicity independent from these different contexts, a pure *C. elegans* CAG model was generated [127]. In this model, CAG repeats are fused with YFP in the polyQ reading frame and expressed in muscle tissue [75]. Although HD is thought of as a purely neurodegenerative disease, patients also have muscle wasting and cardiac defects, making muscle expression of CAG repeats relevant to patients [128, 129]. Expression of CAG in the muscle tissue of *C. elegans* also provides several experimental advantages (large cell size, sensitivity to feeding-based RNAi gene knockdown, sensitive age-dependent phenotypic outputs, etc.) that are not available when CAG is expressed in neurons. The 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 manner in young animals, but the protein transitions from a soluble to an aggregated state as animals age [75]. PolyQ aggregate formation coincides with the onset of motility defects, suggesting a link between polyQ

aggregation and toxicity. However, recent work implies that polyQ aggregation and toxicity are genetically separable events [130]. Therefore, mechanisms other than polyQ aggregation may contribute to toxicity in this model [131]. The toxicity and aggregation of these pure polyQ proteins are strongly influenced by the aging process, as mutants that modify aging (i.e., in insulin/IGF signaling pathway) also modify polyQ aggregation and CAG toxicity [112, 132-134]. 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 aging, 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 heterogeneous aggregation and toxicity depending on the neuron type [135]. 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.

1.5.3 Screening Strategy

Because of the heterogeneous toxicity of CAG in various neuronal subtypes, *C. elegans* muscle cells have provided a more amenable and homogenous cellular context for *in vivo* genetic screens that have ultimately informed 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 knockdowns for enhancers of muscle polyQ aggregation, identifying proteins whose normal function is to oppose polyQ

aggregation [136]. These findings revealed that the breadth of the machinery regulating protein misfolding extends far beyond chaperones and the protein degradation machinery and involves a wide variety of biological processes, including RNA synthesis and processing, protein biosynthesis, and protein trafficking [136] (Figure 1.2). Subsequently, many of the proteins identified in this screen have 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 [136-138].

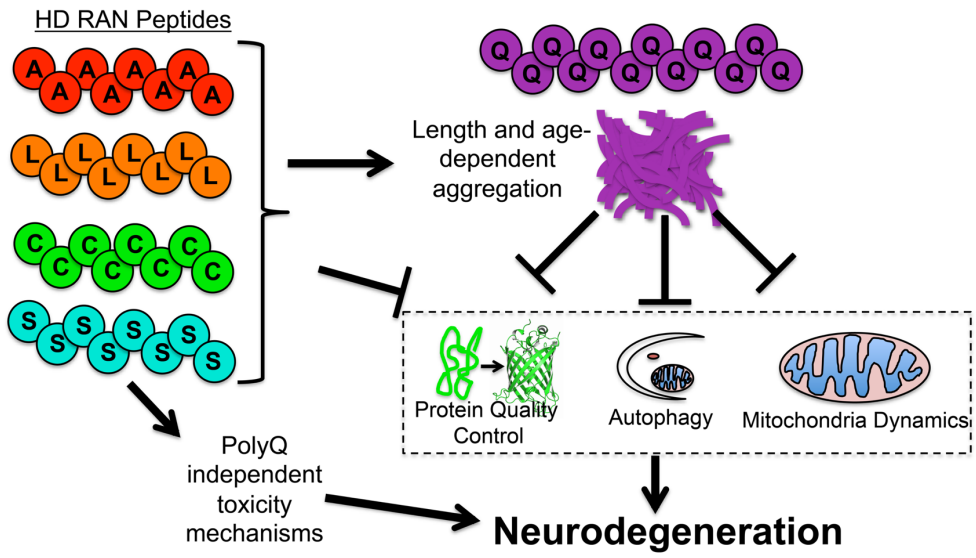


Figure 1.2 Possible Mechanisms of Toxicity of PolyQ and RAN HD PolyPeptides

Defects in protein quality control, autophagy, and mitochondria dynamics have been observed in multiple HD models. It is unclear if one pathway is the driving cause of neurodegeneration, or if their synergistic interactions are required for toxicity. The HD RAN peptides may act through similar pathways as polyQ or through polyQ-independent pathways.

Although the identification of genes that modify muscle-based phenotypes of toxic proteins is a powerful screening tool, it is also important to determine if the modifier is muscle specific or if it also modifies the toxic phenotype in neuronal cells. The neuronal system of *C. elegans* is well defined, and there is a complete connectome map of all 302 neurons in the adult hermaphrodite [139-141]. The position of neurons is conserved between different animals, thus making measurements of neurodegeneration possible at the single neuron level. *C. elegans* neurons also utilize the same neurotransmitters used by human neurons; GABA, acetylcholine, 5-HT, glutamate, and catecholamines [142]. This allows modeling of NDDs in disease-relevant neurons, such as modeling HD in GABAergic neurons or modeling ALS in motor neurons.

C. elegans GABAergic neurons have been used to model multiple types of NDDs as they exhibit robust cellular and phenotypic defects upon degeneration [143, 144]. GABA is an inhibitory neurotransmitter in the human central nervous system [145]. In *C. elegans*, GABA is most often at neuromuscular junctions and can be used as either an inhibitory or an excitatory neurotransmitter [146]. At a *C. elegans* neuromuscular junction, GABA acts as an inhibitor of muscle contraction and counteracts acetylcholine, which causes muscle contractions [146]. Therefore, degeneration of GABAergic neurons has a strong observable phenotype due to disruption in the balance of excitatory acetylcholine and inhibitory GABA signaling at the neuromuscular junction [147]. Defects in GABA signaling lead to overactive acetylcholine signaling and muscle hypercontraction, causing worms to paralyze and appear shrunken from excessive contraction [147]. Loss of GABAergic motor neurons most severely affects directional reversal, since the

coordination between dorsal and ventral muscle activity required for directional changes is lost [148]. Neurodegeneration is easily detectable at the cellular level as well. Motor neurons run along the ventral nerve cord (VNC), a longitudinal nerve tract that extends along the ventral side from the pharynx to the tail of the *C. elegans*. However, the motor neurons need to synapse onto the muscle cells, which are on both the dorsal and ventral sides of the worm. Consequently, motor neurons extend single axons, known as commissures, which synapse onto the muscle cells on the dorsal side of the *C. elegans* (Figure 1.3). These axons are easily observed in live animals by expressing GABAergic-neuron specific GFP reporters and imaging optically transparent animals under fluorescence microscopy [146]. There are 16 discrete GABAergic commissures along the length of a worm that can be observed for neurodegenerative phenotypes such as blebbing or breakage at the level of single axons [146]. Thus, degeneration of GABAergic neurons is easily observable at both the phenotypic and cellular level, making these cells good models for NDDs that affect GABAergic neurons, such as HD, or motor neurons, such as ALS.

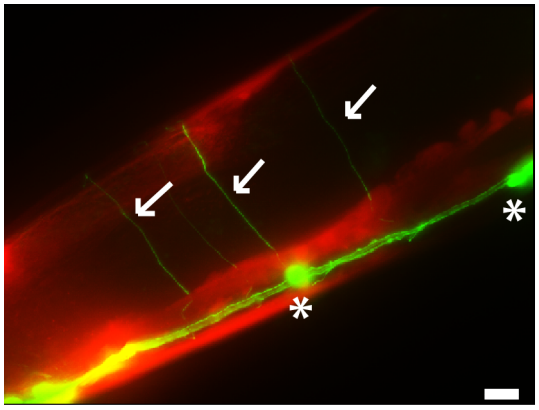


Figure 1.3 Fluorescent Commissure Within a Whole, Living *C. elegans*

Fluorescent image of *C. elegans* with motor neurons in green (*unc-47p::GFP*) and muscle in red (*myo-3p::RFP*). Arrows mark commissures. Asterisks mark cell soma. Ventral nerve cord is along the lower edge of the worm. Scale bar=10 μm.

1.6 ALS/FTD

ALS is an extremely rapidly progressing neurodegenerative disease, with 50% of patients succumbing to the disease within three years of diagnosis [149]. Roughly 95% of ALS cases are sporadic with no family history of ALS [150]. The largest risk factor for ALS is aging [151], with ALS symptoms commonly appearing between ages 40 to 70 [152]. In familial ALS cases, single gene genetic mutations are known for almost 50% of the cases [153]. Depending on the population studied, 40-60% of familial ALS cases are caused by the *C9orf72* G_4C_2 repeat expansion, making it the most common inherited cause of ALS/FTD [8, 9]. The environmental causes of sporadic ALS are poorly

understood. Cigarette smoking is a confirmed environmental risk factor for ALS [154], while lead exposure [155] and pesticide exposure [156] are possible environmental risk factors. In addition, there are recent data linking traumatic brain injury and ALS pathology [157]. Because ALS environmental factors exhibit highly variable penetrance, genetic models of ALS have been predominantly used to study this disease.

1.6.1 ALS Physiology

The first symptom of ALS is typically weakness in a limb or the jaw and rapidly progresses to paralysis and complete loss of voluntary motor control [158]. The paralysis is caused by the degeneration of the upper and lower motor neurons typical of ALS. The upper motor neurons are located in the motor region of the cerebral cortex and innervate the lower motor neurons in the ventral horns of the spinal cord and within brainstem motor nuclei. The upper motor neurons modulate the response of the lower motor neurons, and ALS patients who predominantly have degeneration of the upper motor neurons exhibit muscle stiffness as the muscles are overactivated [159]. The lower motor neurons innervate the muscles, and degradation of lower motor neurons causes loss of reflexes and muscle tone, eventually leading to muscle atrophy.

While ALS is thought of as a pure motility disease, 15% of ALS patients also develop FTD [160]. FTD is a common type of early-onset dementia where the frontal and temporal lobes of the cerebral cortex undergo degeneration. Patients display social dysfunction, such as loss of empathy and decreased social interaction, and they also often have repetitive behaviors and a lack of inhibition [161]. Ten to fifteen percent of FTD patients have a family history of FTD or ALS [160]. In both ALS and FTD, TDP-43

aggregates occur in degenerating neurons, suggesting a shared disease pathway [162]. In addition, some mutations can cause either ALS or FTD, suggesting ALS and FTD are represent phenotypic bookends of a disease spectrum. The most prominent shared mutation, the *C9orf72* repeat expansion, is the most common genetic cause of each disease, and 80% of ALS/FTD patients carry the *C9orf72* repeat expansion [162].

1.6.2 The *C9orf72* G_4C_2 Repeat Expansion

The G_4C_2 repeat expansion in *C9orf72* was associated with ALS in 2011 [8, 9], five years after the linkage of chromosome 9p21 was first identified in a Dutch ALS and FTD kindred [163]. The G_4C_2 repeat expansion is in the first intron of *C9orf72* and is 10–20 repeats in unaffected individuals, but expands to 30-1000s of repeats in ALS patients [8, 9]. The *C9orf72* repeat expansion causes up to 45% of all inherited cases of ALS and 25% of FTD cases [9, 164]. The *C9orf72* repeat expansion is also found in ~6% of sporadic ALS cases [164]. The high frequency of *C9orf72* repeat expansions in ALS/FTD, as well as its association with several other less common neurodegenerative conditions [165], makes the *C9orf72* repeat expansion one of the most common known NDD mutations.

One proposed mechanism for *C9orf72*-G_4C_2 repeat expansion toxicity was loss of function. The G_4C_2 repeat expansion causes a decrease in *C9orf72* RNA expression [8] and C9ORF72 protein levels [166] in patients. C9ORF72 appears to be involved in autophagy initiation [167-170]. However, loss-of-function of C9ORF72 does not appear to be the driving cause of ALS/FTD as *C9orf72*-null murine models develop autoimmune defects, but not ALS/FTD-like symptoms [171-173]. In addition, a patient homozygous

for the *C9orf72*-G_4C_2 repeat expansion did not have a severe form of ALS/FTD [174], as would be expected if the repeat expansion was toxic due to inhibiting *C9orf72* expression. More importantly, loss-of-function point mutations have not been identified in any ALS/FTD patients. Therefore, the *C9orf72*-G_4C_2 repeat expansion is likely toxic through gain-of-function mechanisms.

Indeed, the clinical presentation of *C9orf72* repeat expansion carriers is consistent with an autosomal dominant genetic pattern that could manifest pathologically through at least two nonexclusive molecular mechanisms [162]. First, given that the repeat expansion is within an intron, one hypothesis is that toxicity is mediated through a repeat-containing RNA. In patients, pathologically expanded RNAs form nuclear foci, while non-expanded RNAs do not form nuclear foci [6, 175]. Although the significance of these RNA foci remains unknown, G_4C_2 RNA binds several RNA-binding proteins, including some that regulate the nuclear import/export cycle and nuclear pore complex function [6, 175]. G_4C_2 RNA induced disruption of neuronal nuclear transport may play a pathological role in disease onset and/or severity. A second potential mechanism involves RAN translation in all three reading frames of both the sense and antisense transcript [6], leading to six distinct dipeptide repeat proteins (DPRs) [176]. RAN DPRs appear to be translated from the spliced out G_4C_2-containing intron [89, 177]. However, factors that promote the cytoplasmic localization of the repeat-bearing RNA appear to facilitate RAN translation. Therefore, the observed nuclear RNA foci may serve a protective role and the RAN translation products are toxic [177, 178]. Taken together, significant evidence suggests that specific *C9orf72*-derived RAN DPRs are toxic and may play a pathological role in the development of ALS/FTD.

While RAN DPRs are thought to play a significant role in the toxicity of the *C9orf72* repeat expansion, there is evidence against DPRs having a causal role in disease. For example, in postmortem brain samples from *C9orf72* patients, anti-DPR antibody staining patterns are not well correlated with tissue neurodegeneration [179-181]. One possible explanation for this discrepancy is that cells exhibiting robust RAN translation undergo cell death throughout the progression of ALS, which results in low levels of observable DPRs in postmortem tissue. Supporting this, a murine model has detectable DPRs in brain regions that will degenerate early in the disease course, but the DPRs cannot be detected in those same brain regions late in the disease [10]. A G_4C_2 *Drosophila* model of RNA only toxicity versus RNA+DPR found that RNA alone caused no toxicity, while RNA+DPRs caused strong toxicity [93]. Every model system tested has found the same two DPRs, proline-arginine (PR) and glycine-arginine (GR), to be toxic [10, 95, 98, 182, 183].

1.6.3 *C9orf72* RAN Dipeptide Repeats

Unlike the other DPRs, both sense-derived GR and antisense-derived PR can phase separate, as they are charged due to the arginine amino acids [184]. The separation of oil and water when mixed is a simple type of phase separation. Protein phase separation is more complex as ionic strength, protein concentration, temperature, and post-translational modifications all regulate the phase separation of proteins [185, 186]. There are two types of proteins that phase separate. One type of phase-separating proteins is multi-domain proteins that undergo regulated phase separation depending on the number of bound ligands. The second type is intrinsically disordered proteins with a

low sequence complexity that phase separates independently of ligands [186]. The biological role of phase separation is to concentrate proteins performing a specific task, such as splicing factors in nuclear speckles [187], while still allowing rapid interchange between the structure and the surrounding environment. Other phase-separating cellular structures are stress granules, nuclear pores, and the nucleolus [188-190]. Disrupted phase separation can cause cellular dysfunction. This occurs in multiple NDD models where phase-separated structures lose their fluid characteristics and shift to having a hydrogel nature and eventually form aggregates, trapping the bound proteins [191-194]. Interestingly, *C9orf72*-associated arginine-rich dipeptides localize to multiple phase-separated structures, including the nucleolus, stress granules, and the nuclear pore [195]. PR/GR disrupts the function of these structures [195]. Consistent with these interactions, disruption of nucleocytoplasmic transport genes have been identified in PR/GR suppressor screens performed in both yeast and *Drosophila* [98, 118]. PR/GR also disrupt nucleolar function and prevent stress granule disassembly [195, 196]. It is unknown whether PR/GR interactions with phase-separated domains in general cause neurodegeneration or if PR/GR interactions with a specific phase-separated domain is the main cause. Further screens will help identify the sequence of events that leads to neurotoxicity. *C. elegans* models have previously helped to identify disease pathways and could serve the same role here.

1.7 HD

I also created a *C. elegans* model for CAG RAN products, as CAG/CTG repeat expansions in different genes cause 13 repeat expansion diseases: FECD, HD, HDL2, DM1, SBMA, and SCA Types 1, 2, 3, 6, 7, 8, 12, and 17 [11]. RAN translation has thus far been confirmed in four of these diseases: FECD, HD, DM1, and SCA8 [4, 12, 85]. The CAG repeats encode polyglutamine, which facilitates protein-protein interactions [197]. Polyglutamine is 'sticky' and interacts with other polyglutamine repeats, stabilizing protein binding [197]. However, in repeat expansion diseases the CAG repeat is expanded to ≥40 repeats depending on the disease, and the 'sticky' nature of the polyglutamine becomes toxic, as it forms oligomers (low molecular weight homopolymers) and aggregates [198, 199]. It is currently thought that polyglutamine is the driving cause of toxicity in these diseases [200, 201], even though neurodegeneration is not always linked to polyglutamine aggregation [12]. In HD, degenerating brain regions that lack polyglutamine have significant expression of other CAG/CTG-derived RAN products, strongly suggesting that the RAN products could be contributing to the disease [12]. Most of the CAG/CTG diseases lack treatments, making it imperative to understand the molecular pathways required for CAG RAN toxicity in these diseases.

HD is a monogenic disease caused by a CAG repeat expansion in the first exon of the *HTT* gene [202]. HD is a progressive and incurable neurodegenerative condition that typically occurs in middle to late age [203]. However, up to 10% of patients develop a more aggressive form of the disease called Juvenile Huntington's disease (JHD) [204]. Unaffected individuals have 10-25 CAG repeats in *HTT*, while people with 35-39 CAG repeats in *HTT* are at risk of developing HD in adulthood [205, 206]. Patients with ≥40

repeats will develop HD and typically exhibit symptoms between 30-50 years of age, depending on repeat length and other factors [55]. JHD patients usually have >60 repeats and develop symptoms before the age of 20. In all cases, the disease follows an autosomal dominant inheritance pattern, suggesting that a gain-of-function mechanism underlies both HD and JHD.

1.7.1 HD Physiology

HD causes a wide range of symptoms, including cognitive and motility defects, which vary based on the type of HD. Adult HD patients are diagnosed with HD once they develop motility defects, such as chorea. The motility impairments change to immobility by the end of the disease [202]. HD patients also develop personality changes and undergo cognitive declines up to ten years before they are officially diagnosed with HD [207, 208]. In adult HD, neurodegeneration begins in the striatum and spreads in a predictable pattern to other regions in the basal ganglia, and distally to brain regions such as the frontal cortex [209-211]. The neurodegeneration ultimately leads to death ~20 years after the appearance of motor symptoms [203]. JHD patients have a shorter disease duration of ~15 years after diagnosis [212]. This decrease may be due to delayed diagnosis of JHD, as it is extremely rare [212]. Interestingly, JHD patients do not exhibit chorea or have a spreading pattern of degeneration. Instead of developing chorea, JHD patients present with rigidity, seizures, and severe psychiatric defects [212]. Patients with JHD also have smaller brain sizes with relatively enlarged cerebellum [213] and have degeneration in both the striatum and the cerebellum [214]. The reason for the differences in symptoms and degeneration patterns between HD and JHD is unknown.

Taken together, the clinical data suggest that the increased number of CAG repeats drive a distinct spectrum of symptoms in JHD even though both HD and JHD are caused by similar CAG repeat expansion mutations in the HTT gene.

1.7.2 HTT

The *HTT* gene encodes a large, 348 kilodalton protein that is conserved in metazoans [215]. However, the disease-causing polyQ stretch in the first exon of HTT does not appear in all HTT orthologs [216]. The polyQ repeats are predominantly found in mammals, and the length of the repeat increases with higher species [217]. This may be due to the role of HTT in neurodevelopment, as HTT appears to have key roles in neurodevelopment and heterozygous *HTT* null mice have neuronal defects [218, 219]. Homozygous *HTT* null mice are embryonic lethal, which may be due to other roles of the ubiquitously expressed HTT [219]. In addition to the polyQ domain, HTT has multiple HEAT repeats that act as flexible scaffolding allowing HTT to interact with over 200 proteins [220, 221]. Through these interactions, HTT traffics vesicles and organelles along axons [222]. The polyQ repeat expansion can cause defects in many of HTTs normal roles and interactions [223, 224], which may contribute to HD even if it is not the driving mechanism.

1.7.3 PolyQ

Oligomerizing/aggregating polyQ proteins cause toxicity through many pathways. PolyQ models exhibit disruption of protein folding and degradation [225, 226], inhibition

of autophagy [227], and abnormal mitochondrial dynamics and quality control [228]. PolyQ proteins are thought to be toxic due to low molecular weight homopolymers of polyQ, known as oligomers, while large aggregates exhibit less toxicity and possibly have a protective role [229-231].

The expanded polyQ protein is the focus of most HD research. This is primarily because it exhibits length-dependent toxicity and aggregation that closely matches the course of the disease in HD patients [232]. However, in JHD, the cerebellum and frontal cortex lack detectable polyQ, but still undergo significant neurodegeneration [233]. Some affected brain regions in adult HD, such as the white matter regions of the caudate and putamen of the striatum, also express little to no polyQ protein [233-237]. This suggests that mechanisms other than polyQ toxicity drive toxicity in JHD and contribute to toxicity in adult HD.

1.7.4 CAG RAN Peptides

Recently, Bañez-Coronel and colleagues discovered that expanded CAG repeats in the *mHTT* mRNA, as well as the antisense CUG repetitive mRNA, can undergo RAN translation [233]. RAN translation of the sense CAG repeats, as well as the antisense CUG repeats, produces five polypeptides: polyGlutamine, polyAlanine, polySerine, polyLeucine, and polyCysteine (Figure 1.1). Analysis of postmortem tissue samples shows that RAN polypeptides are abundantly expressed throughout the striatum, frontal cortex, and cerebellum and are qualitatively present at elevated levels in JHD patients (age range 8-23) [12]. RAN polypeptides were also present in similar tissues in adult HD patients, although their levels were lower, and their occurrence was less common

compared to JHD patients [12]. RAN polypeptides were not detected in control patients [233]. Since polyQ levels and aggregation are not major features of JHD, these observations suggest that CAG-derived RAN polypeptides could play a significant but previously unappreciated role in JHD and may be a contributing feature to adult HD. Although RAN polypeptides exhibit a striking correlation with neurotoxicity in JHD and HD, it was not known if these polypeptides confer toxicity in an *in vivo* setting.

1.8 Conclusions

The exact mechanism of toxicity of repeat expansion NDDs is unknown. Initial studies focused on testing mechanisms related to protein gain-of-function toxicity or RNA gain-of-function toxicity. Recently, the discovery of RAN translation has offered another potential cause of toxicity. Determining if and how RAN products cause toxicity in their respective NDD is important to find treatments for these currently untreatable diseases. To aid in determining how the RAN products are toxic, we created RAN models for two different types of repeat expansions, the ALS-associated *C9orf72* G_4C_2 repeat expansion, and the widespread CAG repeat expansion, seen in HD and 12 other CAG/CTG repeat expansion diseases. In both cases, we studied the toxicities of the different RAN products and proceeded to perform a suppressor screen to identify proteins required for the toxicity of the RAN product. Using this approach, we discovered both unique and common cellular pathways that contribute to RAN peptide toxicity.

2.0 Arginine-Rich ALS RAN Dipeptides Exhibit Age-Dependent Toxicity

2.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a rapid, devastating neurodegenerative disease that lacks effective treatments. Within three years of presenting with a slight weakness in a limb, patients typically develop complete loss of voluntary motor control and succumb to this disease [149]. The pathogenic mechanisms that drive ALS progression are unknown. However, cellular dysfunction associated with ALS is commonly studied through investigations of genetic mutations that cause ALS. Most recently, these studies are focused on the most common genetic cause of ALS, a G_4C_2 repeat expansion in the *C9orf72* gene [8, 9]. The *C9orf72* G_4C_2 repeat expansion occurs in the first intron of the *C9orf72* gene. Affected patients carry one normal allele with less than 30 G_4C_2 repeats and one expanded allele with 30-1000s of repeats [8, 9]. The precise number of repeats required to cause disease is difficult to determine due to somatic and germline mosaicism of the repeat. However, repeat lengths of up to 27 do not appear to show somatic instability [238, 239]. These data have led the field to conclude that at least 30 G_4C_2 repeats are required to cause ALS.

The mechanism(s) by which 30 G_4C_2 repeats cause disease is not well understood. One emerging mechanism is the unusual translation of the intronic repeat expansion by a novel type of translation called Repeat-Associated non-AUG (RAN) translation [81, 82]. RAN translation occurs in long, G/C-rich repeats and does not require a start codon [4]. Due to the lack of a start codon, the reading frame of RAN translation is not defined, and

translation can be initiated in any of the three reading frames. Initiation most likely occurs within or very close to the 5' end of the repeat [86], although the precise site of initiation has not been defined in patients. Therefore, RAN translation of the single G_4C_2 RNA repeat gives rise to three distinct dipeptide repeat proteins (DPRs): glycine-arginine (GR), glycine-proline (GP) and glycine-alanine (GA) [81, 82]. The expanded G_4C_2 repeat is also transcribed in the antisense direction [34]. The corresponding antisense RNA is a G_2C_4 repeat and is thus also G/C-rich. Therefore, the antisense RNA also undergoes RAN translation, creating three more DPRs: proline-arginine (PR), proline-alanine (PA) and glycine-proline (GP) [82] (Figure 2.1 A). All six RAN DPRs contain distinct protein sequences downstream of the repeat since they are each in a different reading frame. The most common genetic cause of ALS, a single repeat expansion, thus gives rise to six potentially pathogenic DPRs.

The arginine-rich DPRs, PR and GR, have been identified as the most toxic *C9orf72* DPRs through work in yeast, zebrafish, *Drosophila*, and cells [94, 98, 182, 184, 240]. Because these DPRs are toxic at 50 repeats in a wide range of organisms [94, 98, 195], the DPRs are typically modeled at 50 repeats. The arginine-rich DPRs exhibit several properties that may explain their toxicity. For example, PR/GR can phase separate [195, 241], similar to other neurodegenerative proteins such as TDP-43, FUS, and tau [184, 191]. GR/PR localize to phase-separated structures such as the nucleolus [184], the nuclear pore [242], and stress granules [195]. PR/GR localization disrupts the function of each of these phase-separated structures [98-100, 195, 243]. It is currently unclear if DPR-induced defects in one or more of these structures are the driving cause of the pathogenesis of the arginine-rich dipeptides.

An unbiased genetic screen performed in a multicellular organism could provide unique insights into the primary causes of toxicity due to the arginine-rich DPRs. A screen could determine if specific cellular structures or structural components are required for PR/GR toxicity or if a previously unidentified pathway(s) is more integral to PR/GR toxicity. To date, unbiased screens of PR toxicity have only been performed in single-cell systems, such as yeast [98] and mammalian cells [100]. To determine the genetic requirements for arginine-rich DPRs in a multicellular organism, we created *C. elegans* models of each RAN DPR and characterized the cellular requirements for toxicity. In addition, we studied whether PR/GR toxicity is impacted by aging since the strongest risk factor for ALS is aging, and the role of aging and aging pathways in DPR toxicity was not known. After establishing the cellular requirements for arginine-rich DPR toxicity, we performed a whole-genome RNAi screen to identify the genetic requirements for arginine-rich DPR toxicity. Our screen identified highly conserved genes required for PR toxicity, including known DPR modifiers, such as nuclear transport and RNA binding proteins, and novel genetic pathways, including a nuclear ubiquitination pathway.

2.2 Results

To understand the cellular, temporal, and genetic requirements for toxicity of the arginine-rich dipeptides, we created a *C. elegans* model expressing individual RAN translation products in the absence of repeat-containing RNA [95]. In this model, the nucleotide repeat was eliminated, but the amino acid repeat was maintained by varying the codon usage for the amino acids. The disruption of the G_4C_2 sequence removes

potential RNA secondary structures and possible RNA-based toxicity. The reading frame was defined using a canonical start codon and Kozak sequence. Four *C9orf72*-derived dipeptide sequences (GA, GR, PA, PR; 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 either in muscle cells or in motor neurons (Figure 2.1 A). We generated transgenic animals expressing each individual DPR.

2.2.1 Toxicity of *C9orf72* Dipeptides in Motor Neurons

Consistent with previous studies [94, 98, 182, 184], we found that expression of the arginine-rich DPRs in motor neurons under the *unc-47* promoter was highly toxic in *C. elegans*. Animals expressing arginine-rich DPRs in the motor neurons were viable, consistent with previous studies showing that *unc-47* motor neurons are not required for viability in *C. elegans* [244]. However, motor neuron expression of either $(PR)_{50}$ or $(GR)_{50}$ led to significant reductions in motility, which was measured in thrashing assays (Figure 2.1 B). Motor neurons expressing $(PR)_{50}$ and $(GR)_{50}$ exhibited morphological signs of degeneration commonly observed in other *C. elegans* models of neurodegenerative diseases [245-247], including commissure degeneration and membrane blebbing that were not observed in control animals or in animals expressing either $(PA)_{50}$ or $(GA)_{50}$ (Figure 2.1 C, D). These data show that arginine-rich DPRs cause motility defects and neurodegenerative morphological changes when expressed in *C. elegans* motor neurons.

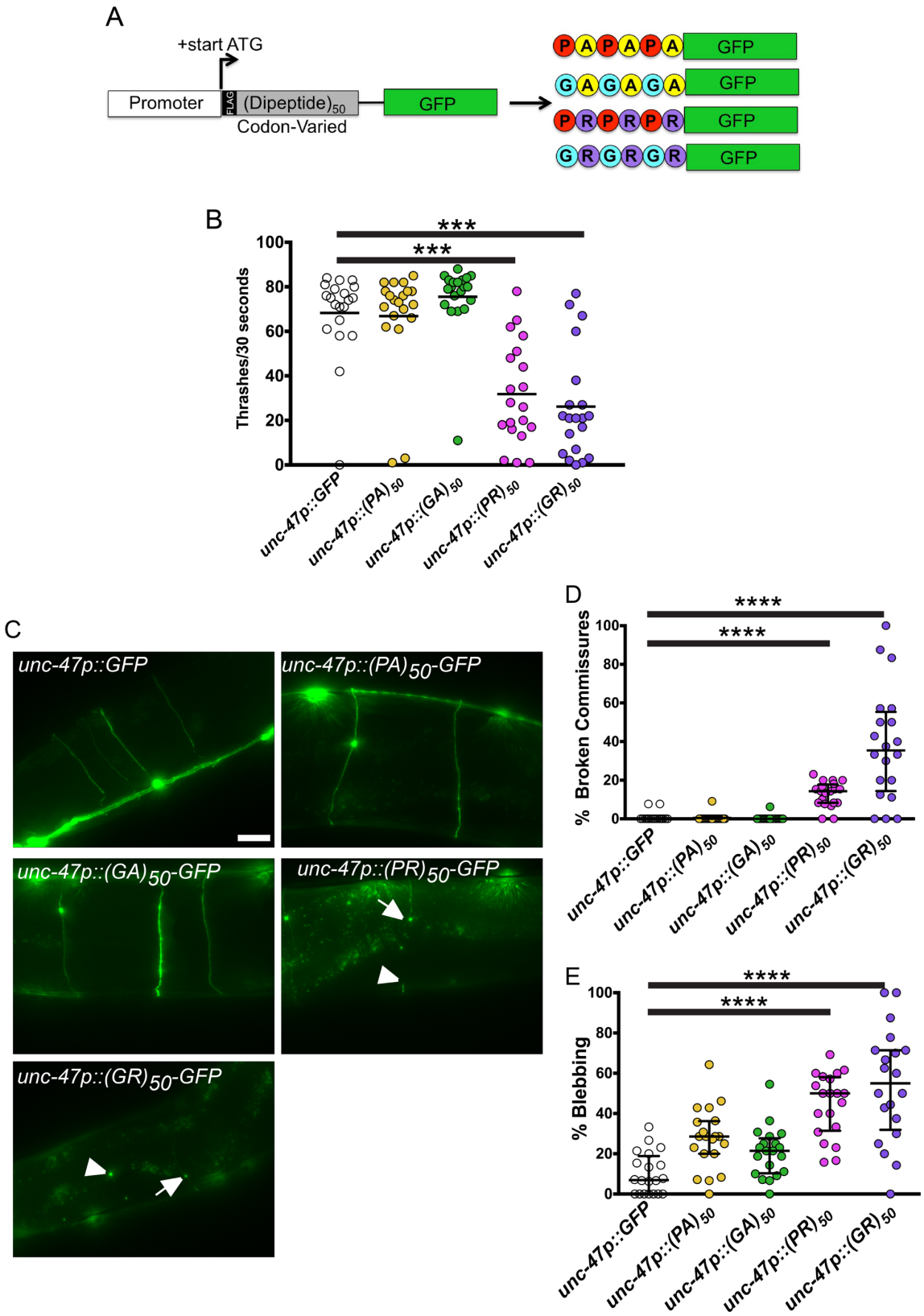

A
+start ATG
Promoter
FLAG
(Dipeptide)50
Codon-Varied
GFP
PAPAPA GFP
GAGAGA GFP
PRPRPR GFP
GRGRGR GFP
B
Thrashes/30 seconds
***
***
unc-47p::GFP
unc-47p::(PA)50
unc-47p::(GA)50
unc-47p::(PR)50
unc-47p::(GR)50
C
unc-47p::GFP
unc-47p::(PA)50-GFP
unc-47p::(GA)50-GFP
unc-47p::(PR)50-GFP
unc-47p::(GR)50-GFP
D
% Broken Commissures
****
****
E
% Blebbing
****
****

Figure 2.1 Arginine-Containing DPRs Are Toxic in Motor Neurons

(A) Molecular strategy for expression of codon-varied DPRs in *C. elegans*. (B) Liquid thrashing quantification of transgenic animals expressing the indicated DPR under the motor neuron-specific *unc-47* promoter. N=20 animals per genotype. Each symbol represents quantification for one animal; the horizontal line indicates the population median. ***P<0.001 versus GFP control (one-way non-parametric ANOVA with Dunn's post hoc test). (C) Representative images of *unc-47*+ motor neurons in animals expressing the indicated DPR and *unc-47p::GFP* to mark the commissures. Arrow points to an example of membrane blebbing. Arrowhead points to examples of commissure breaks. Scale bar=10 μm. (D) Quantification of commissure breaks and (E) membrane blebbing in each of the indicated strains. For each animal, we counted the total number of commissures with blebbing or breaks and divided by the total number of detectable commissures. *Unc-47p::GFP* was used to mark the commissures. N=20 animals per genotype. Each symbol represents quantification for one animal expressed as a percentage. The horizontal line indicates the population median with the interquartile range. ****P<0.0001 versus GFP control (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test). N=20 animals per genotype.

2.2.2 Toxicity of *C9orf72* Dipeptides in Muscle Cells

C. elegans motor neurons are extremely small (2–3 μm in diameter), making cell biological analysis of DPR localization in these cells difficult. Additionally, *C. elegans* neurons are highly resistant to RNA-interference-mediated gene knockdown, which is commonly used to explore mechanisms of protein toxicity in *C. elegans* disease models [136, 248]. To overcome these limitations, we expressed the DPRs in the larger muscle cells under the *myo-3* promoter. Muscle cells regulate motility and are a common site of expression for other *C. elegans* disease models [127, 249, 250]. As we observed in motor neurons, muscle expression of arginine-rich DPRs was highly toxic. Expression of either $(PR)_{50}$ or $(GR)_{50}$ in muscle led to completely penetrant embryonic and larval lethality that could be suppressed by feeding animals *gfp*(RNAi)-expressing bacteria (Figure 2.2 A). *gfp*(RNAi)-sensitive embryonic or larval lethality was not observed in animals expressing either $(PA)_{50}$ or $(GA)_{50}$. $(PR)_{50}$ or $(GR)_{50}$ animals that were removed from *gfp*(RNAi) after embryonic development survived and went on to develop into adults.

However, these DPR-expressing adults exhibited age-dependent paralysis (Figure 2.2 B). $(PR)_{50}$ animals also showed a significantly reduced lifespan that was not observed in $(GR)_{50}$ expressing animals (Figure 2.2 C). The enhanced paralysis observed in $(PR)_{50}$ and $(GR)_{50}$ animals was not due to higher protein levels since *in vivo* imaging and flow cytometry quantification showed that $(PR)_{50}$-GFP and $(GR)_{50}$-GFP fluorescence levels were significantly below those of all other DPRs (Figure 2.3 A, B). It was also not due to higher mRNA expression, since qPCR showed that all DPRs were expressed at similar levels (Figure 2.3 C). The toxicity phenotype required protein production and was not due to any potential RNA effects, since $(PR)_{50}$ transgenic animals lacking a start ATG

exhibited no detectable toxicity (Figure 2.3 D). $(PR)_{50}$ toxicity was dependent on the number of repeats, as animals expressing 5, 15 or 25 repeats did not exhibit embryonic lethality [number of viable transgenic progeny (>L4)/total transgenic progeny after 72h—$(PR)_5$-GFP—139/145; $(PR)_{15}$-GFP—68/75; $(PR)_{25}$-GFP—114/117; P>0.1, Fisher's exact test]. Adults expressing $(PR)_5$-GFP exhibited robust expression that appeared concentrated within almost all nuclei (Figure 2.4 A). Adults expressing both $(PR)_{15}$-GFP and $(PR)_{25}$-GFP expressed much lower levels of GFP when compared with $(PR)_5$-GFP, although the protein that was present also appeared to be concentrated within the nucleus and nucleolus (Figure 2.4 A, B). Adults expressing $(PR)_{25}$ and $(PR)_{50}$, but not those expressing $(PR)_5$ or $(PR)_{15}$, exhibited accelerated age-dependent paralysis (Figure 2.4 C), even though $(PR)_{25}$ was present at dramatically lower levels than $(PR)_5$ (Figure 2.4 B). Together, these studies show that arginine-containing dipeptides exhibit developmental and post-developmental age-dependent toxicity in *C. elegans,* and in the case of the PR dipeptide, this toxicity is repeat-length-dependent.

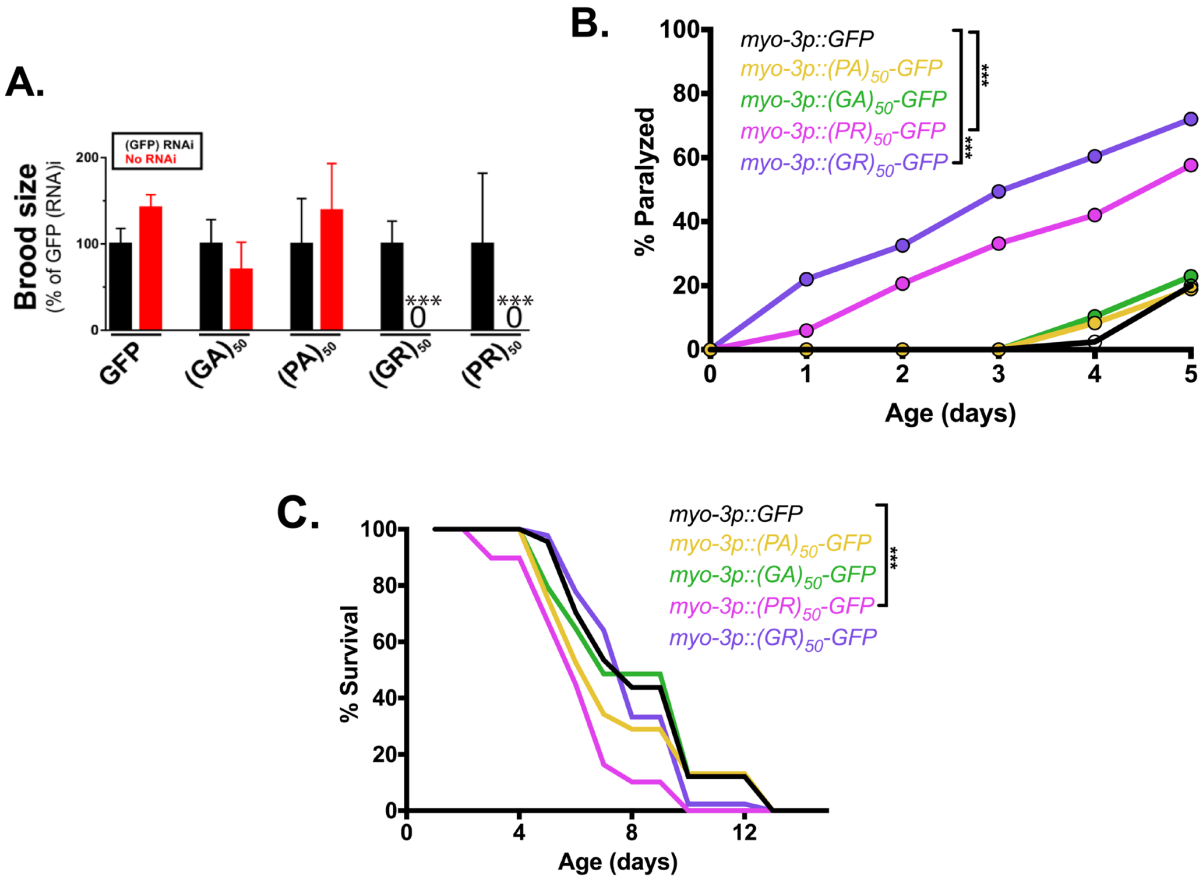


Figure 2.2 Muscle-Expressed Arginine-Rich Dipeptides Are Toxic in *C. elegans*

(A) Brood size for worms expressing each integrated dipeptide protein in the absence or presence of *gfp*(RNAi) at 25°C. Data for each strain are normalized to the mean brood size of animals grown on *gfp*(RNAi). ***P<0.001 (one-way ANOVA with post hoc Tukey's test). (B) Paralysis assay for adult animals raised in the absence of *gfp*(RNAi). N=43–49 animals per genotype. ***P<0.001 (Log-rank test with Bonferroni adjusted P-value). (C) Lifespan for worms expressing each integrated dipeptide protein in the absence of *gfp*(RNAi) at 25°C, N=50 animals per genotype. ***P<0.001 (Log Rank Test with -Bonferroni adjusted P-value)

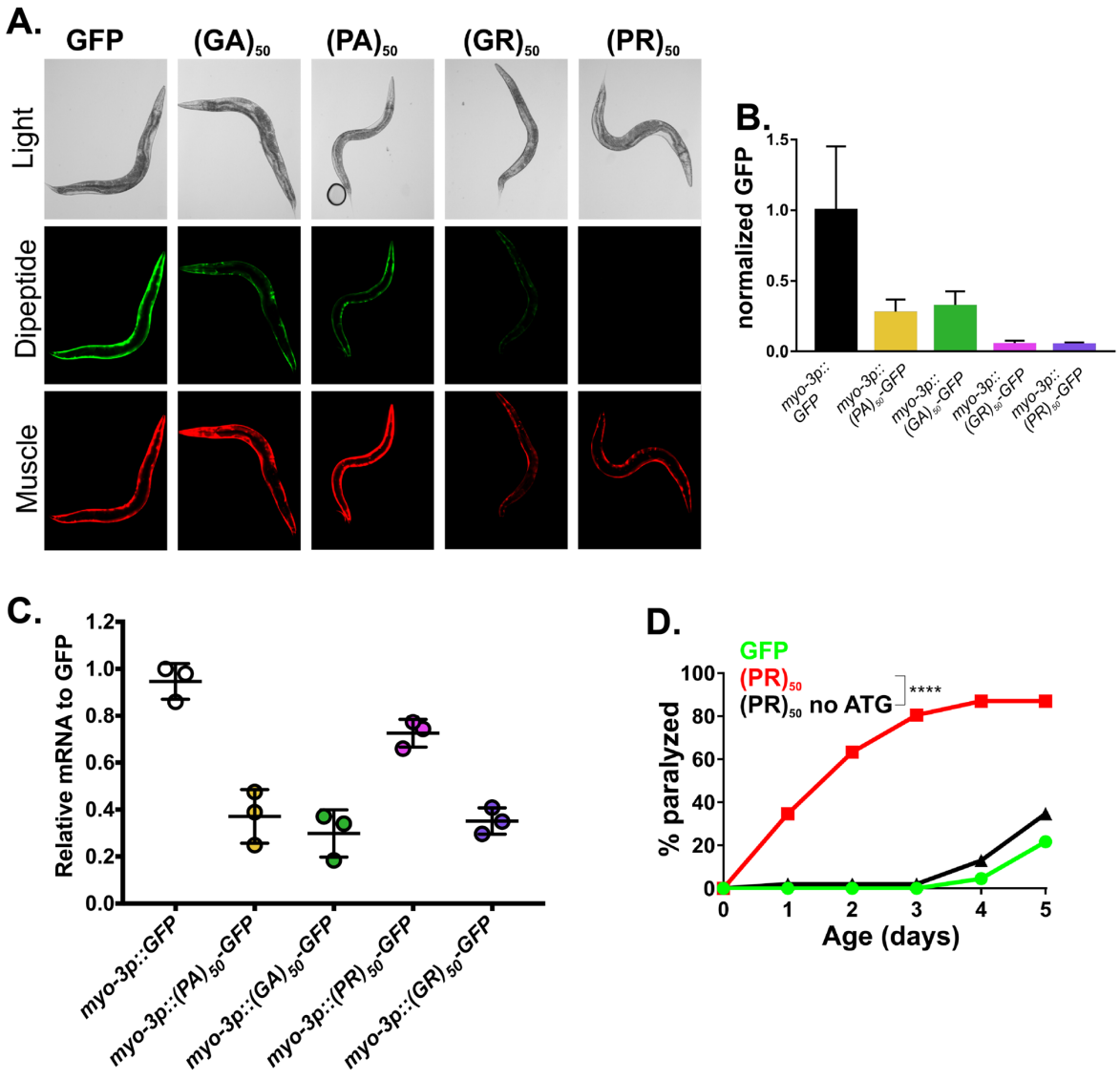


Figure 2.3 Toxicity of PR Was Not Due to Increased Levels of PR or RNA Toxicity

(A) Expression-matched representative images of transgenic animals expressing each dipeptide under the *myo-3* promoter. (B) Average fluorescence of GFP in a worm, as measured by the COPAS Biosorter through measuring the total fluorescence of the worm and dividing it by the time of flight of the worm. (C) Relative mRNA expression of the different dipeptides divided by the control, *myo-3p::GFP*. Actin was used as a control. (D) Paralysis assay for adult animals raised in the absence of *gfp*(RNAi). (N=50/genotype) ***P<0.001 (Log-rank test with Bonferroni adjusted P-value).

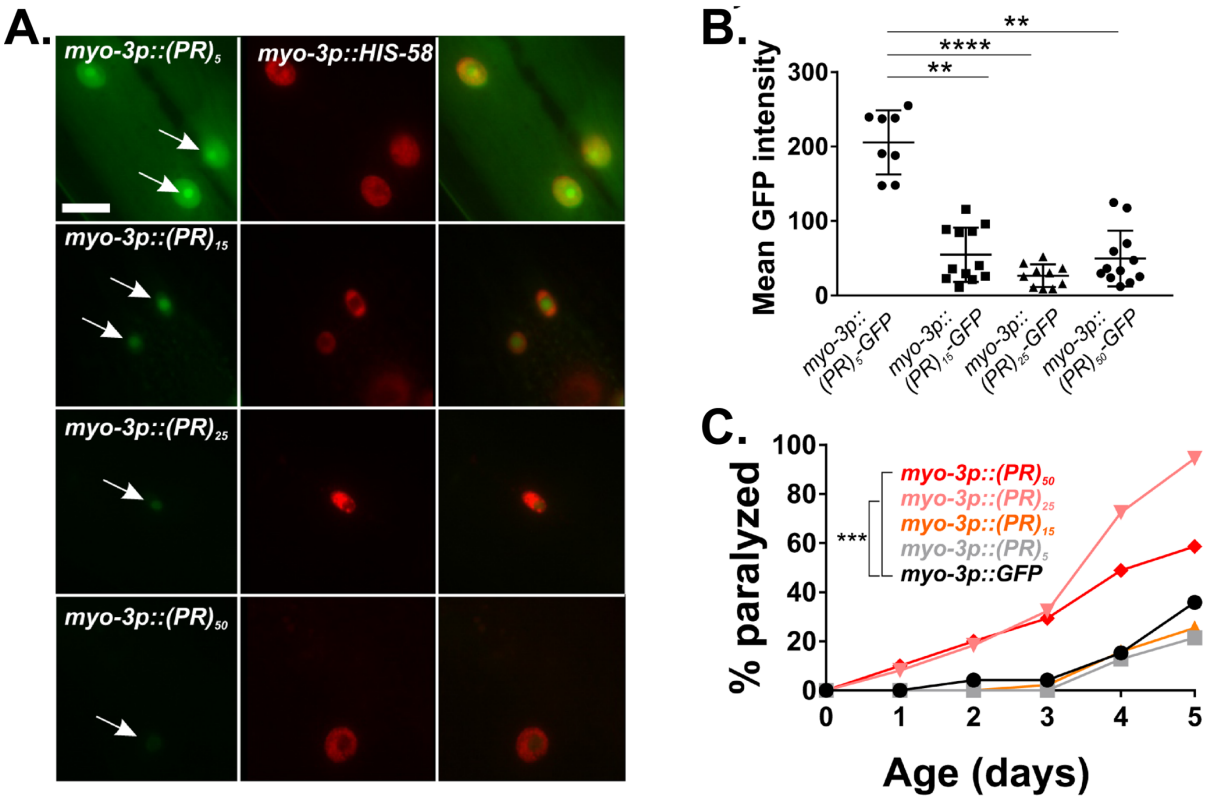


Figure 2.4 The Toxicity of PR is Length Dependent

(A) Fluorescent microscopy of Day 1 adult hermaphrodites expressing $(PR)_5$-GFP, $(PR)_{15}$-GFP, $(PR)_{25}$-GFP or $(PR)_{50}$-GFP (green) in the muscle. Muscle DNA (marked by a *myo-3p::his-58-mCherry* reporter) is in red. Arrows point to sites of nuclear GFP puncta. Scale bar=10 μm. All images were acquired using identical exposure settings. The row labeled '*myo-3p::(PR)*$_{50}$ (enhanced)' was adjusted for brightness and contrast in the GFP channel differently from the other images so that the $(PR)_{50}$-GFP signal is observable. (B) Quantification of the GFP levels from the indicated PR repeat animals. Each point shows the measured value for a single nucleus; the horizontal line indicates the population median with the interquartile range. **P<0.01, ****P<0.0001 versus GFP control (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test). (C) Paralysis assay of animals expressing the indicated number of (PR) repeats under the

control of the *myo-3* promoter. 'Day 0' animals were isolated as L4 stage animals. N=48–50 animals per genotype. ***P<0.001 versus GFP control (Log-rank test with Bonferroni adjusted P-value).

2.2.3 Toxic PR and GR Are Enriched in the Nucleolus

To gain insight into the cell biological properties of $(PR)_{50}$ and $(GR)_{50}$ proteins, we examined their subcellular localization patterns in live animals expressing each DPR in the muscle. Animals expressing both GFP and non-toxic $(PA)_{50}$-GFP exhibited diffuse nuclear and non-nuclear GFP expression that was always excluded from the histone-free nucleolar region (Figure 2.5 A, B). Muscle expressed $(GA)_{50}$-GFP formed intense non-nuclear puncta. In some cases, we noted $(GA)_{50}$-GFP signal in the nucleus, but this nuclear localization was always diffuse and never in puncta (Figure 2.5 A, B). Muscle expressed $(GR)_{50}$-GFP and $(PR)_{50}$-GFP both exhibited substantially lower GFP levels than either of the other DPRs, despite their similar mRNA expression levels (Figure 2.3 C), and appeared to be predominantly localized to the histone-free nucleolar region of the nucleus (Figure 2.5 A, B). Non-nuclear localization was always observed in $(GR)_{50}$ animals but was never observed in $(PR)_{50}$ animals (Figure 2.5 A, B).

The motility of the DPRs that form puncta (GA_{50}, PR_{50}, and GR_{50}) was unclear. Immobile puncta (aggregates) are a common feature of neurodegenerative diseases [46, 199, 251]. To determine whether *C9orf72* DPRs also form protein aggregates, we performed fluorescence recovery after photobleaching (FRAP) to assay their respective motility. Non-nuclear $(GA)_{50}$-GFP exhibited little recovery after FRAP (Figure 2.6 A, B),

suggesting these puncta are highly immobile protein aggregates. In contrast, the nucleolar accumulations of $(GR)_{50}$ and $(PR)_{50}$ exhibited substantial fluorescence recovery (Figure 2.6), which is inconsistent with the known biophysical properties of bona fide aggregates [252]. Since both $(PR)_{50}$ and $(GR)_{50}$ appeared to be localized to the nucleolus, we compared their FRAP dynamics with those of FIB-1, which is homologous to the mammalian nucleolar protein fibrillarin [253]. FIB-1 exhibited recovery kinetics similar to those of $(GR)_{50}$ and $(PR)_{50}$, suggesting that the dynamics of these two DPRs behave similarly to a bona fide nucleolar protein (Figure 2.6 C).

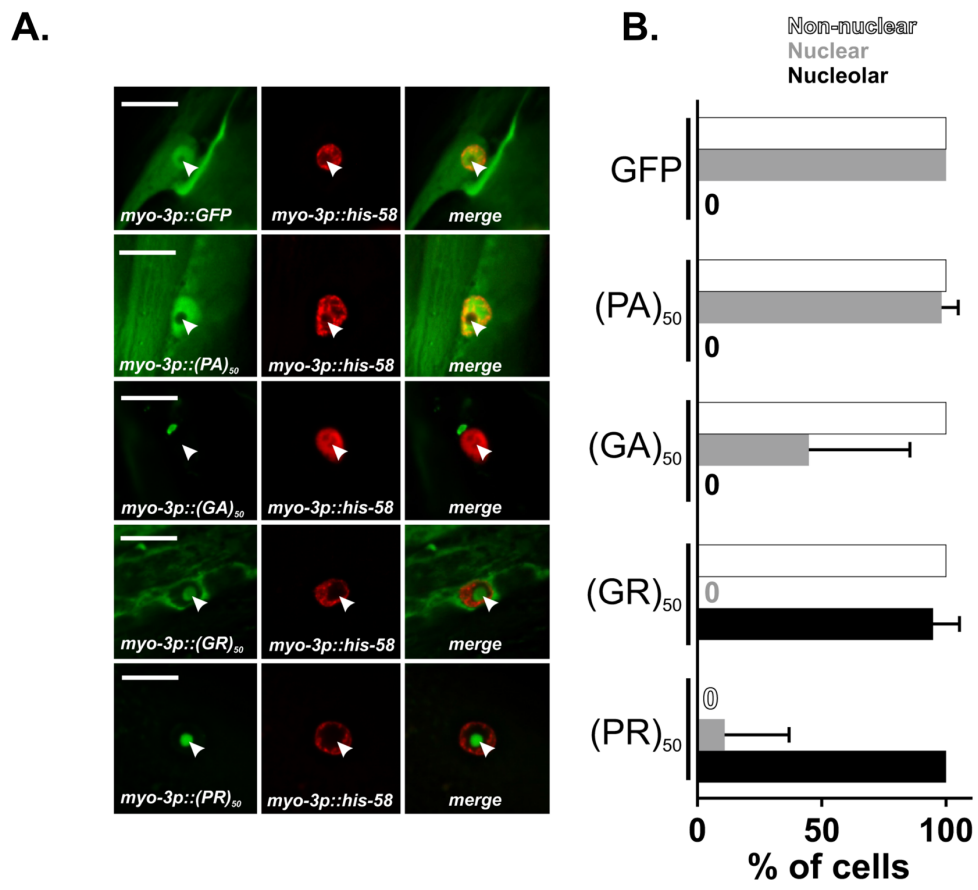


Figure 2.5 $(PR)_{50}$/$(GR)_{50}$ Are Localized to the Nucleolus in *C. elegans* Muscle Cells

(A) Imaging of live, anesthetized Day 1 adult *C. elegans* hermaphrodites expressing the indicated dipeptide. Green shows dipeptide-GFP localization expressed in muscle, red shows histone-tagRFP within the muscle nucleus. Arrowhead points to the histone-free nucleolar region. Scale bar=10 μm. (B) Quantification of DPR–GFP signal that is observed in the nucleoplasm, nucleolus, and non-nuclear cellular compartments. N=55–78 nuclei from five to six animals. Data shown are means ± S.D.

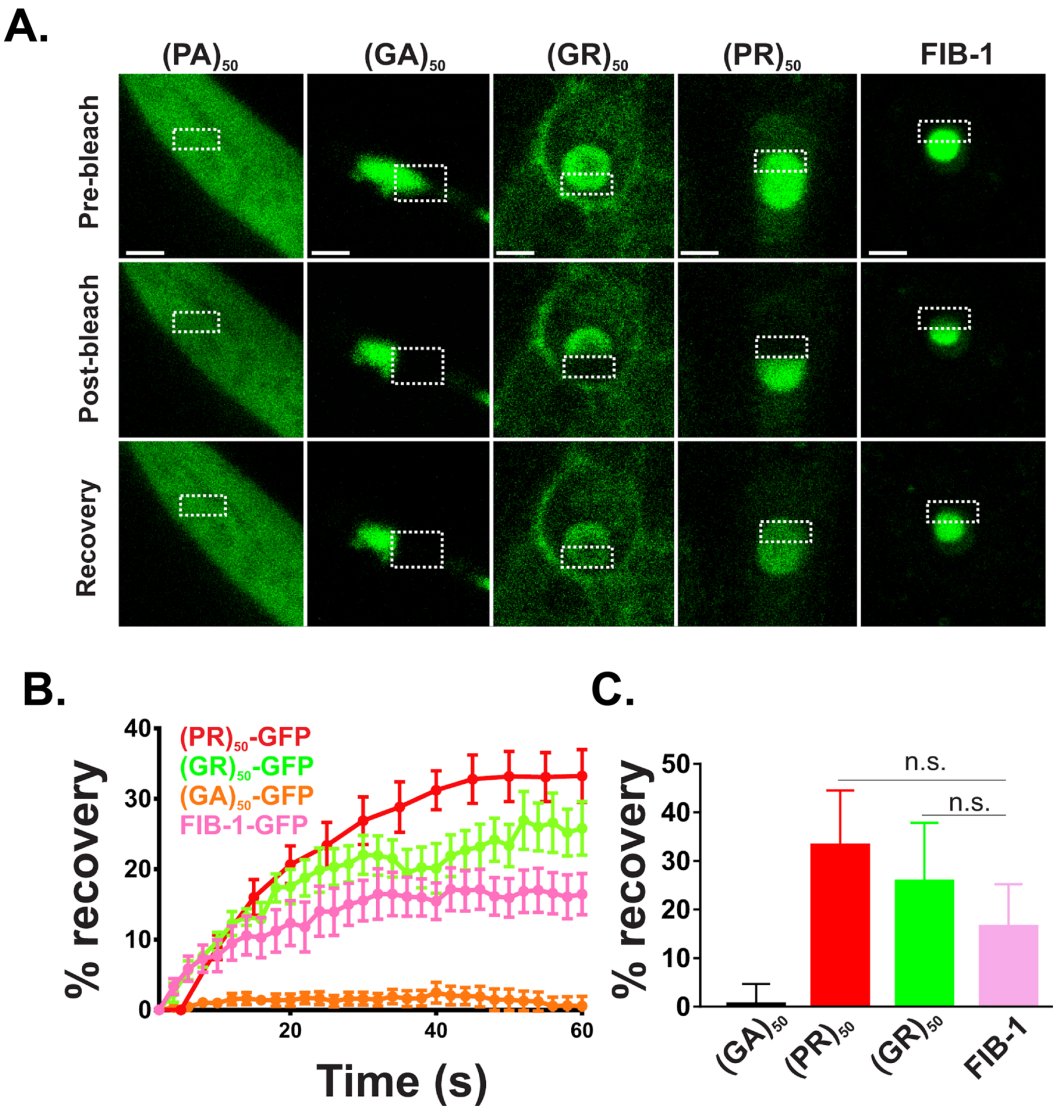


Figure 2.6 Arginine-Rich DPRs Are Not Aggregates

(A) Representative images from FRAP analysis of subcellular localized DPR proteins expressed in muscle. The dashed outline indicates the site of photobleaching and post-bleaching quantification. Recovery images are 150 seconds post-bleach. Scale bar=2 μm. (B) Quantification of FRAP imaging. Data shown are mean ± S.E.M. from 9 to 10 datasets per genotype. (C) Average equilibrium fluorescence recovery after 60 seconds. Data shown are mean ± S.D. from 9 to 10 datasets per genotype. 'n.s.' - versus FIB-1-GFP (one-way ANOVA with a post hoc Dunn's multiple comparison test).

2.2.4 Nuclear Localization is Required for PR and GR Toxicity

Given that the toxic dipeptides, $(PR)_{50}$ and $(GR)_{50}$, were primarily localized to the nucleus in muscle, we asked if nuclear localization was necessary for their toxicity. To test this, we tethered each of the arginine-rich DPRs to a signal sequence-transmembrane domain (SS-TM) tag that restricts the DPR to cellular membranes with the DPR oriented on the cytosolic side of the membrane. In this context, the membrane-localized $(PR)_{50}$ and $(GR)_{50}$ were excluded from the nucleus (Figure 2.7 A). In contrast to soluble $(PR)_{50}$ and $(GR)_{50}$, membrane-localized $(PR)_{50}$ and $(GR)_{50}$ were viable in the absence of *gfp*(RNAi) and showed no age-dependent paralysis (Figure 2.7 B). We further tested whether nuclear localization was sufficient for toxicity by fusing the $(PR)_{50}$ DPR to the coding sequence for the *his-58* gene, which encodes a DNA-binding histone only found in the nucleus. GFP-HIS-58 expressing animals were viable and motile and exhibited strong nuclear GFP expression (Figure 2.7 C, D). However, GFP-HIS-58-$(PR)_{50}$ expressing animals were not viable unless they were cultured under *gfp*(RNAi) conditions [number of viable transgenic progeny (>L4)/total progeny in the absence of GFP after 72h—GFP-HIS-58–348/406; GFP-HIS-58-$(PR)_{50}$–9/56; $P<0.0001$, Fisher's exact test]. Upon removal from *gfp*(RNAi), HIS-58-$(PR)_{50}$ animals exhibited age-dependent paralysis (Figure 2.7 D), and this paralysis occurred with more rapid onset than in $(PR)_{50}$ animals (Figure 2.7 E). Thus, nuclear localization further enhances $(PR)_{50}$ toxicity. GFP fluorescence levels of HIS-58-$(PR)_{50}$ were not higher than GFP fluorescence levels of $(PR)_{50}$ (Figure 2.7 F). Thus, the enhanced HIS-58-$(PR)_{50}$ toxicity was not due to higher expression levels of the transgene. Overall, these data show that nuclear localization is necessary and sufficient for $(PR)_{50}$ toxicity in *C. elegans*.

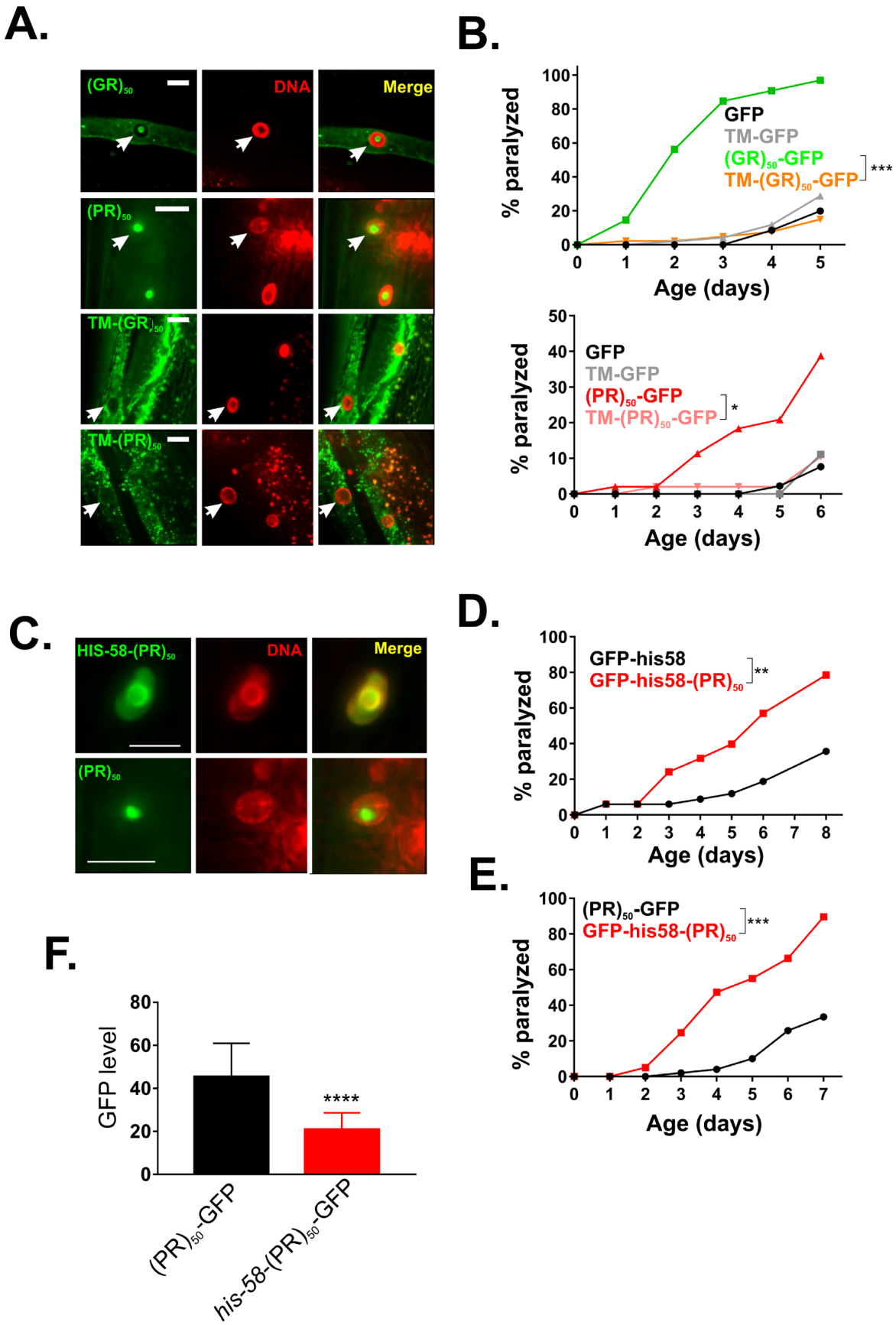

A.
(GR)50
DNA
Merge
(PR)50
TM-(GR)50
TM-(PR)50
B.
% paralyzed
GFP
TM-GFP
(GR)50-GFP
TM-(GR)50-GFP
***
Age (days)
GFP
TM-GFP
(PR)50-GFP
TM-(PR)50-GFP
*
Age (days)
C.
HIS-58-(PR)50
DNA
Merge
(PR)50
D.
% paralyzed
GFP-his58
GFP-his58-(PR)50
**
Age (days)
E.
% paralyzed
(PR)50-GFP
GFP-his58-(PR)50
***
Age (days)
F.
GFP level
****
(PR)50-GFP
his-58-(PR)50-GFP

Figure 2.7 Nuclear Localization of PR/GR is Necessary and Sufficent for Toxicity

(A) Representative images of the indicated soluble or membrane-anchored DPR (green), the HIS-58-mCherry nuclear marker (red), and a merged image. Arrows point to nuclear regions defined by the HIS-58 mCherry signal. Coincident green and red signals in the TM-$(PR)_{50}$ image is intestinal autofluorescence. Scale bar=10 μm. (B) Paralysis assay comparing animals expressing soluble or transmembrane localized $(GR)_{50}$ (top) or $(PR)_{50}$ (bottom). N=48–50 animals per genotype, ***P<0.001, **P<0.01, *P<0.05 (Log-rank test with Bonferroni adjusted P-value). (C) Representative images of HIS-58 anchored $(PR)_{50}$ (green; top) or unanchored $(PR)_{50}$ (green; bottom), HIS-58-mCherry nuclear marker (red) and a merged image. Scale bar=5 μm. (D) Paralysis assay comparing animals expressing nuclear-localized GFP or $(PR)_{50}$. N=30 animals per genotype, **P<0.01 (Log-rank test with Bonferroni adjusted P-value). (E) Paralysis assay comparing soluble $(PR)_{50}$ with HIS-58-$(PR)_{50}$. N=50 animals [$(PR)_{50}$] or 80 animals [HIS-58-$(PR)_{50}$], ***P<0.001 (Log-rank test with Bonferroni adjusted P-value). (F) Quantification of the GFP levels from the indicated genotype animals. Each point shows the measured value for a single nucleus, horizontal line indicates population median with interquartile range. ****P<0.0001 versus $(PR)_{50}$-GFP (student's t-test with non-parametric Mann-Whitney U test).

2.2.5 Age-Associated Toxicity of PR and GR Are Due to Different Mechanisms

The toxicity of $(PR)_{50}$ in an aged cellular environment could be due to the stochastic accumulation of toxic protein levels or toxic protein conformations over long periods of time. Alternatively, the physiological process of aging could lead to a decline in cellular processes that oppose $(PR)_{50}$ toxicity. To differentiate between these two hypotheses, we tested the role of the aging process in the toxicity of $(PR)_{50}$. If $(PR)_{50}$ toxicity is due to the accumulation of toxic protein levels/conformation over time, lifespan-extending conditions should have no effect on the rate of toxicity. However, if the aging process actively regulates $(PR)_{50}$ toxicity, alterations in the rate of aging should enhance (if lifespan shortening) or delay (if lifespan extending) the toxicity of $(PR)_{50}$. The insulin signaling pathway is the most well-characterized pathway that regulates lifespan and youthfulness from worms to humans [111]. Mutations in this pathway are thought to extend lifespan by promoting a youthful cellular physiological state. Therefore, mutations in the insulin signaling pathway can be used to separate chronological and physiological states of aging. In *C. elegans*, the *daf-2* gene encodes the homolog of the insulin/IGF receptor. Loss-of-function mutations in *daf-2* exert their effects via signaling-dependent activation of the FOXO transcription factor DAF-16, as well as the heat shock transcription factor HSF-1 [112]. Mutations in *daf-16* and *hsf-1* suppress phenotypes associated with *daf-2* mutations and accelerate the rate of aging [111].

To determine whether the toxicity of $(PR)_{50}$ is dependent on the chronological age of the cell or the physiological age of the cell, we crossed the $(PR)_{50}$ transgene into the *daf-2*(*e1370*); *daf-16*(*mu86*) double mutant and segregated all possible $(PR)_{50}$ genotypes. If $(PR)_{50}$ toxicity is due to the stochastic accumulation of toxic $(PR)_{50}$ molecules or

conformations over time, then reduced insulin signaling should not alter the age-related phenotypes. However, if $(PR)_{50}$ toxicity is due to aging-induced changes in cellular physiology, then alterations in the rate of aging should slow the initiation and progression of toxicity. We found that in the *daf-2* mutant background, age-dependent declines in $(PR)_{50}$ motility were significantly reduced (Figure 2.8 A). These *daf-2*-dependent phenotypes were suppressed in a *daf-2*; *daf-16* double mutant and were accelerated in the *daf-16* mutant (Figure 2.8 B). Likewise, a mutation in the heat shock transcription factor *hsf-1*, which accelerates the rate of aging [254], enhanced the toxicity of $(PR)_{50}$ (Figure 2.8 C). Interestingly, the protective effects of the *daf-2* mutant were not observed in $(GR)_{50}$ expressing animals (Figure 2.8 A). The protective effects of insulin signaling appear to act downstream from DPR accumulation and nuclear localization, as DPR abundance and subcellular distribution did not appear to be altered in any of these genetic backgrounds (Figure 2.8 D-F). We also found that in motor neurons expressing $(PR)_{50}$, *daf-2* reduced the levels of neurodegeneration-associated membrane blebbing (Figure 2.8 G). These data demonstrate that alterations in the rate of aging impact $(PR)_{50}$ toxicity in both *C. elegans* muscle cells and motor neurons. It also suggests that $(PR)_{50}$ toxicity is dependent on the cellular aging process and is not due to the buildup of toxic protein levels/conformations over time. Additionally, these data imply that $(PR)_{50}$ and $(GR)_{50}$ are toxic through partially distinct mechanisms as enhanced DAF-16/FOXO activation protects $(PR)_{50}$ animals but has no protective effects on $(GR)_{50}$ animals.

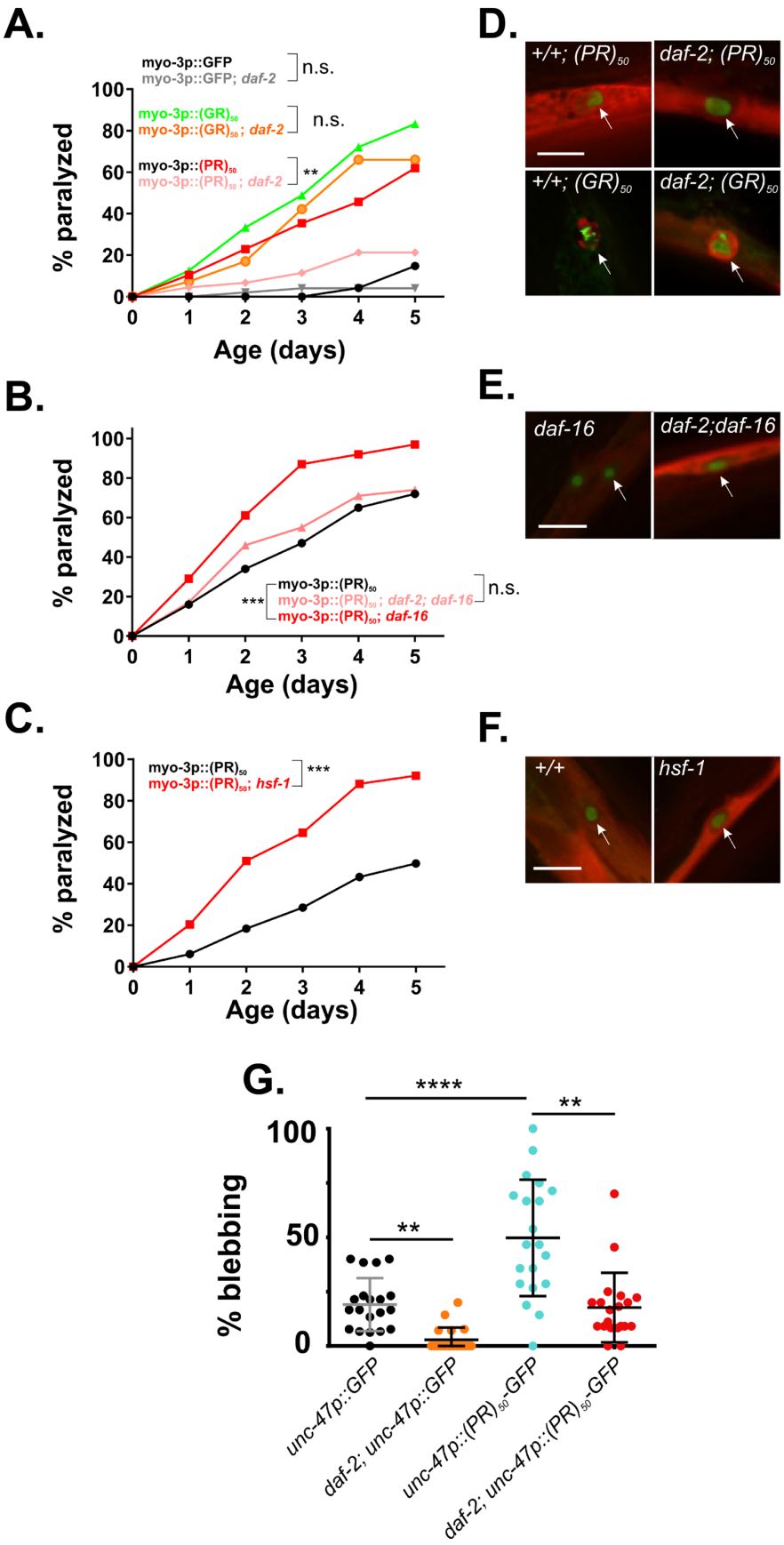

A.
myo-3p::GFP
myo-3p::GFP; daf-2
n.s.
myo-3p::(GR)50
myo-3p::(GR)50; daf-2
n.s.
myo-3p::(PR)50
myo-3p::(PR)50; daf-2
**
% paralyzed
Age (days)
B.
myo-3p::(PR)50
***
myo-3p::(PR)50; daf-2; daf-16
n.s.
myo-3p::(PR)50; daf-16
% paralyzed
Age (days)
C.
myo-3p::(PR)50
myo-3p::(PR)50; hsf-1
***
% paralyzed
Age (days)
D.
+/+; (PR)50
daf-2; (PR)50
+/+; (GR)50
daf-2; (GR)50
E.
daf-16
daf-2;daf-16
F.
+/+
hsf-1
G.
****
**
**
% blebbing
unc-47p::GFP
daf-2; unc-47p::GFP
unc-47p::(PR)50-GFP
daf-2; unc-47p::(PR)50-GFP

Figure 2.8 Altering the Biological Rate of Aging Affects $(PR)_{50}$ Toxicity

(A) Paralysis assay of animals expressing GFP, $(GR)_{50}$ or $(PR)_{50}$ under the control of the *myo-3* promoter in the wild-type or *daf-2(e1370)* background. 'Day 0' animals were isolated as L4 stage animals. N=48–50 animals per genotype. n.s., 'not significant'; **P<0.01 versus GFP control (Log-rank test with Bonferroni adjusted P-value). (B) Paralysis assay of animals expressing $(PR)_{50}$ under the control of the *myo-3* promoter in the wild-type, *daf-16(mu86)*, or *daf-2(e1370); daf-16(mu86)* background. N=48–50 animals per genotype. n.s., 'not significant'; ***P<0.001 versus $(PR)_{50}$ (Log-rank test with Bonferroni adjusted P-value). (C) Paralysis assay of animals expressing $(PR)_{50}$ under the control of the *myo-3* promoter in the wild-type or *hsf-1(sy441)* mutant background. N=48–50 animals per genotype. ***P<0.001 versus wild-type (Log-rank test with Bonferroni adjusted P-value). (D) Fluorescent microscopy images of day 1 adult worms expressing either $(GR)_{50}$ or $(PR)_{50}$ (green) and soluble muscle mCherry (red) in the wild-type or *daf-2(e1370)* mutant background. Arrow points to sites of nuclear DPR accumulation. Scale bar=10 μm. (E) Fluorescent microscopy images of Day 1 adult worms expressing $(PR)_{50}$ (green) and soluble muscle mCherry (red) in the *daf-16(mu86)* or *daf-2(e1370); daf-16(mu86)* mutant background. Arrow points to sites of nuclear DPR accumulation. Scale bar=10 μm. (F) Fluorescent microscopy images of day 1 adult worms expressing $(PR)_{50}$ (green) and soluble muscle mCherry (red) in the wild-type or *hsf-1(sy441)* mutant background. Arrow points to sites of nuclear DPR accumulation. Scale bar=10 μm. (G) Quantification of membrane blebbing in each of the indicated strains. For each animal, we counted the total number of commissures with blebbing and divided by the total number of detectable commissures. N=20 animals per genotype.

Each symbol represents quantification for one animal expressed as a percentage. The horizontal line indicates population median with interquartile range. **P<0.01; ****P<0.0001 versus GFP control (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test). N=20 animals per genotype.

2.2.6 Conserved Genes Suppress $(PR)_{50}$ Toxicity

To identify genes required for $(PR)_{50}$ toxicity, we performed two unbiased screens using muscle-expressed PR. In both screens, we initially selected animals that exhibited enhanced growth, since muscle expression of $(PR)_{50}$ causes larval arrest (Figure 2.9 A). In the first screen, we used an unbiased chemical mutagenesis approach to identify gene mutations that allowed $(PR)_{50}$ growth over a two-week period. While this approach may identify mutations that subtly enhance growth, one major caveat is that we are unable to recover mutations in essential genes that may be required for wild-type growth or reproduction. To address this limitation, we also performed an unbiased whole-genome RNAi screen for suppressors of $(PR)_{50}$ growth arrest over a shorter one-week period. RNAi screens have several advantages over chemical mutagenesis screens. First, the RNAi approach knocks down mRNA levels, versus complete knockout or gain-of-function alleles that can be produced by ENU mutagenesis. Second, RNAi knockdown can be initiated post-developmentally, while gene knockout from an ENU-induced mutation causes alterations in gene function throughout the entirety of life. Finally, the gene knockdown effect from RNAi is enhanced over time, in contrast to animals carrying ENU induced mutations that exhibit the same alterations of gene function between animals. This variable knockdown effect from RNAi is particularly useful when screening

populations of animals since it leads to the generation of animals expressing a range of loss-of-function phenotypes (from partial loss-of-function to complete null). Because of these differences, RNAi screens allow for the identification of genes whose null phenotypes may be lethal or sterile.

The forward mutagenesis screen was performed by treating young animals with N-ethyl-N-nitrosourea (ENU), a DNA mutagen that primarily introduces point mutations [255]. Animals were allowed to self-fertilize for three generations on *gfp*(RNAi) bacteria and were then moved onto *E. coli* OP50 bacteria, a common food source for lab-grown *C. elegans*, which does not repress expression of $(PR)_{50}$. Animals that grew while expressing $(PR)_{50}$ were identified as suppressors. Fifteen suppressors were individually isolated in this manner. However, all fifteen suppressors exhibited strongly reduced transgene expression, based on the expression a *myo-3::RFP* reporter located on the same transgene as the *myo-3p::(PR)$_{50}$-GFP* reporter. Transgene suppression is a commonly observed phenotype in *C. elegans* [256]. To ensure that these mutants caused transgene suppression, we measured the RFP signal using a COPAS BIOSORT, as well as the $(PR)_{50}$-GFP mRNA levels using qPCR. The suppressors had a five-fold reduction in RFP fluorescent signal (Figure 2.9 B). The level of $(PR)_{50}$ mRNA was further reduced, with five of the six tested suppressors having no detectable $(PR)_{50}$ mRNA and one suppressor having a five-fold reduction in levels of $(PR)_{50}$ mRNA (Figure 2.9 C). These data show that all the suppressors identified in our ENU mutagenesis screen were likely transgene suppressors. Therefore, the mechanism of suppression of ENU-isolated suppressors is indirect and not likely to inform us as to specific and conserved mechanisms of toxicity of $(PR)_{50}$.

As an alternative to the ENU suppressor screen, we also performed a genome-wide RNAi screen to identify suppressors of muscle expressed $(PR)_{50}$ toxicity. We employed a commercially available RNAi library to screen 15,865 genes. Eggs were placed on plates seeded with bacteria expressing double-stranded RNA (dsRNA) against the gene of interest. After one week of growth, RNAi suppressors were identified by whether their respective plates contained mobile animals in multiple stages of development. While this screen did identify some transgene suppressors, it also identified several suppressors that did not appear to affect transgene expression levels. We identified 391 initial 'hits' that allowed some growth of $(PR)_{50}$ expressing animals. After six independent rescreens, we identified 12 genes that caused suppression of $(PR)_{50}$ toxicity in at least 4/6 rescreens but did not have strong effects on transgene expression levels (Table 2.1). Seven of the twelve $(PR)_{50}$ suppressors have been previously reported to have loss-of-function phenotypes of 'lethal' or 'sterile', suggesting they would not be isolated in a chemical mutagenesis screen. Remarkably, all of the PR suppressors have strong sequence homology to human genes, and four of the twelve genes were previously identified in PR or GR modifier screens performed in yeast, *Drosophila*, or mammalian cells [98-100], suggesting that mechanisms of $(PR)_{50}$ are evolutionarily conserved. Based on sequence homology, our PR suppressors fell into three distinct functional categories: nuclear transport/RNA binding, ubiquitin-mediated protein degradation, and chromatin regulation. While the lab has pursued these genes further, including demonstrating that the human homologs of these genes also protect against PR toxicity in mammalian cells, I have not been involved in that phase of the project and will not discuss it further. However, this screen illustrates the power of *C.*

elegans genetic approaches for the discovery of new and conserved mechanisms of RAN protein toxicity.

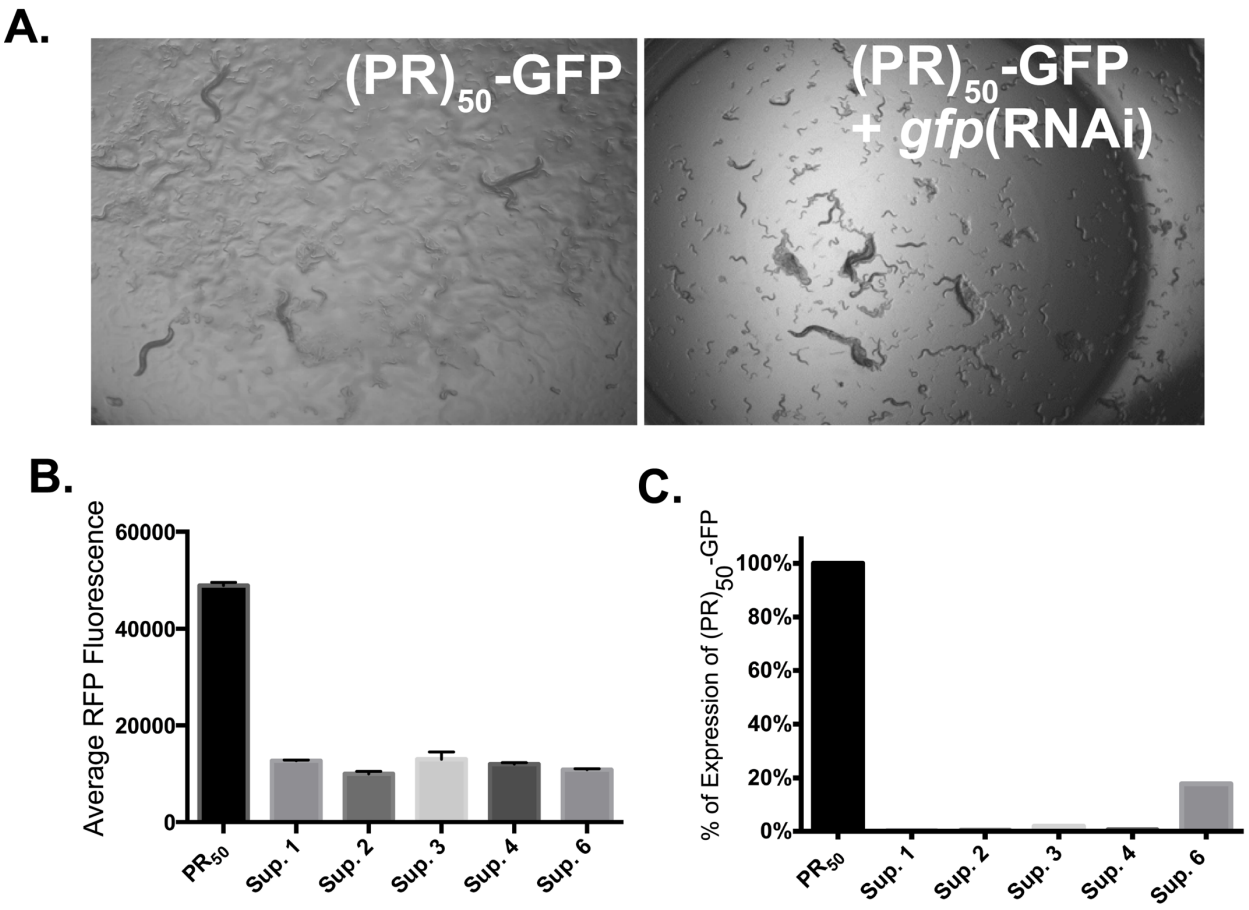


Figure 2.9 Transgene Suppressors Identified in Forward Mutagenesis Screen

(A) Example images of *myo-3p*::$(PR)_{50}$-GFP being expressed and on GFP RNAi to mimic suppression of PR toxicity. (B) Average fluorescence of RFP and (C) GFP in a worm, as measured by the COPAS Biosorter through measuring the total fluorescence of the worm and dividing it by the time of flight of the worm. N=500 animals.

Table 2.1 RNAi Gene Knockdowns that Suppress *myo-3p::(PR)$_{50}$-GFP* Toxicity

C. elegans	Yeast	Fly	Human	Description	Predicted NLS
Nuclear transport / RNA binding					
Rsp-6/C33H5.12	MRD1	X16[2]	SRSF3[1]	RNA binding protein; nuclear localized; homologous to yeast NSR1	No
Ztf-4/T10B11.3	NGR1	CG42458[2]	NCOA5/hnRNPC[1]	contains RRM domain	Yes
C18H9.3	SYH1	Gigyf	GIGYF[1]	yeast homolog SYH1 binds proline-rich sequences and influences nuclear pore distribution	Yes
ima-3/F32E10.4	KAP122[4]	Kap-α3[2,3]	KPNA3[1]	1 of 3 nuclear importin alpha subunits	No
Ergo-1	None	AGO1T1	AGO3	Production of endo-siRNAs	Yes
Ubiquitin-mediated protein degradation					
Bath-43/T16H12.5	None	roadkill	SPOP	Speckle-type POZ protein; may interact with histones	Yes
Ufd-2/T05H10.5	Ufd2	UBE4B/ CG9934	UBE4a	E3 and/or E4 ubiquitin ligase	No
Rpn-12/ZK20.5	RPN12	RPN12	PSMD8	nonATPase regulatory subunit of the 26S proteasome	No
Rpn-9/T06D8.8	RPN9	RPN9	PSMD13	nonATPase regulatory subunit of the 26S proteasome	No
Chromatin regulation					
Met-2 /R05D3.11	MET2	eggless	SETDB1	Histone methyltransferase	Yes
Utx-1/D2021.1	CYC8	Utx	UTX/UTY	Histone demethylase	Yes
Pis-1/T13F2.3	None	PTIP	PAXIP1	Pax transcription factor interacting protein	Yes

NLS=Nuclear Localization Signal

1 – Interacts with both PR and GR in human cells; [195]

2 – Enhanced the toxicity of GR_{50} in *Drosophila* model; [195]

3 – Enhanced the toxicity of PR_{25} in *Drosophila* model; [99]

4 – Suppresses the toxicity of PR_{50} when overexpressed in yeast; [98]

2.3 Discussion

The study of RAN peptides is still in its infancy, with the discovery of RAN translation less than a decade old [4]. Intensive study of the *C9orf72* dipeptides has established that the arginine-rich dipeptides, GR and PR, are consistently toxic in a

variety of systems [98, 99, 183], a finding replicated in our *C. elegans* model. Previously studied systems lack the genetic tools available in *C. elegans* to separate chronological aging from biological aging. We discovered modifying rates of aging had differential effects on $(PR)_{50}$ compared to $(GR)_{50}$. Unbiased screens for genetic suppressors of PR toxicity have been performed in yeast [98] and mammalian cells [100], but these unbiased screens lacked the context of a multicellular organism and may have missed relevant pathways essential for cellular viability or growth. Using our new *C. elegans* model, we bypassed this requirement using a post-developmental RNAi-based approach to screen for suppressors of $(PR)_{50}$ toxicity in a multicellular organism and identified several essential genes involved in previously identified PR toxicity pathways, as well as a novel nuclear ubiquitination pathway that regulates PR toxicity [98-100].

2.3.1 Cellular Localization of Dipeptides Is Linked with Toxicity

The *C9orf72* DPRs exhibit similar toxic and cellular characteristics in *C. elegans* to those observed in other models [98, 195]. In *C. elegans,* $(PR)_{50}$ and $(GR)_{50}$ are toxic in multiple tissues, while $(PA)_{50}$ and $(GA)_{50}$ are not toxic in any tissues. Both PA and GA localize in the cytoplasm, while PR and GR localize in the nucleus. PA is diffuse while GA forms aggregates. Neurodegenerative proteins typically aggregate [251, 257, 258]. Therefore, it was surprising that GA formed aggregates but did not cause toxicity. In other systems, GA dipeptides exhibit mild toxicity at repeat lengths ≥100 [96, 182, 183], and exhibit length-dependent toxicity in human cells [183]. Because we only examined GA at 50 repeats, our studies do not preclude the possibility that GA may also be toxic in *C. elegans* at higher repeat lengths.

We observed that both PR and GR are toxic in multiple tissues and localize to the nucleolus. Nuclear localization appears to be necessary and sufficient for toxicity. PR and GR are not toxic when sequestered from the nucleus and driving PR localization to the nucleus significantly increases PR toxicity. We were unable to generate nuclear enriched GR expressing animals, possible because animals are unable to grow under such toxic conditions. Arginine-rich DPRs are known to localize to the nucleolus [94, 184, 195, 196] and cause nucleolar stress [195, 196]. However, we are the first to establish that PR toxicity requires its localization to the nucleus. The arginine repeats seen in PR and GR mimic the serine-arginine (SR) repeats present in the SR domains of many RNA binding proteins involved in pre-mRNA splicing. One common property of these RNA binding proteins is that the SR domain is highly disordered and promotes de-mixing from the surrounding solution. This process is called phase separation and is a biophysical property of many membraneless organelles, such as the nuclear pore, nucleolus, and nuclear speckles. One of the first discovered characteristics of PR/GR was their ability to phase separate with RNA regulatory proteins in the nucleus [175]. While our studies suggest that such nuclear interactions are critical for PR toxicity, a caveat of these studies is the membrane anchors used to restrain DPRs from the nucleus could disrupt other important properties of PR and GR, such as phase separation or other ordered interactions that require a 3-dimensional space. Previous research found that the toxicity of the arginine-rich dipeptides is correlated with their ability to phase separate [184]. Therefore, in restricting PR and GR to a 2-D membrane, we may have also inhibited the ability of PR and GR to phase separate. To complement this membrane anchoring approach, we attempted to sequester PR and GR away from the nucleus using an nuclear

export signal (NES). However, PR-NES and GR-NES still localized to the nucleus (data not shown).

In mammalian cells, PR and GR also localize to other phase-separated compartments such as nuclear pores, spliceosomes, and stress granules [195, 242]. We did not observe PR and GR localization to any other phase-separated component besides the nucleolus. This could be due to signal intensity, as the GFP signal of $(PR)_{50}$ and $(GR)_{50}$ in worms is extremely low. Another possibility is that high expression of PR and GR in other systems may drive DPR localization to additional phase-separated compartments. However, the extremely low levels of $(PR)_{50}$ and $(GR)_{50}$ expression in our *C. elegans* model may not be sufficient to drive DPR localization to these other membraneless compartments. Localization to phase-separated compartments, such as the nuclear pore, are sometimes inferred from a DPR-induced impairment in nucleocytoplasmic transport or stress granule formation/disaggregation [195, 196]. We did not directly test either of these cellular processes. However, multiple genes involved in nucleocytoplasmic transport are suppressors of PR toxicity in our *C. elegans* model. Therefore, it is likely that PR disrupts nuclear pore function in our model as has been observed in other systems.

Toxic GR and PR DPRs are unique among proteins associated with ALS and other neurodegenerative diseases in that GR/PR do not primarily form protein aggregates. However, they do concentrate within phase separated cellular compartments. Other aggregating mutant proteins associated with different neurodegenerative diseases also impair phase separation, such as tau in Alzheimer's [193], mHTT in HD [194], and FUS [192] and TDP-43 [259] in ALS and FTD. These proteins were initially thought to be toxic

due to the pathological aggregates they formed in their respective neurodegenerative disease. However, aggregation of misfolded proteins can be separated from toxicity of the proteins, suggesting protein aggregates may not be the main cause of toxicity [131]. Instead, low molecular weight homopolymers, known as oligomers [231], are now thought to be a driving cause of toxicity [260]. One way oligomers may be toxic is by disrupting normal phase separation properties [188, 259]. Such a mechanism could link *C9orf72*-associated ALS/FTD to other neurodegenerative conditions typically categorized as protein aggregation diseases.

2.3.2 Repeat-Associated Toxicity of PR

Repeat length is strongly correlated with toxicity in several repeat expansion diseases, such as the polyQ diseases [11]. There have been conflicting reports as to whether the length of the G_4C_2 repeat in *C9orf72* affects the age of onset of the disease [26, 29, 261, 262]. Patients with a family history of ALS or FTD have been identified with a *C9orf72* repeat expansion of only 30 G_4C_2 repeats [9]. However, the diseases may not be solely caused by the *C9orf72* repeat expansion. Other ALS-associated mutations sometimes occur with shorter repeat lengths of G_4C_2 and may act synergistically with the G_4C_2 repeats to enhance their toxicity, thus, masking decreases in toxicity due to a lower repeat length [28]. Another confounding variable is that blood samples are used to determine the number of G_4C_2 repeats in patients. However, the number of G_4C_2 repeats in a blood sample may not match the number of G_4C_2 repeats in the neurons of patients. This is because the repeats can undergo dramatic somatic expansion in the brain [30]. However, some studies suggest this not likely to be common [263]. Somatic expansions

of repeats are common in other diseases [30]. However, the frequency and extent of both somatic and germline expansions appears to be more drastic and variable in *C9orf72* [264, 265].

As our *C. elegans* models contain the dipeptide repeat but not the nucleotide repeat, we solely studied toxicity of repeat length due to the dipeptides. In patients, the length of dipeptides may not be reflective of the length of the G_4C_2 repeat expansion. Since RAN translation does not use a canonical start codon, it is unclear where RAN translation initiates in patient neurons. It is also unknown where RAN translation terminates, as repetitive RNA can cause stalling of translation machinery [266]. Therefore, the entire G_4C_2 repeat may not be translated. Understanding whether dipeptides cause length-dependent toxicity aids in interpreting various disease models as well as determining the pathogenic mechanisms of the dipeptides. The arginine-rich dipeptides exhibit increased toxicity with increased repeat length [94, 183]. A minimum repeat length for GR has been observed in multiple systems [94, 183, 240], but a repeat length where PR is nontoxic had not been identified before our work.

We found that PR is toxic at 25 repeats and 50 repeats, but PR was not toxic at 5 repeats and 15 repeats. At all repeat lengths, the PR protein localized to the nucleolus. The decrease in toxicity was not due to an increase in the amount of PR dipeptide, as the GFP signal was significantly brighter in animals with PR_5 than animals with PR_{50}, suggesting more PR_5-GFP dipeptides were present in the nucleus. While there are more dipeptides, it would take five times more PR_5 dipeptides to have the same molar amount of PR repeats in the nucleus as PR_{25}, so the total molar concentration of PR repeats in the nucleus could determine toxicity. In addition to concentration, the repeat length may

also directly contribute to toxicity. For example, a minimum length of PR repeats might be required for PR to bind to a protein and disrupt its function. Determining if there are structural or organizational differences between nucleoli with PR_5 and nucleoli with PR_{50} could give clues as to how PR_{50} is causing cellular toxicity. In other systems, PR impairs nucleolar structure and overall translation [195, 196, 267]. It would be informative to determine if the minimum PR repeat length required for nucleolar structural impairment and translational inhibition is similar to the minimum PR repeat length required for toxicity.

2.3.3 Age-Associated Toxicity of PR and GR

Another common theme in neurodegenerative diseases is age-associated toxicity. Disease-associated repeat expansions are present from birth; but patients do not develop neurodegeneration until middle to late age. This delay could be due to pathogenesis associated with biological aging. For example, the repeat-derived DPRs may only be toxic when cells are senescent and protein homeostasis pathways are functioning at reduced levels. Another possibility is that the age-dependent toxicity is due to chronological aging, and repeat products slowly accumulate over time until they reach a toxic threshold. Yet another possibility is that RAN translation and subsequent DPR production does not occur until late age. RAN translation is increased by activation of the integrated stress response [86, 89, 90], which is upregulated with age [268]. However, the age requirements for RAN translation have not been studied and will require the production of new tools and reagents that are beyond the scope of my research. Therefore, I focused on determining if DPR toxicity was due to biological or chronological aging.

C. elegans are uniquely situated for studying age-dependent DPR toxicity, as aging has been extensively studied in *C. elegans*. Mutations in the *C. elegans daf-2*/insulin pathway double the lifespan and extend the healthspan of the worms [110, 112]. One of the major effects of activating the *daf-2*/insulin signaling pathway is to delay the proteotoxic consequences of aging, thus making mutants in the *daf-2*/insulin pathway ideal for separating physiological aging from chronological aging [112, 269]. The *daf-2*/insulin pathway has been used to distinguish between toxicity due to chronological aging and toxicity due to biological aging in other worm models of genetic forms of ALS, such as those associated with mutant forms of proteins such as SOD1 [130, 245, 257, 270-272], TDP-43 [143, 273], and FUS [143]. In each of these cases, the worm models express a mutant protein that forms protein aggregates. The toxicity associated with the SOD1 and TDP-43 models is suppressed by delaying biological aging through activation of the *daf-2*/insulin signaling pathway [273, 274]. One way the *daf-2*/insulin signaling pathway is thought to oppose the toxicity of these disease models is via the upregulation of genes that maintain protein folding (i.e. chaperones) or enhance the clearance of misfolded/aggregated proteins (i.e. proteases) [248].

Like SOD1 and TDP-43, we found that $(PR)_{50}$ toxicity can also be suppressed by activating insulin signaling. However, unlike SOD1 and TDP-43, $(PR)_{50}$ did not form detectable protein aggregates. Given that mutations in genes comprising the *daf-2*/insulin signaling pathway can modify all three disease models, this suggests that $(PR)_{50}$ may share common toxic mechanisms with SOD1 and TDP-43, for example, via disruptions in protein homeostasis. Interestingly, insulin signaling did not appear to alter PR abundance

or subcellular localization in our model, suggesting that this pathway may act downstream from PR nuclear functions to provide toxicity protection.

In contrast to $(PR)_{50}$, we found that activation of the *daf-2*/insulin signaling pathway did not delay the toxicity of $(GR)_{50}$. While GR and PR toxicity are usually thought to occur through the same mechanisms, our results are among the first to suggest that there may be some important differences. Such differences could be due to *daf-2* induced changes in DPR expression levels, such as a reduction in $(PR)_{50}$ levels but not $(GR)_{50}$ levels. However, *in vivo* imaging suggests that DPR levels were not significantly altered in the *daf-2* background. Therefore, the precise nature of such differences remains unknown. To date, with the exception of *daf-2*, all of the genes we have identified whose knockdown suppresses $(PR)_{50}$ toxicity also suppress $(GR)_{50}$ toxicity. In the future, genetic screens utilizing animals expressing $(GR)_{50}$ may allow us to identify GR suppressors and define DPR-specific toxicity mechanisms.

2.3.4 Suppressors of PR Toxicity

Through a genome-wide RNAi screen, we identified several pathways required for PR toxicity. These pathways included nucleocytoplasmic transport, mRNA processing, chromatin regulation, and proteasomal protein degradation (Table 2.1). Some of these pathways have been previously linked to PR and GR toxicity by screens in other systems [93, 98-100, 243]. Components of nucleocytoplasmic transport have been identified as modifiers of PR and GR toxicity in yeast, flies, human cells, and now *C. elegans*. Notably, *hnRNPC*/*ztf-4*, *mGIGYF2*/*C18H8.3*, *KPNA3*/*ima-3*, and *SRSF3*/*rsp-6* were identified as modifiers of PR toxicity in our screen as well as in other systems [98, 99, 195, 243]. The

fact that in our blinded genome-wide screen, we discovered modifiers of PR toxicity also identified in other screens reflects both the conserved nature of PR toxicity and the power of the *C. elegans* system to identify conserved modifiers.

We identified the nucleocytoplasmic trafficking pathway as being important for PR toxicity. GR and PR induced nucleocytoplasmic trafficking defects have been reported in multiple systems [99, 242]. Given that the phase-separation properties of nuclear pores allows them selective permeability [189], it is logical dipeptides that localize to phase-separated structures may change the chemical properties of these structures and disrupt nuclear transport [195]. Defects in nucleocytoplasmic transport also occur in other neurodegenerative diseases caused by phase-separating proteins, such as Alzheimer's Disease [275] and HD [214]. Therefore, disrupted nucleocytoplasmic transport could be a general trait of neurodegenerative diseases. However, disrupted nucleocytoplasmic traffic is unlikely to be the only way that PR and GR cause toxicity, as proteins not involved in nucleocytoplasmic transport have also been identified as involved in PR and GR toxicity.

Other cellular pathways required for PR toxicity in our screen and several other screens are chromatin regulation and mRNA processing [98-100, 195, 243]. mRNA processing has been previously implicated in ALS as mutations in the mRNA transport proteins TDP-43 and FUS can cause ALS [276-278]. Aggregates of TDP-43 and FUS are also prominent pathological features of ALS [162, 279]. In general, RNA binding proteins (RBPs) have a higher likelihood of containing disordered regions, which are required for phase separating into stress granules or nuclear speckles [280, 281]. It is currently unclear if either disruption of RBP phase separation or disruption of the role of

RBPs in mRNA processing is a stronger contributor to ALS [282]. Epigenetic modifications such as altered chromatin methylations also occurs in types of ALS not caused by the *C9orf72* repeat expansion [283, 284]. Overall, the effect of chromatin regulation in ALS has received limited attention. However, our identification of chromatin regulators as being critical components of PR toxicity suggests that future studies in this area are warranted.

The fourth pathway that suppressed PR toxicity in our screen was the proteasomal degradation pathway. Accumulation of misfolded proteins is a common theme associated with several neurodegenerative diseases [46, 285]. Previous screens in yeast and human cells have also identified the proteasomal degradation pathway as a suppressor of PR toxicity [98, 100] . Of note, genes that code for two regulatory non-ATPase subunits of the 26S proteasome [100] were identified in our screen, *PSMD8*/*rpn-12* and *PSMD13*/*rpn-9*. While there are 11 additional proteasome regulatory genes and 14 additional core proteasome catalytic subunit genes in *C. elegans*, our screen did not identify any others as being required for PR toxicity. Deletion of *PSMD8* also suppressed PR toxicity in an unbiased CRISPR screen performed in human cells [100]. PR can bind to the proteasome and inhibit proteasomal degradation [286]. Therefore, PR may interact with the proteasome via interaction with RPN9 and/or RPN12. Loss of these proteins may eliminate the ability of PR to interact with the proteasome, likely in the nucleus. Along these lines, we also identified a nuclear E3 ligase adaptor (SPOP/BATH-43), which had not been previously found in a suppressor screen of PR. Interestingly, an interactor of SPOP, RBX1, is a modifier of PR toxicity [100]. SPOP is a substrate adaptor for the CUL3-RING ubiquitin ligase and undergoes phase separation when it interacts with its

substrates [287], which may facilitate SPOP and PR interactions. The fourth gene we found that is involved in proteasomal degradation is *ufd-2*/UBE4B, an E3/E4 ubiquitin ligase [288]. *ufd-2*/UBE4B was one of the strongest suppressors identified in the screen and, like *bath-43*/SPOP, localizes to the nucleus, suggesting that the E3 ligase adaptor SPOP and the E3 ligase UBE4B may interact physically and/or functionally. Interestingly, SPOP is a well-established genetic cause for many prostate, endometrial, and renal cancers [289-291]. Recently, a small molecule inhibitor of SPOP was reported that effectively blocks the growth of SPOP mutant renal cancer cells [292]. In the future, it may be interesting to determine if this drug can also be utilized to inhibit SPOP in a neuronal context and protect against PR toxicity.

2.3.5 Study Limitations

Our model had three main limitations. First, like most other DPR models, our *C. elegans* model utilizes ectopic expression of DPR at levels that are unlikely to mimic physiological DPR levels. However, DPR expression levels in patients have only been determined in postmortem samples, where they are rarely observed in motor neurons and when they are observed, they are present at extremely low levels [179, 180]. However, the rare nature of PR/GR positive motor neurons in postmortem tissue could be due to the death of cells containing the arginine-rich dipeptides. This theory was supported by a recent rodent model where a high concentration of cells contained PR at the onset of disease, but very few cells contained PR near the end of the disease progression and many of the cells in the previously PR-rich region had died [10]. Overcoming this limitation is challenging since low level PR expression does not produce detectable phenotypes in

C. elegans (data not shown). Even in patients, physiological expression levels of DPRs take 30-40 years to cause toxicity. Therefore, overexpression may be a necessary compromise in model systems, including mammalian cell lines, in order to facilitate phenotypes required for discovery of disease mechanisms.

A second limitation of our model is that each dipeptide was expressed individually, whereas dipeptides can be co-expressed in patient cells [34]. A study that co-expressed different dipeptides with GA found that GR or PR could sequester GA into the nucleus [96]. Therefore, co-expressed dipeptides are likely to have reciprocal effects on localization and toxicity. The effects of the G_4C_2 RNA and reduction of the C9ORF72 protein, both of which occur in *C9orf72* repeat expansion patients [8], were also not addressed in our model, both of which could strongly affect the cellular environment and the toxicity of the different dipeptides. C9ORF72 is involved in the autophagosome-lysosome pathway and decreased levels of C9ORF72 cause impaired proteostasis and increase the toxicity of the RAN dipeptides [293]. In the future, examining the toxicity of DPRs in animals carrying a loss-of-function mutation in the *C. elegans C9orf72* homolog *alfa-1* [294] or in the presence of a G_4C_2 transgene [91, 295] may help to address these limitations.

2.3.6 Future Directions

The limitations of our model, in combination with our findings so far, provide a roadmap for future research. There are three future projects that would add to our understanding of *C9orf72* DPR toxicity. The first project is to co-express the dipeptides, most importantly PR and GR, to determine if the dipeptides act synergistically. The only

previous co-expression model co-expressed GA with the other dipeptides [96]. As GA is not toxic in most models, we would instead focus on the commonly toxic dipeptides, PR and GR. We could examine how co-expression of different dipeptides affects cellular localization and the relative toxicity of the dipeptides. Such studies may reveal synergistic interactions between DPRs, which may engage completely distinct toxicity pathways than those identified with our individual PR toxicity model.

The second project would be creating a more accurate *C. elegans* model of C9orf72 toxicity that incorporates DPR toxicity, RNA toxicity and loss of *C9orf72* function. Such a model may allow us to determine the significance of DPR toxicity in the presence of these additional factors. To create such a model, we could insert a G_4C_2 repeat expansion in the first intron of the *C. elegans C9orf72* homolog *alfa-1* to mimic the G_4C_2 repeat expansion in patients. In patients, the G_4C_2 repeat expansion causes a reduction of C9ORF72 protein levels because the repeat reduces *C9orf72* mRNA levels. Thus, a G_4C_2 repeat expansion in the intron of *alfa-1* may cause a reduction of ALFA-1 protein levels, making it a model for loss of C9ORF72 as well as a model for RNA toxicity. Our lab has previously generated a *C. elegans* $(G_4C_2)_{120}$ overexpression model, which undergoes RAN translation (data not shown). However, the repeat is not in the context of either *C9orf72* or *alfa-1*. Placing this repeat within *alfa-1* could create a more holistic perspective to understand the complicated relationship between loss of C9ORF72 function, accumulation of repetitive RNA, and DPRs.

The third possible project is determining how PR impacts proteasome toxicity, which is currently being pursued by other members of the lab. The ubiquitin-proteasome degradation pathway was previously implicated in ALS as knockdown of proteasomes in

murine motor neurons causes symptoms and pathology similar to ALS [49]. Several ALS-causative mutations cause defects in the ubiquitin-proteasome pathway [296, 297]. In our model, RNAi knockdown of either *bath-43*/SPOP or *ufd-2*/UBE4B suppressed PR toxicity in *C. elegans*. Whether knockdown of *bath-43*/SPOP or *ufd-2*/UBE4B also suppresses PR toxicity in human neurons needs to be determined. Given that SPOP is an E3 ligase adaptor and UBE4B is an E3 ligase, on hypothesis is that SPOP and UBE4B interact to mediate PR toxicity. Consistent with this hypothesis, both SPOP and UBE4B are localized to the nucleus [298, 299]. However, SPOP typically acts as an adaptor for cullin-type E3 ligases, and UBE4B is not a cullin-type E3 ligase. This suggests either that SPOP and UBE4B exhibit a novel functional interaction or that other E3 ligases may mediate the role of SPOP in mediating PR toxicity. Whatever the case, the observation that loss of ubiquitin proteasome components prevents PR toxicity is unusual and suggests that the stabilization of one or more pro-survival substrate may mediate protection against PR toxicity. Interestingly, several known SPOP substrates are chromatin regulators, which were also identified in our genetic screen. Testing whether chromatin regulators or other proteins are relevant targets of SPOP and/or UBE4B will be a major goal for future work.

2.3.7 Summary

Our *C. elegans* model of C9orf72-derived RAN dipeptides replicated previous findings in other systems and identified novel requirements for PR and GR toxicity. We established that while PR and GR localize to the nucleus, PR and GR do not entirely share their mechanisms of toxicity. PR toxicity requires an aging cellular environment,

while GR toxicity depends on the chronological age of the cell. An RNAi suppressor screen found twelve conserved suppressors of PR toxicity. Four of the twelve genes had been previously identified as interactors or modifiers of PR and GR toxicity in other systems. All twelve of the PR suppressor genes have human homologues, suggesting that PR toxicity engages highly conserved pathways. Therefore, our *C9orf72* RAN *C. elegans* model established that RAN products can be characterized in *C. elegans* and that *C. elegans* models of RAN products can be used to screen for conserved suppressors of RAN product toxicity.

2.4 Methods

Caenorhabditis elegans strains and culture

Strains were cultured on standard NGM media with *E. coli* OP50 bacteria. Strains expressing $(PR)_{50}$ and $(GR)_{50}$ DPRs were cultured on *gfp*(RNAi) bacteria at 20°C until the experiment, when they were shifted to *E. coli* OP50. The following strains were used; *daf-2(e1370)*, *daf-16(mu86)*, *hsf-1(sy441)*. Standard genetic approaches were utilized to cross mutants into the DPR backgrounds. The homozygous genotype of every strain was confirmed by DNA sequencing of the mutant lesion. Wild-type animals were reisolated in every cross and utilized as the DPR only control in resulting experiments. DPR toxicity studies were carried out using animals grown at 25°C, unless otherwise noted.

Molecular Biology and Transgenics

Codon-varied dipeptide sequences were isolated from previously described plasmids [94] (nucleotide sequences can be found in the Appendix D-Table 4.2). $(PR)_5$, $(PR)_{15}$ and $(PR)_{25}$ codon-varied plasmids were synthesized (GeneArt, ThermoFisher Scientific, Waltham, MA, USA). Dipeptide sequences were isolated as a HindIII/BamHI fragment and subcloned into the *C. elegans* expression vector pPD95.79 to generate dipeptide with a C-terminal GFP tag. Promoters were PCR amplified, incorporating both a 3XFLAG epitope immediately downstream from the start ATG and HindIII sites flanking the fragment. Promoters were subcloned into the DPR-pPD95.79 vectors as HindIII fragments. To make N-terminal GFP–$(PR)_{50}$ and GFP–$(GR)_{50}$ fusions, GFP was PCR amplified and cloned in frame with the start codon of the *myo-3* promoter. The dipeptide sequences were subcloned as a HindIII/BamHI fragment into the pPD49.26 expression vector. The *myo-3*-GFP fragment was then subcloned into the DPR-pPD49.26 as a HindIII fragment.

The membrane-bound DPRs were generated by a PCR fusion method. We first amplified the signal sequence and transmembrane domain from the *pat-3* gene as supplied in vector pPD122.39. The membrane domain was then fused to the *myo-3* promoter in-frame with the start ATG and then subcloned as a HindIII fragment into the DPR-pPD95.79 vectors. The resulting clones produce a membrane-localized DPR with the DPR facing the cytoplasm.

To make the *his-58*–DPR fusions, the *his-58* genomic sequence was PCR amplified and Gibson cloned (New England Biolabs, Ipswich, MA, USA) into the *myo-3p*::GFP sequence downstream of and in-frame with the GFP sequence. The resulting

myo-3p::GFP-his-58 fragment was subcloned as a HindIII fragment into the DPR-pPD49.26 vectors. A list of nucleotide sequences for promoters used can be found in Appendix D-Table 4.4.

Transgenic worms were generated by injecting the DPR construct (20ng/ml) and the *myo-3p*::mCherry pCFJ104 or *myo-3p*::dsRed2 marker plasmid (100ng/ml) into the gonad of wild-type animals. Transgenes were integrated using a standard gamma ray (Cs137) mutagenesis, followed by selection of animals exhibiting 100% transmission of the mCherry marker. Integrated strains were outcrossed six times to wild-type animals. In the case of $(PR)_{50}$ and $(GR)_{50}$, injected animals were maintained on *gfp*(RNAi) plates until the experimental assay was performed. All procedures involving recombinant or synthetic nucleic acid molecules and materials were approved by the University of Pittsburgh Institutional Biosafety Committee.

Thrashing Assay

To measure thrashing, animals were maintained at 25°C on *E. coli* OP50, and transgenic animals for each strain were picked as L4s the day before the experiment. The following day, worms were placed on clean NGM plates and allowed to move freely for 10 min so that most of the bacteria came off the animal before the experiment. Worms were then placed individually into 3 cm-petri dishes containing M9 buffer and allowed to adjust to the new environment for 5 min. The worms were then observed for 30 s and the number of thrashes (reversal of body bend that crosses the midline) was counted. We were unable to score thrashing in *daf-2(e1370)* animals as *daf-2* animals have severely impaired thrashing.

Microscopy

Worms were anesthetized (10mM levamisole) and mounted on agar pads for fluorescence microscopy. Images were collected on either a Leica MZ16FA stereo dissecting scope or a Leica DMI4000 inverted microscope and a Leica DFC 340Fx digital camera (Leica Microsystems, Wetzlar, Germany). Z-stack images were deconvolved using Leica AF6000 software. Unless noted, images within an experiment were collected using the same exposure settings and processed with identical deconvolution parameters.

Commissure Assay

L4 animals of the indicated genotype were isolated at 25°C and imaged 24 hours later as 'Day 1 adults'. All of the strains contained an *unc-47p*::GFP marker to reveal GABA motor neuron morphology. Animals were anesthetized in 10 mM levamisole and Z-series images of GABAergic commissures were collected. Commissure breaks were identified as interruptions in the GFP signal surrounded by dorsal and ventral GFP in the commissures. Blebbing was scored only in the commissures and was identified by the presence of one or more GFP varicosities. For assays involving *daf-2*, we were unable to score either membrane blebbing or commissure breakage in $(GR)_{50}$ expressing animals because the neuronal GFP marker used to score such events was undetectable, likely due to significant neurodegeneration that continued to occur in $(GR)_{50}$; *daf-2* animals.

Brood Size Assay

For brood size assays, at least 10 L4 stage animals were individually picked to *E. coli* OP50-containing NGM plates at 25°C and transferred to a new plate every 24 hour until cessation of egg laying. Each plate was allowed to age for 48–72 hour and the number of animals ≥L4 stage of development was counted. Brood sizes were normalized within each strain to the mean brood size of animals grown on *gfp*(RNAi) bacteria.

Paralysis Assay

For all assays except for that shown in Figure 2.7, gravid DPR expressing transgenic animals were moved from *gfp*(RNAi) to *E. coli* OP50 and allowed to lay eggs for 24 hours. The resulting progeny were allowed to grow up on *E. coli* OP50, permitting DPR accumulation. Ten L4 animals were placed on each of three to five plates (N=30–50 per assay). Each day, animals that failed to move at least half a body length in response to manual stimulation with a platinum wire but were still alive (pharyngeal pumping, movement of less than half a body length) were scored as paralyzed. Animals that died, desiccated, or exhibited internal hatching of progeny were censored from the assay. Each day, mobile animals were transferred to a new plate and paralyzed, dead, and censored animals were removed from the assay.

For the assay in Figure 2.7 D and E, HIS-58-$(PR)_{50}$ progeny from animals moved from *gfp*(RNAi) to *E. coli* OP50 were not viable, presumably owing to the enhanced toxicity of the nuclear localized $(PR)_{50}$ protein. Therefore, for these assays, animals were removed from *gfp*(RNAi) as L4 stage animals, placed on *E. coli* OP50 at 25°C, and motility in the animals moved from *gfp*(RNA) was monitored as described above. Since animals

are removed from the assay once they are scored as paralyzed, it is not appropriate to utilize statistical approaches that compare means and errors between time points (i.e. T-tests, ANOVA tests). Instead, we utilized the Log-rank statistical method, a cumulative statistical approach that compares changes in population sizes over time for a specified endpoint (53). Because this end-point assay is cumulative and not replicative, data points do not contain error bars. For each assay, 2–3 independent trials with 45–50 animals per assay were performed and the results from one representative trial are shown.

Lifespan Assays

For lifespan assays, worms grow at 20°C were picked as L4 and allowed to grow at 20°C until the next day, when 10 young adults were placed on five 3 cm plates. Lifespan assays were performed with *E. coli* OP50 spotted on NGM plates at 25°C. Worms were classified as alive, dead (no movement in response to touch with a wire), or censored (lost or bagged worms) once a day for lifespan assays.

COPAS Fluorescence Quantification

Day 1 adult animals expressing each DPR-GFP transgene under the control of the *myo-3* promoter were washed off plates and analyzed using a COPAS Biosorter. Fluorescent detection settings were identical for all samples. Only animals with time-of-flight measurements from 400-600 (young adults) were used for the analysis of fluorescence.

Quantitative PCR

L1 animals from each of the indicated genotypes were synchronized using the hypochlorite method. Animals were plated on *E. coli* OP50 and allowed to grow at 20°C for ~48 hours. 500 young adult animals were sorted into 1.5 ml tubes using a COPAS BIOSORT. 3 biological replicates were performed for each genotype. Purified total RNA was converted to cDNA and quantitative PCR against GFP was carried out on an ABI7500 qPCR system (Applied Biosystems, Fisher Scientific, Foster City, CA) using the SYBR green method. Samples were normalized against expression for the actin gene *act-2*. Each sample was then normalized against the expression of the GFP only control.

FRAP

FRAP studies were carried out on levamisole anesthetized day 1 adult hermaphrodites using a Leica DMI8 confocal microscope at the University of Pittsburgh Center for Biologic Imaging. Images were captured at 20–30 Hz. Following imaging of baseline fluorescence, a region of interest corresponding to a portion of the nuclear or cytoplasmic foci was photobleached and fluorescence recovery within the photobleached area was monitored over at least 60 seconds. Data were normalized so that the image preceding the photobleach was set to 100% and the first image following the photobleach was set to zero percent.

ENU Forward Mutagenesis Screen

C. elegans L4s were placed in 2.4mM ENU and then gently shaken for 4 hours at room temperature. The worms are then washed 5x with M9 and allowed to recover for

24 hours on NGM plates. After 24 hours, gravid P0s were picked and allowed to lay progeny on (*gfp*)RNAi plates for 6 hours. The P0s were then removed and the plates with the eggs were placed at 20°C. After a week, the starved progeny were washed onto *E. coli* OP50 and placed at 25°C. After a week, the worms still growing at 25°C were picked as possible candidates. To ensure the independence of alleles, only one candidate was picked per plate.

COPAS Sorting of Suppressors

Bleach synchronized adult day 1 adult animals were sorted through the COPAS Biosorter (Union Biometrica, Holliston, MA, USA). The overall RFP signal of the worm and the worm time of flight of were measured. Relative RFP signal was measured as (Overall RFP signal / Time of flight) and used to test for transgene suppression.

RNAi Screen

Genome-wide RNAi screening was performed using a commercially available RNAi feeding library (MRC Geneservice, Cambridge, U.K.). Single colonies were inoculated into 1 ml of LB media containing 25µg/ml carbenicillin and grown overnight with shaking at 37°C. 10 microliters of each culture were spotted on each well of a 24-well plate containing NGM with 20mM NaCl, 1mM IPTG, and 25µg/ml carbenicillin. 15-20 eggs were seeded in each well and the plates were incubated for a week at 20°C. They were then scored on their ability to move and have progeny. The clones were rescreened 2 times and those that are picked up 4/6 times were counted as suppressors of $(PR)_{50}$ toxicity.

Statistical Analysis

Paralysis assays were analyzed using the Kaplan–Meier log-rank function (OASIS) [300, 301]. Comparisons of means were analyzed with either a two-tailed Student's t-test (two groups) or ANOVA (three or more groups) using the Tukey's or Dunn's post-test analysis as indicated in GraphPad Prism7 (GraphPad Software, Inc., La Jolla, CA, USA). P-values of<0.05 were considered significant.

3.0 HD RAN Polypeptides Exhibit PolyQ-Independent Toxicity

3.1 Introduction

Repeat-Associated non-AUG (RAN) translation occurs in four different disease caused by a CAG/CTG repeat expansion: Huntington's disease (HD) [12], spinocerebellar ataxia type 8 (SCA8) [4], Huntington disease-like 2 (HDL2) [4], and Fuchs endothelial corneal dystrophy (FECD) [85]. All four CAG repeat expansions encode identical homopolymeric RAN peptides, although the flanking sequences differ between genes. There are 10 other neurodegenerative diseases caused by a CAG or CTG repeat expansions that are likely to undergo RAN translation, although experimental evidence is currently lacking. Altogether, fourteen distinct diseases could produce RAN translation products from a CAG/CTG repeat. Determining which of the CAG/CTG RAN polypeptides are pathogenic and defining their mechanism(s) of toxicity could expand our understanding of CAG/CTG repeat expansion diseases.

HD is the most common CAG/CTG repeat expansion disease [302, 303]. Patients suffer from this incurable neurodegenerative disease for decades before passing due to secondary complications. HD patients commonly exhibit psychological, cognitive, and motor problems such as depression, rigid thinking, and chorea, respectively [202]. The CAG repeat expansion in the Huntingtin gene, *HTT*, causes all cases of HD [205]. Patients with a CAG repeat expansion ≥40 repeats in *HTT* (also referred to as mutant *HTT* (*mHTT*) in HD patients) begin to exhibit HD-associated motor defects at ~40 years of age [55, 203]. However, the age of onset of HD is inversely correlated with the length

of the CAG repeat [205]. In other words, patients with long CAG repeats develop the disease earlier in life than patients with shorter CAG repeats. Large increases in the length of the CAG repeat arise through impaired replication of the G/C-rich repeats during germline division [304]. Impaired repeat replication leads to intergenerational increases in CAG repeat lengths, or genetic anticipation, which causes an earlier age-of-onset of HD in the offspring relative to the parent [304]. Despite the clear genetic basis for HD, the molecular mechanisms that drive disease severity and outcomes are still poorly understood. Currently, HD and other CAG repeat expansion disease are incurable.

HD is an autosomal dominant monogenic disease. Therefore, most patients have a family history of HD [305] and are aware at an early age that they are at risk of developing HD [306]. These patients can undergo pre-symptomatic genetic testing for HD, but they are not diagnosed as having HD until they develop motor defects [307]. The average post-diagnosis lifespan for HD patients is ~20 years [203]. Prior to diagnosis, HD patients have an increased risk of psychological problems up to ~15 years before motor symptoms appear [207, 208]. Patients typically discover their risk for HD as juveniles, when their parents develop HD. The patients can confirm their HD status at age 18, when they can undergo genetic testing. Patients can display cognitive [208] and psychiatric [207] defects due to HD up to 15 years before diagnosis (~30 years of age). Motility defects normally begin ~40 years of age [203], with patients living 10-20 years after motor problems arise. Since there are no corrective treatments for HD, this disease has a lifelong impact on patients.

Therapeutic strategies for treating HD have focused on the polyglutamine (polyQ) tract within the first exon of the *HTT* gene, which is encoded by the CAG repeat

expansion. The ability of polyQ to aggregate correlates with the length of the polyQ polypeptide. Levels of polyQ aggregation also increase with age [75]. Because of these characteristics, the polyQ aggregates were initially thought to be the cause of toxicity. However, the ability of polyQ to aggregate can be separated from polyQ toxicity [131]. Current studies suggest that polyQ oligomers (low abundance, low molecular weight homopolymers [231]) are stronger contributors to HD then polyQ aggregates (abundant high molecular weight polymers [231]) [308]. However, brain regions such as the caudate white matter lack detectable polyQ aggregates but still undergo apoptosis in HD [12]. In addition, juvenile HD patients have ≥60 CAG repeats and minimal detectable polyQ aggregates in their brains, even though there is widespread loss of neurons [12]. These data clearly suggest that polyQ-independent disease mechanisms also contribute to HD pathology.

One potential polyQ-independent disease mechanism is the production of homopolymeric RAN peptides, which was discovered to occur in HD in a landmark 2015 study [12]. In both adult and juvenile patients, HD RAN products are present in brain regions lacking polyQ that are still undergoing apoptosis. Interestingly, some of these polyQ deficient brain regions have high levels of other RAN peptides [12]. Thus, RAN products appear to contribute to HD in both adults and juvenile patients. HD CAG repeat expansions produce four RAN products: polySerine (polySer), polyLeucine (polyLeu), polyAlanine (polyAla), and polyCysteine (polyCys). Canonical translation produces polyQ. All four RAN peptides and polyQ are weakly toxic at 90 repeats when overexpressed in neuronal cells [12]. Whether or not these peptides are also toxic at more disease-relevant lengths found in the majority of HD patients is not known. Both polySer

and polyGln form protein aggregates in cell culture [12]. Protein aggregates sequester cellular protein folding and degradation resources [310]. Therefore, aggregation of RAN peptides other than polyQ could disrupt protein folding molecular pathways required for HD models. Whether protein folding pathways or other molecular pathways influence the toxicity of RAN peptides is not known.

To address these questions, I created codon-varied GFP-tagged HD RAN homopolymers at both disease-relevant lengths (38 repeats) and highly expanded lengths (90 repeats). I expressed these peptides in multiple cellular settings in *C. elegans.* I found that all RAN peptides, except for polyLeu, formed highly compact aggregates. The only RAN peptide that was toxic in all cellular settings was polyLeu. I found that polyLeu displayed length-associated toxicity and caused significant neurodegeneration *in vivo*. Notably, neurodegeneration was not observed with codon-varied polyGln, suggesting that these two polypeptides cause toxicity through distinct pathways. To better understand the pathways of polyLeu toxicity, I performed an unbiased forward genetic ENU mutagenesis and identified suppressors that ameliorate the toxicity of polyLeu. Whole-genome resequencing and RNAi phenocopy identified several candidate genes required for polyLeu toxicity. The molecular nature of these genes suggests that polyLeu causes toxicity through mechanisms involving the secretory pathway. Identifying the cellular processes involved in polyLeu toxicity gives direction to future research for treatments for HD patients, as well as patients suffering from other diseases caused by CAG repeat expansions.

3.2 Results

3.2.1 HD RAN Model

The CAG repeat expansion in HD could be toxic through the G/C-rich RNA, the polyGln peptide produced through canonical translation, or through one or more of the homopolymeric peptides produced through RAN translation. To focus on the individual toxicity of the RAN peptides, I created *C. elegans* models for each of the five possible CAG RAN products. To individually model the RAN peptides, a previously described codon-variation strategy [95] was employed for each homopolymeric peptide to maintain the amino acid repeats but remove the nucleotide repeats. The codon-varied constructs were also designed to minimize computationally predicted RNA hairpin structures in the resulting RNA, as hairpin structures are thought to be required for RAN translation and could lead to the production of non-relevant RAN products [311]. The individual RAN peptides were instead expressed through canonical AUG-initiated translation and studied at either 38 repeats or 90 repeats (Figure 3.1 A). The 38 repeat peptides are similar to the minimum length of 40 CAG repeats that ensures a patient will develop HD. The 90 repeat peptides mimic the length investigated in a previous paper [12]. The homopolymeric peptides were also tagged at the C-terminus with GFP for *in vivo* visualization of the RAN peptides (Figure 3.1 A). The modeled RAN peptides lacked genetic context (i.e. no flanking human *HTT* sequence) as each RAN peptide is in a different reading frame leading to a unique genetic context for each RAN peptide. Including the genetic contexts of the RAN peptides would limit our ability to assign phenotypes to specific homopolymeric peptides rather than flanking sequences. To

compare the characteristics of the individual RAN peptides to the characteristics of the *mHTT* CAG repeat expansion, I created an expanded CAG repeat with polyGlutamine in-frame with the start codon. Throughout my thesis, I will refer to the CAG-encoded polyGlutamine as “polyQ” and the codon-varied polyGlutamine as “polyGln”. Unlike the codon-varied polyGln, the CAG-encoded polyQ may exhibit either RNA toxicity, toxicity from RAN products besides polyglutamine, or both. To study the functional consequences, I expressed each codon-varied construct in both GABAergic neurons and muscle cells using cell type specific promoters.

3.2.2 GABAergic Neuron Model of HD RAN Polypeptides

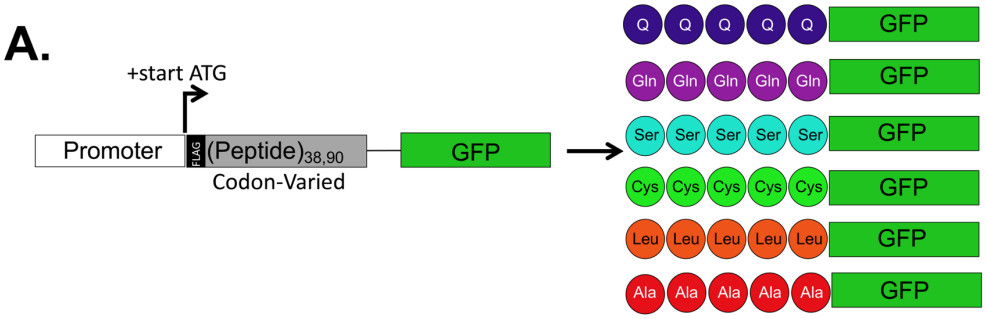
Neurodegeneration due to HD starts in the striatum and most strongly degrades the GABAergic neuronal population in the striatum [312]. HD RAN polypeptides are found in degenerating regions of the striatum that lack polyQ, suggesting they contribute to neurodegeneration [12]. We expressed RAN polypeptides specifically in the GABAergic neurons of *C. elegans* using the *unc-47* promoter, which controls the expression of the vesicular GABA transporter. There are 26 GABAergic neurons in *C. elegans* and 19 of the 26 GABAergic neurons are motor neurons [146]. GABAergic neuron somas are located along the ventral nerve cord and extend single axons, known as commissures, across the dorsal-ventral midline to the dorsal muscle cells. Commissures are commonly used to measure neurodegeneration since they provide single axon resolution [247, 313]. The function of GABAergic neurons can also be assessed using a reversal assay, since loss of GABAergic neurons causes a significant deficit in the ability of *C. elegans* to reverse direction [146]. Quantifying the number, morphology, and function of GABAergic

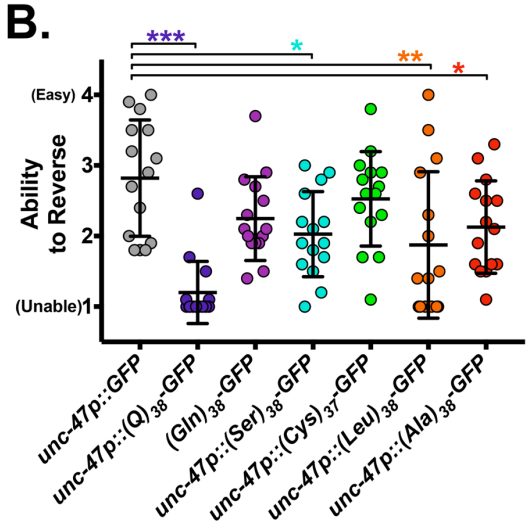
neuron commissures is a useful measure of the neurotoxic potential of proteins expressed in *C. elegans* [95, 143, 144, 247].

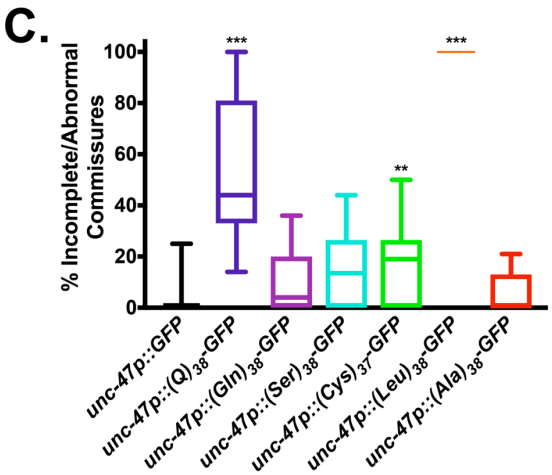
We characterized the effect of the CAG-derived RAN polypeptides on these GABAergic neuron parameters. *$(Q)_{38}$-GFP* worms, which expressed CAG-encoded polyglutamine (and possibly other RAN peptides), strongly inhibited the ability of the worms to reverse when expressed in GABAergic neurons. Surprisingly, *$(Gln)_{38}$-GFP* worms expressing polyglutamine encoded by a codon-varied sequence were comparable to control animals expressing GFP alone. In contrast, *$(Leu)_{38}$-GFP* worms exhibited highly significant movement defects. *$(Ala)_{38}$-GFP* and *$(Ser)_{38}$-GFP* also caused reversal defects (Figure 3.1 B). No defects were observed for *$(Cys)_{37}$-GFP*.

To test if the functional defects were caused by cellular damage, the neuronal commissure structures were examined. Neurodegeneration of commissures causes blebbing (swelling of the cellular membrane), breaks in the commissures, or branching of the commissures [95, 247, 314, 315]. The formation of the commissures was assessed by co-expressing *unc-47p::RFP*, which fills the GABAergic neurons. Commissure morphology was assessed by measuring the percentage of total detectable commissures in the animal that were blebbing, broken, or failed to reach the dorsal side. Consistent with our functional data, animals expressing $(Gln)_{38}$-GFP exhibited no defects in commissure structure (Figure 3.1 C, D). Animals expressing $(Leu)_{38}$-GFP exhibited highly penetrant commissure defects, although the neuron somas were intact. The GABAergic neurons in *$(Leu)_{38}$-GFP* worms traversed the length of the animals instead of the width, suggesting defects in neuron stability and/or axon guidance (Figure 3.1 C, D). *$(Q)_{38}$-GFP* worms also had abnormal commissures, with roughly half of the commissures in each

worm being abnormal or incomplete. $(Cys)_{37}$-GFP did not cause a significant functional defect. However, 19% of the commissures in (*Cys)$_{37}$-GFP* worms were abnormal. $(Ser)_{38}$-GFP caused a slight functional defect but did not cause a significant defect in the form of the commissures (Figure 3.1 C). These data show that the expression of $(Leu)_{38}$-GFP, but not other RAN peptides, is sufficient to cause both structural and functional defects in GABAergic neurons. They also suggest that factors other than polyglutamine contribute to the toxicity of $(Q)_{38}$-GFP, as a codon-varied $(Gln)_{38}$-GFP did not exhibit toxicity.


A.
+start ATG
Promoter
FLAG
(Peptide)38,90
Codon-Varied
GFP
Q Q Q Q Q GFP
Gln Gln Gln Gln Gln GFP
Ser Ser Ser Ser Ser GFP
Cys Cys Cys Cys Cys GFP
Leu Leu Leu Leu Leu GFP
Ala Ala Ala Ala Ala GFP


B.
***
*
**
*
(Easy) 4
3
2
(Unable) 1
Ability to Reverse
unc-47p::GFP
unc-47p::(Q)38-GFP
(Gln)38-GFP
unc-47p::(Ser)38-GFP
unc-47p::(Cys)37-GFP
unc-47p::(Leu)38-GFP
unc-47p::(Ala)38-GFP


C.
% Incomplete/Abnormal Commissures
100
80
60
40
20
0
***
**
***
unc-47p::GFP
unc-47p::(Q)38-GFP
unc-47p::(Gln)38-GFP
unc-47p::(Ser)38-GFP
unc-47p::(Cys)37-GFP
unc-47p::(Leu)38-GFP
unc-47p::(Ala)38-GFP

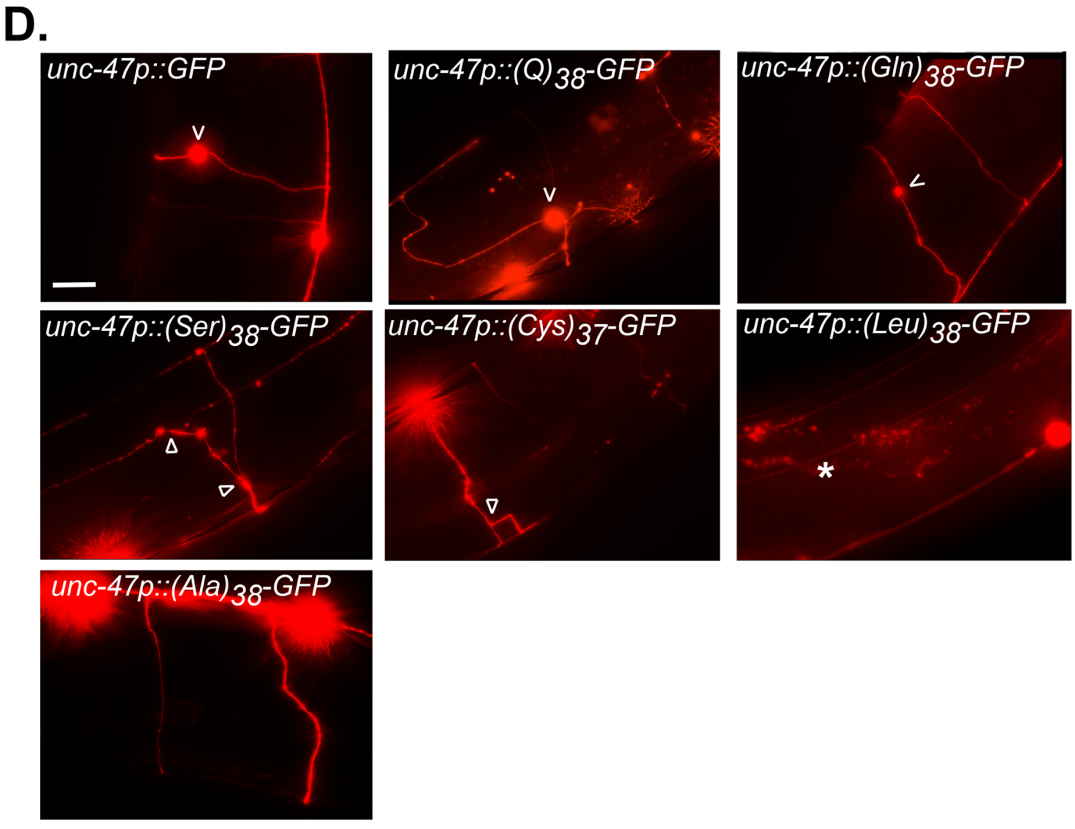

D.
unc-47p::GFP
unc-47p::(Q)38-GFP
unc-47p::(Gln)38-GFP
unc-47p::(Ser)38-GFP
unc-47p::(Cys)37-GFP
unc-47p::(Leu)38-GFP
unc-47p::(Ala)38-GFP

Figure 3.1 PolyLeucine Is the Most Toxic HD RAN Peptide in Neurons

(A) Molecular strategy for expression of codon-varied polypeptide repeats in *C. elegans*. (B) Quantification of the reversal ability of transgenic animals expressing the indicated polypeptide under the GABAergic neuron specific *unc-47* promoter. N=15 animals/genotype. Each symbol represents one animal, the horizontal line is the mean and the bars define the standard deviation. *P<0.05, **P<0.01, ***P<0.001 versus GFP control (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test). (C) Quantification of incomplete or abnormal commissures. For each animal we counted the number of commissures with blebbing, breakage, or that failed to reach the dorsal side, and divided that number by the total number of commissures. N=11-20 animals/genotype. The data is expressed in a box and whisker plot where the whiskers go from min to max. Statistics done by one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test against the GFP control. (D) Representative images of *unc-47*+ motor neurons in animals expressing the described polypeptide. V points to blebs, arrowhead points to branching, and the asterisk is at commissures that fail to reach the dorsal side. Scale bar=10 μm**.**

3.2.3 Muscle Model of HD RAN Polypeptides

C. elegans GABAergic neurons are anatomically small (cell bodies 1-2 micron) and inaccessible to some genetic approaches, such as RNAi-mediated gene knockdown [316]. To gain more insights into RAN peptide cell biology, as well as to generate a platform for RNAi and forward-mutagenesis based genetic suppressor screening, we expressed the HD RAN polypeptides in muscle cells using the *myo-3* promoter, which encodes a muscle-specific myosin heavy chain A. Toxicity in muscle cells can cause larval arrest and/or motility defects [75, 95]. We found that muscle expression of $(Leu)_{38}$-GFP caused highly penetrant larval arrest phenotype (Figure 3.2 A). Muscle expression of $(Ala)_{38}$-GFP, $(Gln)_{38}$-GFP, $(Cys)_{37}$-GFP, and $(Q)_{38}$-GFP caused a weakly penetrant larval arrest phenotype. However, $(Ser)_{38}$-GFP did not induce larval arrest. To facilitate post-developmental expression of each peptide, we initially grew animals on *gfp*(RNAi) and then switched them to *empty vector*(RNAi) to prevent peptide expression during development and permit expression during adulthood [95]. $(Q)_{38}$-GFP caused a significant increase in paralysis of animals as they aged. However, $(Ala)_{38}$-GFP, $(Gln)_{38}$-GFP, and $(Leu)_{38}$-GFP caused no significant enhancement in paralysis (Figure 3.2 B).

Paralysis assays measure strong defects in muscle function. However, the binary nature of this assay is not a sensitive measure of motility. To more quantitatively measure changes in motility, we performed thrashing assays. While only *$(Q)_{38}$-GFP* caused age-dependent paralysis, $(Q)_{38}$-GFP, $(Cys)_{37}$-GFP, $(Leu)_{38}$-GFP, and $(Ala)_{38}$-GFP all caused a decrease in thrashing rates. $(Ala)_{38}$-GFP motility defects were highly penetrant, whereas $(Q)_{38}$-GFP, $(Cys)_{37}$-GFP, and $(Leu)_{38}$-GFP defects were more variable. Similar to motor neuron expression, muscle expression of $(Q)_{38}$-GFP caused a strong motility

defect, but $(Gln)_{38}$-GFP did not cause a motility defect (Figure 3.2 C). As in GABAergic neurons, this suggests that factors other than polyGln contribute to the toxicity of polyQ.

To test if increased repeat length would cause an increase in toxicity, we expressed 90 repeats of codon-varied polyGln, polySer, polyCys, and polyLeu. We were unable to synthesize 90 repeats of polyQ or codon-varied polyAla for unknown technical reasons. Only $(Gln)_{90}$-GFP and $(Leu)_{90}$-GFP caused significant thrashing defects (Figure 3.2 D). Only polyGln had a strong increase in toxicity with increased repeat length, which had been previously observed in *C. elegans* polyQ models [75]. None of the HD polypeptides caused a shortening of lifespan when expressed at 38 repeats (Figure 3.2 E).

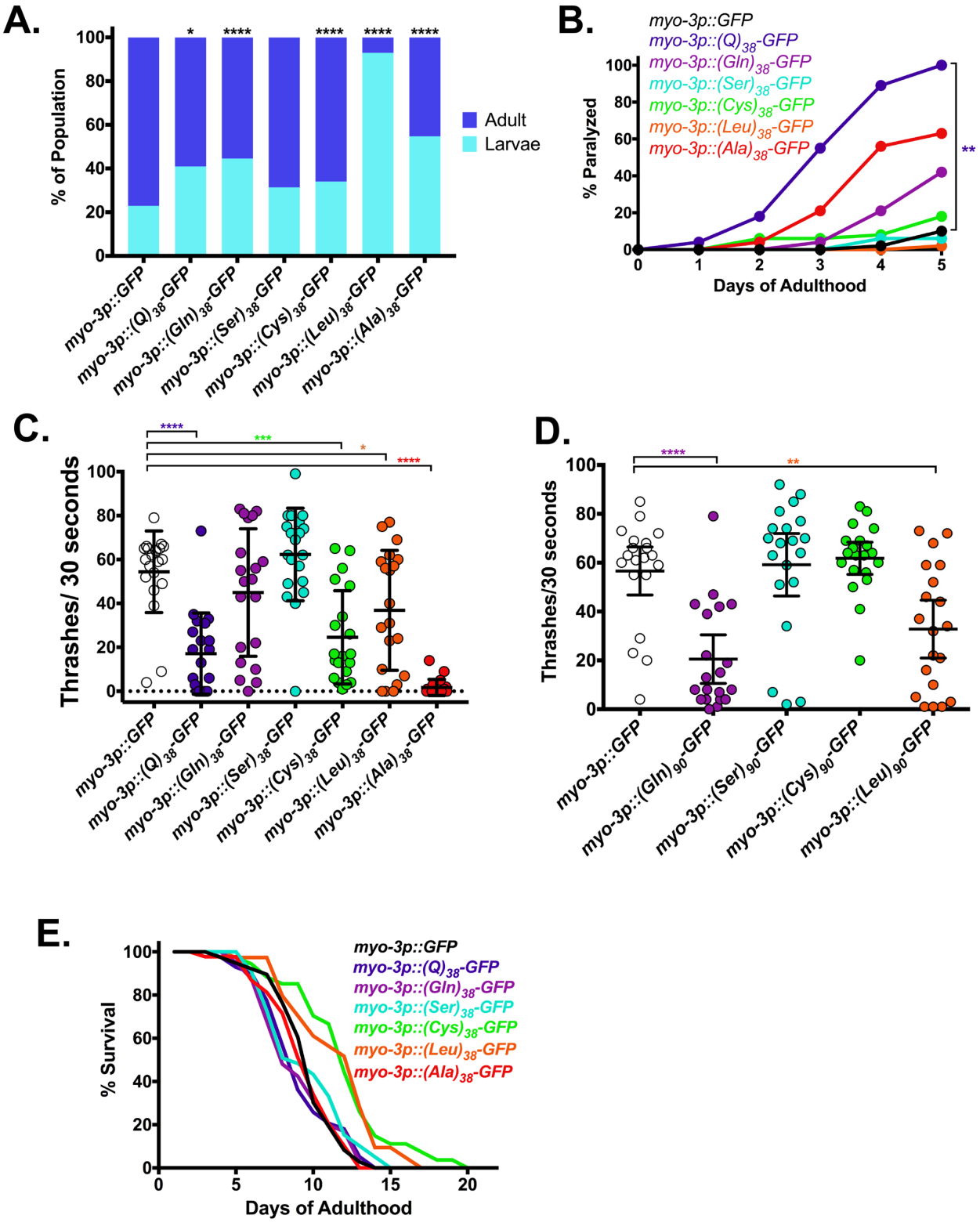


Figure 3.2 The Majority of HD RAN Polypeptides Are Toxic in Muscle

(A) Larval arrest measured by sorting a synchronized population 48 hours after eggs were laid on a plate. The worms were sorted through the COPAS Biosorter and animals with a time of flight ≥ 300 were categorized as adults, animals with a time of flight <300 were categorized as larvae. (N=188-269 worms), *P<0.05, ****P<0.0001, (Kruskal Wallis test of multiple comparisons with a post hoc Dunn's multiple comparison test). (B) Paralysis assay for adult animals raised in the absence of *gfp(*RNAi*).* **P<0.01, (Log Rank Test with Bonferroni-adjusted p-value). (C) Liquid thrashing quantification of transgenic animals raised in the absence of *gfp(*RNAi*)*, expressing the indicating polypeptides under the muscle-specific *myo-3* promoter. N=20 animals/genotype. Each symbol represents one animal, the horizontal line is the mean and the bars represent the standard deviation. *P<0.05, ***P<0.001, ****P<0.0001 versus GFP control (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test). (D) Liquid thrashing quantification of transgenic animals expressing the indicated polypeptides under the muscle-specific *myo-3* promoter (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test). (E) Lifespan measured in transgenic animals raised in the absence of *gfp*(RNAi), expressing the indicating polypeptides under the muscle-specific *myo-3* promoter. N=50 (Log Rank Test with Bonferroni-adjusted P-value).

3.2.4 Localization Patterns of HD RAN Polypeptides

To gain insights into the cell biological properties of the GFP-tagged HD RAN polypeptides, we took advantage of the transparent nature of *C. elegans* and performed live animal fluorescent imaging. Previous studies found that polyQ and polySer form puncta at disease-relevant repeat lengths [12, 75]. However, there is nothing known with regards to the localization properties of the other codon-varied RAN peptides. Both $(Q)_{38}$-GFP and $(Gln)_{38}$-GFP had both diffuse signal and puncta, consistent with previous polyQ models (Figure 3.3 A). $(Ser)_{38}$-GFP formed puncta but did not have diffuse signal (Figure 3.3 A), consistent with other polySer models [12]. $(Ala)_{38}$-GFP and $(Cys)_{37}$-GFP also formed puncta. PolyCys and polySer localization patterns appeared similar. $(Cys)_{90}$-RFP and $(Ser)_{90}$-GFP exhibited strong colocalization, suggesting they are forming on similar structures (Appendix C- Figure 4.1). Unlike the other RAN peptides, the GFP signal for $(Leu)_{38}$-GFP was extremely low and only detectable in a subset of adult muscle cells (vulval muscle cells) (Figure 3.3 A). The low polyLeu signal could be due to localization within a cellular compartment that impairs GFP fluorescence. GFP has impaired folding in oxidizing environments such as the lumen of the endoplasmic reticulum (ER) [317]. Additionally, leucine repeats commonly insert into membranes [318, 319]. Therefore, one possibility is that $(Leu)_{38}$-GFP is in a membrane and the GFP tag is oriented within an oxidizing environment. To test this possibility, I replaced GFP with a more stable version of GFP called superfolded GFP (sfGFP). sfGFP has several point mutations that enhance folding and fluorescence in non-optimal environments, such as the ER [320, 321]. Unlike $(Leu)_{38}$-GFP, $(Leu)_{38}$-sfGFP was observed in adult muscle cell, in addition to the vulval cells, and localized to the periphery of large spherical bodies of unknown origin (Figure

3.3 B). This suggests $(Leu)_{38}$ is membrane-bound and localized to an environment that impairs folding of GFP.

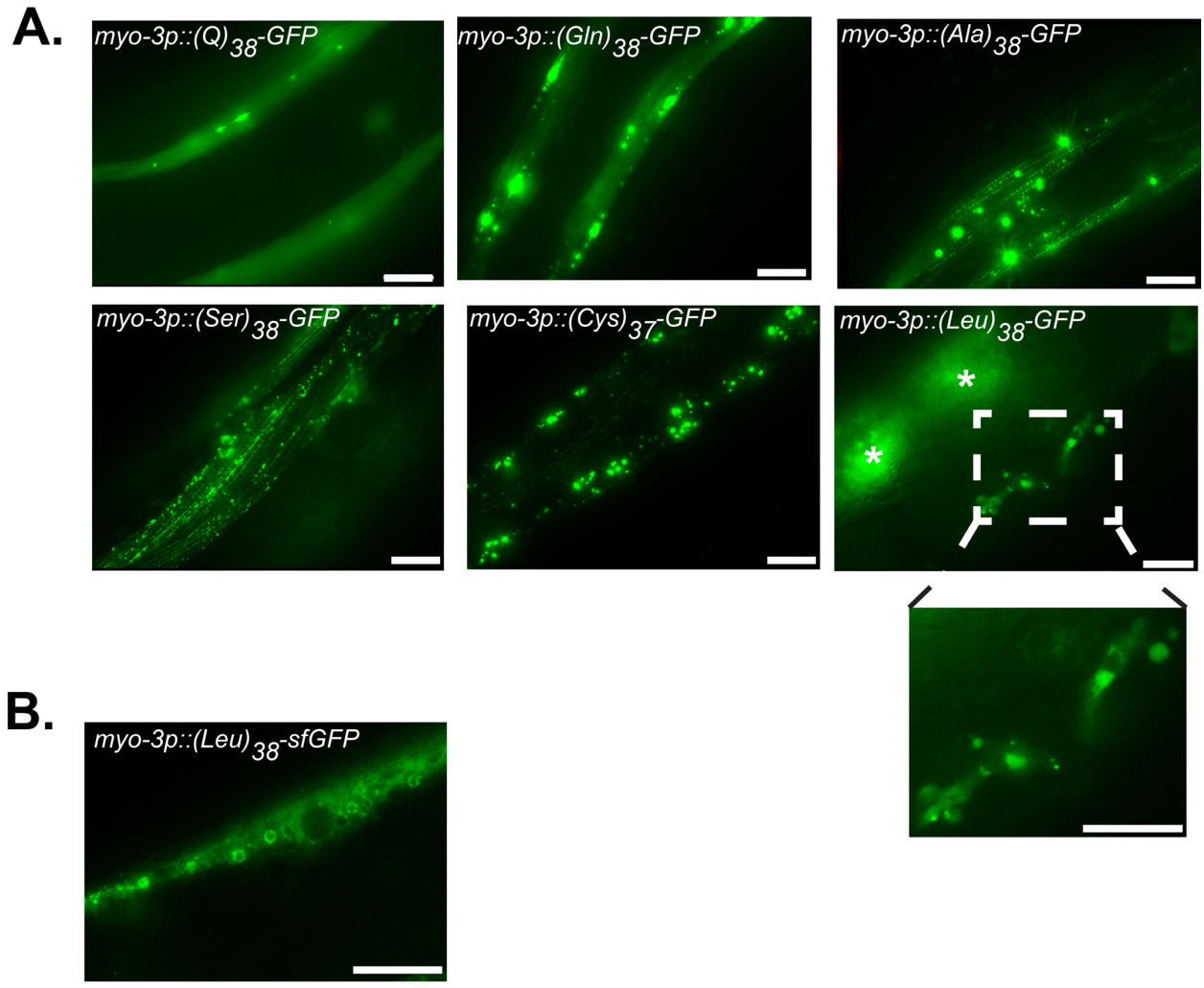


Figure 3.3 Every HD Polypeptide Forms Puncta at 38 Repeats

(A) Representative images of the indicated polypeptide in muscle expressed under *myo-3* promoter. Asterisks mark intestinal autofluorescence. Scale Bar=10 μm. (B) Representative image of *myo-3p*::$(Leu)_{38}$-sfGFP. Scale Bar=10 μm.

3.2.5 All HD RAN Peptides Form Aggregates

All HD RAN products form puncta *in vivo,* which may be aggregates. PolyGln and polySer are known to form aggregates [12, 127]. However, the structure of the puncta containing the other HD RAN products is unknown. The pathogenicity of the HD RAN peptides may be related to type of structures they form in cells. PolyQ aggregates are a common feature of HD patient tissue, and certain modifiers of polyQ aggregation also modify the toxicity of polyQ [310]. To better understand the cellular properties and potential toxic mechanisms of the HD RAN polypeptides, we tested if the HD RAN puncta exhibited restricted diffusion properties, a characteristic of protein aggregates [252]. The standard *in vivo* test for aggregation is Fluorescence Recovery After Photobleaching (FRAP), which measures the rate of diffusion of GFP-labeled molecules within a structure [252]. Soluble GFP molecules rapidly recover most of their fluorescence following FRAP, whereas protein aggregates exhibit slow and limited fluorescent recovery. Previous polyQ models have established that polyQ forms aggregates, so polyQ was used as a positive control for aggregation. We found that polyGln, polySer, polyCys, and polyAla exhibited limited FRAP recovery similar to aggregated polyQ (Figure 3.4). However, polyLeu FRAP recovery was more rapid and extensive, suggesting that polyLeu structures have more freely diffusible molecules within the aggregate and/or more exchange of molecules with the surrounding cytosol. The total recovery of $(Leu)_{38}$ was only 20%, so the majority of polyLeu peptides are immobile in the aggregate (Figure 3.4 C). Every HD polypeptide except for polyLeu formed aggregates with a fluorescent recovery comparable to a polyGln aggregate, including the nontoxic polySer,

emphasizing that polyLeu may be toxic by affecting different cellular pathways then polyglutamine.

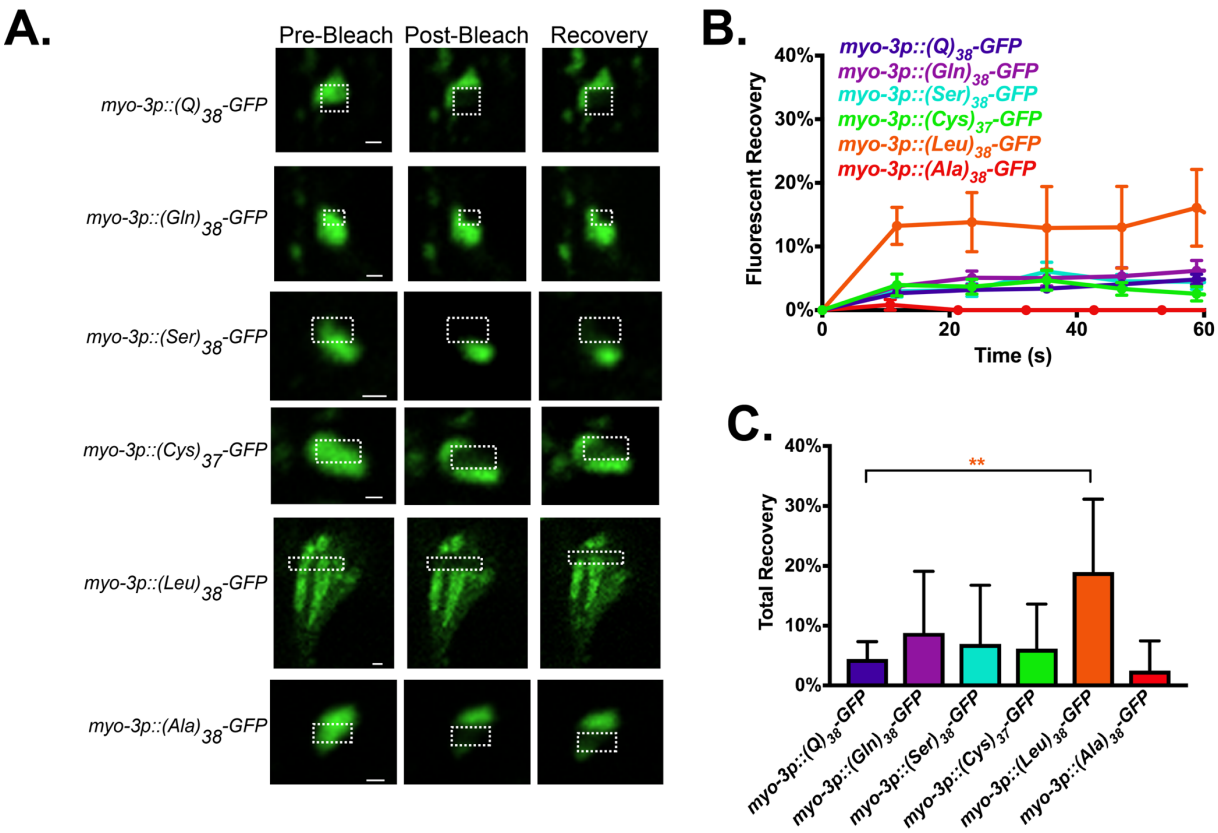


Figure 3.4 PolyLeu Forms Protein Aggregates With Limited Mobility

(A) Representative images from FRAP analysis of subcellular localized polypeptide proteins expressed in muscle. Dashed outline indicates the site of photobleaching and post-bleaching quantification. Recovery images are 60 seconds post-bleach. Scale bar=1 μm. (B) Quantification of FRAP imaging. Data shown are mean ± standard error of the mean (SEM), N=5-17 datasets/genotype). (C) Average equilibrium fluorescence recovery after 60 seconds. Data shown are mean ± SEM for 5-17 datasets. **P<0.01, (one-way ANOVA with a post hoc Dunn's multiple comparison test).

3.2.6 PolyLeu Toxicity is Length Dependent

Like aggregation, length-dependent toxicity is a classic characteristic of CAG repeats. PolyGln is well established to exhibit length-dependent toxicity, which we confirmed using our codon-varied constructs (Figure 3.2 C, D). To determine if polyLeu also exhibits length-dependent toxicity, I expressed polyLeu at 11 repeats, 20 repeats, 29 repeats, and 38 repeats in GABAergic neurons and muscle cells. $(Leu)_{11}$-GFP, $(Leu)_{20}$-GFP and $(Leu)_{29}$-GFP caused no functional defects in GABAergic neurons (Figure 3.5 A). Only $(Leu)_{38}$-GFP caused functional defects (Figure 3.5 A). Therefore, polyLeu exhibits length-dependent toxicity in GABAergic neurons and requires more than 29 repeats to cause toxicity. The change in toxicity could be due to altered localization patterns of polyLeu depending on the length of leucine repeats. We consequently expressed the various lengths of polyLeu in muscle cells to determine whether the different lengths of polyLeu had similar localization patterns. $(Leu)_{11}$-GFP and $(Leu)_{20}$-GFP had distinct localization patterns that appeared localized to membranes, since they outlined but did not fill structures within the cell. However, $(Leu)_{11}$-GFP localizes to a reticular structure that fills the muscle cell and small spherical structures ~1 μm in diameter, while $(Leu)_{20}$-GFP localizes to the periphery of the muscle cell, potentially the plasma membrane, as well as various structures inside the cell that exhibit no consistent shape (Figure 3.5 C). Cellular localization of $(Leu)_{29}$-GFP and $(Leu)_{38}$-GFP was challenging to determine since both lengths of polyLeu appeared to inhibit the fluorescence of the GFP attached to the repeats (Figure 3.5 B). The length-dependent decrease in fluorescence was specific to the polyLeu-bound GFP, as free RFP expressed in the same tissues did not have a length-dependent decrease in fluorescence (Figure 3.5 C). A version of GFP with

enhanced folding (sfGFP) allowed the detection of $(Leu)_{38}$ localization. $(Leu)_{38}$-sfGFP formed hollow spherical bodies that were ~1 μm in diameter, similar in size to the bodies seen with $(Leu)_{11}$-GFP. $(Leu)_{38}$-sfGFP also had a weak signal in the body of the muscle cell, the structure of which was not determined. Therefore, $(Leu)_{11}$-GFP and $(Leu)_{38}$-sfGFP localized to similar hollow, spherical bodies ~1 μm in diameter. $(Leu)_{20}$-GFP had a distinct localization pattern outlining multiple structures in the cell. $(Leu)_{11}$-GFP, $(Leu)_{20}$-GFP, and $(Leu)_{29}$-GFP did not exhibit the larval arrest observed in both $(Leu)_{38}$-GFP and $(Leu)_{38}$-sfGFP. Together, these data suggest that the strong toxicity of $(Leu)_{38}$-GFP is length-dependent, with more than 29 leucine repeats required to cause larval arrest when expressed in muscle, or reversal defects when expressed in motor neurons. The increase in toxicity does not appear to be due to a change in the localization pattern of polyLeu, since non-toxic $(Leu)_{11}$-GFP and $(Leu)_{29}$-GFP exhibited many of the same localization properties as toxic $(Leu)_{38}$-GFP. The length-dependent toxicity of polyLeu, in addition to polyLeu being toxic in multiple tissues, suggests that polyLeu could be a strong contributor to CAG repeat toxicity.

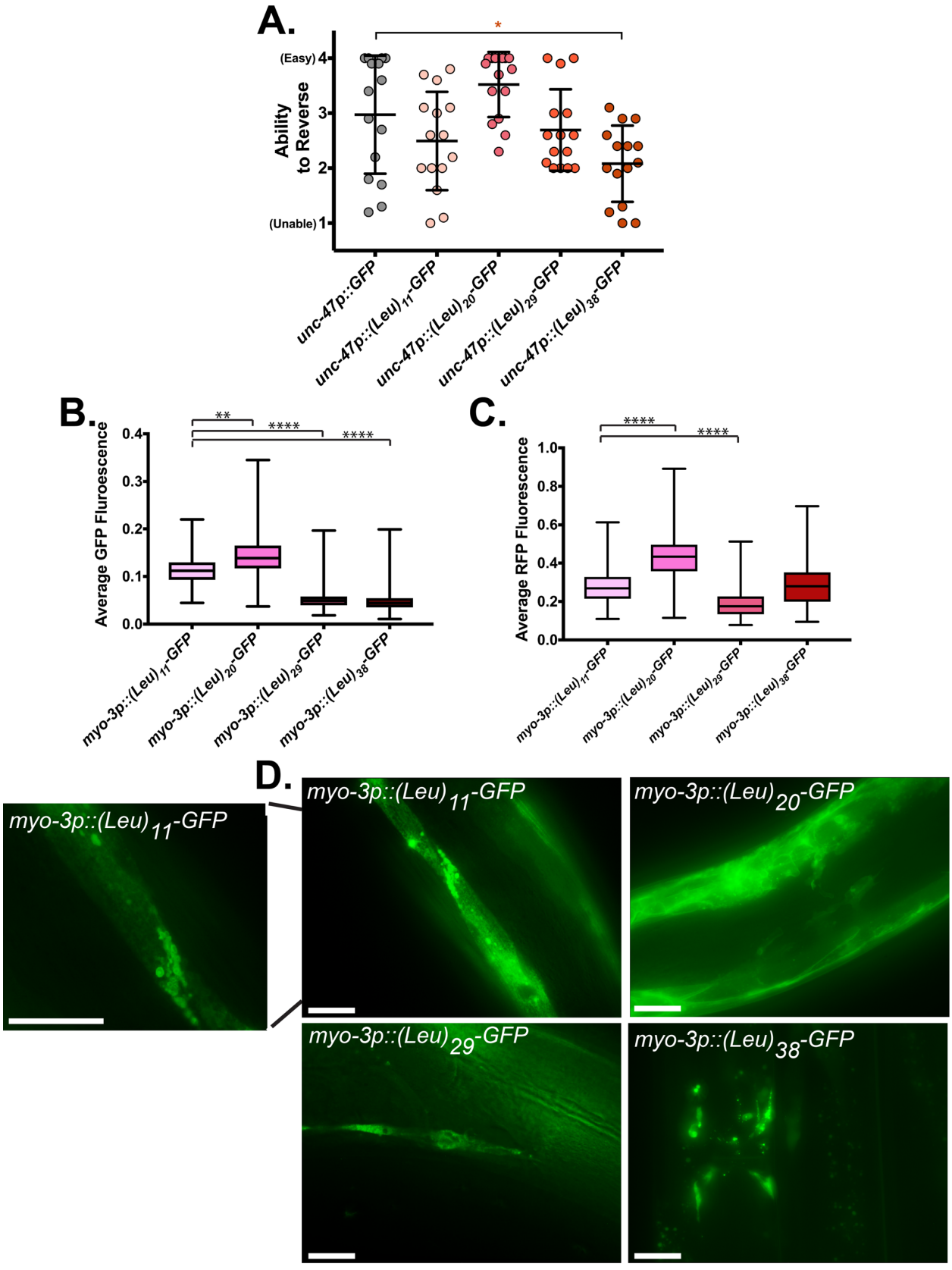

A.
*
(Easy) 4
3
2
(Unable) 1
Ability to Reverse
unc-47p::GFP
unc-47p::(Leu)11-GFP
unc-47p::(Leu)20-GFP
unc-47p::(Leu)29-GFP
unc-47p::(Leu)38-GFP
B.
**
****
****
Average GFP Fluroescence
0.4
0.3
0.2
0.1
0.0
myo-3p::(Leu)11-GFP
myo-3p::(Leu)20-GFP
myo-3p::(Leu)29-GFP
myo-3p::(Leu)38-GFP
C.
****
****
Average RFP Fluorescence
1.0
0.8
0.6
0.4
0.2
0.0
myo-3p::(Leu)11-GFP
myo-3p::(Leu)20-GFP
myo-3p::(Leu)29-GFP
myo-3p::(Leu)38-GFP
D.
myo-3p::(Leu)11-GFP
myo-3p::(Leu)11-GFP
myo-3p::(Leu)20-GFP
myo-3p::(Leu)29-GFP
myo-3p::(Leu)38-GFP

Figure 3.5 Toxicity of PolyLeu Is Length Dependent

(A) Quantification of the reversal ability of transgenic animals expressing the indicated length of polyLeu under the GABAergic neuron specific *unc-47* promoter. N=15 animals/genotype. Each symbol represents one animal, the horizontal line is the medan and the bars represent the standard deviation. *P<0.05 versus GFP control (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test) (B) Average fluorescence of GFP and (C) RFP in single animals, as measured by the COPAS Biosorter through measuring the total fluorescence of the worm and dividing it by the time of flight of the worm. N=149-611. **P<0.01, ***P<0.001, ****P< 0.0001 versus *myo-3p::(Leu)$_{11}$-GFP* control (one-way non-parametric ANOVA with a post hoc Dunn's multiple comparison test) (D) Representative images of the indicated length of polyLeu in muscle expressed under the *myo-3* promoter. The animals are day 1 adults. Scale Bars=20 μm.

3.2.7 Forward Mutagenesis Screen to Identify Suppressors of PolyLeu Toxicity

To identify the molecular pathways required for polyLeu toxicity, we performed a forward mutagenesis screen for suppressors of polyLeu-induced larval arrest. Animals were treated with N-Ethyl-N-Nitrosourea (ENU), which is widely used to induce mutations in forward mutagenesis screens [322-324]. ENU causes alkylation of nucleic acids leading to missense mutations and, unlike other mutagens like ethyl methanesulfonate (EMS), produces a broad spectrum of nucleotide changes [325]. After ENU treatment, animals self-fertilized while growing on (*gfp*)RNAi, which suppressed $(Leu)_{38}$-GFP expression and toxicity. Second generation animals were examined to determine if they carried homozygous recessive mutations that suppressed the toxicity of polyLeu. The animals were removed from (*gfp*)RNAi and placed on *E. coli* OP50, a commonly used laboratory food source for *C. elegans* that allows expression of $(Leu)_{38}$-GFP. Any animals that grew and reproduced over multiple generations carried a genetic suppressor of $(Leu)_{38}$-GFP toxicity (Figure 3.6 A). To identify any suppressors that caused a general decrease in transgene expression, the expression level of another component of the same transgene, *myo-3p::RFP*, was quantified based on RFP fluorescence (Figure 3.6 C). *myo-3p::RFP* is located in the same transgene that expresses $(Leu)_{38}$-GFP, as the two plasmids were co-injected and then integrated into the genome [326]. After selecting for suppressors that did not decrease transgene expression, 25 independent suppressors were isolated (Figure 3.6 B). To determine if similar pathways were likely affected in the different suppressors, the localization of $(Leu)_{38}$-GFP was examined in ten of the identified suppressors. Three distinct localization patterns for $(Leu)_{38}$-GFP were observed in the suppressors (Figure 3.6 D). Two of the suppressors had wild-type expression patterns

of $(Leu)_{38}$-GFP. Six suppressors had increased diffuse signal in the muscle and two suppressors had diffuse signal in the gonads. The variety of localization patterns suggests the screen identified several genes and/or pathways mediating $(Leu)_{38}$-GFP toxicity.

To identify the phenotype causing mutations, we used whole-genome sequencing (WGS) and subsequent RNAi phenocopying. WGS was used to identify nonsynonymous single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) in exons of the suppressors. Several assumptions were made to simplify identification of the relevant mutations: the suppression is due to a single gene, the mutation causes loss of function, and the mutation is homozygous. Homozygous mutations not found in the starting $(Leu)_{38}$-GFP strain were identified in 914 genes (Figure 3.7). Each suppressor had an average of 45.3 ± 12.6 genes containing polymorphisms. While several genes were overrepresented in the screen (*alh-2* (polymorphisms present in 10/25 suppressors), *H03A11.2* (13/25 suppressors), *tag-80* (23/25 suppressors), *ttn-1* (25/25 suppressors), *ZC247.1* (19/25 suppressors)), the same genes were also identified as containing polymorphisms in an unrelated forward mutagenesis screen performed in our lab and appear to be mutated frequently due to their large size. Therefore, we chose to exclude these genes from further analysis. For the remaining genes, we used the Enrichment Analysis Tool on WormBase (www.wormbase.org) to identify protein families overrepresented in our screen. Overrepresented protein families included proteins with ATPase activity, active transmembrane transport, or ATP or GTP binding ability. To test if any of these genes were necessary for larval arrest, I performed a candidate-gene RNAi screen (Figure 3.7). I screened genes from the suppressor screen that were: mutated in

multiple suppressors, had a stop-codon-inducing mutation, or were part of an over-represented protein family. Of the 165 genes that fit these requirements, 108 genes had available RNAi clones for the gene knockdown. These 108 genes were screened by RNAi in two separate screens using three replicates per screen. Five gene knockdowns consistently suppressed $(Leu)_{38}$-GFP induced larval arrest in at least 3 of 5 plates during the two screens (Table 3.1). Many of the candidate genes caused larval arrest when knocked down with RNAi, which is likely why only 5 gene knockdowns phenocopied the suppression of larval arrest. One of the candidates, *F35A5.1*, has the same SNP mutation in 17 suppressors, with no other mutations in *F35A5.1* present in suppressors. The 17 suppressors have different localization patterns of $(Leu)_{38}$-GFP and different levels of suppression of larval arrest, making it unlikely that the mutation in F35A5.1 is the causative mutation. Another gene, *T28A11.17*, is part of a gene family in *C. elegans* with at least eight other genes, some of which contain an L-domain, a protein domain also found in cell-surface receptors such as the insulin receptor (IR) or the epidermal growth factor receptor (EGFR). T28A11.17 is predicted to have two transmembrane domains, suggesting this protein is found in the secretory pathway. However, *T28A11.17* does not have a human homologue and its function is unknown, making its disease relevance unclear. Therefore, we focused on the three remaining genes, each of which were part of a common protein family. SRI-62 is a G-protein coupled receptor (GPCR), PMP-2 is an ATP-binding cassette transporter (ABC transporter), and UBH-4 is a deubiquitinating enzyme (Table 3.2). To determine if loss of these genes caused similar effects, I studied the suppressors thought to have causative mutations in one of the genes.

I characterized the mutations in the identified polyLeu suppressors to determine how the mutations in *sri-62*, *pmp-2*, and *ubh-4* could suppress toxicity. The four suppressors all strongly suppress larval arrest (Figure 3.6 B and Table 3.2). *sri-62* (*dr57*) affects a G-protein-coupled receptor (GPCR) of the serpentine chemoreceptor class. *dr57* converts a glutamine to an early stop-codon mutation (Q85*), which causes *sri-62* truncation or mRNA degradation through nonsense-mediated mRNA decay. The *sri-62* mutant had diffuse $(Leu)_{38}$-GFP in the gonad of animals, which is never observed in wild-type animals. *pmp-2*(*dr59*) affects a predicted ortholog of the human ABC transporter ABCD3, which is involved in peroxisomal import of fatty acids and is associated with some forms of a peroxisome assembly disorder called Zellweger syndrome. *dr59* generates a missense mutation (L457F) within the P-loop domain of *pmp-2* that hydrolyzes ATP. The mutated leucine is also conserved between *C. elegans* and humans, suggesting that loss of PMP-2 ATP hydrolysis and substrate transport may form the basis for $(Leu)_{38}$-GFP suppression. The *pmp-2* mutant altered the $(Leu)_{38}$-GFP expression by causing strong diffuse localization of $(Leu)_{38}$-GFP within muscle. *ubh-4* (*dr58*) and *ubh-4* (*dr60*) both contain the same molecular missense mutation (S6R), although both suppressors were isolated independently (Table 3.1). The mutated serine is not conserved between *C. elegans* and humans and is not within a known catalytic domain, so it is unclear how the mutation impacts UBH-4 function. Both *ubh-4* alleles exhibit a wild-type pattern of $(Leu)_{38}$-GFP localization. Taken together, these data suggest that loss of multiple genes suppress $(Leu)_{38}$-GFP toxicity and the mechanisms of suppression are likely to differ due to the diverse effects of each mutant on $(Leu)_{38}$-GFP localization.

To determine whether *sri-62*, *pmp-2* and *ubh-4* were general or cell-type specific suppressors of polyLeu toxicity, I tested whether these mutants also suppressed $(Leu)_{38}$-GFP toxicity in GABAergic neurons. $(Leu)_{90}$-GFP and GFP were co-expressed in the GABAergic neurons of each suppressor, which causes the loss of normal commissures in wild-type animals (Figure 3.8). We asked whether or not commissure morphology was restored in each of the suppressor mutants. In *ubh-4* (*dr60*), animals had wild-type commissures, suggesting suppression of $(Leu)_{90}$-GFP toxicity in GABAergic neurons (Figure 3.8). Both *pmp-2* (*dr59*) and *sri-62* (*dr57*) had commissure defects similar or worse to the wild-type control, suggesting they were unable to suppress commissure defects (Figure 3.8). These data suggest that *sri-62* and *pmp-2* are only required for polyLeu toxicity in muscle, while *ubh-4* is required for polyLeu toxicity in multiple tissues.

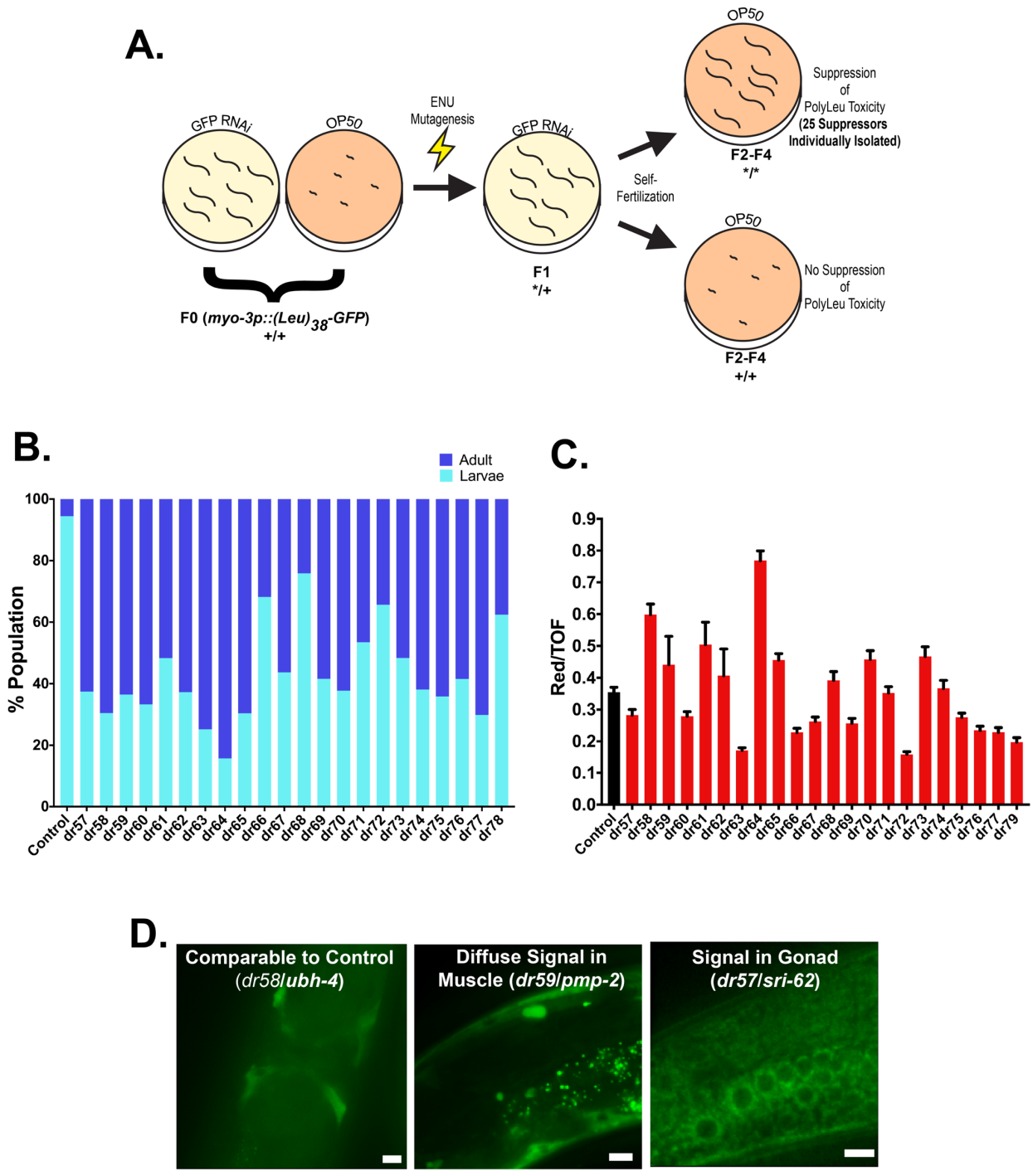


Figure 3.6 Multiple Types of PolyLeu Suppressors Identified in Mutagenesis Screen

(A) Diagram of ENU mutagenesis screen for suppressors of larval arrest caused by *myo-3p::(Leu)$_{38}$-GFP* toxicity. *myo-3p::(Leu)$_{38}$-GFP* animals were treated with ENU mutagenesis and the F2 generation was raised on *gfp(RNAi)* at 20°C to suppress polyLeu toxicity. Plates with the F2 population were allowed to starve and a subset of the starved larvae (F2-F4) were washed onto a new OP50 plate and placed at 25°C. Animals that were able to propagate for several generations had suppressed polyLeu toxicity and one suppressor was isolated from each plate. (B) The worms were sorted through the COPAS Biosorter and animals with a time of flight ≥ 300 were categorized as adults, animals with a time of flight <300 were categorized as larvae. (C) The average RFP levels of the worm were measured by dividing the total RFP signal of a worm by its time of flight (N=211-523). (D) Representative images of the different localization patterns observed in suppressors. V points to vulval muscles. Scale Bar=10 μm.

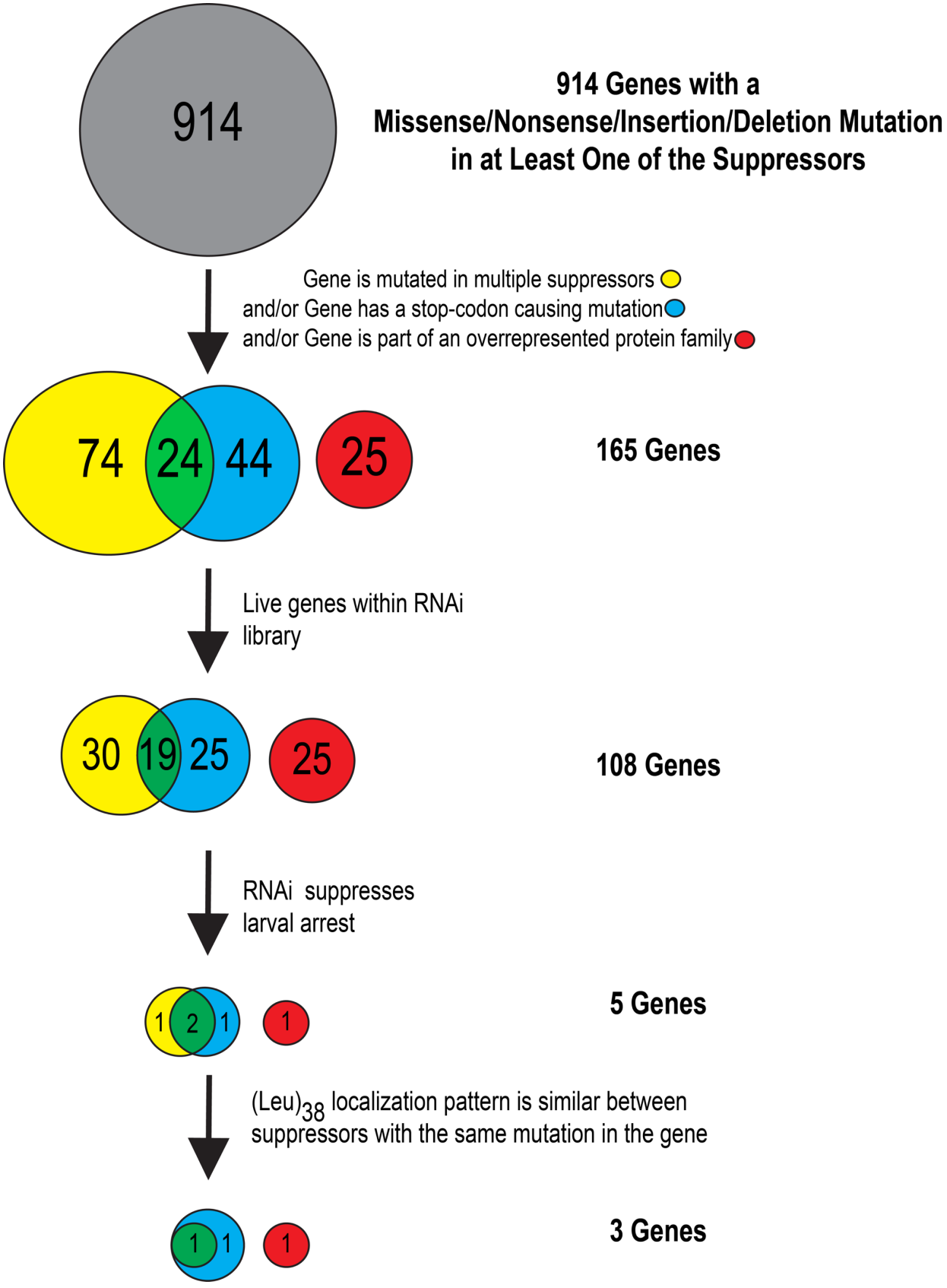

914
914 Genes with a
Missense/Nonsense/Insertion/Deletion Mutation
in at Least One of the Suppressors
Gene is mutated in multiple suppressors
and/or Gene has a stop-codon causing mutation
and/or Gene is part of an overrepresented protein family
74
24
44
25
165 Genes
Live genes within RNAi
library
30
19
25
25
108 Genes
RNAi suppresses
larval arrest
1
2
1
1
5 Genes
(Leu)38 localization pattern is similar between
suppressors with the same mutation in the gene
1
1
1
3 Genes

Figure 3.7 Diagram of Workflow to Identify Genes Required for PolyLeu Toxicity

Twenty-five suppressors were isolated from a forward mutagenesis screen. Whole-genome sequencing of suppressors identified 914 genes containing a nonsynonymous substitution or insertion/deletion in at least one isolated suppressor. A subset of genes were identified that: were mutated in multiple suppressors, had at least one nonsense mutation, or were part of an overrepresented protein family. The Ahringer RNAi library or the ORFeome RNAi library were then checked to determine if they contained RNAi clones for the subset of genes. Genes with RNAi clones in the Ahringer/Orfeome libraries were used in a candidate RNAi screen. Growth was the selection criteria. Starvation-synchronized L1's were seeded onto RNAi plates (2 plates/RNAi strain) and grown at 25°C. After 1 week, the plates were scored blindly for growing and moving larvae. The hits from the first RNAi screen were tested again in a second RNAi screen (3 plates/RNAi strain). RNAi targets that suppressed polyLeu toxicity in 3/5 plates were counted as suppressors. The ENU-induced mutations in the RNAi-confirmed candidate genes were reexamined to determine if duplicated mutations caused similar characteristics (strength of polyLeu suppression and localization of polyLeu-GFP). Genes were considered strong candidates if the ENU mutants had similar phenotypes and RNAi against the gene suppressed polyLeu toxicity.

Table 3.1 RNAi Phenocopied Genes Identified from the Candidate RNAi Screen

Gene	Human Homologue	Group	Suppressors with Mutations	Mutation
F35A5.1	NEFH	Neurofilament	*dr57, dr60, dr61, dr62, dr59, dr63, dr64, dr65, dr67, dr68, dr69, dr71, dr73, dr74, dr75*	E1257K
T28A11.17	None	Unknown	*dr59, dr61, dr64*	V87M
sri-62	None	Receptor	*dr57*	Q85*
ubh-4	*UCHL5*	Ubiquitin	*dr58, dr60*	S6R
pmp-2	*ABCD3*	ABC transporter	*dr59*	L457F

Table 3.2 (*Leu)$_{38}$-GFP* Suppressors with Mutations in RNAi Phenocopied Genes

Suppressor	Suppression	Potential Gene (Allele Effect)	Gene Family	*myo-3p::(Leu)$_{38}$-GFP* Localization
dr57	Strong	*sri-62* (Q85*)	Receptor	Whole Animal
dr59	Strong	*pmp-2* (L457F)	ABC Transporter	Strong Diffuse Signal in Muscle
dr60	Strong	*ubh-4* (S6R)	Ubiquitin Ligase	Unchanged
dr58	Strong	*ubh-4* (S6R)	Ubiquitin Ligase	Unchanged

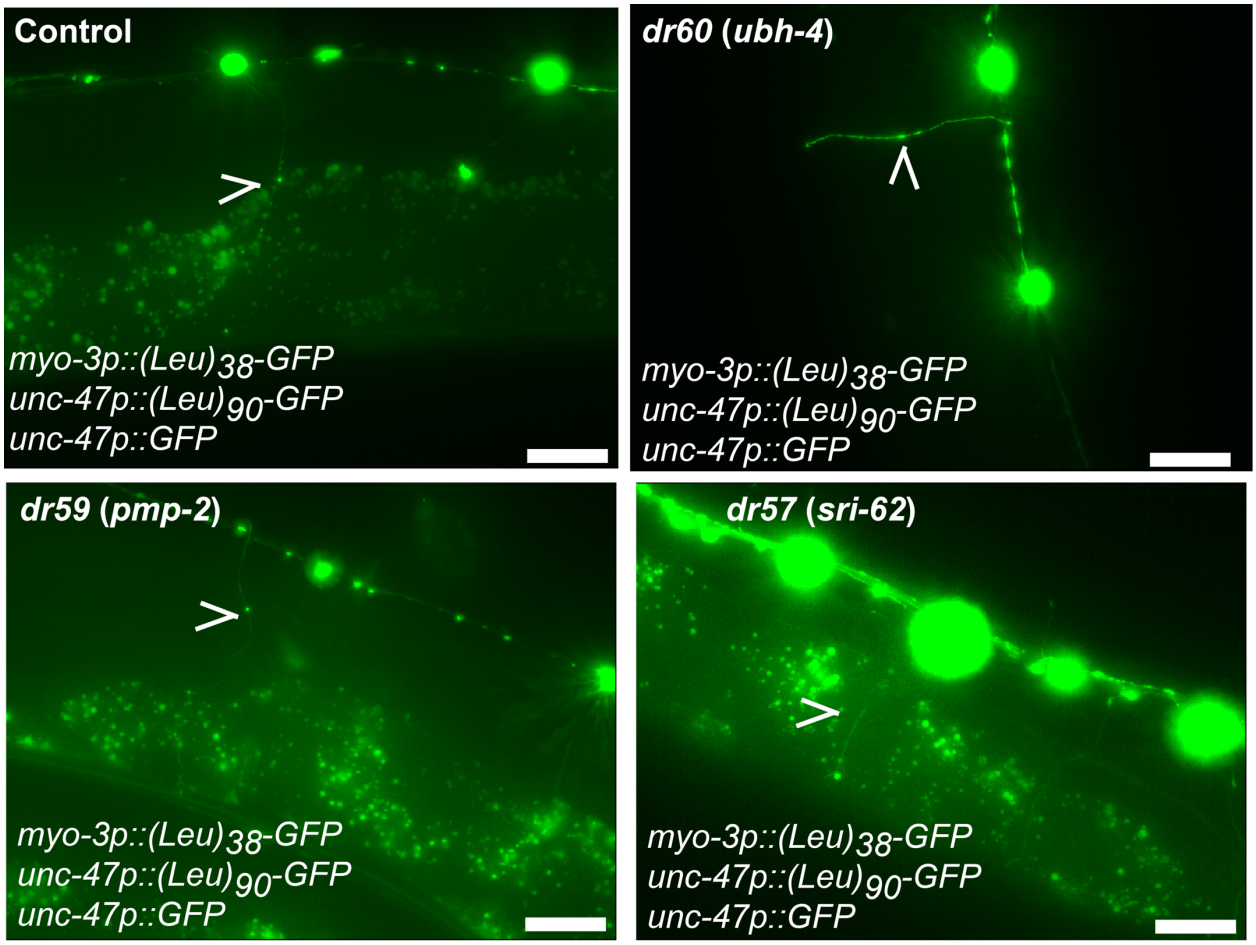


Figure 3.8 *dr60* Suppresses PolyLeu Toxicity in GABAergic Neurons

Representative images of the indicated suppressors with the starting line as a control. V points to the commissures in the image. The green punctate signal in control/*dr55*/*dr53* is autofluorescence. Animals had extrachromosomal expression of *unc-47p::(Leu)$_{90}$-GFP* and *unc-47p::(GFP)* and had *myo-3p::(Leu)$_{38}$-GFP* and *myo-3p::RFP* integrated. The animals are day 1 adults maintained on (*gfp*)RNAi. Scale Bars=20 μm.

3.3 Discussion

In this study, I investigated the cell biological and pathophysiological properties of the newly discovered HD RAN polypeptides. I found that all of the HD RAN peptides formed biophysically defined protein aggregates. However, only a single HD RAN peptide, polyLeu, caused both neuropathological changes and functional defects in the tissues that I studied. Since polyLeu has never before been associated with a disease, polyLeu effector mechanisms are unknown. To identify these mechanisms, I performed an unbiased forward genetic screen for suppressors of polyLeu phenotypes. I identified three candidate genes whose inhibition blocks polyLeu toxicity in muscle. The molecular identity of these genes may provide insight into how polyLeu contributes to CAG diseases like HD. Given that my studies represent the first *in vivo* dissection of CAG-associated RAN peptide properties, I will discuss my findings for each RAN peptide individually, with a particular focus on how my findings integrate with previous reports of the structure and function of each homopolymeric peptide.

3.3.1 PolyCys and PolySer: Weakly Toxic RAN Products

Both polyCys and polySer exhibited similar cell biological and phenotypic effects when expressed in *C. elegans*. PolyCys and polySer formed aggregates at 38 and 90 repeats. They also colocalized when expressed together. The colocalization of polySer and polyCys is not solely due to their polar charge. Glutamine is also a polar amino acid, but polyQ does not colocalize with either polySer or polyCys (Appendix C-Figure 4.2). A possible explanation of the distinct aggregation patterns of polyQ, polySer, and polyCys

could be that polySer and polyCys may localize to a different subcellular environment than polyQ. For example, both Cys and Ser [327], but not Gln, are known targets for post-translational protein palmitoylation, which anchors proteins to cellular membranes [328]. The palmitoylation of Cys residues in another protein, cysteine string protein α, is essential for protein aggregation and is associated with the neurodegenerative disease adult-onset neuronal ceroid lipofuscinosis [329]. Therefore, post-translational modification of polySer and polyCys, but not polyGln, may lead to protein interactions that drive protein aggregation in ways that are distinct from polyGln. Future studies examining the post-translational modification state of the polySer and polyCys protein could provide insights into these and other potential aggregation mechanisms.

Another common property of polyCys and polySer is that expression of either polypeptide caused weak and inconsistently toxic effects. GABAergic expression of $(Cys)_{37}$ caused minor structural but not functional defects, whereas GABAergic expression of $(Ser)_{38}$ caused functional defects, but not structural defects. Muscle expression of $(Cys)_{37}$ caused weak motility defects and larval arrest. However, muscle expression of polySer never caused motility defects or larval arrest. When expressed on their own, polyCys and polySer exhibit little to no phenotypic consequences in *C. elegans.* This was surprising considering a recent study that suggested polySer has significant toxicity in a SCA8 murine model, another CAG repeat expansion disease [330]. However, this model did not express a pure polySer protein. Rather, it expressed the CAG expanded ATXN8 gene, which is the genetic cause of SCA8. polySer is one of several RAN products translated from the SCA8 CAG repeat. However, the other possible RAN products were not studied. In the murine model of SCA8, polySer was thought to be toxic

because RAN translated polySer colocalized with degenerating brain regions that lacked the ATXN8 protein. However, RAN translation products from different reading frames typically occur in the same tissue [12] and can occur in the same cells [6]. Therefore, polySer may be acting as a biomarker for RAN translation, and one or more of the other CAG RAN products are toxic in neurons. My studies strongly suggest that polySer alone does not cause neurodegeneration in the SCA8 murine model. However, polySer may have synergistic interactions with other RAN peptides that enhance toxicity. In the future, it will be important to test this possibility by co-expressing polySer with each of the other RAN products and determining if there is any change in toxicity.

3.3.2 PolyQ and PolyGln: PolyQ Has Polyglutamine-Independent Toxicity

The CAG-encoded polyglutamine (polyQ) model and the CAG/CAA-encoded polyglutamine (polyGln) models exhibit radically different levels of toxicity. GABAergic expression of $(Q)_{38}$ caused substantial structural and functional neuronal defects, but GABAergic expression of $(Gln)_{38}$ caused no detectable changes. Similarly, muscle expression of $(Q)_{38}$ caused extensive motility defects, while muscle expression of $(Gln)_{38}$ caused no motility defects. PolyGln did display length-dependent toxicity, as muscle expression of $(Gln)_{90}$ produced motility defects. However, disease-relevant lengths of polyGln did not cause toxicity, while disease-relevant lengths of polyQ did cause toxicity.

The disparity between the toxicity of these two models of polyglutamine is antithetical to the widely-held belief that polyglutamine is the driving cause of the toxicity in HD. If polyglutamine is the main toxic effect of the CAG repeat, the observed differences between the polyQ model and the polyGln model must be due to

dissimilarities in the biological properties of the polyglutamine expressed in each model. However, the amino acid sequence produced from both models is identical, although both lack surrounding genetic context that could significantly enhance or suppress toxicity levels. In addition, the polyglutamine expressed by both *polyQ-GFP* and *polyGln-GFP* formed aggregates. Therefore, the difference in toxicity between polyQ and polyGln is likely due to the different nucleotide sequences encoding the homopolymer. The CAG repeat in the polyQ model could be toxic through its RNA repeats [331] or possible RAN products [12]. Conversely, the mixed CAG/CAA repeat is not toxic through either its RNA repeats or possible RAN products. The CAG/CAA sequence was designed to minimize the formation of RNA secondary structures, which are required for RNA toxicity [332] and may be required for RAN translation [4]. Therefore, differences in toxicity between polyQ and polyGln animals are likely due to the glutamine-independent causes of toxicity in polyQ, such as the repetitive CAG RNA or RAN products. To determine if RAN products may be contributing to the toxicity of polyQ, RAN translation could be measured by placing GFP in the +1 reading frame or the +2 reading frame of the CAG repeat to detect RAN expression of polySer or polyAla, respectively. In the future, it will be important to re-examine whether or not existing *C. elegans* CAG models are producing RAN peptides in addition to polyGln. The presence of such peptides may dramatically alter the use and interpretation of studies based on these models.

Previous studies have found the nucleotide sequence encoding polyglutamine affects the toxicity of a 'polyQ' model. In both *Drosophila* [333] and cell culture [334], a CAG-encoded polyglutamine was significantly more toxicity than a CAA-encoded polyglutamine. The difference in toxicity between the CAG repeats and CAA repeats was

ascribed to CAG RNA being toxic, as CAG repeats in untranslated regions also cause toxicity [331, 333]. CAA repeats do not undergo RAN translation [4] or cause RNA toxicity [333], so, the CAA-encoded polyglutamine is comparable to the CAG/CAA-encoded polyglutamine used in this study. However, none of the “untranslated” CAG models have been tested for RAN translation. RAN products may be produced in the RNA toxicity models and contribute to their observed phenotypes.

Widely-used models of ‘polyQ’ diseases, such as the murine R6/2 strain and the *C. elegans* AM141 strain, could also be undergoing RAN translation. Therefore, these ‘polyQ’ models may express both polySer and polyAla from the CAG sense RNA strand. If the antisense CUG RNA is also transcribed in these models, RAN translation would additionally produce polyLeu and polyCys, as well as polyAla in a different genetic context. The possibility of glutamine-independent toxicity occurring in common HD models is made clear by the drastic differences in toxicity between my polyQ model and polyGln model. Given the widespread and experimentally proven prevalence of RAN translation in several CAG repeat expansion disorders, referring to these diseases as ‘PolyQ diseases’ may no longer be accurate and our descriptions of these diseases should be modified to include our evolving understanding of the role of the RAN peptides besides PolyQ.

3.3.3 PolyAla: A Possible Contributor to CAG Toxicity

PolyAla may be a strong contributor to CAG repeat toxicity. In *C. elegans*, polyAla was toxic in both muscle and GABAergic neurons. Muscle expression of polyAla caused strong motility defects, and GABAergic expression of polyAla caused functional, but not

structural, defects. $(Ala)_{38}$-GFP also formed aggregates when expressed in muscle cells of *C. elegans*. In humans, expanded polyAla repeats cause aggregation in multiple diseases [335]. The aggregation of the polyAla repeats caused the mislocalization of the polyAla containing proteins and a concordant loss-of-function phenotype [335]. Since polyAla toxicity is mediated by the loss of protein function containing the expanded Ala tract, polyAla toxicity has mainly been studied within its native genetic context. The toxicity of a polyAla peptide without genetic context, as was modeled in our study, has not been thoroughly studied. In addition, because the expanded alanine repeats in the polyAla diseases are on average shorter than 30 repeats, a pure polyAla peptide had not been modeled at an HD relevant length until our work.

A homopolymeric alanine peptide may be toxic through similar mechanisms as the alanine repeat-expansion diseases. A yeast-two hybrid model found that homopolymeric disease-relevant lengths of polyAla self-interact ($(Ala)_{29}$ and $(Ala)_{28}$) [336]. It is unknown if alanine repeat lengths that normally appear in proteins (5-21 repeats) [337] also interact with disease-relevant lengths of alanine repeats. This could be tested by co-expressing $(Ala)_{11}$ and $(Ala)_{38}$. Based on previous research, $(Ala)_{11}$ should be diffuse in our model, while $(Ala)_{38}$ is known to aggregate in our model. If the $(Ala)_{11}$ only aggregates when co-expressed with $(Ala)_{38}$, then $(Ala)_{38}$ may be binding proteins containing similar lengths of polyAla repeats. PolyAla repeats are enriched in transcription factors [337, 338], which need to be localized to the nucleus to perform their function. Therefore, the interaction of the transcription factors containing polyAla repeats with the polyAla aggregates could deplete the cell of available transcription factors. PolyAla aggregates sequestering transcription factors could also explain the difference in toxicity observed between muscle

cells and GABAergic neurons, as the depleted transcription factors may be less important in GABAergic neurons.

3.3.4 PolyLeu Is the Most Toxic HD RAN Product

The most toxic HD RAN product across tissues was polyLeu. Like polyAla, polySer, and polyCys, polyLeu caused functional defects when expressed in GABAergic neurons and motility defects when expressed in muscle. However, polyLeu expression caused highly penetrant phenotypes in both muscle and neurons, which were only weakly observed in the other RAN models. GABAergic expression of polyLeu induced significant morphological defects where every commissure had some sort of structural impairment. Muscle expression of $(Leu)_{38}$-GFP caused larval arrest of ~90% of the population.

The strong toxicity of polyLeu was unexpected, as polyLeu repeat expansions have not been previously linked to a genetic disease. This is in spite of polyLeu repeats being one of the most common single amino acid repeats in the human proteome, with ~1,500 proteins containing polyLeu repeats longer than four leucines [337]. However, all the naturally occurring polyLeu repeats are 11 leucines or shorter, suggesting there is an evolutionary selection against longer polyLeu repeats. The lack of polyLeu repeat expansion diseases could be due to polyLeu repeat expansions disrupting development in humans as it does in *C. elegans*. In mammalian cells, RAN translation, which produce polyLeu, increases activity with activation of the integrated stress response pathway [86, 89, 90], and the integrated stress response pathway is activated with age [268], so RAN translation rates could increase with age. Another possible reason why polyLeu repeat expansions do not typically occur is because polyLeu repeats are not encoded by a

nucleotide repeat. Unlike glutamine, which is encoded by only two codons, leucine is encoded by six codons. However, single amino acid repeats of leucine have an average codon homogeneity of 55%, which is more homogenous than the single amino acid repeats of alanine (37%) [339]. As polyAlanine can undergo pathogenic repeat expansions, it seems likely polyLeu can as well. Previous work has found that polyLeu repeats longer than 30 repeats cause toxicity in cells [340, 341] and in *Drosophila* [342], but all models were created with a CTG repeat that could be undergoing RAN translation. My work is the first evidence that a codon-varied pure polyLeu peptide causes significant cellular toxicity.

The toxicity of polyLeu could be due to the apparent membrane-localization of polyLeu. While the expression pattern of polyLeu changes based on the length of the leucine repeat, every length of polyLeu appears to be membrane-localized. This is consistent with previous findings that polyLeu peptides as short as nine repeats can spontaneously incorporate into lipid membranes [318, 343, 344]. Single amino acid repeats of polyLeu occur in many signal peptides or transmembrane proteins [338]. For this reason, leucine is the most common amino acid in proteins that localize to the Endoplasmic Reticulum (ER), Golgi Apparatus, or vacuoles [345]. Consistent with this, $(Leu)_{11}$-GFP localizes to a tubular network in *C. elegans*, which resembles previous reports of ER [346], and spherical bodies that are ~1.2 μm in diameter, which cluster at the distal regions of the muscle cells. Eleven leucine repeats can act as an anchor signal sequence to target proteins to a membrane [344], so the spherical structures may be part of the endocytic pathway. $(Leu)_{20}$-GFP appears to localize to the plasma membrane and smaller membrane-bound compartments, as $(Leu)_{20}$-GFP marks various structures, some

structures similar to the peripheral shape of a *C. elegans* muscle cell, and other spherical and non-spherical structures. To confirm whether $(Leu)_{20}$-GFP marks the plasma membrane, $(Leu)_{20}$-GFP should be co-localized with a plasma membrane marker, such as glycosylphosphatidylinositol (GPI) [346]. The disease-relevant length of polyLeu, $(Leu)_{38}$-sfGFP, localizes to the boundaries of unidentified spherical structures that are distributed throughout the length of the muscle cell and are similar in size to those seen with $(Leu)_{11}$. The spherical structures containing $(Leu)_{38}$-sfGFP may also be part of the endocytic pathway. $(Leu)_{38}$-sfGFP has a diffuse background signal that may be the ER, based on the localization pattern of $(Leu)_{11}$ and the fact that this signal is only detected with sfGFP-tagged $(Leu)_{38}$, as opposed to GFP-tagged $(Leu)_{38}$. The oxidizing environment of the ER inhibits the folding of GFP, but not the folding of sfGFP, which could explain why $(Leu)_{38}$-GFP had impaired fluorescence. If the spheres are derived directly from the ER, the $(Leu)_{38}$-sfGFP marked spheres should contain ER-resident proteins such as the phosphatidylinositol synthase PISY-1 [347].

Longer repeat lengths of polyLeu disrupt the ability of GFP to fold properly. $(Leu)_{29}$-GFP and $(Leu)_{38}$-GFP have impaired fluorescence due to misfolding of GFP, as a variant of GFP with a higher rate of folding [320], sfGFP, did not have impaired fluorescence when attached to $(Leu)_{29/38}$. There are three possible hypotheses for why longer repeat lengths of polyLeu have impaired fluorescence (Figure 3.9), which make easily testable predictions. The first hypothesis for the impaired folding of $(Leu)_{29/38}$-GFP is that longer leucine repeats such as $(Leu)_{29/38}$ cause a switch in the orientation of polyLeu-GFP compared to shorter leucine repeats such as $(Leu)_{11/20}$. This orientation change causes the C-terminal GFP of $(Leu)_{29/38}$-GFP to be inside the lumen of the

membrane-bound structure, which could be an unfavorable environment for GFP folding. Varying the length of a polyLeu repeat in the signal-anchor sequence of a transmembrane protein can affect the orientation of the protein in the membrane [348]. In addition, membrane-bound organelles such as the ER [321] can suppress GFP fluorescence. The orientation of transmembrane proteins can be studied using glycosylation, which occurs in the ER lumen. Glycosylatable GFP (gGFP) has an N-linked glycosylation site and can be used to determine if the GFP is localized in the ER lumen or the cytoplasm [349, 350]. Cytosol-oriented gGFP will be not be glycosylated and will fluoresce strongly. However, lumen-oriented gGFP will be glycosylated and, thus, will fluoresce weakly [349]. This experiment could identify the orientation of $(Leu)_{38}$-GFP in the membrane.

The second hypothesis for why $(Leu)_{29/38}$-GFP displays impaired folding is that longer repeats of polyLeu, $(Leu)_{29/38}$, localize to structures that disrupt GFP folding, while shorter repeats of polyLeu, $(Leu)_{11/20}$, localize to structures that do not disrupt GFP folding. The localization of a protein can be affected by the length of the transmembrane domain, as seen with the yeast protein UBC6 localizes to the ER, Golgi, or plasma membrane depending on the length of its transmembrane domain [351]. PolyLeu exhibited length-dependent localization patterns with $(Leu)_{11}$ and $(Leu)_{20}$, so $(Leu)_{29/38}$ could also be localized to different cellular regions. If $(Leu)_{38}$ does colocalize with the shorter lengths of polyLeu, $(Leu)_{38}$ is most likely to colocalize with $(Leu)_{11}$ as both $(Leu)_{38}$-sfGFP and $(Leu)_{11}$-GFP localize to spherical cellular structures. The two lengths of polyLeu can be expressed together with different fluorescent tags to determine if they colocalize. Any structures that have $(Leu)_{38}$ but lack $(Leu)_{11}$ can be identified using fluorescent markers for the different endosomal compartments in *C. elegans* [346, 347].

There are established markers for the ER (KDEL [352]), the Golgi (AMAN-2 [353]), the peroxisome (DAF-22 [347]), and other compartments. The identified organelle can be studied to determine if its lumen impairs GFP folding.

The third hypothesis for impaired folding of $(Leu)_{29/38}$-GFP is that higher repeat lengths of polyLeu disrupt the folding environment of the cell, causing a decrease in overall protein folding rates including the folding of GFP. Impaired protein folding occurs in other neurodegenerative models, such as a *C. elegans* $(CAG)_{40}$-YFP model [130] where the $(CAG)_{40}$-YFP sensitizes the worms to other misfolding mutations. Like the $(CAG)_{40}$-YFP model, $(Leu)_{38}$ polypeptides may sequester molecular chaperones away from their normal client proteins. This would deplete the available chaperone levels in either the ER or the cytoplasm, depending on the localization of the misfolded proteins, and increase the misfolding rate of other proteins in that compartment, such as GFP. Impaired folding in the ER causes activation of *binding immunoglobin protein (BiP)*/*hsp-4* [354, 355]. Therefore, one test of this hypothesis would be to determine if *hsp-4* levels are increased in cells expressing $(Leu)_{38}$-GFP. Similar tests could be used to measure impaired protein folding in other compartments, such as the cytoplasm (*hsp-16*) and the mitochondria (*hsp-6*).

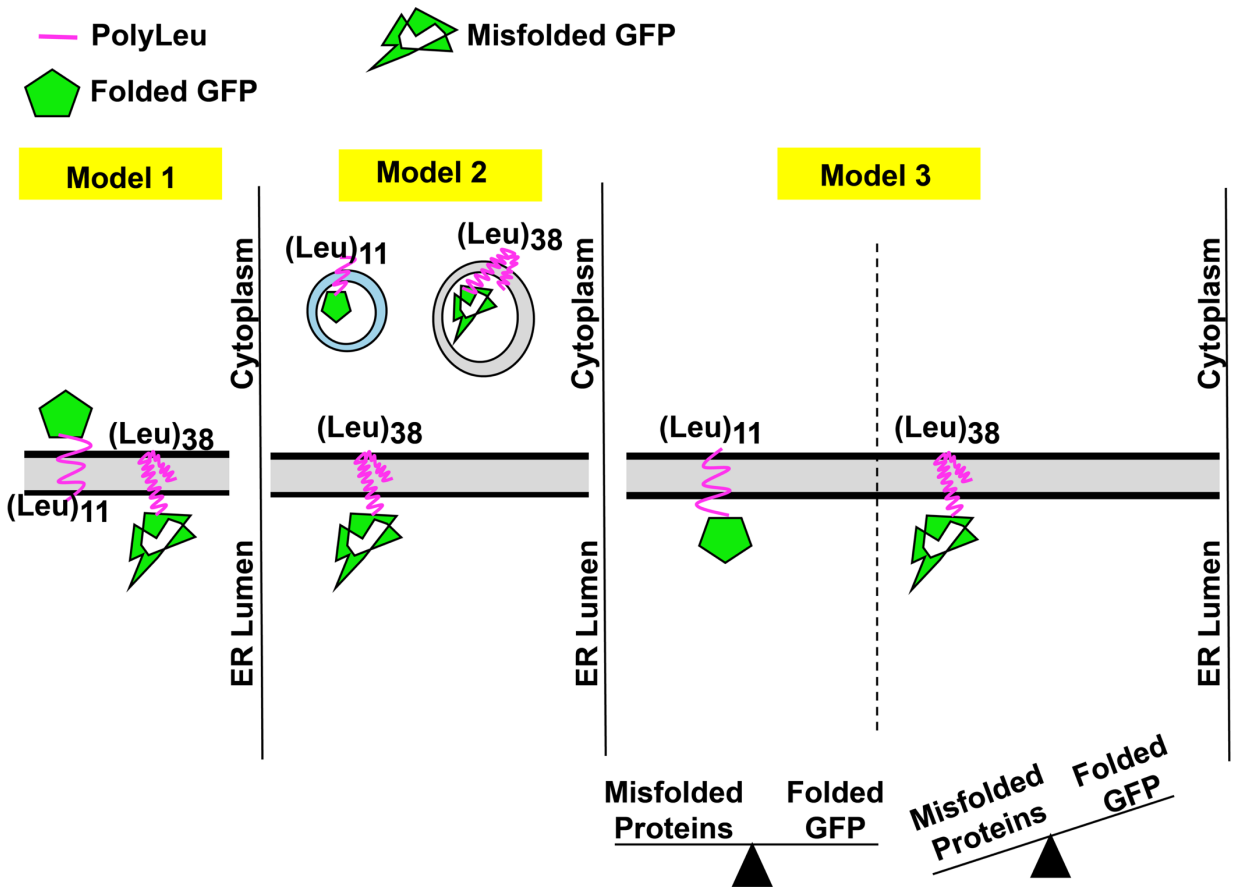


Figure 3.9 Possible Models of PolyLeu Length-Dependent Variation in GFP Folding

Model 1: The shorter polyLeu repeats, such $(Leu)_{11}$, cause the GFP to be oriented in the cytosol where it folds properly and fluoresces, while the longer polyLeu repeats, such as $(Leu)_{38}$, cause the GFP to be oriented in the ER lumen where it is unable to fold properly. Model 2: The length of the polyLeu repeats varies where polyLeu is targeted. $(Leu)_{11}$ is targeted to membrane-bound structures where GFP can fold properly and fluoresce, while $(Leu)_{38}$ is localized in different membrane-bound structures that disrupt the folding and fluorescence of GFP. Model 3: both $(Leu)_{11}$ and $(Leu)_{38}$ have GFP oriented inside the lumen, but GFP folding is impaired in $(Leu)_{38}$ due to an increase in misfolded proteins with $(Leu)_{38}$.

3.3.5 PolyLeu Screen Hits

A forward mutagenesis screen and subsequent RNAi screen for genes required for $(Leu)_{38}$-GFP-induced larval arrest identified three candidates: *sri-62*, *pmp-2*, and *ubh-4*. *sri-62* encodes a serpentine seven-transmembrane GPCR, *pmp-2* encodes an ABC (ATP-binding cassette) transporter, and *ubh-4* encodes a deubiquitinating enzyme.

sri-62 is a nematode-specific gene and is poorly studied. It encodes a 334 amino acid long protein with seven transmembrane domains. As a serpentine receptor, SRI-62 is likely a chemoreceptor and is localized to the plasma membrane. However, both its specific function and its cellular localization is unknown. A study of the expression pattern of *C. elegans* chemoreceptors found *sri-62* expression in the PVT interneuron [356]. The PVT interneuron is important for the spatial arrangement of neurons and secretes netrin to aide neurons in pathfinding [357]. While SRI-62 was not observed in the muscle [356], part of the *sri-62* promoter may have been excluded in the construct used to study *sri-62*. To develop a more accurate understanding of the expression pattern of *sri-62*, it will be important to determine the native expression pattern using methods such as CRISPR/Cas9 tagging of the endogenous *sri-62* gene. SRI-62 is also not likely to be a broad suppressor of neurodegenerative proteins, as reduction in levels of *sri-62* was identified in a previous RNAi screen as an enhancer of tau-induced toxicity in *C. elegans* [358]. In contrast, we identified *sri-62*(RNAi) as a suppressor of polyLeu toxicity. Therefore, how *sri-62* is required for toxicity is likely specific to polyLeu.

Another identified suppressor of polyLeu toxicity, *pmp-2*, also encodes a transmembrane protein. Also, like *sri-62*, *pmp-2* is not well characterized in *C. elegans*. However, it has a strong human homologue, *ABCD3*, that has been extensively studied.

Both *C. elegans* PMP-2 and ABCD3 have four transmembrane domains. PMP-2 likely localizes to the peroxisome [359], as ABCD3 is a peroxisomal protein that imports long fatty acids and branched acyl-CoA into peroxisomes [360]. It is unclear if ABCD3 is primarily translated on cytoplasmic ribosomes or inserted into the ER as ABCD3 has been observed in the ER [361], but ABCD3 can also escape targeting to the ER and instead be translated by free ribosomes [362]. Mutations in ABCD3 are associated with a heterogenous group of peroxisome assembly disorders called Zellweger syndrome [363]. Peroxisome dysfunction has not been previously linked to HD, so it is unclear how the function of PMP-2 is required for polyLeu toxicity.

Other than the fact that they both encode multi-pass transmembrane proteins and are required for polyLeu toxicity, *sri-62* and *pmp-2* have little in common. However, previous studies on polyLeu provides some suggestions as to how these seemingly unrelated proteins may be required for $(Leu)_{38}$ toxicity. For example, a $(Leu)_{24}$ polypeptide capped by two lysines on either end is sufficient to inhibit the function of the sarcoplasmic reticulum calcium pump (SERCA). This peptide was based on the SERCA inhibitor, phospholambin, which contains a leucine-rich transmembrane domain. A similarly constructed polypeptide with only 18 leucines failed to inhibit SERCA [364]. Other research on phospholambin found that the leucine-rich transmembrane domain of phospholambin inhibits SERCA by locking SERCA in an unbound confirmation [365]. Therefore, polyLeu may be toxic in a similar manner as phospholambin by interacting with the hydrophobic transmembrane domains of SRI-62 and PMP-2 and disrupting their ability to function (Figure 3.8). Another possibility is that polyLeu is disrupting the folding of SRI-62 and PMP-2 in the ER, thereby activating the ER unfolded protein response

(UPR^{ER}) and potentially driving ER-associated protein degradation (ERAD) of these and other substrates (Figure 3.10). PolyLeu is more likely to be toxic by disrupting protein folding than to be toxic by disrupting protein function. This is because polyLeu would need to have precise interactions with proteins to disrupt their function, but not their folding. If polyLeu is toxic through disrupting folding of proteins in the ER, $(Leu)_{38}$ expression should cause an increase in misfolded proteins in the ER leading to activation of the UPR^{ER}. UPR^{ER} activation would upregulate the reporter *hsp-4p::GFP* [366], as *hsp-4* is the *C. elegans* homologue of *BiP* and is activated by the UPR^{ER} [367].

The final gene identified in the screen, *ubh-4*, encodes a deubiquitinating enzyme that has a human homologue, *uchl5*. Reducing protein levels of UBH-4 increases the rate of proteasomal degradation [368]. The *ubh-4* homologue in humans, *uchl5*, has been previously linked to CAG repeat expansion diseases. Reducing protein levels of UCHL5 suppresses the aggregation of an ataxin3-polyQ (CAG) model in mammalian cells, which likely also produced polyLeu via RAN translation [368]. Therefore, *ubh-4*/*uchl5* may have conserved effects on polyLeu toxicity. Decreased levels of UBH-4 may suppress polyLeu toxicity through enhanced proteasomal degradation of the misfolded transmembrane proteins, if polyLeu causes toxicity by increasing misfolding of transmembrane proteins (Figure 3.10). Additional studies will be needed to test this model.

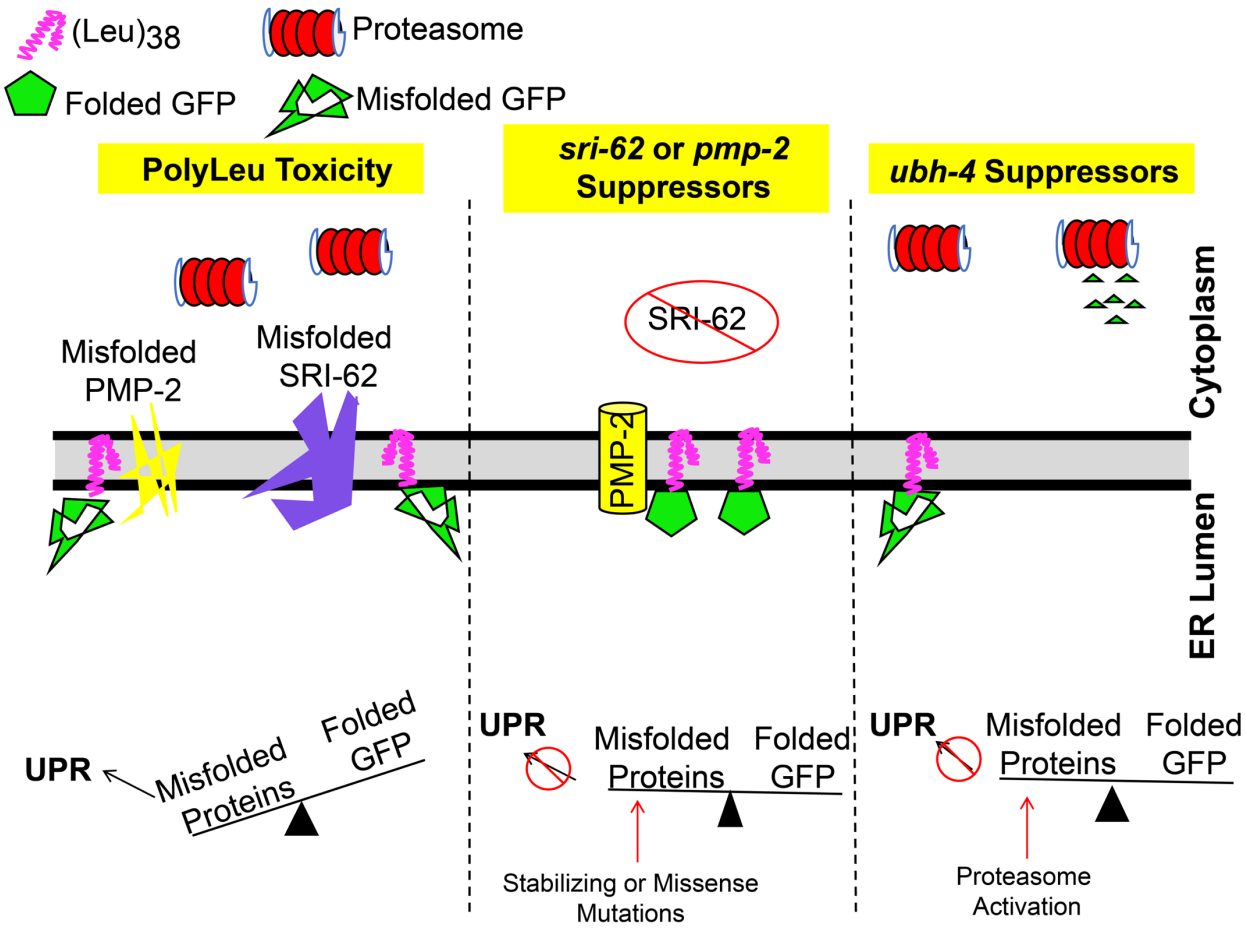


Figure 3.10 Model of PolyLeu Toxicity Through Causing Misfolding in ER

PolyLeu may be toxic by disrupting the folding of transmembrane-domain containing proteins such as PMP-2 and SRI-62. This increase in misfolded proteins would deplete the ER of chaperones and make GFP more likely to misfold, as the ER is already an unfavorable folding environment for GFP. The deletion mutant of *sri-62* would decrease the amount of misfolded proteins in the ER, as SRI-62 no longer needs to fold correctly. The *pmp-2* mutation may stabilize PMP-2 folding or cause degradation of PMP-2, either way causing a decrease in the total amount of misfolded proteins in the ER. Knockdown of *ubh-4* causes activation of the proteasome, so mutations in *ubh-4* may disrupt UBH-4

activity so that the activity of the proteasome is increased. This would increase the degradation of proteins and prevent UPR activation.

3.3.6 Limitations of Model

Our *C. elegans* HD RAN model offers a simplified view of the various HD RAN products, which is advantageous for developing a basic understanding of how each HD RAN product may contribute to HD. However, our model lacks several key pieces of information required for a more nuanced understanding of HD RAN toxicity. In our system, each of the HD RAN products are modeled individually without genetic context. Therefore, we cannot comment on how the HD RAN products interact with each other. It is likely that RAN products are co-expressed in HD since RAN translation occurs in multiple reading frames in the same cell in other RAN diseases [96]. The RAN products may act synergistically when co-expressed. Given that polyAla, polySer, polyCys, and polyGln all formed aggregates, it is reasonable to predict that co-expression of two or more polypeptides may overwhelm the chaperone capacity of the cell and activate relevant unfolded protein response pathways in the cytoplasm and/or the ER. Since polyLeu may cause activation of UPR^{ER}, it is easy to see how the HD RAN polypeptides could act synergistically. To test these models, the individual HD RAN polypeptides need to be co-expressed to determine how the RAN polypeptides are toxic together, not just on their own.

Another limitation of our model is that the HD RAN peptides lack the surrounding genetic context that is present in the *HTT* gene. Previous work modeling CAG repeats in various forms of *HTT* found that certain genetic contexts could greatly increase the CAG

repeat toxicity in mice [369]. With regards to HD RAN translation, the precise translation initiation and termination sites for the various peptides have not yet been described. Work modeling RAN translation of the Fragile X CGG repeat expansion suggests that RAN translation can initiate using near-cognate start codons, which only have one nucleotide different from a start codon [87]. Therefore, new HD RAN models initiating from near-cognate start codons in-frame with the repeat expansions and extending to stop codons downstream of the repeat could prove informative. This would be particularly interesting for polyLeu, as the amino acids on either side of a polyLeu repeat affect the orientation of the polyLeu within a plasma membrane.

The main limitation of our model is also the main advantage; *C. elegans* are not mammals. *C. elegans* allow for rapid discovery of genetic pathways required for toxicity. However, the evolutionary conservation of these pathways needs to be confirmed in either mammalian cells or mice. Therefore, a key experiment that needs to be performed is confirming that the HD polypeptides exhibit similar toxicity in both *C. elegans* and mammalian cells. Previous work by our lab has shown that pathways regulating the toxicity of C9orf72 RAN peptides are highly conserved from *C. elegans* to mammals. Therefore, it seems likely that the pathways regulating the toxicity of HD RAN peptides will also be conserved. Experimental testing of this prediction should be a high future priority.

3.3.7 Conclusion

RAN translation has been confirmed to occur in four different CAG/CTG repeat expansion neurodegenerative diseases, but this is the first work to examine the individual

RAN polypeptides at a disease-relevant length. It has been 26 years since the discovery of the genetic cause of HD in 1993, and soon people with an HD CAG repeat expansion born in the same year of the discovery of the *HTT* CAG repeat expansion could be exhibiting HD symptoms. Treatments need to be found for these patients. Developing a more complete picture of the mechanisms underlying *mHTT* CAG toxicity is a vital part of that goal. Current treatment methods are focusing on knocking down levels of the sense strand of *HTT* RNA. However, polyLeu is produced from the antisense strand of *mHTT.* Therefore, suppression of only the sense strand and sense strand RAN peptides, including polyQ, may not be sufficient to protect these patients from developing HD.

Every HD RAN polypeptide, except for polyGln, exhibited toxicity in the *C. elegans* RAN model when expressed at a disease-relevant repeat length. PolySer, polyCys, polyAla, and polyGln all formed aggregates. PolyLeu caused strong toxicity in both motor neurons and in muscle cells where it caused larval arrest. A forward mutagenesis screen and RNAi screen, along with impaired GFP fluorescence data, suggest that polyLeu may be toxic through disrupting folding of transmembrane proteins in the ER and activating ERAD and the UPR^{ER}. This is consistent with previous “polyQ” models where activation of ERAD was the first stress response detected [370-372]. It was proposed that the activation of ERAD was due to ERAD proteins which were detected in polyQ aggregates. As HTT is produced in the cytoplasm, it has puzzled researchers why it would cause a strong activation of ERAD. One possible explanation is that another RAN peptide, polyLeu, leads to ERAD activation. Additional studies will help to unravel the potential significance of polyLeu and other previously unknown HD RAN peptides.

3.4 Methods

Caenorhabditis elegans strains and culture

Strains were grown on standard NGM media with *gfp*(RNAi) bacteria at 20°C until the generation before the experiment. Animals were picked at the L4 staged and shifted to *E. coli* OP50 and allowed to have progeny at 20°C. The progeny were picked as L4s, kept on *E. coli* OP50 and grown at 25°C. All experiments were performed at 25°C.

Molecular Biology and Transgenics

Codon-varied sequences were synthesized for the 90 repeat polypeptides (Integrated DNA Technologies, Coralville, Iowa, USA). 38 repeat polypeptides were made using "building blocks" which were nucleotide codon-varied sequences for 11 repeats that were synthesized (GeneWiz, South Plainfield, NJ, USA). The "building blocks" could be added to each other by digesting the vector containing a building block with BsmbI and an insert containing a building block with BsaI as previously described [373]. The nucleotide sequences used for the HD polypeptides are listed in Appendix D (Table 4.3) Promoters were produced and cloned in as previously described [95].

Transgenic worms were generated by injecting the DPR construct (20 ng/μl) and the *myo-3p*::mCherry pCFJ104 marker plasmid (100 ng/μl) into the gonad of wild-type animals. Transgenes were integrated using a standard gamma ray (Cs137) mutagenesis, followed by selection of animals exhibiting 100% transmission of the mCherry marker. Integrated strains were outcrossed six times to wild-type animals. Injected animals were maintained on *gfp*(RNAi) plates until the experimental assay was performed. All

procedures involving recombinant or synthetic nucleic acid molecules and materials were approved by the University of Pittsburgh Institutional Biosafety Committee.

Reversal Assays

Animals were maintained on *gfp*(RNAi) at 20°C, until the generation before the experiment. Animals were selected as L4s and placed on *E. coli* OP50 at 20°C to have progeny. 30-40 progeny were picked as L4 animals and moved to 25°C on *E. coli* OP50 bacteria. 24 hours later, the progeny were tested for their ability to reverse (N=20 worms/assay). All assays were done blinded. Animals were lightly tapped on the head with a platinum pick. They were scored from 1-4 on their ability to reverse as previously described [374]. Each worm was scored 5 consecutive times, as wild-type worms began to acclimate to the head tapping at 7 consecutive times.

Commissure Assays

L4 animals of the indicated genotype were isolated at 25°C and imaged 24 hours later as 'Day 1 adults'. All of the strains contained an *unc-47p*::RFP marker to reveal GABAergic motor neuron morphology. Animals were anesthetized in 10 mM levamisole and Z-series images of GABAergic commissures were collected. Commissure breaks were identified as interruptions in the RFP signal surrounded by dorsal and ventral RFP in the commissures. Blebbing was scored only in the commissures and was identified by the presence of one or more RFP varicosities. Abnormal commissures were identified as those exhibiting branching and/or failing to reach the dorsal side of the animal.

Microscopy

C. elegans were anesthetized in 10 mM levamisole for 10 minutes. The animals were then picked onto 3% agarose pads for fluorescence microscopy. Widefield images were collected on a Leica DMi4000 inverted microscope and a Leica DFC 340x digital camera (Leica Microsystems, Wetzlar, Germany). Z-stack images were deconvolved using Leica AF6000 software.

COPAS Experiments

Animals grown at 20°C on (*gfp*)RNAi were placed on *E. coli* OP50 for 6 hours to collect a synchronized brood. The adults were then removed and the plates were placed at 25°C. 48 hours later, the progeny were sorted through a COPAS Biosorter (Union Biometrica, Holliston, MA, USA). Worm time of flight (TOF) was measured. Animals with a TOF≥300 were counted as adults, based on previous experiments. Animals with a TOF below 300 were counted as larvae. For fluorescent measurements (GFP or RFP) fluorescent detection settings were identical for all samples. Average fluorescence was measured as total fluorescence divided by the time of flight.

Paralysis Assays

Gravid animals were moved from *gfp*(RNAi) to *E. coli* OP50 and allowed to lay eggs for 24 hours at 20°C. The resulting progeny were allowed to grow up on *E. coli* OP50, permitting RAN peptide accumulation. Animals were picked as L4s and ten animals were placed on each of five plates (N=50 worms per assay). Each day, animals that failed to move at least half a body length in response to manual stimulation with a

platinum wire but were still alive (pharyngeal pumping, movement of less than half a body length) were scored as paralyzed. Animals that died, desiccated on the plate edges, or exhibited internal hatching of progeny were censored from the assay. Each day, mobile animals were transferred to a new plate and paralyzed, dead, and censored animals were removed from the assay.

Thrashing Assays

Gravid animals were moved from *gfp*(RNAi) to *E. coli* OP50 and allowed to lay eggs for 24 hours at 20°C. The resulting progeny were allowed to grow up on *E. coli* OP50, permitting polypeptide accumulation. 40 transgenic animals for each strain were picked as L4s and moved to an *E. coli* OP50 plate and placed at 25°C. The following day, worms were placed on clean NGM plates and allowed to move freely for 10 minutes so that most of the bacteria came off the animal before the experiment. Worms were placed individually into 3 cm petri dishes containing M9 buffer and allowed to adjust to the new environment for 5 minutes. The worms were observed for 30 seconds and the number of thrashes (reversal of body bend that crosses the midline) was counted.

Lifespan Assays

Worms grown at 20°C were picked as L4 and allowed to grow at 20°C until the next day, when 10 young adults were placed on five 3 cm plates. Lifespan assays were performed with *E. coli* OP50 spotted on NGM plates at 25°C. Worms were classified as alive, dead (no movement in response to touch with a wire), or censored (lost or bagged worms) once per day for lifespan assays.

FRAP Microscopy

Confocal fluorescence images were captured on a Nikon A1plus confocal microscope through an Apo 60x/ 1.4NA Oil objective lens. The microscope was operated on the NIS-Elements AR version 5.02 software platform. GFP or RFP were excited at 488 or 561, respectively. Images were captured every 11 seconds to avoid bleaching over the course of imaging. Following imaging of baseline fluorescence, a region of interest corresponding to a portion of the puncta was photobleached and fluorescence recovery within the photobleached area was monitored over at least 60 seconds. Data were normalized so that the image preceding the photobleach was set to 100% and the first image following the photobleach was set to zero percent. Imaging conditions over the time course of the experiment caused minimal loss of signal, suggesting an absence of photobleaching during the monitoring period. FRAP analysis was performed using Fiji software [375]. X-y drift was corrected using the "Correct 3d Drift" plugin [376].

ENU Suppressor Screen

myo-3p::(Leu)$_{38}$-GFP L4s were placed in 2.4mM ENU and then gently shaken for 4 hours at room temperature. The worms are then washed 5x with M9 and allowed to recover for 24 hours on NGM plates. After 24 hours, gravid P0s were picked and allowed to lay progeny on (*gfp*)RNAi plates for 6 hours. The P0s were then removed and the plates with the eggs were placed at 20°C. After a week, the starved progeny were washed onto *E. coli* OP50 and placed at 25°C. After a week, the worms still growing at 25°C were picked as possible candidates. To ensure independence of alleles, only one candidate

was selected per plate. The picked worms were allowed to propagate at 25°C. Worms that continued to propagate and had detectable RFP under the microscope were picked as suppressors.

Whole-Genome Sequencing

Genomic DNA was isolated from each of the suppressor lines using the PureGene Core Kit A according to the manufacturer's instructions. Genomic DNA libraries were prepared by BGI Genomics (Cambridge, MA USA) using Next-Generation Sequencing (HiSeq Xten). Deep sequencing reads were analyzed using Galaxy (usegalaxy.org). Both snps and indels were isolated as described in Appendix E. Only homozygous snps/indel which did not occur in the starting strain and were within genes were considered as possible causative mutations.

RNAi Screen

Bacteria were grown from the Ahringer or Orfeome library for the screen. Bacteria were seeded on (RNAi) NGM in 10cm petri dishes and allowed to grow for 3 days at room temperature. *myo-3p*::$(Leu)_{38}$-GFP animals were bleached as previously described [377], onto a clean NGM plate and left for 18 hours at 20°C so the progeny would synchronize by arresting as L1s. The progeny were washed off the plate with M9 media, and rinsed 2-3 times with M9 media to remove dauer pheromones. ~30 L1 animals were seeded onto each plate. The plates were placed at 25°C and left for 1 week and then blindly scored based on the presence or absence of moving larvae.

Statistical Analysis

Comparison of means were analyzed with ANOVA using Dunn's post-test analysis in GraphPadPrism 7 (Graphpad Software, Inc., LaJolla, CA, USA). The plateau was measured using one phase association in GraphPadPrism 7.

4.0 HD/ALS RAN Models Summary

We have created and characterized two different RAN peptide models, one for ALS-associated GGGGCC repeat expansion, and one for HD-associated CAG repeat expansions. The most toxic ALS associated dipeptides are PR and GR which localize to phase-separated compartments such as the nucleolus. The most toxic HD polypeptide is polyLeu, which interacts with membranes. While initially the mechanism of toxicity for these two different RAN products appears distinct, there may be similarities. An RNAi screen for genetic suppressors of PR toxicity identified multiple genes whose products likely phase separate, based on the prevalence of low complexity domains. However, proteins in the ubiquitin-proteasome pathway were also identified. PR appears to be toxic through its localization to phase-separated compartments and its interactions with the proteins localized there. Similarly, polyLeu may be toxic by localizing to lipid membranes and either inhibiting folding of proteins with transmembrane domains, or inhibiting the function of the transmembrane domains. An unbiased screen in polyLeu identified a regulator of ubiquitin-mediated protein degradation (*ubh-4*) as required for polyLeu toxicity. *ubh-4* expression can modify proteasome activity, suggesting that the proteasome may be important for suppressing polyLeu toxicity. Therefore, the cellular dysfunction caused by PR/GR and polyLeu may be due to their disruption of cellular compartments. The unbiased screens performed in the *C. elegans* ALS RAN model and the *C. elegans* HD RAN model have yielded molecular regulations of toxicity in both diseases.

Appendix A - Abbreviations Table

ABC transporter	ATP-binding cassette transporter
ALS	Amyotrophic Lateral Sclerosis
C	cytosine
C9	*Chromosome 9 open reading frame 72*
C9orf72	*Chromosome 9 open reading frame 72*
C. elegans	*Caenorhabditis elegans*
CCD	Cleidocranial dysplasia
CCHS	Congenital central hypoventilation
Cftr	cystic fibrosis transmembrane conductance regulator
DM1	myotonic dystrophy type 1
DM	myotonic dystrophy type 2
DMPK	DM1 protein kinase
DRPLA	dentatorubral-pallidoluysian atrophy
DPR	dipeptide repeat
E. coli	*Escherichia coli*
EGFR	epidermal growth-factor receptor
EMS	ethyl methanesulfonate
ENU	N-ethyl-N-nitrosourea
EPM1	Myoclonus progressive epilepsy of Unverricht and Lundborg
ER	endoplasmic reticulum
FECD	Fuchs endothelial corneal dystrophy
FMR1	fragile X mental retardation 1
FMR2	Fragile XE syndrome mental retardation
FRAP	fluorescent recovery after photobleaching
FTDA	Friedreich ataxia
FTD	frontotemporal dementia

FXS	fragile X syndrome
FXTAS	fragile X-associated tremor/ataxia syndrome
G	guanine
G_4C_2	GGGGCC
GABA	γ-aminobutryic acid
gGFP	glycosylatable GFP
GP	glycine-proline
GPCR	G-protein-coupled receptor
GPI	glycosylphosphatidylinositol
GR	glycine-arginine
HD	Huntington's disease
HDAC	histone deacetylase
HDL2	Huntington disease-like 2
HFG	Hand-foot-genital syndrome
HPE5	Holoprosencephaly 5
HTT	*huntingtin*
IR	Insulin receptor
JHD	Juvenile Huntington's disease
mHTT	*mutant huntingtin*
MRGH	mental retardation, growth hormone deficiency
N.S.	not significant
NES	nuclear exit signal
NLS	nuclear localizaiton signal
NDD	neurodegenerative disease
OPMD	oculopharyngeal muscular dystrophy
PA	proline-alanine
polyQ	polyGlutamine (encoded by CAG repeats)
polyAla	polyAlanine
polyCys	polyCysteine

polyGln	polyGlutamine (encoded by codon-varied repeats)
polyLeu	polyLeucine
polySer	polySerine
PR	proline-arginine
pre-mRNA	precursor messenger RNA
RAN	Repeat-Associated non-AUG Translation
ROS	reactive oxygen species
SBMA	spinal-bulbar muscular atrophy
SCA8	spinocerebellar ataxia type 8
SCA17	spinocerebellar ataxia type 17
sfGFP	superfolded GFP
SOD1	superoxide dismutase 1
SPD	syndactyly
UTR	untranslated region
VNC	ventral nerve cord
WGS	Whole-Genome Sequencing
XLMR	X-linked mental retardation

Appendix B – Repeat Expansion Diseases

Table 4.1 List of All Known Repeat Expansion Diseases

Disease	Repeat	Repeat Location	Encode	Suspected Mechanism of Toxicity	RAN Translation	Age Onset	NDD?
ALS/FTD	GGGGCC	Intron	--	RAN/RNA GOF	Yes	Yes	Yes
SBMA	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
SCA1	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
SCA2	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
SCA3	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
SCA6	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
SCA7	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
SCA8	CTG/CAG	UTR/Exon	PolyQ	RNA GOF/Protein GOF	Yes	Yes	Yes
HD	CAG	Exon	PolyQ	Protein GOF	Yes	Yes	Yes
HDL2	CTG	UTR	--	RNA GOF	?	Yes	Yes
SCA12	CAG	UTR	--	Unk.	?	Yes	Yes
SCA17	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
DRPLA	CAG	Exon	PolyQ	Protein GOF	?	Yes	Yes
DM1	CTG	UTR	--	RNA GOF	Yes	Yes	Yes
DM2	CTG/CCTG	Intron	--	RNA GOF	Yes	Yes	Yes
FXTAS	CGG	UTR	--	RAN GOF	Yes	Yes	Yes
SCA31	TGGAA	Intron	--	RAN GOF/RNA GOF	Yes	Yes	Yes
FRDA	GAA	Intron	--	LOF	?	Yes	Yes
SCA10	ATTCT	Intron	--	RNA GOF	?	Yes	Yes
EPM1	CCCCGCCCCGCG	Promoter	--	LOF	?	Yes	Yes
OPMD	GCG	Exon	polyAla	Protein GOF*	?	Yes	No

Table 4.1 (continued)

FECD	CTG	Intron	--	RNA GOF	Yes	Yes	No
FMR1	CGG	UTR	--	LOF	?	No	No
FMR1	CGG	UTR	--	LOF	?	No	No
XLMR	GCG	Exon	polyAla	LOF*	?	No	No
SPD	GCG	Exon	polyAla	protein GOF*	?	No	No
CCD	GCG	Exon	polyAla	LOF*	?	No	No
HPE5	GCG	Exon	polyAla	LOF*	?	No	No
HFG	GCG	Exon	polyAla	LOF*	?	No	No
FOXL2 (BPES)	GCG	Exon	polyAla	LOF*	?	No	No
Multiple Skeletal Dysplasia	GAC	Exon	polyAla	LOF	?	No	No
FRAXA	CGG	UTR	--	LOF	?	No	No
FRAXE	CCG	UTR		LOF	?	No	No
FRA11b Jacobsen syndrome	CCG	UTR	--	DNA GOF	?	No	No
CCHS	GCN	Exon	polyAla	Unk.*	?	No	No
MRGH	GCN	Exon	polyAla	LOF, polyAla	?	No	No

*=Other mutations in the same gene cause the same disease.

-- = Does not undergo canonical translation.

GOF=gain of function. LOF=loss of function. NDD=Neurodegenerative Disease

Unk.= Unknown

ALS=amyotrophic lateral sclerosis. CCD=Cleidocranial dysplasia. CCHS=Congenital Central Hypoventilation. FTD=frontotemporal dementia. HD=Huntington's disease. HDL2=Huntington disease-like 2. DM1=Myotonic Dystrophy Type 1. DM2=Myotonic

Dystrophy Type 2. DRPLA=Dentatorubral-Pallidoluysian Atrophy. EPM1=Myoclonus Progressive Epilepsy of Unverricht and Lundborg. FECD=Fuchs endothelial corneal dystrophy. FMR1=Fragile X Syndrome of mental retardation. FMR2=Fragile XE syndrome Mental retardation. FRDA=Friedreich Ataxia. FXTAS=Fragile X-Associated Tremor/Ataxia Syndrome. HFG=Hand-foot-genital syndrome. HPE5=Holoprosencephaly 5. MRGH=Mental retardation, growth hormone deficiency. OPMD=Oculopharyngeal Muscular Dystrophy. SBMA=spinal-bulbar muscular atrophy. SCA=spinocerebellar ataxia. SPD=Syndactyly. XLMR=X-linked mental retardation.

Appendix C – Colocalization Data

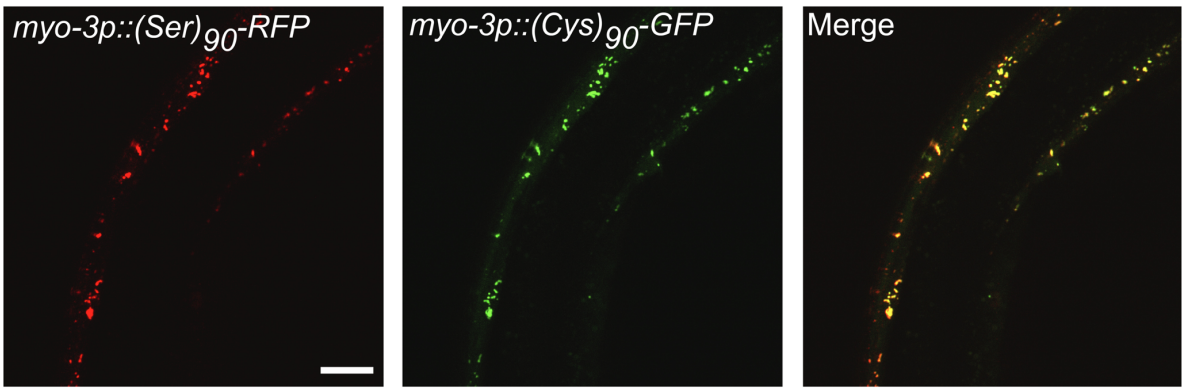


Figure 4.1 PolySer and PolyCys Colocalize

Day 1 adult animals grown at 20°C. Animals express *myo-3p::(Ser)$_{90}$-RFP* and *myo-3p::(Cys)$_{90}$-GFP.* PolySer and PolyCys exhibited strong colocalization. Immobilized in levamisole and imaged using a Leica DMI8 confocal microscope at the University of Pittsburgh Center for Biologic Imaging. Size bar is 20 μm.

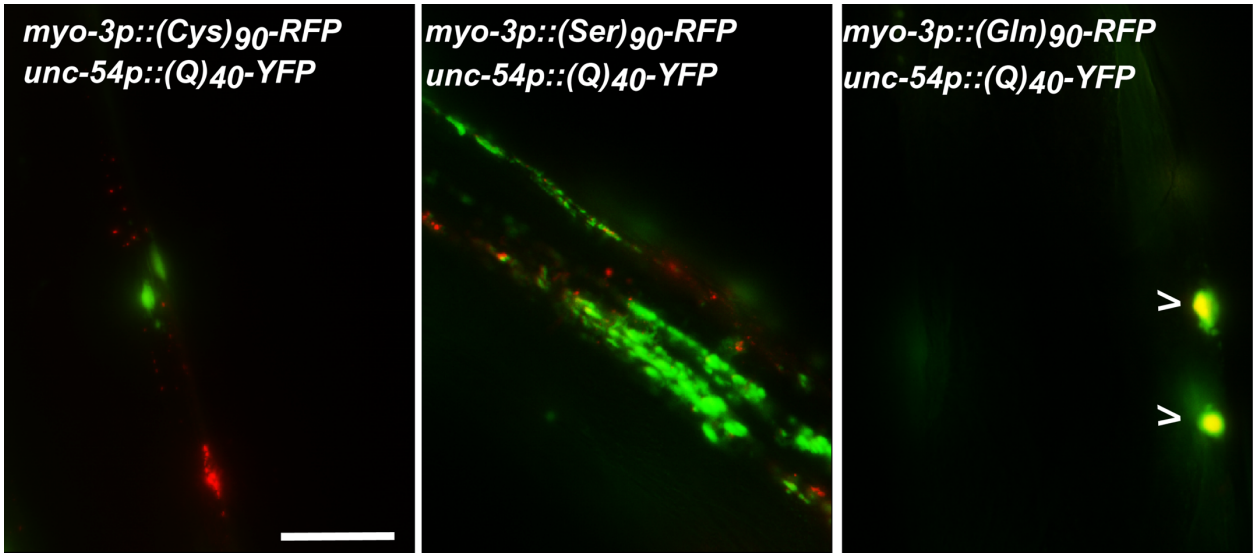


Figure 4.2 PolySer and PolyCys Do Not Colocalize with PolyQ

Day 1 adult animals grown at 20°C. AM141 animals (integrated with *unc-54p::(Q)$_{40}$-YFP*) express extrachromosomal *myo-3p::(Ser)$_{90}$-RFP*, *myo-3p::(Cys)$_{90}$-RFP, or myo-3p::(Gln)$_{90}$-RFP*. Codon-varied (Gln)$_{90}$ colocalizes with the CAG-repeat derived (Q)$_{40}$. Both (Ser)$_{90}$ and (Cys)$_{90}$ fail to colocalize with (Q)$_{40}$. The animals were immobilized in levamisole and imaged using a Leica DMI4000 inverted microscope and a Leica DFC 340Fx digital camera (Leica Microsystems, Wetzlar, Germany). The arrow points to colocalized signal. Size bar is 30 μm.

Appendix D - Plasmid Sequences

Table 4.2. Nucleotide Sequences of Codon-Varied DPRs

$(PA)_{50}$
aagcttgggccggccccggctccggcacctgcgcccgctccagcaccggcgcctgctcccgcaccagctcctgcacccgcgcca gcacccgctccagcgcctgctcctgcaccggctcccgcgccagcacctgctccggcgcccgcaccagctcctgccccggccccg gctccggcacctgcgcccgctccagcaccggcgcctgctcccgcaccagctcctgcacccgcgccagcacccgctccagcgcct gctcctgcaccggctcccgcgccagcacctgctccggcgcccgcaccagctcctgccccgggttggtaccgagctc**ggatcc**
$(GA)_{50}$
aagcttcccggggccggtgccggagcaggcgctggtgcgggcgccggagctggtgcaggcgcgggagccggtgctggcgcag gagcgggtgcaggcgcgggtgccggagctggcgcaggagcgggtgctggcgccggagcaggtgcgggcgctggggccgggg ccggtgccggagcaggcgctggtgcgggcgccggagctggtgcaggcgcgggagccggtgctggcgcaggagcgggtgcag gcgcgggtgccggagctggcgcaggagcgggtgctggcgccggagcaggtgcgggcgctggggccggctggtaccgagctc**g gatcc**
$(PR)_{50}$
aagcttggccggccccggccgcgtcctcgcccacgaccgagacctcgtccacgcccgcgacctagaccacggccgcgccctcg accaagacctcgcccgagaccacgacctcgtccgcgcccaagacctcgaccgcgtccacgccctagaccgcgaccccggccc cggccgcgtcctcgcccacgaccgagacctcgtccacgcccgcgacctagaccacggccgcgccctcgaccaagacctcgccc gagaccacgacctcgtccgcgcccaagacctcgaccgcgtccacgccctagaccgcgaccccgggggtaccgagctc**ggatc c**
$(PR)_{25}$
aagcttccccggccgcgtcctcgcccacgaccgagacctcgtccacgcccgcgacctagaccacggccgcgccctcgaccaag acctcgcccgagaccacgacctcgtccgcgcccaagacctcgaccgcgtccacgccctagaccgcgaccccgggggtaccgag ctc**ggatcc**
$(PR)_{15}$
aagcttccgcgccctcgaccaagacctcgcccgagaccacgacctcgtccgcgcccaagacctcgaccgcgtccacgccctag accgcgaccccgggggtaccgagctc**ggatcc**
$(PR)_{5}$
aagcttccgcgtccacgccctagaccgcgaccccgggggtaccgagctc**ggatcc**
$(GR)_{50}$
aagcttgcccggggccggggccgtggtcgcggacgaggcagaggtcgtggacgcggccgaggtagaggacgtggccgcggt cgaggaagaggtcgcggcagaggacgaggtcgtggccgcggaagaggtcgaggccgtggacgcggtagaggccgaggccg gggccggggccgtggtcgcggacgaggcagaggtcgtggacgcggccgaggtagaggacgtggccgcggtcgaggaagag gtcgcggcagaggacgaggtcgtggccgcggaagaggtcgaggccgtggacgcggtagaggccgaggccggcggtaccgag ctc**ggatcc**

Red-*HindIII* and *BamHI* cloning sites

Table 4.3 Nucleotide Sequences of Codon-Varied HD RAN Polypeptides

$(Ala)_{38}$
aagcttcgggtctcagcagccgcagcggccgcagcagccgcggcagccgcagcggccgcagcagccgcggca gccgcagcggccgcagcagccgcggcagccgcagcggccgcagcagccgcggcagcgagac**ggatcc**
$(Cys)_{37}$
aagcttcgggtctcatgttgctgttgctgctgttgctgttgttgttgctgttgctgctgttgctgttgttgttgctgttgctgctgttgct gttgttgttgctgttgctgctgttgctgttgttgtttgagac**ggatcc**
$(Ser)_{38}$
aagcttcgggtctcatcttcctcatcgtcctcatcgtcttcctcttcctcatcgtcctcatcgtcttcctcttcctcatcgtcctcat cgtcttcctcttcctcatcgtcctcatcgtcttcctctttgagac**ggatcc**
$(Leu)_{38}$
aagcttcgggtctcattattgcttctcctgttacttctcctattattgcttctcctgttacttctcctattattgcttctcctgttacttctc ctattattgcttctcctgttacttctcctattattgagac**ggatc**
$(Gln)_{38}$
aagcttcgggtctcacagcaacaacagcagcaacagcagcaacagcaacaacagcagcaacagcagcaacag caacaacagcagcaacagcagcaacagcaacaacagcagcaacagcagcaacagcagagac**ggatc**
$(Q)_{38}$
aagcttcgggtctcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagagac**ggatc**
$(Leu)_{11}$
aagcttcgggtctcattattgcttctcctgttacttctcctattattgagac**ggatc**
$(Leu)_{20}$
aagcttcgggtctcattattgcttctcctgttacttctcctattattgcttctcctgttacttctcctattattgagac**ggatc**
$(Leu)_{29}$
aagcttcgggtctcattattgcttctcctgttacttctcctattattgcttctcctgttacttctcctattattgcttctcctgttacttctc ctattattgagac**ggatc**
$(Leu)_{90}$
aagcttttattgcttctcctactgttattgcttctcctactgttattgcttctcctactgttattgcttctcctactgttattgcttctccta ctgttattgcttctcctactgttattgcttctcctactgttattgcttctcctactgttattgcttctcctactgttattgcttctcctactgtt attgcttctcctactgttattgcttctcctactgttattgcttctcctactgttattgcttctcctactgttattgcttctcctactgc**gga tc**
$(Cys)_{90}$
aagctttgttgctgttgctgctgttgttgctgttgctgttgctgttgctgctgttgttgctgttgctgttgctgttgctgctgttgttgctg ttgctgttgctgttgctgctgttgttgctgttgctgttgctgttgctgctgttgttgctgttgctgttgctgctgttgctgttgctgttgctgttg ctgttgctgttgctgctgttgttgctgttgctgttgctgttgctgctgttgttgctgttgctgttgctgttgctgctgttgttgctgttgcc **ggatc**
$(Ser)_{90}$
aagctttcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcg tcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcct catcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgtcttcctcatcgt cttccc**ggatc**

Table 4.3 (continued)

$(Gln)_{90}$
aagcttcaacagcagcaacaacagcaacagcagcaacagcaacaacagcaacagcagcaacagcaacaaca gcagcaacaacagcaacagcagcaacagcaacaacagcaacagcagcaacagcaacaacagcagcaacaa cagcaacagcagcaacagcaacaacagcaacagcagcaacagcaacaacagcagcaacaacagcaacagc agcaacagcaacaacagcaacagcagcaacagcaacagcaacaacagcaacagcagcaacagcaac**ggat c**

Red-*HindIII* and *BamHI* cloning sites

Table 4.4 Promoter Sequences and Targeting Sequences

***Myo-3p* - FLAG**
aagcttgtttgatgaaaaccaatgaaacaagtgattatagtctctgttttcgttaattttgaattttgcttgataaggctgcaac aaagatcaggttgacatatattttcagtaatttattttaacctgtactctatcactgccggctataataagttcttgaataaaat aatttttcccgacaaaacatgagtatttctttcgaaaataaaagtgcaggctaattagagattattctgtaattaactgcataa tttgtcacgtgccatagttttacattccactacgtcatagttcttaaaatactaatctcctgaaaatagaagtaggtgaagaa agtttaattatcagttctaaaatgacaattgatctttggaatatgttctgaaactaccgatcattgaacagatgctatttgaatg atatagaattgtatatttgcaatttctgaaacgcgttcttaaaggcacacagattaattcaaaagggtctggccgcaaaaa ggtttatggtggccgattttgagttttgtgtgtgattgctttttcacaatcagtgttttcaggattatgtgatgaactagatcttcaa gtttcgttacatttcatatgttttcggaactcacgaagtacatattgggtattgtgctcaaaaaattcagcaatcagcttcgctc cgctgactttagaacccaaaaaaatagtatggccaaactgactgtgttacgatcatttcaatttttcaatacatatttaagat ttctaagagtaagaaggtcaaaaactgttctggaatacatatatatttttcaggttacaattagtcaaaaagtgcactgaa atatacgttttaatttcacgaataacccaattagttcaatgtatttttggtcaaccaacgttaaagtttggcttccaaccaattat catttctgatcaaccacaatgttttttctttatctgcaagttaatttttattttatccagatgtttggcatatttttcaattcttcactag cgcccacttcttgcacttccggcgccctgaatctaatgcatctgttgcaagaattgaaagaccaatcaacacattgttttctt cacgagatactgaagaaaatgaataaaaacagagaaaaagagccatgtgattagtgacaactgttgctaacagata ccatagcttggacttggtacgtgatggcaacgtatgggtcaacaaaaatgattgcagaggggtgcaaaacagtcaa gtcgagaaaatatgaaaaacagaaaacaaagaacagaaaaatgggtttgagagtcagtataatttataaaagaaa aattgtacatagaaattaaccatttttgtagaagaagttatttttcaagcatcgttaaaaattattcaaagcaccttatttcatat ttaattttaaacatggttaaatgaacaacacggtgcgcaatcaggaaaacttgaaatctgaaactgttgttgtgatcttcttc gcaactgttcagatagcactagtgtaatgttaagagtgcgcgaatataatggaatataatggatcacacctcctgccatc aggtaaacgtctctgttatcacatatttccaactattaaatttttaccttttacagttttacatttttttgaaaaaagtaacttttttgtc ttcaaaatccctgacgaaaatatcaaatattttaatcgagactgcagaggaaccgattgatgatttggaaaatccagcttt acctgtgtaagaactgaaaagtttcataaccctagggtattcccagttacattccccactggctaacaatagcacccagtt tttcatcacctttcttcaaatttctcggcgatttgttaaaaacaaaatttgtgtcccttctctgatatctctatgtctctaaacacaa gttcatcggaaaacgaaggagggtaggtgttggttgggctcccgaagtgaaaatagaagagcaagaatagaatatta gagagagagtgcagagagggcgggatagctcccgggattccgttttcttcttctttatcttcaacgatgatgtgtgtgcgtg ttgtatagattctgttgctccccccacaactcgctccgaaggctcaatacaattcaattgatattggaggagagcctaccgg agtgggaggataagaagaaacataagaagaagaagaagaagaagcatgcttctggtttttgatgctatgaaaacgg cacaaaaagatgattgaggtccctttttcaatacttctctcatctttcaaatcccattgaaacctaaaacttctcaccacgct ttaccattgttctccaaaaacttatagcaatgtctataactttttttatctctgaaaagcagtgttccatttttctttttcctatttttatttc aattgtttctcacatttcgtttggattctttgcttgtcaaccagcttcttcttccacttttaccgtctaattttcagggcagggagcc atcaaacccacgaccactagatccatctagaaATG<u>gattacaaggatgacgacgataag</u>**aagctt**
***Myo-3p*-signal sequence-transmembrane-FLAG**
aagcttgtttgatgaaaaccaatgaaacaagtgattatagtctctgttttcgttaattttgaattttgcttgataaggctgcaac aaagatcaggttgacatatattttcagtaatttattttaacctgtactctatcactgccggctataataagttcttgaataaaat aatttttcccgacaaaacatgagtatttctttcgaaaataaaagtgcaggctaattagagattattctgtaattaactgcataa tttgtcacgtgccatagttttacattccactacgtcatagttcttaaaatactaatctcctgaaaatagaagtaggtgaagaa agtttaattatcagttctaaaatgacaattgatctttggaatatgttctgaaactaccgatcattgaacagatgctatttgaatg atatagaattgtatatttgcaatttctgaaacgcgttcttaaaggcacacagattaattcaaaagggtctggccgcaaaaa ggtttatggtggccgattttgagttttgtgtgtgattgctttttcacaatcagtgttttcaggattatgtgatgaactagatcttcaa gtttcgttacatttcatatgttttcggaactcacgaagtacatattgggtattgtgctcaaaaaattcagcaatcagcttcgctc

Table 4.4 (continued)

cgctgactttagaacccaaaaaaatagtatggccaaactgactgtgttacgatcatttcaatttttcaatacatatttaagat ttctaagagtaagaaggtcaaaaactgttctggaatacatatatatttttcaggttacaattagtcaaaaagtgcactgaa atatacgttttaatttcacgaataacccaattagttcaatgtatttttggtcaaccaacgttaaagtttggcttccaaccaattat catttctgatcaaccacaatgttttttctttatctgcaagttaatttttattttatccagatgtttggcatatttttcaattcttcactag cgcccacttcttgcacttccggcgccctgaatctaatgcatctgttgcaagaattgaaagaccaatcaacacattgttttctt cacgagatactgaagaaaatgaataaaaacagagaaaaagagccatgtgattagtgacaactgttgctaacagata ccatagcttggacttggtacgtgatggcaacgtatgggtcaacaaaaatgattgcagaggggtgcaaaacagtcaa gtcgagaaaatatgaaaaacagaaaacaaagaacagaaaaatgggtttgagagtcagtataatttataaaagaaa aattgtacatagaaattaaccatttttgtagaagaagttatttttcaagcatcgttaaaaattattcaaagcaccttatttcatat ttaattttaaacatggttaaatgaacaacacggtgcgcaatcaggaaaacttgaaatctgaaactgttgttgtgatcttcttc gcaactgttcagatagcactagtgtaatgttaagagtgcgcgaatataatggaatataatggatcacacctcctgccatc aggtaaacgtctctgttatcacatatttccaactattaaatttttaccttttacagttttacatttttttgaaaaaagtaacttttttgtc ttcaaaatccctgacgaaaatatcaaatattttaatcgagactgcagaggaaccgattgatgatttggaaaatccagcttt acctgtgtaagaactgaaaagtttcataaccctagggtattcccagttacattccccactggctaacaatagcacccagtt tttcatcacctttcttcaaatttctcggcgatttgttaaaaacaaaatttgtgtcccttctctgatatctctatgtctctaaacacaa gttcatcggaaaacgaaggagggtaggtgttggttgggctcccgaagtgaaaatagaagagcaagaatagaatatta gagagagagtgcagagagggcgggatagctcccgggattccgttttcttcttctttatcttcaacgatgatgtgtgtgcgtg ttgtatagattctgttgctcccccacaactcgctccgaaggctcaatacaattcaattgatattggaggagagcctaccgg agtgggaggataagaagaaacataagaagaagaagaagaagaagcatgcttctggtttttgatgctatgaaaacgg cacaaaaagatgattgaggtccctttttcaataccttctctcatctttcaaatcccattgaaacctaaaacttctcaccacgct ttaccattgttctccaaaaacttatagcaatgtctataacttttttatctctgaaaagcagtgttccatttttcttttttcctatttttatttc aattgtttctcacatttcgtttggattctttgcttgtcaaccagcttcttcttccacttttaccgtctaattttcagggcagggagcc atcaaacccacgaccactagatccatctagaaATG***ccaccttcaacatcattgctgctcctcgcagcacttcttc cattcgctttaccagcaagcgattggaagactggagaagtcactccggataatgcaacagtctgggtcagaa aacataaagattgtcctccacctgtccctgtgctcgcaattgtgctcggagtcattgcgggtatcgtaatcctcg gaattcttctcttgttgctctggaaattgctcacagtacttcatgatagatccgaggtaccggtagaaaaa***gatta caaggatgacgacgataag**aagctt**
***Myo-3p*-GFP-FLAG-his-58**
aagcttgtttgatgaaaaccaatgaaacaagtgattatagtctctgttttcgttaattttgaattttgcttgataaggctgcaac aaagatcaggttgacatatattttcagtaatttattttaacctgtactctatcactgccggctataataagttcttgaataaaat aattttcccgacaaaacatgagtatttctttcgaaaataaaagtgcaggctaattagagattattctgtaattaactgcataa tttgtcacgtgccatagttttacattccactacgtcatagttcttaaaatactaatctcctgaaaatagaagtaggtgaagaa agtttaattatcagttctaaaatgacaattgatctttggaatatgttctgaaactaccgatcattgaacagatgctatttgaatg atatagaattgtatatttgcaatttctgaaacgcgttcttaaaggcacacagattaattcaaaagggtctggccgcaaaaa ggtttatggtggccgattttgagttttgtgtgtgattgctttttcacaatcagtgttttcaggattatgtgatgaactagatcttcaa gtttcgttacatttcatatgttttcggaactcacgaagtacatattgggtattgtgctcaaaaaattcagcaatcagcttcgctc cgctgactttagaacccaaaaaaatagtatggccaaactgactgtgttacgatcatttcaatttttcaatacatatttaagat ttctaagagtaagaaggtcaaaaactgttctggaatacatatatatttttcaggttacaattagtcaaaaagtgcactgaa atatacgttttaatttcacgaataacccaattagttcaatgtatttttggtcaaccaacgttaaagtttggcttccaaccaattat catttctgatcaaccacaatgttttttctttatctgcaagttaatttttattttatccagatgtttggcatatttttcaattcttcactag cgcccacttcttgcacttccggcgccctgaatctaatgcatctgttgcaagaattgaaagaccaatcaacacattgttttctt cacgagatactgaagaaaatgaataaaaacagagaaaaagagccatgtgattagtgacaactgttgctaacagata ccatagcttggacttggtacgtgatggcaacgtatgggtcaacaaaaatgattgcagaggggtgcaaaacagtcaa

Table 4.4 (continued)

gtcgagaaaatatgaaaaacagaaaacaaagaacagaaaaatgggtttgagagtcagtataatttataaaagaaa aattgtacatagaaattaaccatttttgtagaagaagttatttttcaagcatcgttaaaaattattcaaagcaccttatttcatat ttaattttaaacatggttaaatgaacaacacggtgcgcaatcaggaaaacttgaaatctgaaactgttgttgtgatcttcttc gcaactgttcagatagcactagtgtaatgttaagagtgcgcgaatataatggaatataatggatcacacctcctgccatc aggtaaacgtctctgttatcacatatttccaactattaaatttttaccttttacagttttacatttttttgaaaaaagtaacttttttgtc ttcaaaatccctgacgaaaatatcaaatattttaatcgagactgcagaggaaccgattgatgatttggaaaatccagcttt acctgtgtaagaactgaaaagtttcataaccctagggtattcccagttacattccccactggctaacaatagcacccagtt tttcatcacctttcttcaaatttctcggcgatttgttaaaaacaaaatttgtgtcccttctctgatatctctatgtctctaaacacaa gttcatcggaaaacgaaggagggtaggtgttggttgggctcccgaagtgaaaatagaagagcaagaatagaatatta gagagagagtgcagagagggcgggatagctcccgggattccgttttcttcttctttatcttcaacgatgatgtgtgtgcgtg ttgtatagattctgttgctcccccacaactcgctccgaaggctcaatacaattcaattgatattggaggagagcctaccgg agtgggaggataagaagaaacataagaagaagaagaagaagaagcatgcttctggtttttgatgctatgaaaacgg cacaaaaagatgattgaggtccctttcaataccttctctcatctttcaaatcccattgaaacctaaaacttctcaccacgct ttaccattgttctccaaaaacttatagcaatgtctataactttttatctctgaaaagcagtgttccatttttctttttcctatttttatttc aattgtttctcacatttcgtttggattctttgcttgtcaaccagcttcttcttccacttttaccgtctaattttcagggcagggagcc atcaaacccacgaccactagatccatctagaaATGagtaaaggagaagaacttttcactggagttgtcccaattcttgt tgaattagatggtgatgttaatgggcacaaattttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttа cccttaaatttatttgcactactggaaaactacctgttccatgggtaagtttaaacatatatatactaactaaccctgattattt aaattttcagccaacacttgtcactactttctgttatggtgttcaatgcttctcgagatacccagatcatatgaaacggcatg acttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatattttcaaagatgacgggaactacaagac acgtaagtttaaacagttcggtactaactaaccatacatatttaaattttcaggtgctgaagtcaagtttgaaggtgatacc cttgttaatagaatcgagttaaaaggtattgattttaaagaagatggaaacattcttggacacaaattggaatacaactat aactcacacaatgtatacatcatggcagacaaacaaaagaatggaatcaaagttgtaagtttaaacatgattttactaa ctaactaatctgatttaaattttcagaacttcaaaattagacacaacattgaagatggaagcgttcaactagcagaccatt atcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtccacacaatctgccctttcga aagatcccaacgaaaagagagaccacatggtccttcttgagtttgtaacagctgctgggattacacatggcatggatga actatacaaagattacaaggatgacgacgataag**atgccaccaaagccatctgccaagggagccaagaaggc** **cgccaagaccgtcgttgccaagccaaaggacggaaagaagagacgtcatgcccgcaaggaatcgtactc** **cgtctacatctaccgtgttctcaagcaagttcacccagacaccggagtctcctccaaggccatgtctatcatga** **actccttcgtcaacgatgtattcgaacgcatcgcttcggaagcttcccgtcttgctcattacaacaaacgctcaa** **cgatctcatcccgcgaaattcaaaccgctgtccgtttgattctcccaggagaacttgccaagcacgccgtgtct** **gagggaaccaaggccgtcaccaagtacacttccagcaagaagctt**
***Unc-47p*-FLAG**
aagcttcgtgcatcttcaaagatttggtataattatgtagagtgaaaaaatagtggaaaatagataaatttttgaaaatatt tgggctccatccagagtccatttttccggagtccactcaaatagttttcagaaaaattggacttgaaagatgtagagctttgt tgcatacagaattagaataattaaaaattgggtagaacattttcttcgtaaattttcagagttaatattggcaaagactgtttg atttgcacttttgactggaaaaaaaggctatagtccacagaaaactattttccagaaagtttatttttttccaaatttttaattgta caataaagaattaacatgaagatggaaaatgctcaattttcagtaaaaaattacaaaaatatctataaacaagtaaact actacaactaccaaaaagaatgaaaacaaacaaattggaagtaatgcgaaacaaaaaatgaatgagcgatctgct gagctctctctctctctctttccatctttctctctatatcttcatcttttgctagcagccacttttggtgtgaagggtgtagaaga aagagaaagagagaccgaaacataaagacatagacaacacggtcttcctccacttctatacacactcttctgcttcttct tctcctgctggctttcaagcaacggacacggccgaaagcagaagccgtgcagaaaaaaagaagaagaagaaaaa

Table 4.4 (continued)

agaaaagagctgtggccacacaaacaaatggatgttgtctctgtagcaatggcggttggctccgcgagagcgccgga
ggggaaattgggagacaaaaagttaatgcacttgacaatacagcgatataattcagatttactgtgacgggaagattgt
tgacatggggaataggaagaatgaagatctcaacgttttcggaacagttgtacgatatgattttccagggtactaaaaca
ttaaaactgtcaaaggggtacagcatgttgtcatcacttcaaacttttcaatgcataaatttcttgtcagatctgatcctctata
cttgaagtttataccgataaaggttttcaacagattttgaagcatttgaatcaggggtgcctgggtaaacctatcatacgaa
actgccaatttgtcctgtgaatcgtatctattaaaacaagactgaaaataactaaacttctacgtcaaaaagttgacaaa
acaactttcttggaaaaaaagctgttgggatcccggaacagtcgaaagtcggtggcaagcgccgaactgctgacggt
ctaaccggggcacaaatcaggggtgagcggcaaacgatttttccggcaaatcggcaaatcggcaaattgccaatatt
gaaatacccggcaaatcggtaaatagccggaattgaaaatttccggcaaactggtaaaccgcaaattgctgatttgcc
gaatttgccgggaagacggcaattgccaaacatattcggcaaattgtggttttgcacttttttggaaatttcagaatttcaatc
ggcaaattgtgcacatcctatgaattttcctacatctattttgaaaagtaagcaaattctatgaaaatatctaaagaaaaat
ggaaaaaattttcaaaaaggcacagttttaagtgtttccgtctaataaaaaaatccccctaaacacttccggcaaattgat
gttcggcaaatggcaaatcggaaacttgccgaaaattacagtttccggtaaatcggcaaaccggcaaactgcctgaat
tgaaaagttccgtcaaatcggcaaaccgacaacacccctggcacaaatgatggacatactgaggcaatttgccggttt
tccaattgcaggaaattttcaattccggcagtgtgccgatttgccggaaattttaattcaggcaaattgccgatttcccgattt
cccgatttgccggaaaaaatcgtttgccgcccacccctgggtctgaaccttgattgttacaaaacatttttagctctttggag
aaataaaatgaatctcgtaaaatttaattgacgaggacgatattagctgtctctttagaccaaattcagaaaaaaagaa
agaatacttcccaaatttccggtccctctctcgttttttttgccaataaactcactatagtcgctggttcccccctattcacattta
ttctaccaatccatcagtggaaccagaaaaaagaagagcctttcggtttggagagtagggtctaataatcccccgtgct
cttcaaatcattgtgccaacacacagacacactttatgtgtgctcacacacacacgctatttgaagagcgaagacgacg
acgacgcattcagagctcttttccacgaaatttgctccatctttccacaatctgtctttcctgtgagacgacagcgtcacattt
atttcattacagATGgattacaaggatgacgacgataag**aagctt**

Red – HindIII cloning sites
Underline – FLAG sequence
ATG – start codon
Green – GFP sequence
Bold italics – Signal sequence-transmembrane domain
Bold – *his-58* sequence

Appendix E – Galaxy Workflow

Workflow and restraints used to extract novel snps and indels from sequencing information on Galaxy (https://usegalaxy.org)

1. Input datasets
    a. *C. elegans* FASTA reference genome
    b. FASTQ forward reads
    c. FASTQ reverse reads
    d. SnpEff database

2. Map with BWA for Illumina
    a. Library is paired-end
    b. Maximum edit distance (aln -n)=0
    c. Fraction of missing alignments given 2% uniform base error rate (aln -n)=0.04
    d. Maximum number of gap opens (aln -o)=0
    e. Maximum number of gap extensions (aln -e)=-1
    f. Disallow long deletion within 16 bp towards the 3' end (aln -d)
    g. Disallow insertion/deletion within 5 bp towards the end (aln -i)
    h. Number of first subsequences to take as seed (aln -l)=-1
    i. Maximum edit distance in the seed (aln -k)= 2
    j. Mismatch penalty (aln -M)=3
    k. Gap open penalty (aln -O)=11
    l. Gap extension penalty (aln -E)=4
    m. Disable iterative search (aln -N)=False
    n. Maximum number of alignments to output in the XA tag for reads paired properly (samse/sampe-n)=3
    o. Maximum number of alignments to output in the XA tag for disconcordant read pairs (excluding singletons (sampe -N))=10
    p. Maximum insert size for a read pair to be considered as being properly (sampe-a)=500
    q. Maximum occurrences of a read for pairing (sampe -o)=100000

3. Filter Sequence Alignment/Map (SAM)
    a. The read is unmapped
    b. Do not set the states for this flag

4. SAM to Binary Alignment/MAP (BAM)

5. Add or Replace Read Groups

6. Unified Genotyper
    a. Binding for reference-ordered datas
        i. Identify SNPs
        ii. Minimum phred-scaled confidence threshold at which variants not at ‘trigger’ track sites should be called=30.0
        iii. Minimum phred-scaled confidence threshold at which variants not at ‘trigger’ track sites should be emitted=30.0
        iv. Basic GATK options and basic analysis options
7. Identify Snps
    a. Input/output format=VCF
    b. Genetic code for sequence= standard
    c. No upstream/downstream intervals (0 bases)
    d. Do not show downstream/intergenic/intron/upstream/5’UTR/3’UTR changes
    e. Filter out specific effects
    f. Filter out and do not report
        i. Synonymous coding
        ii. Synonymous start
        iii. Synonymous stop

Bibliography

1. Collaborators, G.B.D.N., *Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016.* The Lancet Neurology, 2019. **18**(5): p. 459-480.

2. *World Population Ageing 2017- Highlights*, in *United Nations*, P.D. Department of Economic and Social Affairs, Editor. 2017.

3. de Koning, A.P.J., et al., *Repetitive elements may comprise over two-thirds of the human genome.* PLoS genetics, 2011. **7**(12): p. e1002384-e1002384.

4. Zu, T., et al., *Non-ATG–initiated translation directed by microsatellite expansions.* Proc Natl Acad Sci U S A, 2011. **108**(1): p. 260-265.

5. Chung, D.W., et al., *A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression.* Human Molecular Genetics, 2011. **20**(17): p. 3467-3477.

6. Zu, T., 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**: p. E4968-77.

7. Cho, D.H., et al., *Antisense Transcription and Heterochromatin at the DM1 CTG Repeats Are Constrained by CTCF.* Molecular Cell, 2005. **20**(3): p. 483-489.

8. DeJesus-Hernandez, M., et al., *Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS.* Neuron, 2011. **72**(2): p. 245-256.

9. Renton, Alan E., et al., *A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD.* Neuron, 2011. **72**(2): p. 257-268.

10. Zhang, Y.-J., et al., *Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity.* Science, 2019. **363**(6428): p. eaav2606.

11. Rudich, P., Lamitina, Todd, *Models and Mechanisms of Repeat Expansion Disorders: A Worm's Eye View.* Journal of Genetics, 2018. **97**: p. 667-677.

12. Bañez-Coronel, M., et al., *RAN Translation in Huntington Disease.* Neuron, 2015. **88**: p. 667-77.

13. Saccon, R.A., et al., *Is SOD1 loss of function involved in amyotrophic lateral sclerosis?* Brain 2013. **136**(Pt 8): p. 2342-2358.

14. Lin, Y., et al., *R loops stimulate genetic instability of CTG.CAG repeats.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(2): p. 692-697.

15. Pearson, C.E., et al., *Slipped‐strand DNAs formed by long (CAG)·(CTG) repeats: slipped‐out repeats and slip‐out junctions.* Nucleic Acids Research, 2002. **30**(20): p. 4534-4547.

16. Mosbach, V., L. Poggi, and G.-F. Richard, *Trinucleotide repeat instability during double-strand break repair: from mechanisms to gene therapy.* Current Genetics, 2019. **65**(1): p. 17-28.

17. Williams, G.M. and J.A. Surtees, *MSH3 Promotes Dynamic Behavior of Trinucleotide Repeat Tracts In Vivo.* Genetics, 2015. **200**(3): p. 737.

18. Lee, J.-M., et al., *Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease.* Cell, 2015. **162**(3): p. 516-526.

19. Cohen, H., et al., *Increased instability of human CTG repeat tracts on yeast artificial chromosomes during gametogenesis.* Molecular and cellular biology, 1999. **19**(6): p. 4153-4158.

20. Brocklebank, D., et al., *Repeat instability in the 27-39 CAG range of the HD gene in the Venezuelan kindreds: Counseling implications.* American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics, 2009. **150B**(3): p. 425-429.

21. Sundararajan, R. and C.H. Freudenreich, *Expanded CAG/CTG Repeat DNA Induces a Checkpoint Response That Impacts Cell Proliferation in Saccharomyces cerevisiae.* PLOS Genetics, 2011. **7**(3): p. e1001339.

22. Falush, D., et al., *Measurement of mutational flow implies both a high new-mutation rate for Huntington disease and substantial underascertainment of late-onset cases.* American journal of human genetics, 2001. **68**(2): p. 373-385.

23. Goldberg, Y.P., et al., *Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects.* Nature Genetics, 1993. **5**(2): p. 174-179.

24. Zühlke, C., et al., *Mitotic stability and meiotic variability of the (CAG)n repeat in the Huntington disease gene.* Human Molecular Genetics, 1993. **2**(12): p. 2063-2067.

25. Myers, R.H., et al., *Factors related to onset age of Huntington disease.* The American Journal of Human Genetics, 1982. **34**(3): p. 481-488.

26. van Blitterswijk, M., et al., *Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study.* The Lancet Neurology, 2013. **12**(10): p. 978-988.

27. Byrne, S., et al., *Intermediate repeat expansion length in C9orf72 may be pathological in amyotrophic lateral sclerosis.* Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 2014. **15**(1-2): p. 148-150.

28. Erzurumluoglu, E., et al., *The association between repeat number in C9orf72 and phenotypic variability in Turkish patients with frontotemporal lobar degeneration.* Neurobiology of Aging, 2019. **76**: p. 216.e1-216.e7.

29. Dols-Icardo, O., et al., *Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia.* Human Molecular Genetics, 2013. **23**(3): p. 749-754.

30. Swami, M., et al., *Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset.* Human Molecular Genetics, 2009. **18**(16): p. 3039-3047.

31. Telenius, H., et al., *Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm.* Nature Genetics, 1994. **6**(4): p. 409-414.

32. Verkerk, A.J., et al., *Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome.* Cell, 1991. **65**.

33. Spada, A.R.L., et al., *Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy.* Nature, 1991. **352**(6330): p. 77-79.

34. Mori, K., et al., *The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS.* Science, 2013. **339**: p. 1335-8.

35. Katayama, S., et al., *Antisense Transcription in the Mammalian Transcriptome.* Science, 2005. **309**(5740): p. 1564.

36. He, Y., et al., *The antisense transcriptomes of human cells.* Science (New York, N.Y.), 2008. **322**(5909): p. 1855-1857.

37. Matsui, K., et al., *Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes.* Hepatology, 2008. **47**(2): p. 686-697.

38. Hawkins, P.G. and K.V. Morris, *Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5.* Transcription, 2010. **1**(3): p. 165-175.

39. Beltran, M., et al., *A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition.* Genes & development, 2008. **22**(6): p. 756-769.

40. Faghihi, M.A., et al., *Evidence for natural antisense transcript-mediated inhibition of microRNA function.* Genome biology, 2010. **11**(5): p. R56-R56.

41. Yu, W., et al., *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.* Nature, 2008. **451**(7175): p. 202-206.

42. Mikaeili, H., et al., *FAST-1 antisense RNA epigenetically alters FXN expression.* Scientific Reports, 2018. **8**(1): p. 17217.

43. Lodato, M.A., et al., *Aging and neurodegeneration are associated with increased mutations in single human neurons.* Science, 2018. **359**(6375): p. 555.

44. Gilmer, L.K., et al., *Age-related changes in mitochondrial respiration and oxidative damage in the cerebral cortex of the Fischer 344 rat.* Mechanisms of ageing and development, 2010. **131**(2): p. 133-143.

45. David, D.C., et al., *Widespread protein aggregation as an inherent part of aging in C. elegans.* PLoS biology, 2010. **8**: p. e1000450.

46. Neumann, M., et al., *Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis.* Science, 2006. **314**(5796): p. 130.

47. Hara, T., et al., *Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice.* Nature, 2006. **441**(7095): p. 885-889.

48. Komatsu, M., et al., *Loss of autophagy in the central nervous system causes neurodegeneration in mice.* Nature, 2006. **441**(7095): p. 880-884.

49. Tashiro, Y., et al., *Motor Neuron-specific Disruption of Proteasomes, but not Autophagy, Replicates Amyotrophic Lateral Sclerosis.* The Journal of Biological Chemistry, 2012.

50. Geisler, S., et al., *PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1.* Nature Cell Biology, 2010. **12**: p. 119.

51. Shen, W.-C., et al., *Mutations in the ubiquitin-binding domain of OPTN/optineurin interfere with autophagy-mediated degradation of misfolded proteins by a dominant-negative mechanism.* Autophagy, 2015. **11**(4): p. 685-700.

52. Seo, H., et al., *Proteasome Activator Enhances Survival of Huntington's Disease Neuronal Model Cells.* PLOS ONE, 2007. **2**(2): p. e238.

53. Chondrogianni, N., et al., *20S proteasome activation promotes life span extension and resistance to proteotoxicity in Caenorhabditis elegans.* FASEB Journal, 2015.

54. Cummings, C.J., et al., *Over-expression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA1 mice.* Human Molecular Genetics, 2001. **10**(14): p. 1511-1518.

55. Andrew, S.E., et al., *The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease.* Nature Genetics, 1993. **4**(4): p. 398-403.

56. Nethisinghe, S., et al., *PolyQ Tract Toxicity in SCA1 is Length Dependent in the Absence of CAG Repeat Interruption.* Frontiers in Cellular Neuroscience, 2018. **12**: p. 200.

57. Li, J.-L., et al., *A Genome Scan for Modifiers of Age at Onset in Huntington Disease: The HD MAPS Study.* The American Journal of Human Genetics, 2003. **73**(3): p. 682-687.

58. Nicolas, G., et al., *Juvenile Huntington disease in an 18-month-old boy revealed by global developmental delay and reduced cerebellar volume.* American Journal of Medical Genetics Part A, 2011. **155**(4): p. 815-818.

59. Udd, B. and R. Krahe, *The myotonic dystrophies: molecular, clinical, and therapeutic challenges.* The Lancet Neurology, 2012. **11**(10): p. 891-905.

60. Reiner, A., I. Dragatsis, and P. Dietrich, *Genetics and neuropathology of Huntington's disease.* International Review of Neurobiology, 2011. **98**: p. 325-372.

61. Landwehrmeyer, G.B., et al., *Huntington's disease gene: Regional and cellular expression in brain of normal and affected individuals.* Annals of Neurology, 1995. **37**(2): p. 218-230.

62. McGeer, E.G. and P.L. McGeer, *Duplication of biochemical changes of Huntington's chorea by intrastriatal injections of glutamic and kainic acids.* Nature, 1976. **263**(5577): p. 517-519.

63. Schwarcz, R. and J.T. Coyle, *Neurochemical sequelae of kainate injections in corpus striatum and substantia nigra of the rat.* Life Sciences, 1977. **20**(3): p. 431-436.

64. Suzuki, N., et al., *The mouse C9ORF72 ortholog is enriched in neurons known to degenerate in ALS and FTD.* Nature Neuroscience, 2013. **16**: p. 1725.

65. Floeter, M.K., et al., *Disease progression in C9orf72 mutation carriers.* Neurology, 2017. **89**(3): p. 234-241.

66. Shu, L., et al., *The Association between C9orf72 Repeats and Risk of Alzheimer's Disease and Amyotrophic Lateral Sclerosis: A Meta-Analysis.* Parkinson's disease, 2016. **2016**: p. 5731734-5731734.

67. Hensman Moss, D.J., et al., *C9orf72 expansions are the most common genetic cause of Huntington disease phenocopies.* Neurology, 2014. **82**(4): p. 292-299.

68. Lorenzo, L.-E., et al., *Differential expression of GABAA and glycine receptors in ALS-resistant vs. ALS-vulnerable motoneurons: possible implications for selective vulnerability of motoneurons.* European Journal of Neuroscience, 2006. **23**(12): p. 3161-3170.

69. Tsvetkov, A.S., et al., *Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration.* Nature Chemical Biology, 2013. **9**: p. 586.

70. Costa, V. and L. Scorrano, *Shaping the role of mitochondria in the pathogenesis of Huntington's disease.* The EMBO journal, 2012. **31**(8): p. 1853-1864.

71. Kong, J. and Z. Xu, *Massive Mitochondrial Degeneration in Motor Neurons Triggers the Onset of Amyotrophic Lateral Sclerosis in Mice Expressing a Mutant SOD1.* The Journal of Neuroscience, 1998. **18**(9): p. 3241.

72. Nunomura, A., et al., *Oxidative Damage Is the Earliest Event in Alzheimer Disease.* Journal of Neuropathology & Experimental Neurology, 2001. **60**(8): p. 759-767.

73. Wilde, G.J.C., et al., *Differential Vulnerability of the CA1 and CA3 Subfields of the Hippocampus to Superoxide and Hydroxyl Radicals In Vitro.* Journal of Neurochemistry, 1997. **69**(2): p. 883-886.

74. Mankodi, A., et al., *Myotonic Dystrophy in Transgenic Mice Expressing an Expanded CUG Repeat.* Science, 2000. **289**(5485): p. 1769.

75. Morley, J.F., et al., *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, 2002. **99**: p. 10417-22.

76. Jana, N.R., et al., *Co-chaperone CHIP Associates with Expanded Polyglutamine Protein and Promotes Their Degradation by Proteasomes.* The Journal of Biological Chemistry, 2005. **280**(12): p. 11635-11640.

77. Hipp, M.S., et al., *Indirect inhibition of 26S proteasome activity in a cellular model of Huntington's disease.* The Journal of Cell Biology, 2012. **196**(5): p. 573.

78. Ladd, P.D., et al., *An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals.* Human Molecular Genetics, 2007. **16**(24): p. 3174-3187.

79. Wilburn, B., et al., *An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice.* Neuron, 2011. **70**(3): p. 427-440.

80. Moseley, M.L., et al., *Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8.* Nature Genetics, 2006. **38**: p. 758.

81. Ash, P.E.A., et al., *Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS.* Neuron, 2013. **77**: p. 639-46.

82. Mori, K., et al., *Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins.* Acta Neuropathologica, 2013. **126**(6): p. 881-893.

83. Zu, T., et al., *RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2.* Neuron, 2017. **95**: p. 1292-1305.e5.

84. Todd, Peter K., et al., *CGG Repeat-Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome.* Neuron, 2013. **78**: p. 440-455.

85. Soragni, E., et al., *Repeat-Associated Non-ATG (RAN) Translation in Fuchs' Endothelial Corneal Dystrophy.* Investigative ophthalmology & visual science, 2018. **59**(5): p. 1888-1896.

86. Green, K.M., et al., *RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response.* Nature Communications, 2017. **8**(1): p. 2005.

87. Kearse, M.G., et al., *CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins.* Molecular Cell, 2016. **62**: p. 314-322.

88. Feng, Y., et al., *Translational suppression by trinucleotide repeat expansion at FMR1.* Science, 1995. **268**(5211): p. 731.

89. Cheng, W., et al., *C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation.* Nature Communications, 2018. **9**(1): p. 51-51.

90. Westergard, T., et al., *Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress.* EMBO Molecular Medicine, 2019. **11**(2): p. e9423.

91. Kramer, N.J., et al., *Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts.* Science, 2016. **353**: p. 708-12.

92. Tran, H., et al., *Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS.* Neuron, 2015. **87**(6): p. 1207-1214.

93. Freibaum, B.D., et al., *GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport.* Nature, 2015. **525**: p. 129-133.

94. Wen, X., 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): p. 1213-1225.

95. Rudich, P., et al., *Nuclear localized C9orf72-associated arginine-containing dipeptides exhibit age-dependent toxicity in C. elegans.* Human Molecular Genetics, 2017.

96. Lee, Y.-B., et al., *C9orf72 poly GA RAN-translated protein plays a key role in amyotrophic lateral sclerosis via aggregation and toxicity.* Human Molecular Genetics, 2017. **26**(24): p. 4765-4777.

97. Yamakawa, M., et al., *Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS.* Human Molecular Genetics, 2014. **24**(6): p. 1630-1645.

98. Jovičić, A., et al., *Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS.* Nature Neuroscience, 2015. **18**: p. 1226.

99. Boeynaems, S., et al., *Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD.* Scientific Reports, 2016. **6**: p. 20877.

100. Kramer, N.J., et al., *CRISPR–Cas9 screens in human cells and primary neurons identify modifiers of C9ORF72 dipeptide-repeat-protein toxicity.* Nature Genetics, 2018. **50**(4): p. 603-612.

101. Gövert, F. and S.A. Schneider, *Huntington's disease and Huntington's disease-like syndromes.* Current Opinion in Neurology, 2013. **26**: p. 420-427.

102. Kim, D.-K., T.H. Kim, and S.-J. Lee, *Mechanisms of aging-related proteinopathies in Caenorhabditis elegans.* Experimental & Molecular Medicine, 2016. **48**: p. e263.

103. Brenner, S., *The Genetics of Caenorhabditis elegans* Genetics, 1974. **77**(1): p. 71.

104. Mello, C.C., et al., *Efficient gene transfer in C. elegans: extrachromosomal maintenance and integration of transforming sequences.* The EMBO journal, 1991. **10**(12): p. 3959-3970.

105. Lai, C.H., et al., *Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics.* Genome research, 2000. **10**(5): p. 703-713.

106. Sin, O., H. Michels, and E.A.A. Nollen, *Genetic screens in Caenorhabditis elegans models for neurodegenerative diseases.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2014. **1842**(10): p. 1951-1959.

107. Chen, X., et al., *Dual sgRNA-directed gene knockout using CRISPR/Cas9 technology in Caenorhabditis elegans.* Scientific reports, 2014. **4**: p. 7581-7581.

108. Gao, A.W., et al., *Forward and reverse genetics approaches to uncover metabolic aging pathways in Caenorhabditis elegans.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2018. **1864**(9, Part A): p. 2697-2706.

109. Taguchi, A., L.M. Wartschow, and M.F. White, *Brain IRS2 Signaling Coordinates Life Span and Nutrient Homeostasis.* Science, 2007. **317**(5836): p. 369.

110. Kenyon, C., et al., *A C. elegans mutant that lives twice as long as wild type.* Nature, 1993. **366**(6454): p. 461-464.

111. Murphy, C.T., Hu P.J., *Insulin/insulin-like growth factor sginaling in C. elegans* WormBook, 2013. **The *C. elegans* Research Community**.

112. Hsu, A.-L., C.T. Murphy, and C. Kenyon, *Regulation of Aging and Age-Related Disease by DAF-16 and Heat-Shock Factor.* Science, 2003. **300**(5622): p. 1142.

113. Ilieva, H., M. Polymenidou, and D.W. Cleveland, *Non–cell autonomous toxicity in neurodegenerative disorders: ALS and beyond.* The Journal of Cell Biology, 2009. **187**(6): p. 761.

114. Ewbank, J.J. and N. Pujol, *Local and long-range activation of innate immunity by infection and damage in C. elegans.* Current Opinion in Immunology, 2016. **38**: p. 1-7.

115. Cappellano, G., et al., *Immunity and inflammation in neurodegenerative diseases.* American Journal of Neurodegenerative Disease, 2013. **2**(2): p. 89-107.

116. Donnelly, Christopher J., et al., *RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention.* Neuron, 2013. **80**(2): p. 415-428.

117. Haeusler, A.R., et al., *C9orf72 nucleotide repeat structures initiate molecular cascades of disease.* Nature, 2014. **507**: p. 195-200.

118. Zhang, K., et al., *The C9orf72 repeat expansion disrupts nucleocytoplasmic transport.* Nature, 2015. **525**: p. 56.

119. Van Assche, R., et al., *In vitro aggregating β-lactamase-polyQ chimeras do not induce toxic effects in an in vivo Caenorhabditis elegans model.* Journal of Negative Results in Biomedicine, 2017. **16**: p. 14.

120. Faber, P.W., et al., *Polyglutamine-mediated dysfunction and apoptotic death of a Caenorhabditis elegans sensory neuron.* Proc Natl Acad Sci U S A, 1999. **96**: p. 179-84.

121. Bates, E.A., et al., *Differential Contributions of Caenorhabditis elegans Histone Deacetylases to Huntingtin Polyglutamine Toxicity.* The Journal of Neuroscience, 2006.

122. Pandey, U.B., et al., *HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS.* Nature, 2007. **447**: p. 860.

123. Jia, H., et al., *Histone deacetylase (HDAC) inhibitors targeting HDAC3 and HDAC1 ameliorate polyglutamine-elicited phenotypes in model systems of Huntington's disease.* Neurobiology of disease, 2012. **46**(2): p. 351-361.

124. Parker, J.A., et al., *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, 2001. **98**: p. 13318-23.

125. Chalfie, M. and J. Sulston, *Developmental genetics of the mechanosensory neurons of Caenorhabditis elegans.* Developmental Biology, 1981. **82**(2): p. 358-370.

126. Lejeune, F.-X., et al., *Large-scale functional RNAi screen in C. elegans identifies genes that regulate the dysfunction of mutant polyglutamine neurons.* BMC Genomics, 2012. **13**: p. 91.

127. Satyal, S.H., et al., *Polyglutamine aggregates alter protein folding homeostasis in Caenorhabditis elegans.* Proc Natl Acad Sci U S A, 2000. **97**: p. 5750-5.

128. Melkani, G.C., *Huntington's Disease-Induced Cardiac Disorders Affect Multiple Cellular Pathways.* Reactive oxygen species (Apex, N.C.), 2016. **2**(5): p. 325-338.

129. Mielcarek, M., et al., *Dysfunction of the CNS-heart axis in mouse models of Huntington's disease.* PLoS Genetics, 2014. **10**(8): p. e1004550-e1004550.

130. Gidalevitz, T., et al., *Destabilizing Protein Polymorphisms in the Genetic Background Direct Phenotypic Expression of Mutant SOD1 Toxicity.* PLOS Genetics, 2009. **5**(3): p. e1000399.

131. Gidalevitz, T., et al., *Natural genetic variation determines susceptibility to aggregation or toxicity in a C. elegans model for polyglutamine disease.* BMC Biology, 2013. **11**: p. 100.

132. Morley, J.F. and R.I. Morimoto, *Regulation of longevity in Caenorhabditis elegans by heat shock factor and molecular chaperones.* Molecular biology of the cell, 2004. **15**(2): p. 657-664.

133. Teixeira-Castro, A., et al., *Neuron-specific proteotoxicity of mutant ataxin-3 in C. elegans: rescue by the DAF-16 and HSF-1 pathways.* Human Molecular Genetics, 2011. **20**: p. 2996-3009.

134. Moronetti Mazzeo, L.E., et al., *Stress and aging induce distinct polyQ protein aggregation states.* Proc Natl Acad Sci U S A, 2012. **109**: p. 10587-92.

135. Brignull, H.R., et al., *Polyglutamine proteins at the pathogenic threshold display neuron-specific aggregation in a pan-neuronal Caenorhabditis elegans model.* The Journal of Neuroscience, 2006. **26**: p. 7597-606.

136. Nollen, E.A.A., et al., *Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation.* Proc Natl Acad Sci U S A, 2004. **101**: p. 6403-8.

137. Kitamura, A., et al., *Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state.* Nature Cell Biology, 2006. **8**: p. 1163.

138. Teuling, E., et al., *Modifiers of mutant huntingtin aggregation.* PLoS Currents, 2011.

139. Varshney, L.R., et al., *Structural Properties of the Caenorhabditis elegans Neuronal Network.* PLOS Computational Biology, 2011. **7**(2): p. e1001066.

140. White John, G., et al., *The structure of the nervous system of the nematode Caenorhabditis elegans.* Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1986. **314**(1165): p. 1-340.

141. Hall, D.H. and R.L. Russell, *The posterior nervous system of the nematode Caenorhabditis elegans: serial reconstruction of identified neurons and complete pattern of synaptic interactions.* The Journal of Neuroscience, 1991. **11**(1): p. 1.

142. Brownlee, D.J.A. and I. Fairweather, *Exploring the neurotransmitter labyrinth in nematodes.* Trends in Neurosciences, 1999. **22**(1): p. 16-24.

143. Vaccaro, A., et al., *Mutant TDP-43 and FUS Cause Age-Dependent Paralysis and Neurodegeneration in C. elegans.* PLOS ONE, 2012. **7**(2): p. e31321.

144. Kraemer, B.C., et al., *Neurodegeneration and defective neurotransmission in a Caenorhabditis elegans*

model of tauopathy. Proc Natl Acad Sci U S A, 2003. **100**(17): p. 9980.

145. McCormick, D.A., *GABA as an inhibitory neurotransmitter in human cerebral cortex.* Journal of Neurophysiology, 1989. **62**(5): p. 1018-1027.

146. Schuske, K., A.A. Beg, and E.M. Jorgensen, *The GABA nervous system in C. elegans.* Trends in Neurosciences, 2004. **27**(7): p. 407-414.

147. McIntire, S.L., et al., *The GABAergic nervous system of Caenorhabditis elegans.* Nature, 1993. **364**(6435): p. 337-341.

148. Jorgensen, E.M. and M.L. Nonet, *Neuromuscular junctions in the nematode C. elegans.* Seminars in Developmental Biology, 1995. **6**(3): p. 207-220.

149. del Aguila, M.A., et al., *Prognosis in amyotrophic lateral sclerosis.* Neurology, 2003. **60**(5): p. 813.

150. Byrne, S., et al., *Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis.* Journal of Neurology, Neurosurgery & Psychiatry, 2011. **82**(6): p. 623.

151. Alonso, A., et al., *Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study.* European journal of neurology, 2009. **16**(6): p. 745-751.

152. Logroscino, G., et al., *Incidence of amyotrophic lateral sclerosis in Europe.* Journal of neurology, neurosurgery, and psychiatry, 2010. **81**(4): p. 385-390.

153. Zou, Z.-Y., et al., *Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis.* Journal of Neurology, Neurosurgery & Psychiatry, 2017. **88**(7): p. 540.

154. Armon, C., *Smoking may be considered an established risk factor for sporadic ALS.* Neurology, 2009. **73**(20): p. 1693-1698.

155. Wang, M.-D., et al., *A meta-analysis of observational studies of the association between chronic occupational exposure to lead and amyotrophic lateral sclerosis.* Journal of occupational and environmental medicine, 2014. **56**(12): p. 1235-1242.

156. Malek, A.M., et al., *Pesticide exposure as a risk factor for amyotrophic lateral sclerosis: A meta-analysis of epidemiological studies: Pesticide exposure as a risk factor for ALS.* Environmental Research, 2012. **117**: p. 112-119.

157. McKee, A.C., et al., *TDP-43 proteinopathy and motor neuron disease in chronic traumatic encephalopathy.* Journal of neuropathology and experimental neurology, 2010. **69**(9): p. 918-929.

158. Wijesekera, L.C. and P.N. Leigh, *Amyotrophic lateral sclerosis.* Orphanet journal of rare diseases, 2009. **4**: p. 3-3.

159. Grad, L.I., et al., *Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS).* Cold Spring Harbor Perspectives in Medicine, 2017. **7**(8).

160. Ringholz, G.M., et al., *Prevalence and patterns of cognitive impairment in sporadic ALS.* Neurology, 2005. **65**(4): p. 586.

161. McKhann, G.M., et al., *Clinical and Pathological Diagnosis of Frontotemporal Dementia: Report of the Work Group on Frontotemporal Dementia and Pick's Disease.* JAMA Neurology, 2001. **58**(11): p. 1803-1809.

162. Ling, S.-C., M. Polymenidou, and Don W. Cleveland, *Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis.* Neuron, 2013. **79**(3): p. 416-438.

163. Vance, C., et al., *Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2,Äì21.3.* Brain, 2006. **129**(4): p. 868-876.

164. Majounie, E., et al., *Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study.* The Lancet Neurology, 2012. **11**(4): p. 323-330.

165. Cooper-Knock, J., P.J. Shaw, and J. Kirby, *The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype.* Acta Neuropathologica, 2014. **127**(3): p. 333-345.

166. Waite, A.J., et al., *Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion.* Neurobiology of aging, 2014. **35**(7): p. 1779.e5-1779.e13.

167. Webster, C.P., et al., *The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy.* The EMBO Journal, 2016. **35**(15): p. 1656-1676.

168. Sellier, C., et al., *Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death.* The EMBO journal, 2016. **35**(12): p. 1276-1297.

169. Ugolino, J., et al., *Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling.* PLOS Genetics, 2016. **12**(11): p. e1006443.

170. Sullivan, P.M., et al., *The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway.* Acta Neuropathologica Communications, 2016. **4**(1): p. 51.

171. O'Rourke, J.G., et al., *C9orf72 is required for proper macrophage and microglial function in mice.* Science, 2016. **351**(6279): p. 1324.

172. Burberry, A., et al., *Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease.* Science Translational Medicine, 2016. **8**(347): p. 347ra93.

173. Atanasio, A., et al., *C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice.* Scientific Reports, 2016. **6**: p. 23204.

174. Fratta, P., et al., *Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia.* Acta neuropathologica, 2013. **126**(3): p. 401-409.

175. Lee, Y.-B., et al., *Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic.* Cell Reports, 2013. **5**(5): p. 1178-1186.

176. Cleary, J.D. and L.P.W. Ranum, *Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders.* Current Opinion in Genetics & Development, 2014. **26**: p. 6-15.

177. Hautbergue, G.M., et al., *SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits.* Nature Communications, 2017. **8**: p. 16063.

178. Mori, K., et al., *Reduced hnRNPA3 increases C9orf72 repeat RNA levels and dipeptide-repeat protein deposition.* EMBO Reports, 2016. **17**(9): p. 1314.

179. Mackenzie, I.R., et al., *Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations.* Acta Neuropathologica, 2013. **126**: p. 859-879.

180. Davidson, Y.S., et al., *Brain distribution of dipeptide repeat proteins in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72.* Acta Neuropathologica Communications, 2014. **2**.

181. Schludi, M.H., et al., *Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing.* Acta Neuropathologica, 2015. **130**: p. 537-55.

182. Mizielinska, S., et al., *C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins.* Science

, 2014. **345**(6201): p. 1192-1194.

183. Bennion Callister, J., et al., *Modelling C9orf72 dipeptide repeat proteins of a physiologically relevant size.* Human Molecular Genetics, 2016. **25**(23): p. 5069-5082.

184. Kwon, I., et al., *Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells.* Science, 2014. **345**(6201): p. 1139-1145.

185. Mitrea, D.M. and R.W. Kriwacki, *Phase separation in biology; functional organization of a higher order.* Cell Communication and Signaling, 2016. **14**(1): p. 1.

186. Alberti, S., *Phase separation in biology.* Current Biology, 2017. **27**(20): p. R1097-R1102.

187. Hennig, S., et al., *Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles.* The Journal of cell biology, 2015. **210**(4): p. 529-539.

188. Molliex, A., et al., *Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization.* Cell, 2015. **163**.

189. Schmidt, H.B. and D. Görlich, *Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity.* eLife, 2015. **4**: p. e04251.

190. Brangwynne, C.P., T.J. Mitchison, and A.A. Hyman, *Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes.* Proc Natl Acad Sci U S A, 2011. **108**.

191. Elbaum-Garfinkle, S., *Matter over mind: Liquid phase separation and neurodegeneration.* The Journal of Biological Chemistry, 2019.

192. Patel, A., et al., *A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.* Cell, 2015. **162**(5): p. 1066-1077.

193. Wegmann, S., et al., *Tau protein liquid-liquid phase separation can initiate tau aggregation.* The EMBO Journal, 2018. **37**(7): p. e98049.

194. Peskett, T.R., et al., *A Liquid to Solid Phase Transition Underlying Pathological Huntingtin Exon1 Aggregation.* Molecular Cell, 2018. **70**(4): p. 588-601.e6.

195. Lee, K.-H., et al., *C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles.* Cell, 2016. **167**(3): p. 774-788.e17.

196. Tao, Z., et al., *Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity.* Human Molecular Genetics, 2015. **24**(9): p. 2426-2441.

197. Schaefer, M.H., E.E. Wanker, and M.A. Andrade-Navarro, *Evolution and function of CAG/polyglutamine repeats in protein–protein interaction networks.* Nucleic Acids Research, 2012. **40**(10): p. 4273-4287.

198. Legleiter, J., et al., *Mutant Huntingtin Fragments Form Oligomers in a Polyglutamine Length-dependent Manner in Vitro and in Vivo.* Journal of Biological Chemistry, 2010. **285**(19): p. 14777-14790.

199. Scherzinger, E., et al., *Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo.* Cell, 1997. **90**(3): p. 549-58.

200. Adegbuyiro, A., et al., *Proteins Containing Expanded Polyglutamine Tracts and Neurodegenerative Disease.* Biochemistry, 2017. **56**(9): p. 1199-1217.

201. Chung, C.G., H. Lee, and S.B. Lee, *Mechanisms of protein toxicity in neurodegenerative diseases.* Cellular and Molecular Life Sciences, 2018. **75**(17): p. 3159-3180.

202. De Souza, R.A.G. and B.R. Leavitt, *Neurobiology of Huntington's Disease*, in *Behavioral Neurobiology of Huntington's Disease and Parkinson's Disease*, H.H.P. Nguyen and M.A. Cenci, Editors. 2015, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 81-100.

203. Foroud, T., et al., *Differences in duration of Huntington's disease based on age at onset.* Journal of Neurology, Neurosurgery & Psychiatry, 1999. **66**(1): p. 52.

204. Quarrell, O., et al., *The Prevalence of Juvenile Huntington's Disease: A Review of the Literature and Meta-Analysis.* PLoS currents, 2012. **4**: p. e4f8606b742ef3-e4f8606b742ef3.

205. MacDonald, M.E., et al., *A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.* Cell, 1993. **72**(6): p. 971-983.

206. Rubinsztein, D.C., et al., *Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats.* American journal of human genetics, 1996. **59**(1): p. 16-22.

207. Duff, K., et al., *Psychiatric Symptoms in Huntington's Disease before Diagnosis: The Predict-HD Study.* Biological Psychiatry, 2007. **62**(12): p. 1341-1346.

208. Duff, K., et al., *Mild cognitive impairment in prediagnosed Huntington disease.* Neurology, 2010. **75**(6): p. 500-507.

209. Glass, M., M. Dragunow, and R.L.M. Faull, *The pattern of neurodegeneration in Huntington's disease: a comparative study of cannabinoid, dopamine, adenosine and GABAA receptor alterations in the human basal ganglia in Huntington's disease.* Neuroscience, 2000. **97**(3): p. 505-519.

210. Hobbs, N.Z., et al., *The progression of regional atrophy in premanifest and early Huntington's disease: a longitudinal voxel-based morphometry study.* Journal of Neurology, Neurosurgery & Psychiatry, 2010. **81**(7): p. 756.

211. Poudel, G.R., et al., *Network spread determines severity of degeneration and disconnection in Huntington's disease.* Human Brain Mapping, 2019. **0**(0).

212. Ribaï, P., et al., *Psychiatric and Cognitive Difficulties as Indicators of Juvenile Huntington Disease Onset in 29 Patients.* JAMA Neurology, 2007. **64**(6): p. 813-819.

213. Tereshchenko, A., et al., *Brain structure in juvenile-onset Huntington disease.* Neurology, 2019. **92**(17): p. e1939.

214. Grima, J.C., et al., *Mutant Huntingtin Disrupts the Nuclear Pore Complex.* Neuron, 2017. **94**(1): p. 93-107.e6.

215. Takano, H. and J.F. Gusella, *The predominantly HEAT-like motif structure of huntingtin and its association and coincident nuclear entry with dorsal, an NF-kB/Rel/dorsal family transcription factor.* BMC neuroscience, 2002. **3**: p. 15-15.

216. Li, Z., et al., *A Putative Drosophila Homolog of the Huntington's Disease Gene.* Human Molecular Genetics, 1999. **8**(9): p. 1807-1815.

217. Tartari, M., et al., *Phylogenetic Comparison of Huntingtin Homologues Reveals the Appearance of a Primitive polyQ in Sea Urchin.* Molecular Biology and Evolution, 2007. **25**(2): p. 330-338.

218. White, J.K., et al., *Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion.* Nature Genetics, 1997. **17**(4): p. 404-410.

219. O'Kusky, J.R., et al., *Neuronal degeneration in the basal ganglia and loss of pallido-subthalamic synapses in mice with targeted disruption of the Huntington's disease gene.* Brain Research, 1999. **818**(2): p. 468-479.

220. Andrade, M.A. and P. Bork, *HEAT repeats in the Huntington's disease protein.* Nature Genetics, 1995. **11**(2): p. 115-116.

221. Ratovitski, T., et al., *Huntingtin protein interactions altered by polyglutamine expansion as determined by quantitative proteomic analysis.* Cell cycle (Georgetown, Tex.), 2012. **11**(10): p. 2006-2021.

222. Schulte, J. and J.T. Littleton, *The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology.* Current trends in neurology, 2011. **5**: p. 65-78.

223. Hong, Y., et al., *Mutant Huntingtin Impairs BDNF Release from Astrocytes by Disrupting Conversion of Rab3a-GTP into Rab3a-GDP.* The Journal of Neuroscience, 2016. **36**(34): p. 8790.

224. Kaltenbach, L.S., et al., *Huntingtin interacting proteins are genetic modifiers of neurodegeneration.* PLoS genetics, 2007. **3**(5): p. e82-e82.

225. Nollen, E.A., et al., *Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation.* Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6403-8.

226. Satyal, S.H., et al., *Polyglutamine aggregates alter protein folding homeostasis in Caenorhabditis elegans.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5750-5.

227. Ravikumar, B., et al., *Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease.* Nat Genet, 2004. **36**(6): p. 585-95.

228. Berendzen, K.M., et al., *Neuroendocrine Coordination of Mitochondrial Stress Signaling and Proteostasis.* Cell, 2016. **166**(6): p. 1553-1563 e10.

229. Leitman, J., F. Ulrich Hartl, and G.Z. Lederkremer, *Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress.* Nature Communications, 2013. **4**: p. 2753.

230. Kim, Yujin E., et al., *Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors.* Molecular Cell, 2016. **63**(6): p. 951-964.

231. Hoffner, G. and P. Djian, *Monomeric, oligomeric and polymeric proteins in huntington disease and other diseases of polyglutamine expansion.* Brain sciences, 2014. **4**(1): p. 91-122.

232. Snell, R.G., et al., *Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease.* Nat Genet, 1993. **4**(4): p. 393-7.

233. Banez-Coronel, M., et al., *RAN Translation in Huntington Disease.* Neuron, 2015. **88**(4): p. 667-77.

234. Herndon, E.S., et al., *Neuroanatomic profile of polyglutamine immunoreactivity in Huntington disease brains.* J Neuropathol Exp Neurol, 2009. **68**(3): p. 250-61.

235. Bohanna, I., et al., *Diffusion tensor imaging in Huntington's disease reveals distinct patterns of white matter degeneration associated with motor and cognitive deficits.* Brain Imaging Behav, 2011. **5**(3): p. 171-80.

236. Fennema-Notestine, C., et al., *In vivo evidence of cerebellar atrophy and cerebral white matter loss in Huntington disease.* Neurology, 2004. **63**(6): p. 989-95.

237. Reading, S.A., et al., *Regional white matter change in pre-symptomatic Huntington's disease: a diffusion tensor imaging study.* Psychiatry Res, 2005. **140**(1): p. 55-62.

238. Fratta, P., et al., *Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion.* Neurobiology of aging, 2015. **36**(1): p. 546.e1-546.e5467.

239. Nordin, A., et al., *Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD.* Human Molecular Genetics, 2015. **24**(11): p. 3133-3142.

240. Swaminathan, A., et al., *Expression of C9orf72-related dipeptides impairs motor function in a vertebrate model.* Human Molecular Genetics, 2018. **27**(10): p. 1754-1762.

241. Boeynaems, S., et al., *Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics.* Molecular Cell, 2017. **65**(6): p. 1044-1055.e5.

242. Shi, K.Y., et al., *Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export.* Proc Natl Acad Sci U S A, 2017. **114**(7): p. E1111.

243. Chai, N. and A.D. Gitler, *Yeast screen for modifiers of C9orf72 poly(glycine-arginine) dipeptide repeat toxicity.* FEMS Yeast Research, 2018. **18**(4).

244. McIntire, S.L., E. Jorgensen, and H.R. Horvitz, *Genes required for GABA function in Caenorhabditis elegans.* Nature, 1993. **364**(6435): p. 334-337.

245. Li, J., K.-x. Huang, and W.-d. Le, *Establishing a novel C. elegans model to investigate the role of autophagy in amyotrophic lateral sclerosis.* Acta Pharmacologica Sinica, 2013. **34**: p. 644.

246. Aggad, D., et al., *TDP-43 Toxicity Proceeds via Calcium Dysregulation and Necrosis in Aging Caenorhabditis elegans Motor Neurons.* The Journal of Neuroscience, 2014. **34**(36): p. 12093.

247. Earls, L.R., et al., *Coenzyme Q protects Caenorhabditis elegans GABA neurons from calcium-dependent degeneration.* Proc Natl Acad Sci U S A, 2010. **107**(32): p. 14460.

248. Cohen, E., et al., *Opposing Activities Protect Against Age-Onset Proteotoxicity.* Science, 2006. **313**(5793): p. 1604.

249. Gidalevitz, T., et al., *Progressive disruption of cellular protein folding in models of polyglutamine diseases.* Science, 2006. **311**: p. 1471-4.

250. Bank, E.M., et al., *Structural and physiological phenotypes of disease-linked lamin mutations in C. elegans.* Journal of Structural Biology, 2012. **177**(1): p. 106-112.

251. Lee, V.M.Y., M. Goedert, and J.Q. Trojanowski, *Neurodegenerative Tauopathies.* Annual Review of Neuroscience, 2001. **24**(1): p. 1121-1159.

252. Brignull, H.R., et al., *Modeling Polyglutamine Pathogenesis in C. elegans*, in *Methods in Enzymology*. 2006, Academic Press. p. 256-282.

253. Allen, A.K., J.E. Nesmith, and A. Golden, *An RNAi-based suppressor screen identifies interactors of the Myt1 ortholog of Caenorhabditis elegans.* G3 (Bethesda, Md.), 2014. **4**(12): p. 2329-2343.

254. Garigan, D., et al., *Genetic Analysis of Tissue Aging in Caenorhabditis elegans: A Role for Heat-Shock Factor and Bacterial Proliferation.* Genetics, 2002. **161**(3): p. 1101.

255. Pastink, A., et al., *Sequence analysis of N-ethyl-N-nitrosourea-induced vermilion mutations in Drosophila melanogaster.* Genetics, 1989. **123**(1): p. 123-129.

256. Hsieh, J., et al., *The RING finger/B-Box factor TAM-1 and a retinoblastoma-like protein LIN-35 modulate context-dependent gene silencing in Caenorhabditis elegans.* Genes & Development, 1999. **13**(22): p. 2958-2970.

257. Wang, J., et al., *An ALS-Linked Mutant SOD1 Produces a Locomotor Defect Associated with Aggregation and Synaptic Dysfunction When Expressed in Neurons of Caenorhabditis elegans.* PLOS Genetics, 2009. **5**(1): p. e1000350.

258. Martindale, D., et al., *Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates.* Nature Genetics, 1998. **18**: p. 150-154.

259. Conicella, A.E., et al., *ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain.* Structure (London, England : 1993), 2016. **24**(9): p. 1537-1549.

260. Kayed, R., et al., *Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis.* Science, 2003. **300**(5618): p. 486.

261. Gijselinck, I., et al., *The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter.* Molecular Psychiatry, 2015. **21**: p. 1112.

262. Van Mossevelde, S., et al., *Relationship between C9orf72 repeat size and clinical phenotype.* Current Opinion in Genetics & Development, 2017. **44**: p. 117-124.

263. Pamphlett, R., et al., *Can ALS-Associated C9orf72 Repeat Expansions Be Diagnosed on a Blood DNA Test Alone?* PLOS ONE, 2013. **8**(7): p. e70007.

264. Xi, Z., et al., *Jump from pre-mutation to pathologic expansion in C9orf72.* The American Journal of Human Genetics, 2015. **96**(6): p. 962-970.

265. Beck, J., et al., *Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population.* The American Journal of Human Genetics, 2013. **92**(3): p. 345-353.

266. Usdin, K., *The biological effects of simple tandem repeats: lessons from the repeat expansion diseases.* Genome research, 2008. **18**(7): p. 1011-1019.

267. Moens, T.G., et al., *C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is rescued by eIF1A.* Acta Neuropathologica, 2019. **137**(3): p. 487-500.

268. Segev, Y., D.M. Michaelson, and K. Rosenblum, *ApoE ε4 is associated with eIF2α phosphorylation and impaired learning in young mice.* Neurobiology of Aging, 2013. **34**(3): p. 863-872.

269. Scerbak, C., et al., *Insulin signaling in the aging of healthy and proteotoxically stressed mechanosensory neurons.* Frontiers in genetics, 2014. **5**: p. 212.

270. Li, J., et al., *Human superoxide dismutase 1 overexpression in motor neurons of Caenorhabditis elegans causes axon guidance defect and neurodegeneration.* Neurobiology of Aging, 2014. **35**(4): p. 837-846.

271. Thompson, M.L., et al., *TorsinA rescues ER-associated stress and locomotive defects in C. elegans models of ALS.* Disease Models & Mechanisms, 2014. **7**(2): p. 233-243.

272. Oeda, T., et al., *Oxidative stress causes abnormal accumulation of familial amyotrophic lateral sclerosis-related mutant SOD1 in transgenic Caenorhabditis elegans.* Human Molecular Genetics, 2001. **10**(19): p. 2013-2023.

273. Zhang, T., et al., *TDP-43 neurotoxicity and protein aggregation modulated by heat shock factor and insulin/IGF-1 signaling.* Human Molecular Genetics, 2011. **20**(10): p. 1952-1965.

274. Boccitto, M., T. Lamitina, and R.G. Kalb, *Daf-2 signaling modifies mutant SOD1 toxicity in C. elegans.* PloS one, 2012. **7**(3): p. e33494-e33494.

275. Eftekharzadeh, B., et al., *Tau Protein Disrupts Nucleocytoplasmic Transport in Alzheimer's Disease.* Neuron, 2018. **99**(5): p. 925-940.e7.

276. Gitcho, M.A., et al., *TDP-43 A315T mutation in familial motor neuron disease.* Annals of neurology, 2008. **63**(4): p. 535-538.

277. Kwiatkowski, T.J., et al., *Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis.* Science, 2009. **323**(5918): p. 1205.

278. Vance, C., et al., *Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6.* Science (New York, N.Y.), 2009. **323**(5918): p. 1208-1211.

279. Mackenzie, I.R.A., R. Rademakers, and M. Neumann, *TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia.* The Lancet Neurology, 2010. **9**(10): p. 995-1007.

280. Järvelin, A.I., et al., *The new (dis)order in RNA regulation.* Cell Communication and Signaling, 2016. **14**(1): p. 9.

281. Srivastava, A., S. Ahmad, and M.M. Gromiha, *Deciphering RNA-Recognition Patterns of Intrinsically Disordered Proteins.* International journal of molecular sciences, 2018. **19**(6): p. 1595.

282. Butti, Z. and S.A. Patten, *RNA Dysregulation in Amyotrophic Lateral Sclerosis.* Frontiers in genetics, 2019. **9**: p. 712-712.

283. Jimenez-Pacheco, A., et al., *Epigenetic Mechanisms of Gene Regulation in Amyotrophic Lateral Sclerosis*, in *Neuroepigenomics in Aging and Disease*, R. Delgado-Morales, Editor. 2017, Springer International Publishing: Cham. p. 255-275.

284. Belzil, V.V., R.B. Katzman, and L. Petrucelli, *ALS and FTD: an epigenetic perspective.* Acta Neuropathologica, 2016. **132**(4): p. 487-502.

285. Warrick, J.M., et al., *Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in Drosophila.* Cell, 1998. **93**(6): p. 939-49.

286. Gupta, R., et al., *The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome.* eNeuro, 2017. **4**(1): p. ENEURO.0249-16.2017.

287. Bouchard, J.J., et al., *Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments.* Molecular Cell, 2018. **72**(1): p. 19-36.e8.

288. Murayama, Y., T. Ogura, and K. Yamanaka, *Characterization of C-terminal adaptors, UFD-2 and UFD-3, of CDC-48 on the polyglutamine aggregation in C. elegans.* Biochemical and Biophysical Research Communications, 2015. **459**: p. 154-160.

289. Barbieri, C.E., et al., *Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer.* Nature genetics, 2012. **44**(6): p. 685-689.

290. Le Gallo, M., et al., *Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes.* Nature genetics, 2012. **44**(12): p. 1310-1315.

291. Liu, J., et al., *Analysis of Drosophila Segmentation Network Identifies a JNK Pathway Factor Overexpressed in Kidney Cancer.* Science, 2009. **323**(5918): p. 1218.

292. Guo, Z.-Q., et al., *Small-Molecule Targeting of E3 Ligase Adaptor SPOP in Kidney Cancer.* Cancer Cell, 2016. **30**(3): p. 474-484.

293. Shi, Y., et al., *Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons.* Nature Medicine, 2018. **24**: p. 313.

294. Therrien, M., et al., *Deletion of C9ORF72 results in motor neuron degeneration and stress sensitivity in C. elegans.* PloS one, 2013. **8**(12): p. e83450-e83450.

295. Burgess, R.W., et al., *Forward Genetic Screen in Caenorhabditis elegans Suggests F57A10.2 and acp-4 As Suppressors of C9ORF72 Related Phenotypes.* Mol. Neurosci, 2016. **9**.

296. Deng, H.-X., et al., *Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia.* Nature, 2011. **477**(7363): p. 211-215.

297. Cheroni, C., et al., *Functional alterations of the ubiquitin-proteasome system in motor neurons of a mouse model of familial amyotrophic lateral sclerosis.* Human molecular genetics, 2009. **18**(1): p. 82-96.

298. Ackermann, L., et al., *E4 ligase–specific ubiquitination hubs coordinate DNA double-strand-break repair and apoptosis.* Nature Structural &Amp; Molecular Biology, 2016. **23**: p. 995.

299. Marzahn, M.R., et al., *Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles.* The EMBO journal, 2016. **35**(12): p. 1254-1275.

300. Han, S.K., et al., *OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research.* Oncotarget, 2016. **7**(35): p. 56147-56152.

301. Yang, J.-S., et al., *OASIS: online application for the survival analysis of lifespan assays performed in aging research.* PloS one, 2011. **6**(8): p. e23525-e23525.

302. Schöls, L., et al., *Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis.* The Lancet Neurology, 2004. **3**(5): p. 291-304.

303. Rawlins, M.D., et al., *The Prevalence of Huntington's Disease.* Neuroepidemiology, 2016. **46**(2): p. 144-153.

304. Ranen, N.G., et al., *Anticipation and instability of IT-15 (CAG)n repeats in parent-offspring pairs with Huntington disease.* The American Journal of Human Genetics, 1995. **57**(3): p. 593-602.

305. Kay, C., et al., *The molecular epidemiology of Huntington disease is related to intermediate allele frequency and haplotype in the general population.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2018. **177**(3): p. 346-357.

306. Williams, J.K., et al., *Caregiving by teens for family members with Huntington disease.* Journal of family nursing, 2009. **15**(3): p. 273-294.

307. McCusker, E.A. and C.T. Loy, *Huntington Disease: The Complexities of Making and Disclosing a Clinical Diagnosis After Premanifest Genetic Testing.* Tremor and other hyperkinetic movements (New York, N.Y.), 2017. **7**: p. 467-467.

308. Sun, C.-S., et al., *Conformational switch of polyglutamine-expanded huntingtin into benign aggregates leads to neuroprotective effect.* Scientific reports, 2015. **5**: p. 14992.

309. Becher, M.W., et al., *Intranuclear Neuronal Inclusions in Huntington's Disease and Dentatorubral and Pallidoluysian Atrophy: Correlation between the Density of Inclusions andIT15CAG Triplet Repeat Length.* Neurobiology of Disease, 1998. **4**(6): p. 387-397.

310. Silva, M.C., et al., *A Genetic Screening Strategy Identifies Novel Regulators of the Proteostasis Network.* PLoS Genet, 2011. **7**.

311. Wang, Z.-F., et al., *The Hairpin Form of r(G4C2)exp in c9ALS/FTD Is Repeat-Associated Non-ATG Translated and a Target for Bioactive Small Molecules.* Cell Chemical Biology, 2019. **26**(2): p. 179-190.e12.

312. Graveland, G.A., R.S. Williams, and M. DiFiglia, *Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease.* Science, 1985. **227**(4688): p. 770.

313. Garcia, S.M., et al., *Neuronal signaling modulates protein homeostasis in Caenorhabditis elegans post-synaptic muscle cells.* Genes & Development, 2007. **21**: p. 3006-16.

314. Nix, P., et al., *Axon Regeneration Genes Identified by RNAi Screening in C. elegans.* The Journal of Neuroscience, 2014. **34**(2): p. 629.

315. González-Hunt, C.P., et al., *Exposure to Mitochondrial Genotoxins and Dopaminergic Neurodegeneration in Caenorhabditis elegans.* PLOS ONE, 2014. **9**(12): p. e114459.

316. Kennedy, S., D. Wang, and G. Ruvkun, *A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans.* Nature, 2004. **427**(6975): p. 645-649.

317. Jain, R.K., et al., *Oligomerization of green fluorescent protein in the secretory pathway of endocrine cells.* The Biochemical Journal, 2001. **360**(Pt 3): p. 645-649.

318. Kuroiwa, T., et al., *Systematic analysis of stop-transfer sequence for microsomal membrane.* The Journal of Biological Chemistry, 1991. **266**(14): p. 9251-9255.

319. Chen, H. and D.A. Kendall, *Artificial Transmembrane Segments.: Requirements for Stop Transfer and Polypeptide Orientation* The Journal of Biological Chemistry, 1995. **270**(23): p. 14115-14122.

320. Pédelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature Biotechnology, 2005. **24**: p. 79.

321. Aronson, D.E., L.M. Costantini, and E.L. Snapp, *Superfolder GFP is fluorescent in oxidizing environments when targeted via the Sec translocon.* Traffic (Copenhagen, Denmark), 2011. **12**(5): p. 543-548.

322. De Stasio, E.A. and S. Dorman, *Optimization of ENU mutagenesis of Caenorhabditis elegans.* Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 2001. **495**(1): p. 81-88.

323. Thompson, O., et al., *The million mutation project: A new approach to genetics in Caenorhabditis elegans.* Genome Research, 2013. **23**(10): p. 1749-1762.

324. Acevedo-Arozena, A., et al., *ENU mutagenesis, a way forward to understand gene function.* Annual Review of Genomics and Human Genetics 2008. **9**: p. 46-69.

325. Flibotte, S., et al., *Whole-genome profiling of mutagenesis in Caenorhabditis elegans.* Genetics, 2010. **185**(2): p. 431-441.

326. Fire, A., *Integrative transformation of Caenorhabditis elegans.* The EMBO Journal, 1986. **5**(10): p. 2673-2680.

327. Takada, R., et al., *Monounsaturated Fatty Acid Modification of Wnt Protein: Its Role in Wnt Secretion.* Developmental Cell, 2006. **11**(6): p. 791-801.

328. Rocks, O., et al., *An Acylation Cycle Regulates Localization and Activity of Palmitoylated Ras Isoforms.* Science, 2005. **307**(5716): p. 1746.

329. Diez-Ardanuy, C., et al., *A cluster of palmitoylated cysteines are essential for aggregation of cysteine-string protein mutants that cause neuronal ceroid lipofuscinosis.* Scientific reports, 2017. **7**(1): p. 10-10.

330. Ayhan, F., et al., *SCA8 RAN polySer protein preferentially accumulates in white matter regions and is regulated by eIF3F.* The EMBO Journal, 2018. **37**(19): p. e99023.

331. Hsu, R.-J., et al., *Long Tract of Untranslated CAG Repeats Is Deleterious in Transgenic Mice.* PLOS ONE, 2011. **6**(1): p. e16417.

332. García-López, A., et al., *In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models.* Proc Natl Acad Sci U S A, 2011. **108**(29): p. 11866-11871.

333. Li, L.-B., et al., *RNA toxicity is a component of ataxin-3 degeneration in Drosophila.* Nature, 2008. **453**(7198): p. 1107-1111.

334. Bañez-Coronel, M., et al., *A Pathogenic Mechanism in Huntington's Disease Involves Small CAG-Repeated RNAs with Neurotoxic Activity.* PLOS Genetics, 2012. **8**(2): p. e1002481.

335. Albrecht, A.N., et al., *A molecular pathogenesis for transcription factor associated poly-alanine tract expansions.* Human Molecular Genetics, 2004. **13**(20): p. 2351-2359.

336. Oma, Y., et al., *Interactions between homopolymeric amino acids (HPAAs).* Protein science : a publication of the Protein Society, 2007. **16**(10): p. 2195-2204.

337. Kumar, A.S., D.T. Sowpati, and R.K. Mishra, *Single Amino Acid Repeats in the Proteome World: Structural, Functional, and Evolutionary Insights.* PLOS ONE, 2016. **11**(11): p. e0166854.

338. Pelassa, I. and F. Fiumara, *Differential Occurrence of Interactions and Interaction Domains in Proteins Containing Homopolymeric Amino Acid Repeats.* Frontiers in Genetics, 2015. **6**(345).

339. Mularoni, L., R. Guigó, and M.M. Albà, *Mutation patterns of amino acid tandem repeats in the human proteome.* Genome biology, 2006. **7**(4): p. R33-R33.

340. Dorsman, J.C., et al., *Strong aggregation and increased toxicity of polyleucine over polyglutamine stretches in mammalian cells.* Human Molecular Genetics, 2002. **11**: p. 1487-1496.

341. Oma, Y., et al., *Comparative analysis of the cytotoxicity of homopolymeric amino acids.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2005. **1748**(2): p. 174-179.

342. van Eyk, C.L., et al., *Comparative toxicity of polyglutamine, polyalanine and polyleucine tracts in Drosophila models of expanded repeat disease.* Human Molecular Genetics, 2011. **21**(3): p. 536-547.

343. Gurezka, R., et al., *A Heptad Motif of Leucine Residues Found in Membrane Proteins Can Drive Self-assembly of Artificial Transmembrane Segments.* The Journal of Biological Chemistry, 1999. **274**(14): p. 9265-9270.

344. Whitley, P., et al., *A 12-Residue-long Polyleucine Tail Is Sufficient to Anchor Synaptobrevin to the Endoplasmic Reticulum Membrane.* The Journal of Biological Chemistry, 1996. **271**(13): p. 7583-7586.

345. Cascarina, S.M. and E.D. Ross, *Proteome-scale relationships between local amino acid composition and protein fates and functions.* PLoS computational biology, 2018. **14**(9): p. e1006256-e1006256.

346. Rolls, M.M., et al., *Targeting of Rough Endoplasmic Reticulum Membrane Proteins and Ribosomes in Invertebrate Neurons.* Molecular Biology of the Cell, 2002. **13**(5): p. 1778-1791.

347. Thomas, B.J., et al., *CemOrange2 fusions facilitate multifluorophore subcellular imaging in C. elegans.* PLOS ONE, 2019. **14**(3): p. e0214257.

348. Wahlberg, J.M. and M. Spiess, *Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain.* The Journal of Cell Biology, 1997. **137**(3): p. 555-562.

349. Lee, H., et al., *Glycosylatable GFP as a compartment-specific membrane topology reporter.* Biochemical and Biophysical Research Communications, 2012. **427**(4): p. 780-784.

350. Lee, H. and H. Kim, *Membrane topology of transmembrane proteins: determinants and experimental tools.* Biochemical and Biophysical Research Communications, 2014. **453**(2): p. 268-276.

351. Yang, M., et al., *The Transmembrane Domain of a Carboxyl-terminal Anchored Protein Determines Localization to the Endoplasmic Reticulum.* The Journal of Biological Chemistry, 1997. **272**(3): p. 1970-1975.

352. Lee, Z.Y., et al., *Compartmentalization of the endoplasmic reticulum in the early C. elegans embryos.* The Journal of Cell Biology, 2016. **214**(6): p. 665.

353. Broekhuis, J.R., et al., *SQL-1, homologue of the Golgi protein GMAP210, modulates intraflagellar transport in C. elegans.* Journal of Cell Science, 2013. **126**(8): p. 1785.

354. Kozutsumi, Y., et al., *The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins.* Nature, 1988. **332**(6163): p. 462-464.

355. Shen, X., et al., *Complementary Signaling Pathways Regulate the Unfolded Protein Response and Are Required for C. elegans Development.* Cell, 2001. **107**(7): p. 893-903.

356. Vidal, B., et al., *An atlas of Caenorhabditis elegans chemoreceptor expression.* PLoS Biology, 2018. **16**(1): p. e2004218-e2004218.

357. Ren, X.-C., et al., *Role of netrin UNC-6 in patterning the longitudinal nerves of Caenorhabditis elegans.* Journal of Neurobiology, 1999. **39**(1): p. 107-118.

358. Kraemer, B.C., et al., *Molecular pathways that influence human tau-induced pathology in Caenorhabditis elegans.* Human Molecular Genetics, 2006. **15**(9): p. 1483-1496.

359. *pmp-2*. 2019.

360. Imanaka, T., et al., *Characterization of the 70-kDa Peroxisomal Membrane Protein, an ATP Binding Cassette Transporter.* The Journal of Biological Chemistry, 1999. **274**(17): p. 11968-11976.

361. Geuze, H.J., et al., *Involvement of the endoplasmic reticulum in peroxisome formation.* Molecular biology of the cell, 2003. **14**(7): p. 2900-2907.

362. Sakaue, H., et al., *The N-terminal motif of PMP70 suppresses cotranslational targeting to the endoplasmic reticulum.* The Journal of Biochemistry, 2016. **159**(5): p. 539-551.

363. Gärtner, J., H. Moser, and D. Valle, *Mutations in the 70K peroxisomal membrane protein gene in Zellweger syndrome.* Nature Genetics, 1992. **1**(1): p. 16-23.

364. Afara, M.R., et al., *Rational Design of Peptide Inhibitors of the Sarcoplasmic Reticulum Calcium Pump.* Biochemistry, 2006. **45**(28): p. 8617-8627.

365. Akin, B.L., et al., *The Structural Basis for Phospholamban Inhibition of the Calcium Pump in Sarcoplasmic Reticulum.* The Journal of Biological Chemistry, 2013.

366. Hou, N.S., et al., *Activation of the endoplasmic reticulum unfolded protein response by lipid disequilibrium without disturbed proteostasis in vivo.* Proc Natl Acad Sci U S A, 2014. **111**(22): p. E2271-E2280.

367. Calfon, M., et al., *IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA.* Nature, 2002. **415**(6867): p. 92-96.

368. Matilainen, O., et al., *Insulin/IGF-1 Signaling Regulates Proteasome Activity through the Deubiquitinating Enzyme UBH-4.* Cell Reports, 2013. **3**(6): p. 1980-1995.

369. Yu, Z.-X., et al., *Mutant Huntingtin Causes Context-Dependent Neurodegeneration in Mice with Huntington's Disease.* The Journal of Neuroscience, 2003. **23**(6): p. 2193.

370. Duennwald, M.L. and S. Lindquist, *Impaired ERAD and ER stress are early and specific events in polyglutamine toxicity.* Genes & Development, 2008. **22**(23): p. 3308-3319.

371. Kalathur, R.K.R., et al., *The unfolded protein response and its potential role in Huntington's disease elucidated by a systems biology approach.* F1000Research, 2016. **4**: p. 103-103.

372. Shenkman, M., H. Eiger, and Z. Lederkremer Gerardo, *Genesis of ER Stress in Huntington's Disease*, in *Endoplasmic Reticulum Stress in Diseases*. 2015.

373. Scior, A., et al., *Directed PCR-free engineering of highly repetitive DNA sequences.* BMC Biotechnology, 2011. **11**: p. 87-87.

374. Yanik, M.F., et al., *Functional regeneration after laser axotomy.* Nature, 2004. **432**(7019): p. 822-822.

375. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nature Methods, 2012. **9**: p. 676.

376. Parslow, A., A. Cardona, and R.J. Bryson-Richardson, *Sample drift correction following 4D confocal time-lapse imaging.* Journal of visualized experiments : JoVE, 2014(86): p. 51086.

377. Steiernagle, T., *Maintenance of C. elegans* 1999, The *C. elegans* Research Community: http://www.wormbook.org.