The carboxyl-terminus of RAN translated GGGGCC nucleotide repeat expansions modulates toxicity in models of ALS/FTD.

Authors: Fang He[1,2*], Brittany N. Flores[1], Amy Krans[1], Michelle Frazer[1,3], Sam Natla[1], Sarjina Niraula[2], Olamide Adefioye[2], Sami J. Barmada[1], and Peter K Todd[1,4*]

Affiliations:

1) Department of Neurology, University of Michigan, Ann Arbor, MI, USA
2) Department of Biological and Health Sciences, Texas A&M University, Kingsville, TX, USA
3) Neuroscience Graduate Program, University of California at Los Angeles, Los Angeles, CA, USA
4) Veterans Affairs Medical Center, Ann Arbor, MI, USA

Contact:

Fang.he@tamuk.edu

700 University Blvd

Kingsville, TX 78363

USA

Petertod@umich.edu

4005 BSRB

109 Zina Pitcher Place

Ann Arbor, MI 48109-2200

USA

Key words: *C9ORF72*, Amyotrophic Lateral Sclerosis, Dipeptide repeat proteins (DPRs), repeat-associated non-AUG translation (RAN), Carboxyl Terminus, Drosophila.

Abstract:

An intronic hexanucleotide repeat expansion in *C9ORF72* causes familial and sporadic amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). This repeat is thought to elicit toxicity through RNA mediated protein sequestration and repeat-associated non-AUG (RAN) translation of dipeptide repeat proteins (DPRs). We generated a series of transgenic *Drosophila* models expressing GGGGCC (G_4C_2) repeats either inside of an artificial intron within a GFP reporter or within the 5' UTR of GFP placed in different downstream reading frames. Expression of 484 intronic repeats elicited minimal alterations in eye morphology, viability, longevity, or larval crawling but did trigger RNA foci formation, consistent with prior reports. In contrast, insertion of repeats into the 5' UTR elicited differential toxicity that was dependent on the reading frame of GFP relative to the repeat. Greater toxicity correlated with a short and unstructured carboxyl terminus in the glycine-arginine (GR) RAN protein reading frame. This change in C-terminal sequence triggered nuclear accumulation of all three RAN DPRs. A similar differential toxicity and dependence on the GR carboxyl terminus was observed when repeats were expressed in rodent neurons. The presence of the native carboxyl-termini across all three reading frames was partly protective. Taken together, these findings suggest that carboxyl terminal sequences outside of the repeat region may alter the behavior and toxicity of dipeptide repeat proteins derived from GGGGCC repeats.

Introduction

Nucleotide repeat expansions are a common cause of neurodegenerative disease. Transcribed repetitive sequences can elicit toxicity through at least two possible mechanisms. First, repeats as RNA can bind to and sequester specific proteins and preclude them from performing their normal functions [1]. Alternatively, repeats can be translated into proteins that accumulate and elicit toxicity in model systems [2]. Recently, it was recognized that translation of these repetitive sequences can occur in the absence of an AUG start codon through a process known as repeat associated non-AUG (RAN) translation [3,4]. As such, repeat expansions located outside of protein coding open reading frames can also produce toxic proteins.

In 2011, two groups simultaneously identified GGGGCC repeat expansions as a cause of amyotrophic lateral sclerosis and frontotemporal dementia (C9ALS/FTD) [5,6]. This mutation is common, explaining upwards of 40% of all familial cases of ALS and FTD and 10% of all sporadic cases of these conditions. Most people have only a few repeats, with a cutoff for normal defined as less than 22-28 [7]. Pathologic expansions are usually hundreds to thousands of repeats. The repeat resides the first intron of *C9ORF72*, a gene that is highly expressed in the brain but whose function remains unclear, though evidences indicate the encoded protein may act as guanine exchange factors for activating Rab proteins [8–10]. It is bi-directionally transcribed, such that both G_4C_2 and C_4G_2 repeat RNAs are produced, and RNA foci from both transcripts are observed in patient-derived cells and tissues [11–13]. Moreover, products of RAN translation are produced from both of these transcripts, leading to 6 different dipeptide repeat (DPR) containing proteins [11,12,14,15]. Repeat expansions also impact *C9ORF72* transcription, isoform choice, and splicing [5,14,16–18].

A series of model systems have provided insights into C9ALS/FTD pathogenesis. *Drosophila* models demonstrate repeat-associated toxicity when the repeat is placed within the 5' untranslated region (UTR; hereafter referred to as 5' leader) of a transgene such as GFP [19]. In contrast, long repeats interrupted by stop codons elicit little toxicity [20,21]. Flies expressing dipeptide repeat proteins via AUG initiated translation and independent of G_4C_2 repeat sequences are also toxic when expressed in flies and cells in some, but not all reading frames. Specifically, toxicity appears greatest with

expression of glycine-arginine and proline-arginine repeat proteins in *Drosophila,* with evidence for a role of glycine-alanine proteins in mammalian neurons and model systems [20–25]. Mouse models using interventricular adenoviral delivery of 66 or 149 G_4C_2 repeats in isolation exhibit RNA foci, RAN translation, neurodegeneration, and motor phenotypes [26]. Taken together, these studies do not preclude a role for the repeat RNA in disease toxicity, but suggest that RAN translation directly contributes to neurodegeneration in C9ALS/FTD.

Less attention has thus far been placed on the native sequence context of the repeat. The initial screening of expansion of G_4C_2 repeats near ALS loci found that though such repeats are quite common in human genome, the expansion of G_4C_2 repeats were only detectable in *C9ORF72* gene but not in other ALS loci genes [27], indicating that the context of the expanded repeats at *C9ORF72* gene are quite unique for expanded repeat toxicity. The sequence context may well be critical to determining the final location of the repeat RNA, its interactions with translational machinery, and the RNA binding proteins with which it interfaces. Tran et al. addressed this issue in *Drosophila* by providing evidence that (a) repeats located within introns are less toxic than repeats placed in 5'm^7G capped and polyadenylated transcripts, and (b) this enhanced toxicity was associated with increased DPR production [21,28]. However, sequence context may also alter the final translated peptides produced by RAN translation and this could influence their toxicity. For example, in Huntington disease, expression of a *huntingtin* (*HTT*) isoform containing exon 1 alone with the CAG repeat produces a highly toxic and aggregate-prone factor that accumulates in patient brains [29,30]. More recently, Sellier et al. demonstrated a role for the carboxy-terminus of FMRpolyG, a RAN translated protein derived from a CGG repeat expansion in the Fragile X gene, as critical for repeat toxicity in models of Fragile X-associated Tremor Ataxia Syndrome (FXTAS) [31]. We therefore sought to explore whether the repeat location within a transcript, and the sequences 3' to the repeat that might influence the final RAN translated products, have any influence on GGGGCC repeat toxicity in disease models.

Here we show that expression of up to 484 interrupted G_4C_2 repeats within a *Drosophila* intron exhibit minimal toxicity, largely consistent with prior studies [20,21,28].

In contrast, relatively short (28) repeats placed in the 5' UTR of GFP elicited marked toxicity in *Drosophila* in some, but not all, reading frames relative to a downstream reporter. This differential toxicity correlated with the size and content of the carboxyl terminus of the GR reading frame- with larger carboxyl termini or GFP fusions leading to low toxicity. The relative toxicity observed in these *Drosophila* models correlates with nuclear accumulation of GR DPR-containing proteins and is recapitulated in rodent neurons. The inclusion of the native carboxyl terminus provides some neuroprotective effects. Taken together, these findings suggest that repeat context has a significant influence on its toxicity and identifies a new role for the carboxyl terminus in modulating DPR toxicity across model systems.

Materials and Methods

Construction of mammalian transfection vectors

Constructs for mammalian expression experiments s were generated using by digesting GFP $(G_4C_2)_{71}$ pcDNA3.2 (a gift from Christopher Shaw) with XbaI andinserted into pcDNA3.1 containing GFP. One to 2 nucleotides were inserted upstream of the GFP to result in each reading frame. Using NheI and PmeI restriction enzymes, the $(G_4C_2)_{71}$ GFP fragment was inserted into pGW for neuronal expression. To generate plasmids with or without the native C-terminus, a small fragment was inserted using AscI and MfeI, replacing GFP in pGW. The ATG-V5 fragments were inserted upstream of the repeat using Acc65I and NotI in pGW.

Drosophila Fly Stocks/Genetics

All flies were maintained with standard food and culture conditions at 25˚C, while all crosses were made and maintained at 29˚C unless otherwise stated. Fly lines from the Bloomington Stock Center are: GMR-GAL4 (#8605), and UAS-GFP (multiple lines). Actin5C-GAL4/CyO (a gift from Zhe Han), motor neuron specific OK6-GAL4 (a gift from Cathy Collins), RU-486 inducible Tubulin-GAL4 (tub5) driver line (gift from Scott Pletcher).

The sequences for intronic and 5' leader fly lines are detailed in Supplemental Tables 1. Briefly, for intronic repeat lines, a mini intron from the fly *Prospero* gene [32] was inserted into the middle of GFP, synthesized commercially (Genewiz, NJ), and cloned into the EcoRI-XbaI site of pUAST. An intronic fragment with either 3 or 28 G_4C_2 repeats along with 40nt of upstream and 260nt of downstream intronic sequence from human *C9ORF72* was PCR amplified according to [33] from genomic DNA and inserted into the intron. The 21 repeat tiling fragment was generated using primer NotI-C9 AnchorR+PspOMI tiling (Supplemental Table 2). The 28 repeat insert was generated using primer NotI-C9R+XhoI-C9F directly from genomic DNA of an ALS patient with 28 repeats. Repeat blocks of $(G_4C_2)_{21}$ bracketed by NotI and PspOMI restriction sites were then serially inserted next to this repeat, producing intermediate 49, 70, 91, and 121 repeat containing constructs. This element was then concatamerized to generate lines with 242 (GFP-iC_{242}) or 484 (GFP-iC_{484}) interrupted repeats.

5' leader repeat lines were generated by cloning the same 28 repeat sequence described above with approximately 30nt on either side of intronic sequence into the 5'UTR of pEGFP-N1 (Takara Bio USA) in the 0+ reading frame. The repeat and GFP were then subcloned into the NotI site in pUAST. To induce frameshifts, annealing primers were inserted into the AgeI site between the repeat and the AUG of GFP (Supplemental Table 2). pUAST vectors carrying the respective inserts were used to generate transgenic lines by standard p-element insertion (BestGene, CA).

For the fly lines with or without native C-terminus, the sequence flanking the $(G_4C_2)_{69}$ repeats in constructs for neuron transfection were digested by KpnI and XbaI, and the inserts were cloned to the vector pUAST-AttB (Drosophila Genomics Resource Center, IN) digested with the same set of enzymes. The resulting pUAST-AttB vectors were then used for site-specific transgenesis using PhiC31 integrase technique to locus AttP40 [34] (BestGene, CA). All constructs were verified by Sanger sequencing and agarose gel sizing (Supplemental Figure S1A).

Eye phenotype imaging and quantification

Representative fly eye images were taken by Leica M125 stereomicroscope and photographed with a Leica DFC425 digital camera as previously described [35]. Eye morphology of 1-2 day post eclosion flies was quantitatively scored as previously described [36]. Briefly, we used the following criteria: supernumerary inter-ommatidial bristles, abnormal bristle orientation, ommatidium fusion, ommatidium pitting, disorganization of ommatidial array, retinal collapse. The presence of each feature will be given 1 point, and additional 2 points were given if more than 5% of the eyes were affected, or 4 points if more than 50% of eyes were affected. Higher score means the eyes are more degenerated. Over 100 flies were scored per genotype in at minimum of three independent crosses. Scores were calculated and are presented as mean ± SEM.

Fly viability and eclosion rates

Analysis of eclosion rates were performed as described [37]. Each transgenic line was crossed to Actin5C-GAL4 (act5C-GAL4) ubiquitous driver balanced over a marker chromosome (CyO), on standard food at 29˚C, if the transgene elicited no toxicity, then 50% of progeny should have the CyO marker and 50% express the transgene. Over 100 flies of each genotype were scored over multiple crosses. The relative percent progeny carrying the transgene were expressed as a percent of total eclosed flies. Analysis of viability post eclosion was performed as described [38].

Larval crawling assay

Transgenic lines were crossed to the motor neuron specific driver line OK6-GAL4 at 29˚C, and the 3rd instar larvae with desired genotypes were collected. Crawling assay was performed with slight modification from [39]: 3rd instar larvae were collected by adding 20% sucrose on the food, briefly washed with MilliQ water, and then maintained in a water droplet until being analyzed. The larvae were accommodated at room temperature on 2% agarose gel in 10cm petri dish for 1min, and then the crawling distance in the next 1min was recorded and processed by ImageJ. At least 50 larvae from the appropriate genotype were recorded and pooled as the same genotype.

***Drosophila* Lifespan Assay**

The UAS transgenic lines were crossed to Tub5-GAL4 (tub5) Geneswitch driver flies on standard food absent of RU-486 at 29˚C and adult offspring of the desired genotypes were collected 2-3 days after eclosion and transferred to standard fly food without yeast granules containing 200μM RU-486. The flies were transferred to fresh food with drug every 3-4 days. Each genotype started with at least 4 vials of 25 flies/vial (2 vials of males and 2 vials of females) and the survival was determined daily or every other day for 50 days or until all flies were dead. For each genotype, at least 2 individual lines were examined.

Generation of specific dipeptide antibodies

The rabbit polyclonal antibodies were generated by Abclonal (Cambridge, MA). Generation of (GA) and (GR) specific antibodies was performed as previously described [40]. To detect GP, we raised antisera against a synthetic $(GP)_6$ peptide. All antibodies were affinity purified prior to usage.

Western Blotting

Protein quantification was as previously described and imaged on standard film [41,42]. Briefly, fly head or cell lysates were lysed in RIPA buffer with protease inhibitors (Roche) and passed through a 27-gauge needle to shear DNA. Equal amounts of protein were run on a 12% SDS polyacrylamide gel. After transfer to PVDF membrane, blots were incubated with the following antibodies: monoclonal mouse anti- GFP (Sigma, MO; clone 7.1 and 13.1, 1:1000), mouse anti-Tubulin (DHSB, IA; clone E7, 1:1,000), mouse anti-V5 (Thermo Fisher, CA; 1:1000), rabbit anti-GA (1:100), rabbit anti-GP (1:5000), rabbit anti-GR (1:5000) or mouse anti-β-Actin (Sigma, MO; 1:5000). At least three independent experiments were performed and scanned films were processed and quantified using ImageJ software.

In Situ Hybridization

1-3d post eclosion fly heads from each genotype crossed to GMR-GAL4 driver were isolated, immediately frozen in OCT media, and then cryosectioned to 10μm. Transverse sections were then fixed with 4% paraformaldehyde in 1x PBS for 15min,

washed 3x in 1x PBS, and permeabilized with 2% acetone for 5min. *In situ* hybridization was performed as follows: sections were prehybridizied with 50% formamide in 2x SSC for 30min in room temperature and then hybridized with 500μl hybridization solution in a light proof box for overnight in 37˚C. The hybridization solution is as follows: 0.6ng/ml Cy5 labeled 2'-O-Me-$(CCCCGG)_5$ RNA probe (IDT DNA Technologies, IA), 0.02% BSA, 132mg/ml yeast RNA (Sigma, MO), 1μl RNAse inhibitor (Sigma), 50% formamide in 2x SSC. After hybridization, slides were washed with 3x 30min of 50% formamide in 0.5x SSC at 56˚C, followed by 3x10min wash of 0.5x SSC wash at room temperature. After washes, samples were dried for 5min, incubated with 100μl Prolong Gold with DAPI (Thermo Fisher, MA) for 1 hour, and examined on an Olympus FV1000 confocal microscope with identical laser settings for each slide. Images were overlaid and quantitated in ImageJ software.

Cell based Immunofluorescence

COS-7 cells were maintained 37°C in 5% CO_2 incubators. Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% Pen-Strep was used as culture media. Cells were transfected using Lipofectamine LTX with Plus Reagent (Thermo Fisher, MA) using manufacturer's protocol. For immunofluorescent detection of GFP, cells were cultured on 4-well chamber slides. 48 hours after transfection, cells were fixed with 4% paraformaldehyde for 15min, washed, permeabilized 0.1% triton X-100, blocked with 5% normal goat serum in 1x PBS containing 0.1% triton X-100, and incubated with GFP and GA (1:100), GP (1:500), or GR (1:500) antibodies overnight. Slides were washed and probed with AlexaFluor 488 labeled goat anti-mouse (Thermo Fisher, MA; 1:500) and Alexa Fluor 555 labeled goat anti-rabbit (Thermo Fisher, MA; 1:500) antibodies and visualized with an Olympus epifluorescence microscope with Slidebook 5.5 software with identical fluorescent settings for each slide.

RNA isolation and qRT-PCR

Total RNA was extracted from 15 fly heads of each genotype using Trizol (ThermoFisher) and quantitated by Nano-Drop (Thermo Fisher, MA). 1μg of total RNA

were then reverse-transcribed to cDNA by iScript cDNA synthesis kit (Bio-Rad, CA). The primers used are detailed in Supplemental Table 2. PCR analysis was performed using the iQ SYBR Green Supermix on a myIQ Single Color qPCR system (BioRad, CA). All runs included a standard dilution curve representing 2x to 0.02x of the RNA concentration utilized for all primer sets to ensure linearity. Equivalent efficiency of individual primer sets was confirmed prior to data analysis. The level of *GFP* was normalized to *RPL32* mRNA for each sample run and expressed as a ratio of levels to GFP-iC3 lines (fold control expression) unless otherwise stated. All samples were run in triplicate in each qPCR run and all data represent at least three independent experiments.

Primary neuron cultures & transfection

All mammalian rodent work was approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Michigan. Primary mixed cortical neurons were dissected from embryonic day 19-20 Long-Evans rat pups and cultured at 6.0×10^5 cells/ml in 96 well cell culture plates as previously described [43–45]. Brains of a single litter were combined to maximize cell counts, resulting in neurons from a mixed population of male and female pups. Neurons were cultured in NEUMO photostable medium containing SOS supplement (Cell Guidance Systems, MO) at 37°C in 5% CO_2. After 4 days in culture, neurons were transfected with 0.2μg DNA (total) and 0.5μL Lipofectamine 2000 (Thermo Fisher, CA) per well in a 96-well plate, as described previously [45]. Neurons were incubated with the DNA/Lipofectamine 2000 mix for 20min at 37°C before rinsing. The remainder of the protocol followed the manufacturer's protocol.

Longitudinal fluorescence microscopy

Automated longitudinal fluorescence microscopy began 24 hours post-transfection for 10 days, as previously described [43–45]. Images were acquired using an inverted microscope (Nikon Instruments, NY) with a 20x objective lens, a PerfectFocus system, a Lambda XL Xenon lamp (Sutter Instruments, CA) with 5mm liquid light guide (Sutter Instruments, CA), and either an Andor iXon3 897 EMCCD

camera or Andor Zyla4.2 (+) sCMOS camera. All stage, shutter, and filter wheel movements were done using a custom code written in μManager, ImageJ [45].

Statistical Analysis

For most fly comparisons, scores were analyzed using a non-parametric Kruskal-Wallis ANOVA with Dunn's correction for multiple comparisons. A Log-rank (Mantel-Cox) Chi square test was performed for *Drosophila* survival rates. Error bars represent the standard error of the mean except for proportion numbers, where the error bars represent the 95% confidence interval.

Results

Generation of an intronic repeat model of C9ALS/FTD in *Drosophila*

To create an intronic repeat model of C9ALS/FTD, we inserted the first intron of the fly *Prospero* gene [32] into the sequence of GFP in a fashion that retained the canonical splice site donor and acceptor sequences (Figure 1A). We then inserted either three G_4C_2 repeats or concatamers of 28 and 21 G_4C_2 repeat units with short interruptions into the intron, up to a maximal repeat length of 484 repeats (Supplemental Figure S1, Table S1). After initial studies demonstrated no significant phenotype associated with 28 or 49 intronic repeats (not shown), we focused our analysis on lines expressing 3 (iC_3), 242 (iC_{242}), or 484 (iC_{484}) intronic repeats.

The G_4C_2 repeat as RNA can form a strong G-quadruplex secondary structure that might interfere with RNA metabolism and splicing [18,46,47]. To determine if the repeat taken out of its native context is capable of eliciting alterations in mRNA splicing, we measured the total, unspliced, and spliced mRNA from fly lines containing 3, 242, or 484 intronic repeats (Figure 1B). There was no significant difference in the abundance of spliced *GFP* RNA across repeat sizes (Figure 1B). Consistent with this, the amount of GFP protein expressed in the intronic repeat flies was comparable to that seen in flies expressing *GFP* alone and did not decline with larger repeat sizes (Figure 1C). In contrast, there was a significant repeat-length dependent increase in unspliced pre-mRNA containing the G_4C_2 repeat when normalized to *RPL32* (Figure 1B). Additionally,

RNA foci were observed in flies expressing long G_4C_2 repeats but not short repeats (Figure 1D).

We utilized the UAS-GAL4 system to target expression of the intronic repeats to different tissues. Expression of GFP-iC_3, GFP-iC_{242}, or GFP-iC_{484} in the eye using a GMR-GAL4 driver elicited modest toxicity in GFP-iC_{242} eye phenotypes compared to a control line expressing GFP alone using a standardized toxicity scale [35] (Figure 1E). No significant eye phenotypes were observed in GFP-iC_{484} expressing lines. Similarly, activating ubiquitous expression during development using an Actin5C-GAL4 driver elicited no differences in eclosion rates between expanded intron lines and the control GFP line (Figure 1F). We also selectively activated expression of the intronic repeat transgenes in motor neurons using an OK6-GAL4 driver line and evaluated larval crawling as a motor phenotype. We observed no difference between short and long intronic repeats compared to flies expressing *GFP* alone (Figure 1G). To evaluate whether a reduction in lifespan might be elicited in adulthood, we activated transgene expression in adult male flies after eclosion using a Geneswitch-Tubulin-GAL4 driver by addition of RU-486 to the fly food. We observed no significant impact of intronic repeat expression on viability (Figure 1H).

The role of repeat location in GGGGCC toxicity

To determine if the repeat location within a transcript might impact its toxicity, we inserted the same 28 G_4C_2 repeat cassette used to generate the intronic repeats into the 5' leader of GFP rather than an intron (Figure 2A). No start codons were present between the transcription start site and the repeat, although the ATG of GFP was retained 3' to the repeat. We introduced small frameshift mutations between the repeat and the open reading frame of GFP so that it acted as a reporter for RAN translation in the 0+ (GA), 1+ (GP), or 2+ (GR) reading frames, respectively.

When the G_4C_2 repeat was expressed in an intronic context, levels of GFP protein were similar between short repeats and expanded repeats. However, placing the repeat in the 5' leader of *GFP* dramatically affected overall *GFP* expression. A higher molecular weight GFP fusion protein product was observed by western blot in the 0+ (GA) frame, but undetectable in the GP (+1) or GR (+2) reading frames (Figure 2B).

Expression of all three constructs led to detectable RNA foci (Figure 2C). The presence and abundance of RNA foci did not directly predict toxicity in these flies, similar to prior observations [28].

When expressed in fly ommatidia, 5' leader repeat fly lines elicited greater toxicity than was observed in the intronic repeat fly lines across a range of measures, consistent with a prior publication [28]. However, surprisingly, there were significant differences in the toxicity observed that was dependent on the reading frame in which the GFP reporter was placed. Specifically, 1+ $(G_4C_2)_{28}$ GFP fly lines elicited a marked degeneration of the eye (Figure 3A) and to a lesser extent in 2+ $(G_4C_2)_{28}$ GFP flies. Developmental toxicity, as assessed by eclosion rates after ubiquitous expression, and a severe larval motor phenotype observed with isolated motor neuron expression, largely mirrored toxicity findings observed in the eye (Figure 3B-C). Moreover, rapid declines in viability after transgene activation in adulthood were observed in all 5' leader flies, with the most robust toxicity in 1+ $(G_4C_2)_{28}$ GFP fly lines (Figure 3D).

In an attempt to explain these differences in phenotype, we compared transgene mRNA expression across lines and observed no correlation between toxicity and abundance of the transgene expression (Supplemental Figure 1B). We next analyzed three independent lines for each transgene to control for effects of transgene insertion sites. Phenotypes for different lines of the same transgene were very similar despite significant differences across transgenes (Supplemental Figure 1C). Of note, the DNA sequence 5' to the repeat in all three transgene constructs were identical, precluding contributions from an upstream initiation event in only one construct as a driving force in the observed phenotypic differences (Supplemental Table 1).

Accumulation and localization of DPR protein correlates with toxicity

To determine what factors might be triggering the differential toxicity between these *Drosophila* lines, we used a series of new and recently developed [40] antibodies against the three sense strand DPRs: GA, GR, and GP. These polyclonal antibodies recognize epitope tagged versions of the indicated repeat proteins both by western blot and by immunocytochemistry in transfected cells (Supplemental Figure 2A-B). These

antibodies also stain perinuclear inclusions in cerebellar slices from ALS patients with *C9ORF72* repeat expansions but not in healthy controls (Supplemental Figure 2C).

We utilized these antibodies to determine the relative abundance and distribution of each DPR in all of the transgenic *Drosophila* lines described. We reasoned that the epitope tags provide information on a single reading frame, but that RAN translation occurs in all 3 reading frames and thus the relative abundance of each product might be informative for their roles in toxicity. Using transverse ommatidial sections derived from *Drosophila* expressing each transgene (UAS-GFP, GFP-iC_3, GFP-iC_{484}, and 0+, 1+, or 2+ $(G_4C_2)_{28}$ GFP lines) crossed to GMR-GAL4, we performed immunocytochemistry against each of the three DPRs using these new antibodies. No staining was observed in lines expressing GFP in isolation or in the driver line alone. However, staining was consistently present with both GA and GR antibodies in the intronic fly models, with significantly more staining at larger repeat sizes (Figure 4). In contrast, GP staining was not reliably observed from the intronic fly lines. In 5' UTR lines there were marked differences in the staining for both GR and GP between lines, with the greatest staining present in the +1 $(G_4C_2)_{28}$ GFP lines that exhibited the greatest toxicity (Figure 4C-D). In addition, this same line exhibited a marked increase in both total nuclear staining and in the nuclear-cytoplasmic ratio for all three DPR proteins (Figure 4D), suggesting a correlation between both abundance and cellular localization of these proteins and the observed toxicity in *Drosophila* [20,25,28,48].

Prior studies have demonstrated that carboxyl terminal sequences and epitope tags can influence the toxicity associated with translated repeat expansions such as polyglutamine proteins [30]. More recently, the carboxyl terminus of the RAN translated protein FMRpolyG that is generated from GGC repeats in FXTAS was found to influence the relative toxicity of this protein through aberrant protein-protein interactions [31]. We therefore analyzed the C-termini of the transgenes used to generate the different exonic flies (Table 1). In the native *C9ORF72* gene, each DPR reading frame has 30-55 amino acids after the repeat prior to the presence of a stop codon. Such C-termini were detectable in ALS patients carrying the expanded G_4C_2 repeats according to previous studies [13]. In our exonic flies, 2 reading frames have C-termini over 200

amino acids (one of these frames is GFP) while the third is much shorter (15-21 amino acids).

Based on these observations, we hypothesized that the shorter C-termini in the GR reading frame allowed for stabilization of this product within cells that favored its redistribution into the nucleus and a subsequent activation of toxicity. To explore this idea further, we utilized longitudinal microscopy [31,43–45] to track survival of individual neurons expressing sequences similar to those used in the fly experiments (Figure 5A). Primary rodent cortical neurons were transfected with plasmids encoding mApple, to visualize neuronal soma, and each of the 5' leader G_4C_2 constructs. These cells were imaged at 24h intervals over a 10d period, and the time of death for individual neurons determined programmatically using a set of sensitive criteria (rounding or blebbing of the cell body or neurites, loss of fluorescence) validated in previous studies [49,50]. Relative survival in each condition was compared by Cox proportional hazards analysis. As we observed in *Drosophila,* expression of an expanded G_4C_2 repeat in the 5' UTR of GFP was toxic in neurons. The relative toxicity of each construct was again dictated by the reading frame of the GFP tag, with toxicity being greatest in those with 1+ G_4C_2 GFP or 2+ G_4C_2 GFP, mirroring what was observed in *Drosophila* (Figure 5A).

To evaluate the potential role of different carboxyl terminal sequences on toxicity, we generated a new series of expression vectors that take into account the length and sequence of the C-terminus. One set of vectors have the native C-terminus (CT) for all three reading frames. Another set of vectors was generated with shortened C-terminal regions in all three reading frames (ΔCT) (Figure 5B). The G_4C_2 repeat with the native C-terminal sequence showed increased toxicity with an increasing repeat size (Figure 5B, solid green vs solid blue line). However, when the native C-terminus was removed and replaced with a short C-terminus, toxicity was significantly increased (Figure 5B, dashed lines vs solid lines). These $(G_4C_2)_{69}$ repeats with the native C-terminus (CT) or with a short C-terminus (ΔCT) were also introduced to the same locus (AttP40) in *Drosophila* using pUAST-AttB vectors to achieve equal expression levels. When expressed in the eye, inclusion of the the C-terminus partially suppressed the repeat-associated toxicity (Figure 5C).

In an attempt to determine which C-terminal reading frame was most critical for toxicity, we introduced a series of AUG driven V5 tags into the 5' UTR upstream of the repeat in different reading frames. Enhancing production of GA or GR from these repeats significantly boosted the toxicity of these constructs to a greater degree than placement of an AUG codon in the GP reading frame. However, replacing the native C-terminal sequences with stop codons only enhanced repeat toxicity when the AUG codon was placed in the GR reading frame (Figure 6C-E). Taken together, these data suggest that the native C-terminus may mitigate toxicity arising from RAN translation, and in particular toxicity related to GR DPR proteins.

Discussion

How hexanucleotide repeat expansions in *C9ORF72* elicit neurodegeneration is a topic of intense research [51,52]. Work by multiple groups has established the presence of both nuclear RNA foci and RAN translation derived DPR proteins in tissues and cells from affected patients. Initial studies largely relied on expression of the repeat out of its normal sequence context or production of DPR proteins via AUG-initiated translation. These studies importantly demonstrate that even relatively short RNA repeats and small DPRs can be toxic [19,20,22–26,53–55]. The position of the GGGGCC repeat within an mRNA is a major factor in determining its relative toxicity in *Drosophila,* with repeats residing in an intron exhibiting minimal toxicity while even short repeats in the 5' leader elicit significant toxicity [28]. In this work we observe that sequences 3' to the repeat also modulate repeat toxicity by altering the carboxyl termini of DPRs generated via RAN translation. These findings suggest a specific role for C-terminal sequences in modulating repeat toxicity in C9ALS/FTD while making it clear that repeat sequence context needs to be carefully considered in future studies of disease pathogenesis.

Sequences located in coding regions or within the 5' or 3' UTR are typically processed to capped and polyadenylated mRNAs and rapidly trafficked out of the nucleus, where they interface with a series of RNA binding proteins and the translational machinery. In contrast, intronic sequences are typically spliced out during transcription, formed into RNA lariats and targeted for debranching and degradation in the nucleus.

As such, one potential explanation for the observed differences in toxicity between intronic and 5' leader sequence repeats is the distribution of the repeat containing transcripts within the cell and their potential for translation. *In vitro* and cell based studies using reporter systems suggest that while GGGGCC repeats can be translated from intronic or bicistronic transcripts, their production is most robust when present in capped and polyadenylated mRNAs [56–60]. Consistent with this prediction, the abundance of specific RAN translation products from relatively short 5' leader repeats were much greater than that observed from larger intronic repeats (Figures 1-4 and [28]). Moreover, the relative abundance and distribution of these DPR proteins was predictive of toxicity, while the total RNA abundance and RNA foci counts were similar across repeat containing lines, consistent with prior studies [21,28]. While studies in mammalian or human model systems at endogenous expression levels are needed to fully arbitrate the relative roles of RNA and protein in GGGGCC repeat toxicity, our work is consistent with the emerging consensus supporting a significant role of RAN translation in C9ALS/FTD pathogenesis.

In contrast, our finding that the frame in which the C-terminal epitope tag resides was strongly predictive of repeat toxicity is surprising. This finding was quite robust, as we observed this same relationship across multiple different *Drosophila* lines with distinct insertion sites (Supplemental Figure S1) and also when these same constructs were expressed in rodent neurons (Figure 6). The "GFP frame" where the toxicity was greatest would create a fusion with a GP dipeptide repeat protein—the least toxic DPR based on multiple studies using AUG driven expression. Thus, it could be that "out of frame" untagged products are largely driving toxicity while the GFP tag is actually suppressing toxicity derived from certain DPRs. In this context, it is intriguing that the small sequence changes we introduced below the repeat to shift the reading frame resulted in a short disordered 20 amino acid C-terminal tail in the GR reading frame of the repeat in the 1+ $(G_4C_2)_{28}$ GFP flies with the most severe phenotypes, with longer C-terminal tails in the GR frame in both the 2+ $(G_4C_2)_{28}$ GFP and 0+$(G_4C_2)_{28}$GFP flies (Figure 3A, Table 1, and Supplemental Tables 1 and 2). This C-terminal change correlated with both more aggregates of GR DPR in 1+ $(G_4C_2)_{28}$ GFP flies and greater nuclear localization of both the GR and other DPRs.

Given that isolated expression of GR dipeptide repeats in the absence of the RNA repeat exhibits marked toxicity in *Drosophila,* rodent and cell based model systems [20,25,53,61,62], changes in its relative solubility, stability or distribution may well influence the toxicity elicited by G_4C_2 repeats. In contrast, the relative lack of toxicity associated with the 0+ $(G_4C_2)_{28}$ GFP transgene may also be partially explained by a stabilizing effect of a GFP tag on the GA DPR. This GA-GFP fusion product remains largely cytoplasmic and soluble in our assays, which may impede GR mediated toxicity in *Drosophila* [25]. Regardless, our results suggest that the C-terminal region of RAN derived proteins influences their relative toxicity and cellular distribution: a finding with significant implications for design of future model systems at both this and other disease causing repeat expansions.

In rodent neurons, we observed a protective effect when the native C-terminus is retained in all three reading frames. Based on studies where we serially enhanced expression of each DPR reading frame by inserting an AUG codon in a strong Kozak context above the GGGGCC repeat element, this protective effect seems to be largely driven by the GR reading frame, as elimination of the C-termini in GR driven constructs enhanced toxicity while doing so in GA or GP driven constructs had no significant or even a mildly protective effect. Given the propensity for GR DPRs to phase separate, this native C-terminus may influence the behavior of these peptides when generated in patients. Future studies will be needed to determine if and how such surrounding sequences influence the biochemical behavior of DPRs in general, and GR DPRs in particular.

In summary, our findings demonstrate that the position of the repeat within a transcript as well as its immediate surrounding sequences can have a major impact on its relative toxicity, and on the distribution and abundance of different RAN translated proteins. Future work will be needed to determine the selective effects that these sequence modifications have on the biophysical properties of RAN translated proteins and their exact roles in disease pathogenesis in ALS and FTD. When coupled with recent findings in FXTAS [31] and Spinocerebellar type 36 [63,64], these findings suggest that similar analyses in other repeat disorders are likely warranted.

List of Abbreviations

ALS: Amyotrophic lateral sclerosis; FTD: frontotemporal dementia; RAN translation: repeat-associated non-AUG initiated translation; DPR: Dipeptide repeats; UTR: untranslated region; FXTAS: fragile X-associated tremor/ataxia syndrome; GA:glycine-alanine; GP: glycine-proline; GR: glycine-arginine

Declarations:

Ethics approval and consent to participate

All mammalian rodent work was approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Michigan.

Consent for publication

Not applicable.

Availability of data and material

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was funded by NIH R01NS086810, R01NS099280, the Department of Veterans Affairs BLRDs 1I21BX001841 and I01BX003231, and a pilot grant from ALS Association to PKT. Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions:

FH and PKT designed the study. FH, MF, AK, BF, SN, SN, and OA conducted the experiments. FH, AK, and PKT wrote the manuscript. All authors edited the manuscript.

Acknowledgements:

We thank Scott Pletcher, Zhe Han, and Cathy Collins for sharing fly lines. We thank Michael Sutton and the MBNI at Michigan for use of their Confocal Microscope. Reagents generated in the Drosophila Genome research center were supported by NIH Grant 2P40OD010949.

References:

1. Todd PK, Paulson HL. RNA-mediated neurodegeneration in repeat expansion disorders. Ann Neurol. 2010;67:291–300.

2. Zoghbi HY, Orr HT. Glutamine Repeats and Neurodegeneration. Annu Rev Neurosci. 2000;23:217–47.

3. Cleary JD, Ranum LPW. Repeat associated non-ATG (RAN) translation: New starts in microsatellite expansion disorders. Curr Opin Genet Dev. 2014;26:6–15.

4. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, et al. Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A. 2011;108:260–5.

5. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron. 2011;72:245–56.

6. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011;72:257–68.

7. Woollacott IOC, Mead S. The C9ORF72 expansion mutation: Gene structure, phenotypic and diagnostic issues. Acta Neuropathol. 2014;127:319–32.

8. Farg MA, Sundaramoorthy V, Sultana JM, Yang S, Atkinson RAK, Levina V, et al. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum Mol Genet. 2014;23:3579–95.

9. Levine TP, Daniels RD, Gatta AT, Wong LH, Hayes MJ. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics. 2013;29:499–503.

10. Zhang D, Iyer LM, He F, Aravind L. Discovery of novel DENN proteins: Implications for the evolution of eukaryotic intracellular membrane structures and human disease.

Front Genet. 2012;3:283.

11. Zu T, Liu Y, Bañez-Coronel M, Reid T, Pletnikova O, Lewis J, 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:E4968-4977.

12. Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PEA, Caulfield T, et al. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. 2013;126:829–44.

13. Mori K, Arzberger T, Grässer FA, Gijselinck I, May S, Rentzsch K, et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. 2013;126:881–93.

14. Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science (80-). 2013;339:1335–8.

15. Ash PEA, Bieniek KF, Gendron TF, Caulfield T, Lin W-L, Dejesus-Hernandez M, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron. 2013;77:639–46.

16. Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, et al. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: A gene identification study. Lancet Neurol. 2012;11:54–65.

17. Liu EY, Russ J, Wu K, Neal D, Suh E, McNally AG, et al. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. 2014;128:525–41.

18. Haeusler AR, Donnelly CJ, Periz G, Simko EAJ, Shaw PG, Kim MS, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature. 2014;507:195–200.

19. Xu Z, Poidevin M, Li X, Li Y, Shu L, Nelson DL, et al. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc Natl Acad Sci U S A. 2013;110:7778–83.

20. Mizielinska S, Grönke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, et al. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science (80-). 2014;345:1192–4.

21. Moens TG, Mizielinska S, Niccoli T, Mitchell JS, Thoeng A, Ridler CE, et al. Sense and antisense RNA are not toxic in Drosophila models of C9orf72-associated ALS/FTD. Acta Neuropathol. 2018;135:445–57.

22. May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, et al. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol. 2014;128:485–503.

23. Zhang YJ, Jansen-West K, Xu YF, Gendron TF, Bieniek KF, Lin WL, et al. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. Acta Neuropathol. 2014;128:505–24.

24. Wen X, Tan W, Westergard T, Krishnamurthy K, Markandaiah SS, Shi Y, et al. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate invitro and invivo neuronal death. Neuron. 2014;84:1213–25.

25. Yang D, Abdallah A, Li Z, Lu Y, Almeida S, Gao FB. FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. Acta Neuropathol. 2015;130:525–35.

26. Chew J, Gendron TF, Prudencio M, Sasaguri H, Zhang YJ, Castanedes-Casey M, et al. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. Science (80-). 2015;348:1151–4.

27. He F, Jones JM, Figueroa-Romero C, Zhang D, Feldman EL, Goutman SA, et al. Screening for novel hexanucleotide repeat expansions at ALS-and FTD-associated loci.

Neurol Genet. 2016;2:e71.

28. Tran H, Almeida S, Moore J, Gendron TF, Chalasani UD, Lu Y, et al. Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. Neuron. 2015;87:1207–14.

29. Ross CA, Tabrizi SJ. Huntington's disease: From molecular pathogenesis to clinical treatment. Lancet Neurol. 2011;10:83–98.

30. Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, et al. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci U S A. 2013;110:2366–70.

31. Sellier C, Buijsen RAM, He F, Natla S, Jung L, Tropel P, et al. Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron. 2017;93:331–47.

32. Scamborova P, Wong A, Steitz JA. An Intronic Enhancer Regulates Splicing of the Twintron of Drosophila melanogaster prospero Pre-mRNA by Two Different Spliceosomes. Mol Cell Biol. 2004;24:1855–69.

33. Meisler MH, Grant AE, Jones JM, Lenk GM, He F, Todd PK, et al. C9ORF72 expansion in a family with bipolar disorder. Bipolar Disord. 2013;15:326–32.

34. Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat Genet. 2008;40:476–83.

35. Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron. 2013;78:440–55.

36. Pandey UB, Nie Z, Batlevi Y, McCray BA, Ritson GP, Nedelsky NB, et al. HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. Nature. 2007;447:859–63.

37. Lanson NA, Maltare A, King H, Smith R, Kim JH, Taylor JP, et al. A Drosophila model of FUS-related neurodegeneration reveals genetic interaction between FUS and TDP-43. Hum Mol Genet. 2011;20:2510–23.

38. Todd PK, Oh SY, Krans A, Pandey UB, Di Prospero NA, Min KT, et al. Histone deacetylases suppress cgg repeat-induced neurodegeneration via transcriptional silencing in models of Fragile X Tremor Ataxia Syndrome. PLoS Genet. 2010;6:e1001240.

39. Nichols CD, Becnel J, Pandey UB. Methods to assay Drosophila behavior. J Vis Exp. 2012;61:3795.

40. Flores BN, Dulchavsky ME, Krans A, Sawaya MR, Paulson HL, Todd PK, et al. Distinct c9orf72-associated dipeptide repeat structures correlate with neuronal toxicity. PLoS One. 2016;e0165084.

41. He F, Krans A, Freibaum BD, Paul Taylor J, Todd PK. TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. Hum Mol Genet. 2014;23:5036–51.

42. Iliff AJ, Renoux AJ, Krans A, Usdin KU, Sutton MA, Todd PK. Impaired activity-dependent FMRP translation and enhanced mGluR-dependent LTD in fragile X premutation mice. Hum Mol Genet. 2013;22:1180–92.

43. Flores BN, Li X, Malik AM, Martinez J, Beg AA, Barmada SJ. An Intramolecular Salt Bridge Linking TDP43 RNA Binding, Protein Stability, and TDP43-Dependent Neurodegeneration. Cell Rep. 2019;27:1133–50.

44. Malik AM, Miguez RA, Li X, Ho YS, Feldman EL, Barmada SJ. Matrin 3-dependent neurotoxicity is modified by nucleic acid binding and nucleocytoplasmic localization. Elife. 2018;7:e35977.

45. Weskamp K, Safren N, Miguez R, Barmada S. Monitoring neuronal survival via longitudinal fluorescence microscopy. J Vis Exp. 2019;143:10.3791/59036.

46. Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, et al. The

C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. Elife. 2016;5.

47. Reddy K, Zamiri B, Stanley SYR, Macgregor RB, Pearson CE. The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J Biol Chem. 2013;288:9860–6.

48. Schludi MH, May S, Grässer FA, Rentzsch K, Kremmer E, Küpper C, et al. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol. 2015;130:537–55.

49. Arrasate M, Finkbeiner S. Automated microscope system for determining factors that predict neuronal fate. Proc Natl Acad Sci U S A. 2005;102:3840–5.

50. Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature. 2004;431:805–10.

51. Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. Nat Rev Neurol. 2018;14:544–58.

52. Gitler AD, Tsuiji H. There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS. Brain Res. 2016;1647:19–29.

53. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science (80-). 2014;345:1139–45.

54. Tao Z, Wang H, Xia Q, Li K, Li K, Jiang X, et al. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. Hum Mol Genet. 2015;24:2426–41.

55. Yamakawa M, Ito D, Honda T, Kubo KI, Noda M, Nakajima K, et al. Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. Hum Mol Genet. 2015;24:1630–45.

56. Green KM, Glineburg MR, Kearse MG, Flores BN, Linsalata AE, Fedak SJ, et al. RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. Nat Commun. 2017;8:2005.

57. Cheng W, Wang S, Mestre AA, Fu C, Makarem A, Xian F, et al. C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. Nat Commun. 2018;9:51.

58. Sonobe Y, Ghadge G, Masaki K, Sendoel A, Fuchs E, Roos RP. Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. Neurobiol Dis. 2018;116:155–65.

59. Tabet R, Schaeffer L, Freyermuth F, Jambeau M, Workman M, Lee CZ, et al. CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. Nat Commun. 2018;9:152.

60. Westergard T, McAvoy K, Russell K, Wen X, Pang Y, Morris B, et al. Repeat-associated non- AUG translation in C9orf72- ALS / FTD is driven by neuronal excitation and stress. EMBO Mol Med. 2019;11:e9423.

61. Lopez-Gonzalez R, Lu Y, Gendron TF, Karydas A, Tran H, Yang D, et al. Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron. 2016;92:383–91.

62. Zhang YJ, Gendron TF, Ebbert MTW, O'Raw AD, Yue M, Jansen-West K, et al. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. Nat Med. 2018;24:1136–42.

63. Todd TW, McEachin ZT, Chew J, Burch AR, Jansen-West K, Tong J, et al. Hexanucleotide Repeat Expansions in c9FTD/ALS and SCA36 Confer Selective Patterns of Neurodegeneration In Vivo. Cell Rep. 2020;In press.

64. McEachin ZT, Gendron TF, Raj N, García-Murias M, Banerjee A, Purcell RH, et al. Chimeric Peptide Species Contribute to Divergent Dipeptide Repeat Pathology in

c9ALS/FTD and SCA36. Neuron. 2020;In press.

Table 1: RAN Translation Peptides in G_4C_2 repeat transgenic flies and plasmids

Genotype	In $(GA)_n$ frame	In $(GP)_n$ frame	In $(GR)_n$ frame
C9ORF72 sense strand	$(GA)_n$ Cterminal 31 AAs $(GA)_n$WSGRARGRARG GAAVAVPAPAAAEAQ AVASG*	$(GP)_n$ Cterminal 34 AAs $(GP)_n$GRGRGGPGGGP GAGLRLRCLRPRRRR RRRWRVGE*	$(GR)_n$ Cterminal 54 AAs $(GR)_n$GVVGAGPGAGPGR GCGCGACARGGGGAGG GEWVSEEAASWRVAVW GSAAGKRRG*
0+ $(G_4C_2)_{28}$ GFP	$(GA)_{28}$ Cterminal 253 AAs (GFP) $(GA)_{28}$GAWSGRARDP PVAT-GFP*	$(GP)_{28}$ Cterminal 15 AAs *$(GP)_{28}$GRGRGGPGIHRSPPW*	$(GR)_{28}$ Cterminal 258 AAs $(GR)_{28}$GVVGAGPGST... AAYNHYK*
1+ $(G4C2)_{28}$ GFP	$(GA)_{28}$ Cterminal 264 AAs $(GA)_{28}$GAWSGRARDP... KRPRL*	$(GP)_{28}$ Cterminal 258 AAs (GFP) *$(GP)_{28}$GRGRGGPGIHRS DLEPVAT-GFP*	$(GR)_{28}$ Cterminal 20 AAs $(GR)_{28}$GVVGAGPGSTGQIS NRSPPW*
2+ $(G4C2)_{28}$ GFP	$(GA)_{28}$ Cterminal 21 AAs $(GA)_{28}$GAWSGRARDPP VRSRNRSPPW*	$(GP)_{28}$ Cterminal 264 AAs *$(GP)_{28}$GRGRGGPGIH.... KRPRL*	$(GR)_{28}$ Cterminal 258 AAs (GFP) $(GR)_{28}$GVVGAGPGSTGQ ISKPVAT-GFP*
ATG-V5 GA/GP/GR CT	$(GA)_{71}$ Cterminal 39 AAs $(GA)_{71}$GRGRVGAPWSG RARGRARGGAAVAVP APAAAEAQAVASG*	$(GP)_{71}$ Cterminal 40 AAs $(GP)_{71}$VVEGWARRGRG GPGGGPGAGLRLRCLR PRRRRRRRWRVGE*	$(GR)_{71}$ Cterminal 60 AAs $(GR)_{71}$SWKGGRAVVGAGP GAGPGRGCGCGACARGG GGAGGGEWVSEEAASWR VAVWGSAAGKRRG*
ATG-V5 GA/GP/GR ΔCT	(GA) Cterminal 9 AAs $(GA)_{71}$GRGRVGAPE*	$(GP)_{71}$ Cterminal 10 AAs $(GP)_{71}$VVEGWARLSE*	$(GR)_{71}$ Cterminal 7 AAs (GR)71 SWKGGRA*

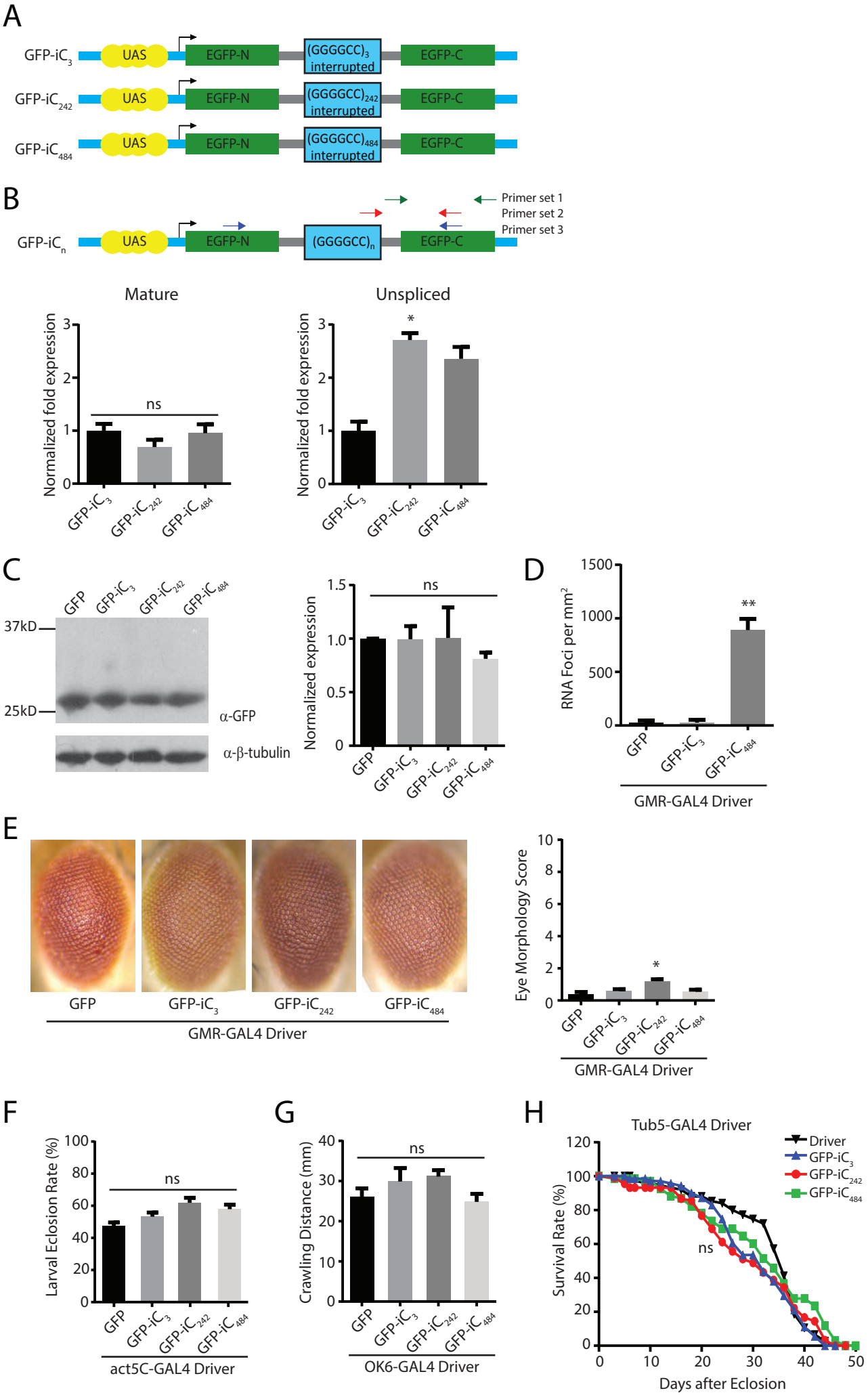
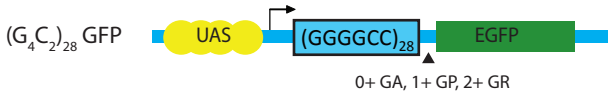
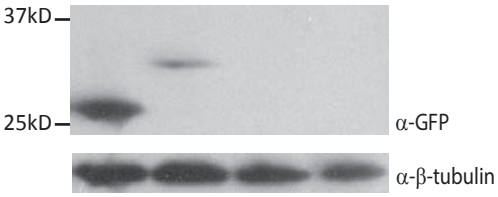

A
GFP-iC3
UAS
EGFP-N
(GGGGCC)3 interrupted
EGFP-C
GFP-iC242
(GGGGCC)242 interrupted
GFP-iC484
(GGGGCC)484 interrupted
B
GFP-iCn
(GGGGCC)n
Primer set 1
Primer set 2
Primer set 3
Mature
Unspliced
Normalized fold expression
ns
*
C
GFP
37kD
25kD
α-GFP
α-β-tubulin
Normalized expression
D
RNA Foci per mm2
**
GMR-GAL4 Driver
E
Eye Morphology Score
F
Larval Eclosion Rate (%)
act5C-GAL4 Driver
G
Crawling Distance (mm)
OK6-GAL4 Driver
H
Tub5-GAL4 Driver
Driver
Survival Rate (%)
Days after Eclosion

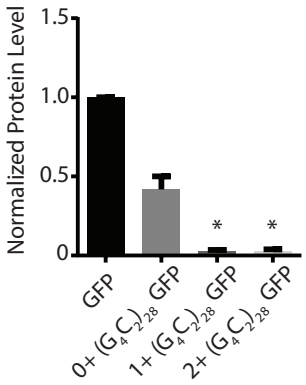
Figure 1: Intronic G_4C_2 repeats influence splicing and elicit RNA foci in *Drosophila.* A) Schematic of constructs used to generate transgenic fly lines. The *Prospero* fly intron 1 (gray lines) was introduced into UAS-GFP (blue lines). Either three G_4C_2 repeats or serial $(G_4C_2)_{21\text{-}28}$ repeat units separated by 14nt interruptions were serially inserted into the intron. B) Locations of the primer pairs used for measuring unspliced, spliced, and total mRNA are show in the schematic. Quantification of the expression of the mature (left) and unspliced (right) GFP product in the indicated fly lines (right). C) Western blot of lysates from heads of G_4C_2 intronic repeat flies (left) with quantification of GFP normalized to beta-tubulin (right). D) Quantification of RNA foci in the retina of the indicated intronic fly. E) Representative eye phenotypes from flies of the indicated genotypes crossed to gmr-GAL4 to drive expression in developing ommatidia (left) and quantification of eye phenotype scores (right). F) Quantification of the number of progeny carrying the transgene after flies of the indicated genotypes were crossed to a ubiquitous driver (act5c-GAL4). G) Quantification of the distance crawled by 3rd instar larvae from the crosses of the indicated fly genotypes to a larval motor neuron specific driver (ok6-GAL4). H) Flies carrying the indicated transgenes were crossed to a Tubulin Geneswitch driver (tub5) to allow ubiquitous expression post eclosion. Adult male flies were placed on food containing RU-486 to activate gene expression and their viability was tracked over 50 days. Graphs represent means ± SEM. * $p<0.05$; ** $p<0.01$ by Kruskal-Wallis after Dunn's correction for multiple comparisons.

A


$(G_4C_2)_{28}$ GFP
UAS
$(GGGGCC)_{28}$
EGFP
0+ GA, 1+ GP, 2+ GR

B


37kD
25kD
α-GFP
α-β-tubulin


Normalized Protein Level
1.5
1.0
0.5
0
*
*
GFP
GFP
GFP
GFP
0+ $(G_4C_2)_{28}$
1+ $(G_4C_2)_{28}$
2+ $(G_4C_2)_{28}$

C

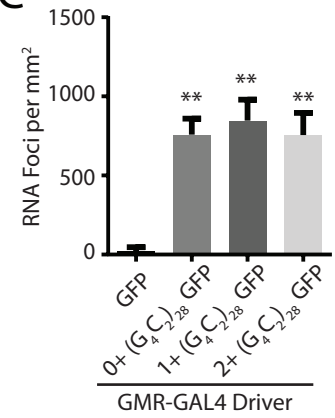

RNA Foci per mm^2
1500
1000
500
0
**
**
**
GFP
GFP
GFP
GFP
0+ $(G_4C_2)_{28}$
1+ $(G_4C_2)_{28}$
2+ $(G_4C_2)_{28}$
GMR-GAL4 Driver

Figure 2: *Drosophila* 5'UTR G_4C_2 repeat models. A) Schematic of constructs used to generate transgenic fly lines. $(G_4C_2)_{28}$ repeats were inserted into the 5'UTR of GFP without an upstream start codon in all three reading frames relative to the GFP reading frame. B) Western blot (left) of lysates from heads of G_4C_2 exonic repeat flies normalized to beta-tubulin (right). C) Quantification of the presence of RNA foci in the retina of the indicated fly lines with. Graphs represent means ± SEM. * $p<0.05$; ** $p<0.01$ by Kruskal-Wallis after Dunn's correction for multiple comparisons.

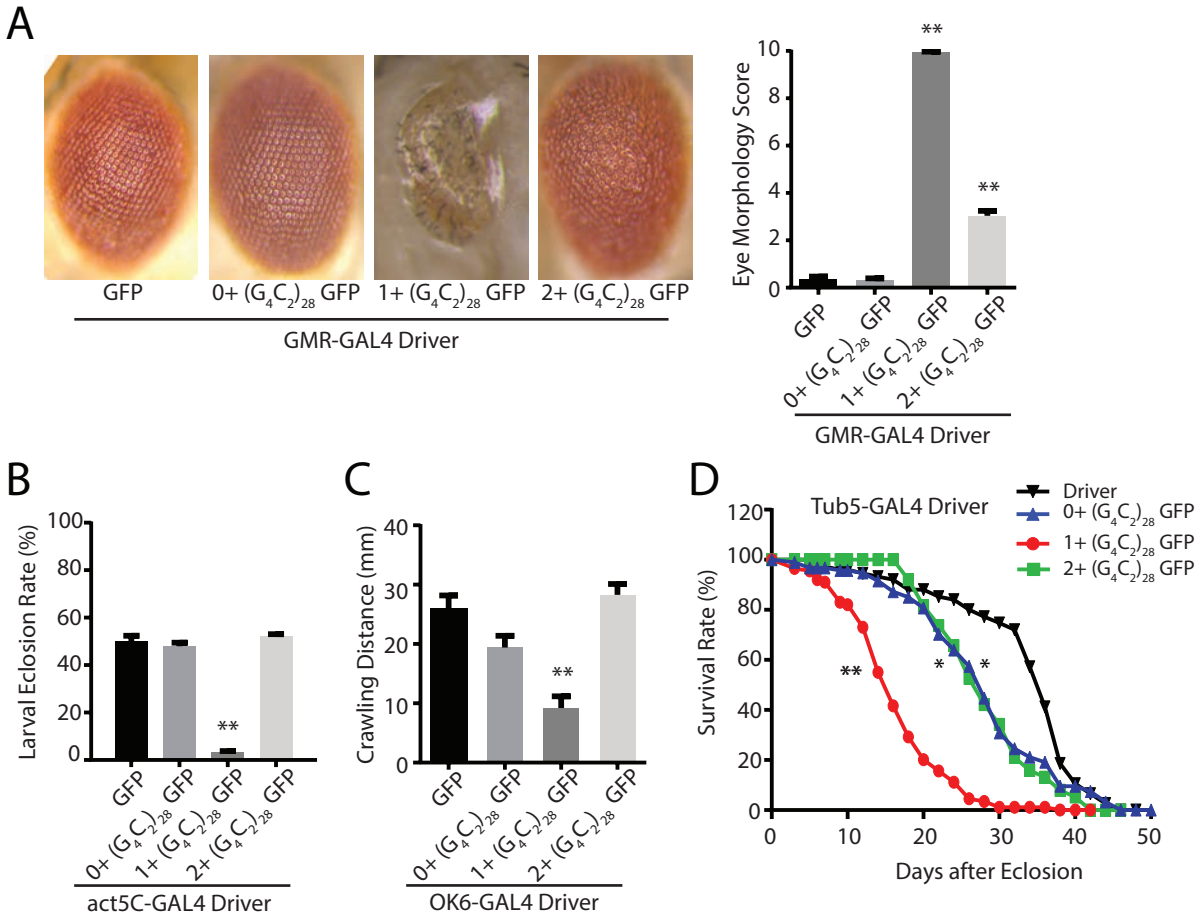
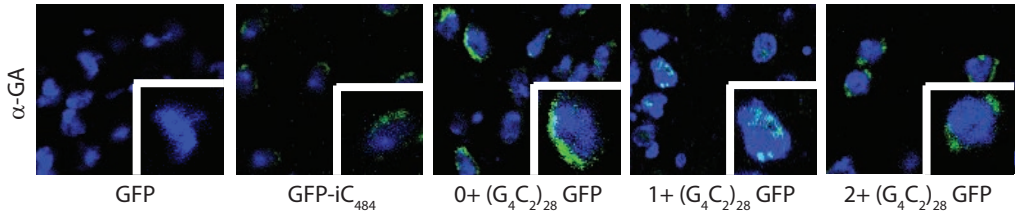
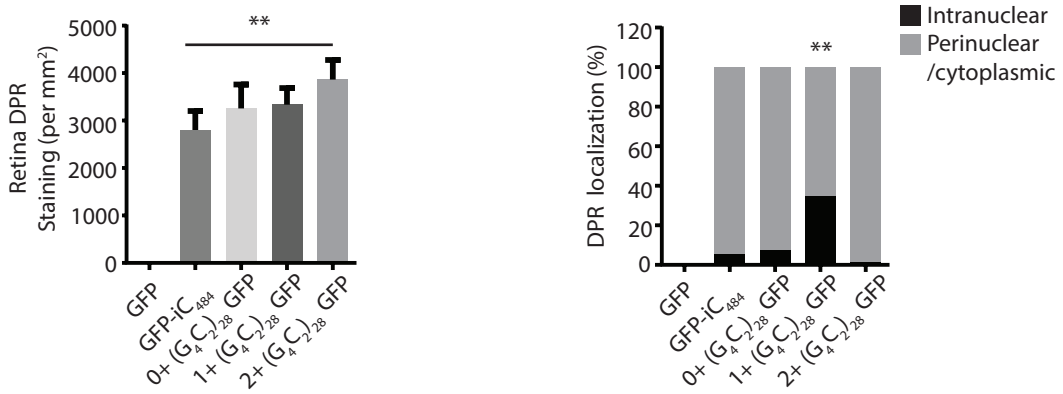

A
GFP
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP
GMR-GAL4 Driver
Eye Morphology Score
**
**
B
Larval Eclosion Rate (%)
**
act5C-GAL4 Driver
C
Crawling Distance (mm)
**
OK6-GAL4 Driver
D
Tub5-GAL4 Driver
Driver
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP
Survival Rate (%)
**
*
*
Days after Eclosion

Figure 3: 5'UTR G_4C_2 repeats with different Carboxyl termini elicit differential toxicity in *Drosophila* A) Representative eye phenotypes from flies of the indicated genotypes crossed to GMR-GAL4 to drive expression in developing ommatidia (right) and quantification of eye phenotype scores (right). B) The number of progeny carrying the transgene was determined after flies of the indicated genotypes were crossed to a ubiquitous driver (act5c-GAL4). C) Quantification of the distance crawled by 3rd instar larvae from the crosses of the indicated fly genotypes to a larval motor neuron specific driver (ok6-GAL4). D) Flies carrying the indicated transgenes were crossed to a Tubulin Geneswitch driver (tub5) to allow ubiquitous expression post eclosion. Adult male flies were placed on food containing RU-486 to activate gene expression and their viability was tracked over 50 days. Graphs represent means ± SEM. ** $p<0.01$ by Kruskal-Wallis after Dunn's correction for multiple comparisons. Log-rank (Mantel-Cox) test for survival, * $p<0.05$, ** $p<0.01$.

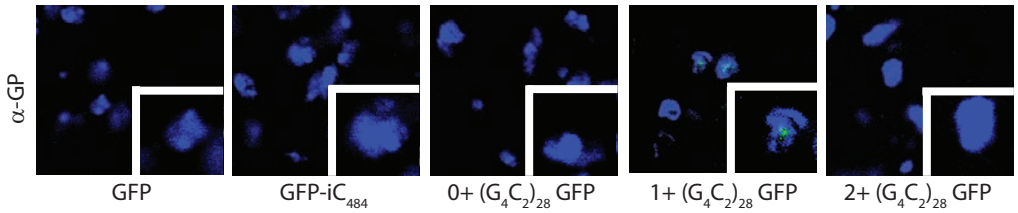
A


α-GA
GFP
GFP-iC484
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP

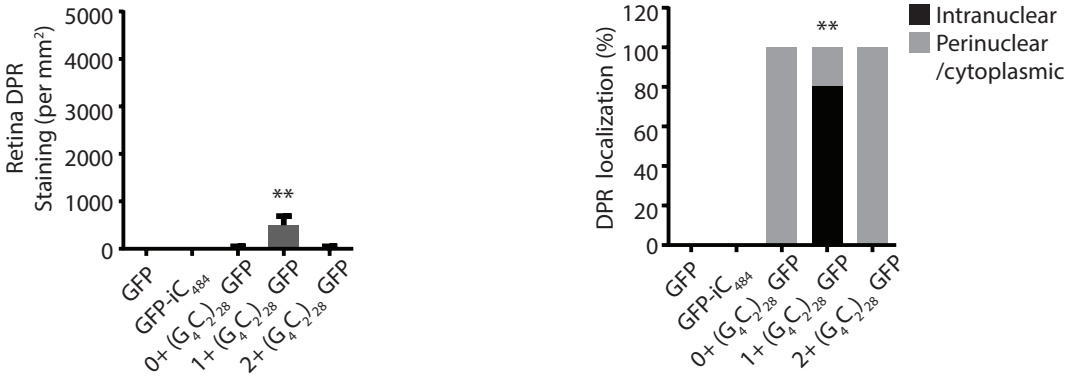
B


Retina DPR Staining (per mm2)
5000
4000
3000
2000
1000
0
**
GFP
GFP-iC484
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP
DPR localization (%)
120
100
80
60
40
20
0
Intranuclear
Perinuclear /cytoplasmic

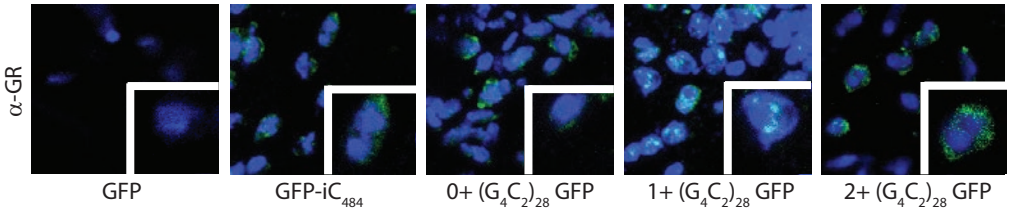
C


α-GP
GFP
GFP-iC484
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP

D


Retina DPR Staining (per mm2)
5000
4000
3000
2000
1000
0
**
GFP
GFP-iC484
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP
DPR localization (%)
120
100
80
60
40
20
0
Intranuclear
Perinuclear /cytoplasmic

E


α-GR
GFP
GFP-iC484
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP

F

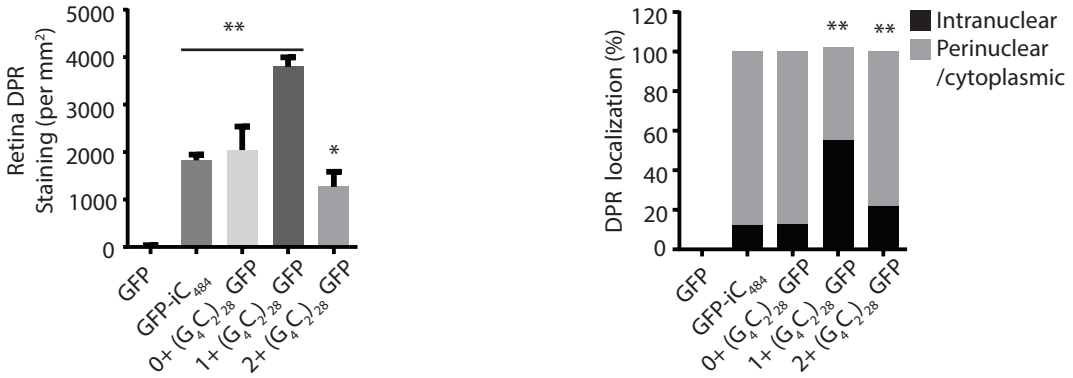
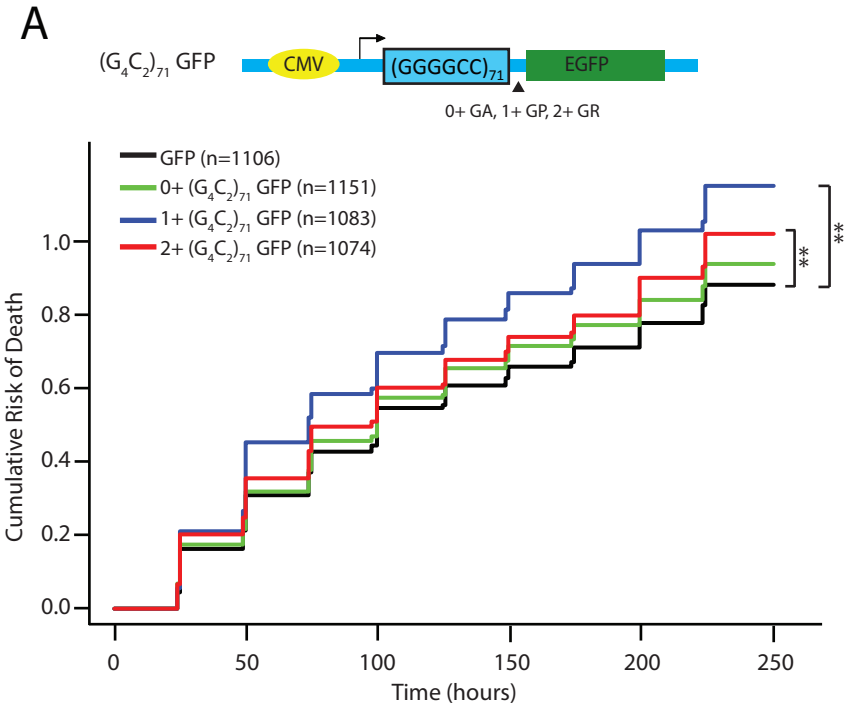
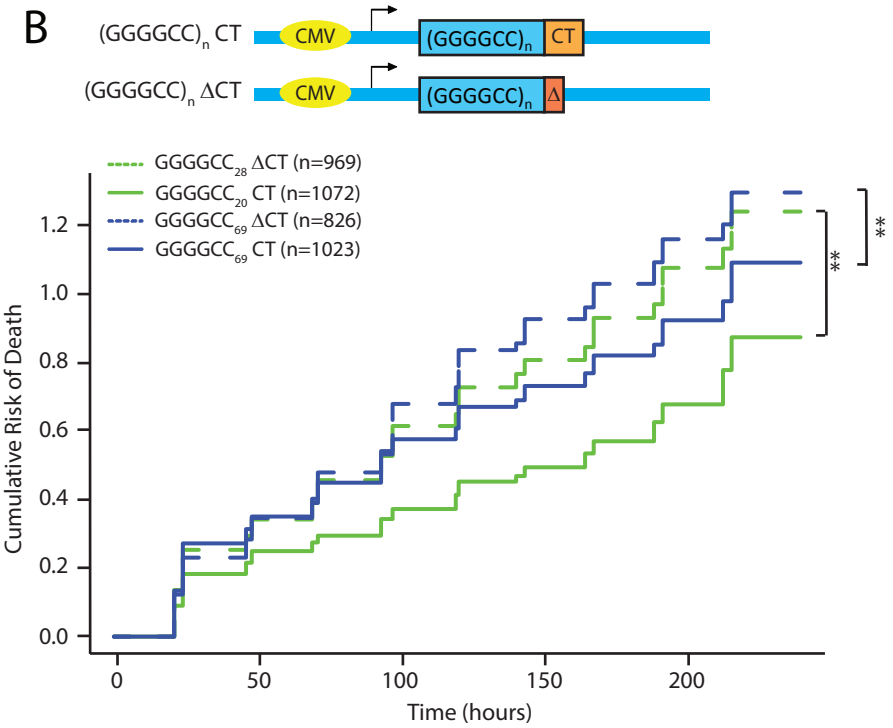

Retina DPR Staining (per mm2)
5000
4000
3000
2000
1000
0
**
*
GFP
GFP-iC484
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP
DPR localization (%)
120
100
80
60
40
20
0
**
**
Intranuclear
Perinuclear /cytoplasmic

Figure 4: G_4C_2 repeats toxicity is associated with nuclear accumulation of RAN translation peptides. A) Representative images of transverse retinal sections from *Drosophila* of the indicated genotypes probed with antibody against GA dipeptide repeats. B) Quantification of total GA repeat staining by immunofluorescence (left) and the percentile of cells with significant nuclear accumulation of GA DRP staining for the indicated genotypes (right). C) Representative Immunofluorescence images of *Drosophila* retinal sections probed with antibodies against the GP dipeptide repeats. D) Quantification of total GP repeat staining by immunofluorescence (left) and the percentile of cells with significant nuclear accumulation of GP DRP staining for the indicated genotypes (right). E) Representative Immunofluorescence image of transverse retinal sections from *Drosophila* of the indicated genotypes probed with antibody against GR dipeptide repeats. F) Quantification of total GR repeat staining by immunofluorescence (left) and the percentile of cells with significant nuclear accumulation of GR DRP staining for the indicated genotypes. Graphs represent means ± SEM. * $p<0.05$; ** $p<0.01$ by Kruskal-Wallis after Dunn's correction for multiple comparisons.

A


$(G_4C_2)_{71}$ GFP
CMV
$(GGGGCC)_{71}$
EGFP
0+ GA, 1+ GP, 2+ GR
GFP (n=1106)
0+ $(G_4C_2)_{71}$ GFP (n=1151)
1+ $(G_4C_2)_{71}$ GFP (n=1083)
2+ $(G_4C_2)_{71}$ GFP (n=1074)
Cumulative Risk of Death
Time (hours)
**
**

B


$(GGGGCC)_n$ CT
CMV
$(GGGGCC)_n$
CT
$(GGGGCC)_n$ ΔCT
CMV
$(GGGGCC)_n$
Δ
$GGGGCC_{28}$ ΔCT (n=969)
$GGGGCC_{20}$ CT (n=1072)
$GGGGCC_{69}$ ΔCT (n=826)
$GGGGCC_{69}$ CT (n=1023)
Cumulative Risk of Death
Time (hours)
**
**

C

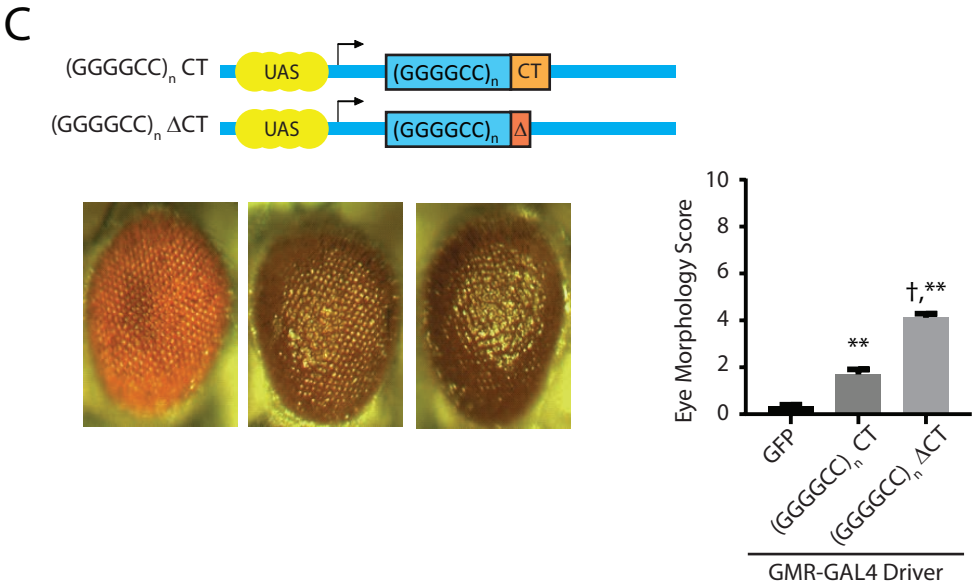

$(GGGGCC)_n$ CT
UAS
$(GGGGCC)_n$
CT
$(GGGGCC)_n$ ΔCT
UAS
$(GGGGCC)_n$
Δ
Eye Morphology Score
**
†,**
GFP
$(GGGGCC)_n$ CT
$(GGGGCC)_n$ ΔCT
GMR-GAL4 Driver

Figure 5: Native Carboxyl terminal sequences reduce GGGGCC repeat toxicity in rodent neurons and in flies. A) Schematic of the plasmids transfected into neurons. Cumulative risk of death in each reading frame when the GGGGCC repeat is in the 5' leader sequence of GFP. B) Schematic of the plasmids transfected into neurons. Cumulative risk of death of neurons with normal and expanded G_4C_2 repeats with and without the native Cterminus. ** $p<0.01$ by Cox proportional hazard analysis. C) Schematic of the plasmids used to generate transgenic fly lines. Representative eye phenotypes from flies of the indicated genotypes crossed to gmr-GAL4 to drive expression in developing ommatidia (left) and quantification of eye phenotype scores (right). Graphs represent means ± SEM.** $p<0.01$ by Kruskal-Wallis after Dunn's correction for multiple comparisons compared to control group. † $p<0.01$ by Kruskal-Wallis after Dunn's correction for multiple comparisons compared to $(G_4C_2)_{69}$ CT group.

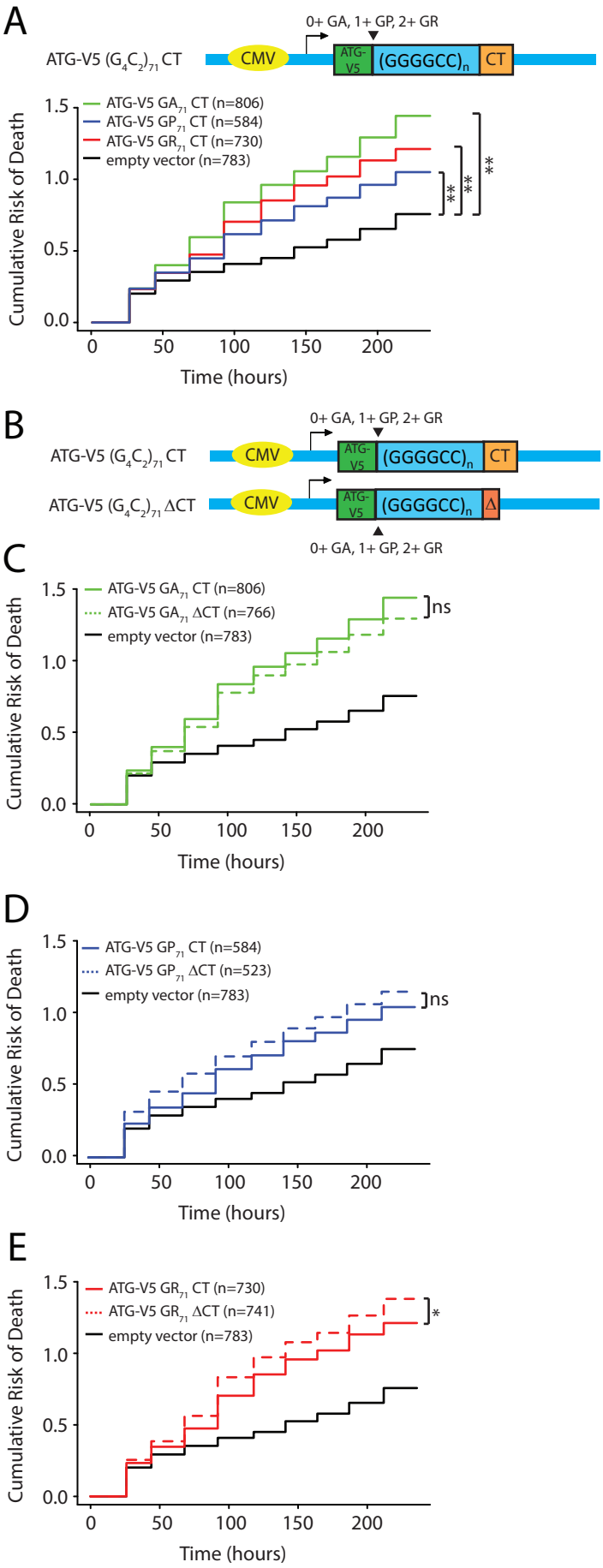
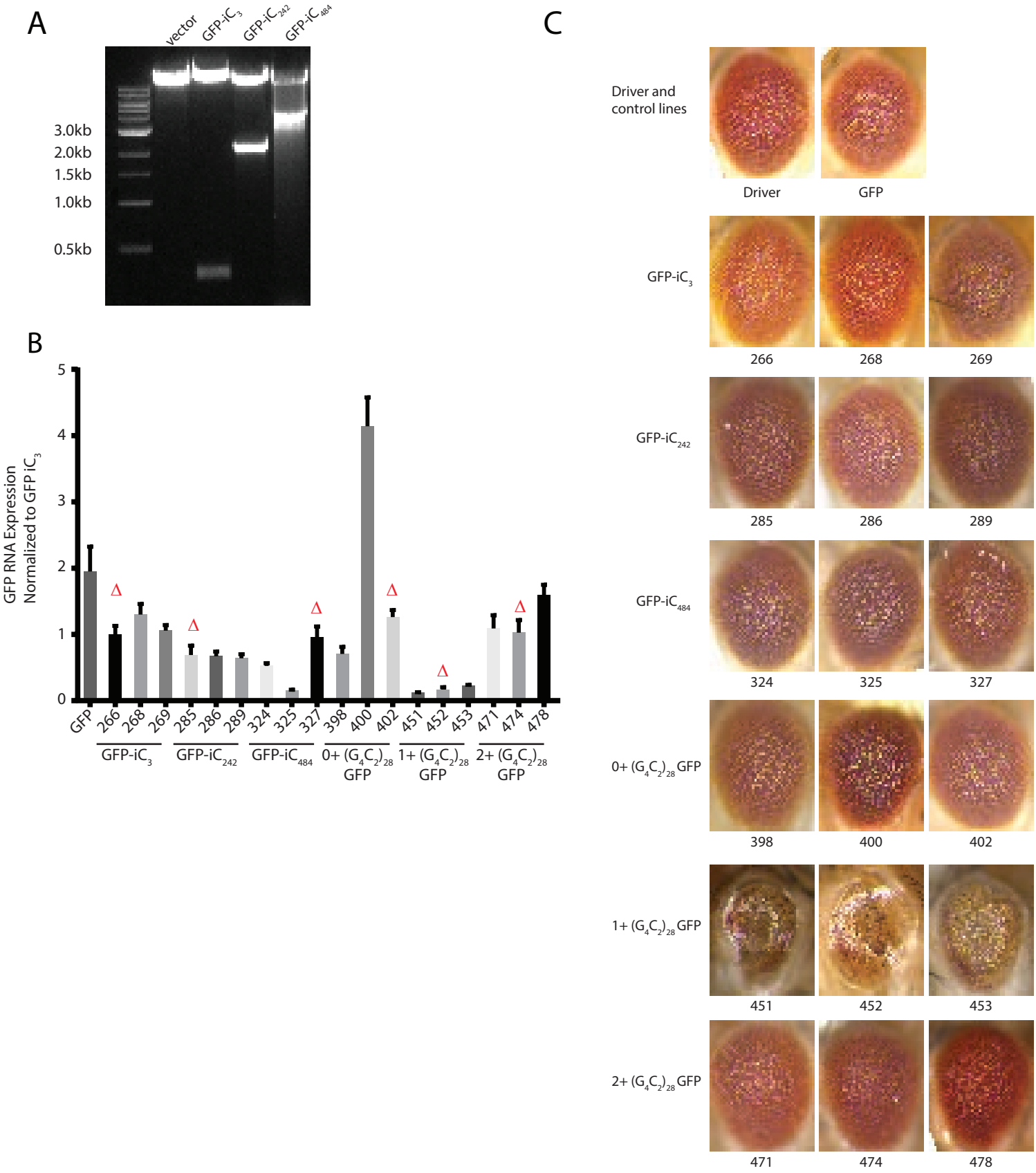
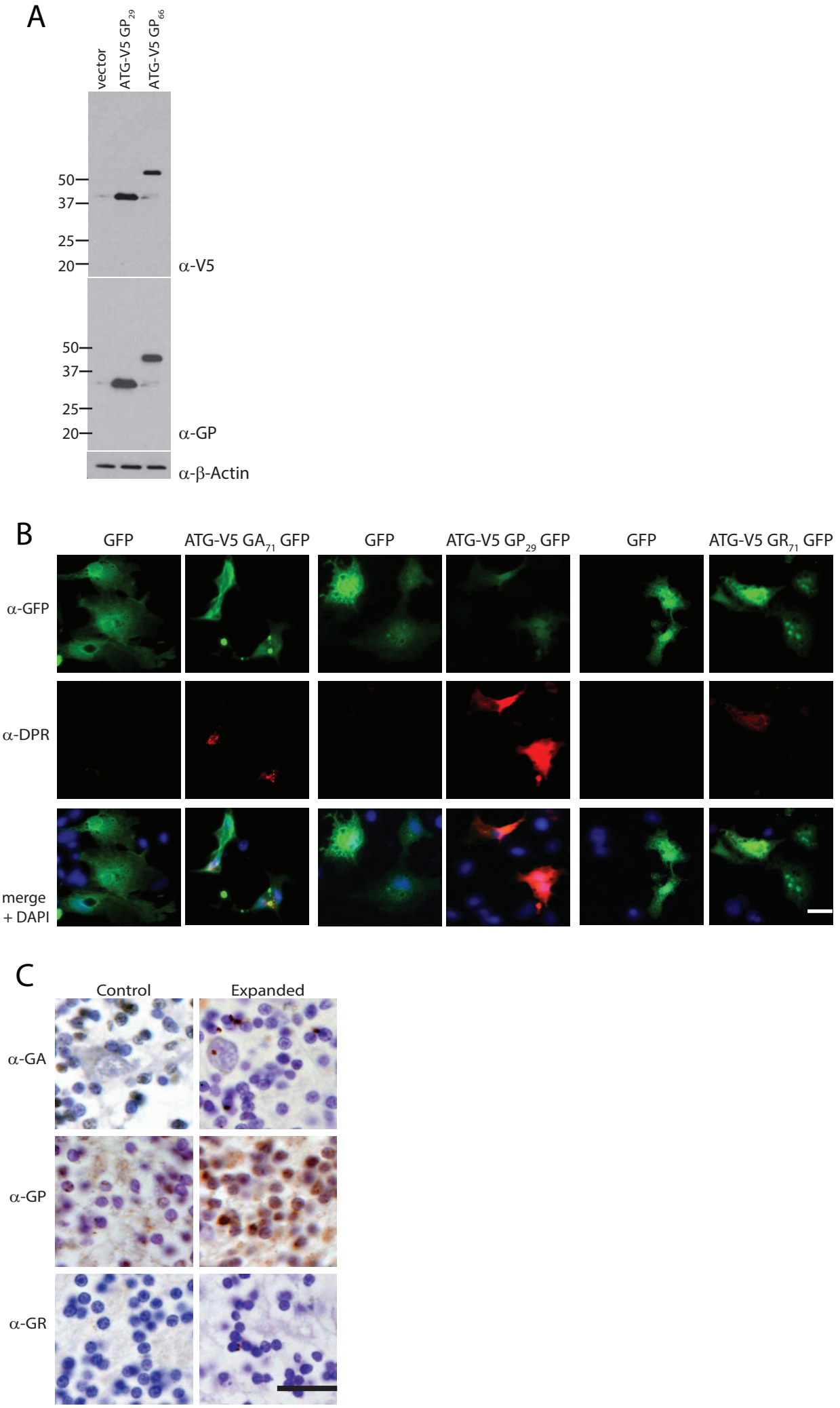

A
ATG-V5 (G4C2)71 CT
CMV
0+ GA, 1+ GP, 2+ GR
ATG-V5
(GGGGCC)n
CT
ATG-V5 GA71 CT (n=806)
ATG-V5 GP71 CT (n=584)
ATG-V5 GR71 CT (n=730)
empty vector (n=783)
**
**
**
Cumulative Risk of Death
Time (hours)
B
ATG-V5 (G4C2)71 CT
ATG-V5 (G4C2)71 ΔCT
Δ
C
ATG-V5 GA71 CT (n=806)
ATG-V5 GA71 ΔCT (n=766)
empty vector (n=783)
ns
D
ATG-V5 GP71 CT (n=584)
ATG-V5 GP71 ΔCT (n=523)
empty vector (n=783)
ns
E
ATG-V5 GR71 CT (n=730)
ATG-V5 GR71 ΔCT (n=741)
empty vector (n=783)
*

Figure 6: Native Carboxyl terminal sequences decrease GR dipeptide repeat elicited toxicity in rodent neurons. A) Schematic of the plasmid for each individual reading frame with the native C-terminal sequence transfected into neurons with the cumulative risk of death below. B) Schematic of the plasmids for each individual reading frame with different C-terminal sequences transfected into neurons. C) Cumulative risk of death of neurons transfected with GA repeats with expanded repeats with different C-termini. D) Cumulative risk of death of neurons transfected with GP repeats with expanded repeats with different C-termini. E) Cumulative risk of death of neurons transfected with GR repeats with expanded repeats with different C-termini.
*$p<0.05$ and ** $p<0.01$ by Cox proportional hazard analysis.


A
vector
GFP-iC3
GFP-iC242
GFP-iC484
3.0kb
2.0kb
1.5kb
1.0kb
0.5kb
B
GFP RNA Expression
Normalized to GFP iC3
5
4
3
2
1
0
Δ
GFP
266
268
269
285
286
289
324
325
327
398
400
402
451
452
453
471
474
478
GFP-iC3
GFP-iC242
GFP-iC484
0+ (G4C2)28 GFP
1+ (G4C2)28 GFP
2+ (G4C2)28 GFP
C
Driver and control lines
Driver
GFP
GFP-iC3
266
268
269
GFP-iC242
285
286
289
GFP-iC484
324
325
327
0+ (G4C2)28 GFP
398
400
402
1+ (G4C2)28 GFP
451
452
453
2+ (G4C2)28 GFP
471
474
478

Supplemental figure 1: Repeat size, RNA expression, and eye phenotypes for intronic and exonic G_4C_2 repeat flies. A) DNA agarose gel validating repeat size of intronic repeat containing vectors. B) GFP mRNA expression from transgenic fly lines. Lines with comparable GFP mRNA levels (Δ) were used for major assays conducted. C) Representative eye images from multiple independent lines for each genotype.


A
vector
ATG-V5 GP29
ATG-V5 GP66
50
37
25
20
α-V5
50
37
25
20
α-GP
α-β-Actin
B
GFP
ATG-V5 GA71 GFP
GFP
ATG-V5 GP29 GFP
GFP
ATG-V5 GR71 GFP
α-GFP
α-DPR
merge
+ DAPI
C
Control
Expanded
α-GA
α-GP
α-GR

Supplemental Figure 2: Characterization of dipeptide repeat antibodies. Dipeptide repeat (DPR) antibodies specifically recognize G_4C_2 containing plasmids by western blot (A) and immunocytochemistry (B). DPR positive aggregates are visible in patient cerebellar tissue. Scale bar of B = 20μm.

Table S1: Sequences for G_4C_2 Repeat-containing pUAST Vectors

Name	Sequence	Comments
GFP-intron	GAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGT GTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGT TCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA CCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGC AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGgtgagtttccacctagtc caccgtgtgcagcagaaacgagacagagagagagagagagagagacggggagaaaagtagaagta gtagtagtggtagaaaggaa**gcggccgcctcgag**ggcgcgccactagtgctagcggtacccccttagatctat ctataccccctctctccctctgtggctctgccccacacggacatggaaatttggaaagacaatcgaccatccatcc gacacacatatctcattcatataccctatatctataacgccaccagGCATCGACTTCAAGGAGGAC GGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTC TATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG AGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG GTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA GCTGTACAAGTAATCTAGA	Backbone for all GFP-intronic constructs. Uppercase letters are exons while lowercase are artificial introns. Green highlight indicates interrupted GFP. Red text is the intronic MCS. Restriction sites NotI and XhoI are bolded.
NotI-$(GGGGCC)_{3 \text{ or } 21}$-XhoI	**gcggccgc**TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG CGTGGTCGGGGCGGGCCCGGGGGCGGGCCCGGGGCGGGGCTGCGGTTG CGGTGCCTGCGCCCGCGGCGGCGGAGGCGCAGGCGGTGGCGAGTGGGTG AGTGAGGAGGCGGCATCCTGGCGGGTGGCTGTTTGGGGTTCGGCTGCCGG GAAGAGGCGCGGGTAGAAGCGGGGGCTCTCCTCAGAGCTCGACGCATTTTT ACTTTCCCTCTCATTTCTCTGACCGAAGCTGGGTGTCGGGCTTTCGCCTCTA GCGACTGGTG**ctcgag**	Insert for short G_4C_2 repeats. The repeat region is highlighted yellow. The NotI and XhoI sites are red bolded text.
NotI-$(GGGGCC)_{28}$-XhoI	**gcggccgc**CCGCAGCCTGTAGCAAGCTCTGGAACTCAGGAGTCGCGCGCTAG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCC GGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCGTGGTCGGGGC*GGGCCC* GGGGGCGGGCCCGGGGCGGGGCTGCGGTTGCGGTGCCTGCGCCCGCGG CGGCGGAGGCGCAGGCGGTGGCGAGTGGGTGAGTGAGGAGGCGGCATCC TGGCGGGTGGCTGTTTGGGGTTCGGCTGCCGGGAAGAGGCGCGGGTAGAA GCGGGGGCTCTCCTCAGAGCTCGACGCATTTTTACTTTCCCTCTCATTTCTC TGACCGAAGCTGGGTGTCGGGCTTTCGCCTCTAGCGACTGGTG**ctcgag**	Insert with 28 G_4C_2 repeats is highlighted yellow. The PspOMI site used future self-insertion to make C_{242} and C_{484} is italicized.
NotI-$(GGGGCC)_{21}$-PspOMI	**gcggccgc**TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG GGCCGGGGCCGGGGCGTGGTC*GGGCCC*	Building block of 21 G_4C_2 repeats to generate C_{49}, C_{70}, C_{91} and C_{121}.

Table S2: Primers used for cloning, In Situ, and qPCR

For C9ORF72 Intronic repeat cloning

Primer name	Primer sequence
NotI-C9 AnchorR	5'-CCAGCGGCCGCTACGCATCCAGTTTGAGACGGGGGCC-3'
XhoI-C9 F	5'-CCCTCGAGCACCAGTCGCTAGAGGCGAAAGCCCG-3'
NotI-C9R	5'-CCAGCGGCCGCCCGCAGCCTGTAGCAAGC-3'
PspOMI-tiling	5'-AAAGGGCCCGACCACGCCCCGGCC-3'

For G_4C_2 exonic repeat cloning

Primer name	Primer sequence
Adapter1	5'-ACCGGTCAGATCTCGAACCGGT-3'
Adapter2	5'-ACCGGTCAGATCTCGAAACCGGT-3'

For In situ analysis

Primer name	Primer sequence
Probe	5'-/5Cy5N/mGmGmCmCmCrCrGrGrCrCrCrCrGr GrCrCrCrCrGrGrCrCrCrCrGmGmCmCmCmC-3'

For qRT-PCR of GFP in flies

GFP unspliced forward	5'-ACTAGTGCTAGCGGTACCCCTTAGATC-3'
GFP spliced forward	5'-TCTTCTTCAAGGACGACGGCAACTAC-3'
GFP reverse (for spliced+unspliced)	5'-GTACTCCAGCTTGTGCCCCAGGATGT-3'
RPL32 forward	5'-GCCCAGCATACAGGCCCAAG-3'
RPL32 reverse	5'-AAGCGGCGACGCACTCTGTT-3'