C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins

Sarah Mizielinska[#1], **Sebastian Grönke**[#2], **Teresa Niccoli**[#3], **Charlotte E. Ridler**[1], **Emma L. Clayton**[1], **Anny Devoy**[1], **Thomas Moens**[1,3], **Frances E. Norona**[1], **Ione O.C. Woollacott**[1], **Julian Pietrzyk**[1], **Karen Cleverley**[1], **Andrew J. Nicoll**[1,4], **Stuart Pickering-Brown**[6], **Jacqueline Dols**[2], **Melissa Cabecinha**[3], **Oliver Hendrich**[2], **Pietro Fratta**[1,5], **Elizabeth M.C. Fisher**[1,5], **Linda Partridge**[2,3,†], and **Adrian M. Isaacs**[1,†]

[1]Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK

[2]Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Strasse 9b, 50931 Cologne, Germany

[3]Department of Genetics, Evolution and Environment, Institute of Healthy Ageing, UCL, Darwin Building, Gower Street, London WC1E 6BT, UK

[4]MRC Prion Unit, UCL Institute of Neurology, London WC1N 3BG, UK

[5]MRC Centre for Neuromuscular Disease, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK

[6]Institute of Brain, Behaviour and Mental Health, Faculty of Human and Medical Sciences, University of Manchester, AV Hill Building, Oxford Road, Manchester, M13 9PT, UK

[#] These authors contributed equally to this work.

Abstract

An expanded GGGGCC repeat in *C9orf72* is the most common genetic cause of frontotemporal dementia and amyotrophic lateral sclerosis. A fundamental question is whether toxicity is driven by the repeat RNA itself and/or by dipeptide repeat proteins generated by repeat-associated, non-ATG translation. To address this question we developed in vitro and in vivo models to dissect repeat RNA and dipeptide repeat protein toxicity. Expression of pure repeats in *Drosophila* caused adult-onset neurodegeneration attributable to poly-(glycine-arginine) proteins. Thus expanded repeats promoted neurodegeneration through neurotoxic proteins. Expression of individual dipeptide repeat proteins with a non-GGGGCC RNA sequence showed both poly-(glycine-arginine) and poly-(proline-arginine) proteins caused neurodegeneration. These findings are consistent with a dual toxicity mechanism, whereby both arginine-rich proteins and repeat RNA contribute to *C9orf72*-mediated neurodegeneration.

Frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) are adult-onset, neurodegenerative diseases associated with personality change, language dysfunction and

[†]Corresponding authors. a.isaacs@prion.ucl.ac.uk; l.partridge@ucl.ac.uk.

progressive muscle weakness. These syndromes overlap genetically and pathologically, and can also co-occur in individuals, and within families (1). An intronic GGGGCC hexanucleotide repeat expansion in *C9orf72* is the most common genetic cause of both FTD and ALS (C9FTD/ALS) (2–4), and can be found in patients diagnosed with all common neurodegenerative diseases (5). Healthy individuals carry fewer than 33 hexanucleotide repeats, with 2 repeats being the most common, but C9FTD/ALS cases carry between 400 and 4400 repeats (2, 5, 6).

The repeat expansion could cause disease by three possible mechanisms: i) toxic sense and/or antisense repeat RNA species that sequester key RNA-binding proteins, ii) toxic dipeptide repeat (DPR) proteins, generated by repeat-associated, non-ATG (RAN) translation, or iii) reduced expression of *C9orf72*. The absence of a severe phenotype in a homozygous *C9orf72* mutation case (7), and the lack of *C9orf72* coding mutations (8) argue against loss-of-function as a primary mechanism. Neuronal aggregates of RNA, termed RNA foci, generated from both sense and antisense repeat transcripts are frequent in C9FTD/ALS patient brain (9–13). The GGGGCC repeat can be translated in all sense and antisense frames, two of which encode the same DPR, resulting in five DPR proteins, all of which form inclusions in widespread brain regions (10, 12, 14–18). It is therefore of fundamental importance to understand the contributions of repeat RNA and DPR proteins to *C9orf72*-mediated neurodegeneration.

A major obstacle in the investigation of large expanded repeats is that they are inherently unstable. We used recombination-deficient *E. coli* and a cloning strategy termed recursive directional ligation (19) to sequentially build seamless pure repeats from small GGGGCC repeat units (fig. S1). This allowed generation of a stable range of pure repeats from 3 to a maximum of 103 (Fig. 1A). To dissect repeat RNA and DPR protein toxicity, we generated "RNA-only" repeats, using our cloning strategy to insert interruptions containing stop codons in all sense and antisense frames. In models of other non-coding repeat expansion disorders, interruptions comprising 4 to 11 percent of the total repeat sequence confer stability while maintaining pathogenicity in vitro and in vivo (20–22). One of three 6-base-pair interruptions, each containing one stop codon in the sense and one in the antisense direction, were inserted every 12 GGGGCC repeats, resulting in a stop codon for all 6 (sense and antisense) frames, and interruptions that comprised 8% of the total sequence (fig. S2). We generated stop codon-interrupted RNA-only repeats equivalent in length to our pure repeats and longer RNA-only repeats up to ~288 (Fig. 1A). GGGGCC repeat RNA forms a stable tertiary structure termed a G-quadruplex (23). Circular dichroism showed that the RNA-only repeats formed RNA G-quadruplexes similarly to pure repeat RNA (Fig. 1B), showing that the interruptions did not affect the tertiary structure of the RNA. To investigate the formation of RNA foci, constructs were expressed in human neuroblastoma cells. RNA fluorescence in situ hybridisation (FISH) showed that formation of RNA foci was length-dependent for both pure and RNA-only repeats which, at equivalent length, had the same propensity to form foci (Fig 1C, D).

In order to differentiate between repeat RNA and DPR protein toxicity in vivo, we generated lines of the fruit fly *Drosophila melanogaster* carrying a range of our pure and RNA-only repeats under the UAS promoter, integrated into the same genomic location to ensure

equivalent expression levels. When expressed specifically in the adult fly, the different repeats expressed sense transcripts at comparable levels and of the expected sizes, but no antisense transcripts (fig. S3A). RNA FISH showed that pure and RNA-only repeats were both able to generate RNA foci in *Drosophila* (fig. S4). Immunoblotting using an anti-poly-(GR) antibody (Fig. 2A) or an anti-poly-(GP) antibody (fig. S5B) showed that, as expected, the pure repeats generated DPR proteins and the RNA-only repeats did not. Expression of both 36 and 103 pure repeats in the eye caused eye degeneration, whereas 36, 108 and ~288 RNA-only repeats had no effect under the same conditions (Fig. 2B). The toxicity of the pure repeats was thus attributable to the presence of DPR proteins. Increasing expression levels of the pure repeats, by increasing the temperature (24), led to lethality from both 36 and 103 repeats (Fig. 2C, fig S6) but for the RNA-only repeats had no effect, again demonstrating that the pure repeats caused lethality through the production of DPR proteins.

The GMR-Gal4 driver is expressed throughout development. However, ALS and FTD are adult-onset diseases. To circumvent developmental effects, we confined expression of the repeat constructs to adult neurons, using the inducible elav-GeneSwitch driver. Expression of 36 and 103 repeats killed all flies by 30 days post-eclosion. 36, 108 and ~288 RNA-only repeats had no effect, showing that the neurotoxicity of the pure repeats was attributable to DPR protein production (Fig. 2D). To confirm this, we reduced protein synthesis in the 36 and 103 pure repeat-expressing flies with a sub-lethal dose of cycloheximide, which ameliorated the reduction in lifespan caused by the pure repeats (Fig. 2E), again showing that toxicity was attributable to DPR proteins.

To assess whether DPR protein expression alone was sufficient for toxicity, "protein-only" constructs were generated, by using alternative codons to those found within the GGGGCC repeat. We compared the two arginine-containing DPR proteins, glycine-arginine (GR) and proline-arginine (PR), with two neutral DPR proteins, proline-alanine (PA) and glycine-alanine (GA). When constructs containing 36 dipeptide repeats (equivalent to 36 pure GGGGCC repeats) were expressed in the fly eye, the arginine-containing DPR proteins GR and PR caused eye degeneration and lethality, while GA and PA DPR proteins had no effect (Fig. 3A, C). Thus, the arginine-containing DPR proteins induced toxicity. We next generated longer protein-only sequences, of equivalent length to 103 pure repeats. Expression of $(PR)_{100}$ or $(GR)_{100}$ caused eye degeneration and increased lethality, while $(PA)_{100}$ and $(GA)_{100}$ had no effect (Fig. 3B, C). Expression of $(PR)_{100}$ and $(GR)_{100}$ in adult neurons caused a substantial decrease in survival (Fig. 3D); a late-onset reduction in survival was also observed in $(GA)_{100}$-expressing flies, while $(PA)_{100}$ had no effect. Expression levels varied among the individual protein-only constructs, but did not correlate with toxicity (fig. S3C), which was therefore attributable to the arginine-rich sequences. Thus, the highly basic arginine-containing DPR proteins drove *C9orf72* GGGGCC repeat toxicity in *Drosophila* neurons.

Our data identified GR and PR DPR proteins as the predominant toxic protein species, although all five DPR proteins form inclusions in affected brain regions. Similarly, the distribution of poly-(GA) inclusions does not correlate well with neurodegeneration (25). The presence of arginine in both of the highly toxic DPR species suggests a common pathological mechanism, perhaps attributable to their basic nature or a common structural

motif. Restricted expression of *C9orf72* to specific neuronal populations (26), or a deficit in the affected neurons' ability to clear these particular proteins, may explain why these highly toxic proteins cause selective neurodegeneration. In patients all five DPR proteins may be produced in a single neuron. While our findings indicate that toxicity is driven by the arginine-rich DPR proteins it remains possible that high focal levels of the other DPR proteins could contribute to cytotoxicity.

We have been able to separate RNA and DPR toxicity associated with *C9orf72* GGGGCC repeats and, surprisingly, our data suggests that the major toxic species are the DPR proteins. However, the DPR protein toxicity that we observed from over-expression of pure repeats does not rule out an additional contribution of RNA toxicity. Several lines of evidence suggest a toxic role of repeat RNA. In C9FTD patient brain, RNA foci are most abundant in the frontal cortex, which has the greatest degree of neuronal loss, and frontal cortex RNA foci burden correlates with age at onset in C9FTD cases (9). GGGGCC repeats also sequester several RNA-binding proteins, which could lead to toxicity (13, 27–31). However, modelling RNA toxicity may require longer repeats that are closer to the pathological range seen in disease, possibly because a toxic threshold of repeat number must be crossed. A continuing conundrum is why the same expanded repeat can cause either pure FTD or pure ALS. Our data raise the possibility that the different patient phenotypes could be caused by differences in the relative contributions of RNA or protein-mediated toxicity within distinct neuronal subtypes. A further prediction from this hypothesis is that genetic variants that affect RAN translation or DPR protein levels may also contribute to disease penetrance.

Supplementary Materials

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and Notes

1. Lillo P, Hodges JR. Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. J Clin Neurosci. 2009; 16:1131–1135. [PubMed: 19556136]

2. Dejesus-Hernandez M, et al. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron. 2011; 72:245–256. [PubMed: 21944778]

3. Renton AE, et al. A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. Neuron. 2011; 72:257–268. [PubMed: 21944779]

4. Majounie E, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. 2012; 11:323–330. [PubMed: 22406228]

5. 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. Am J Hum Genet. 2013; 92:345–353. [PubMed: 23434116]

6. 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. Lancet Neurol. 2013; 12:978–988. [PubMed: 24011653]

7. Fratta P, et al. Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. Acta Neuropathol. 2013; 126:401–409. [PubMed: 23818065]

8. Harms MB, et al. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Neurobiol Aging. 2013; 34:2234–2239. [PubMed: 23597494]

9. Mizielinska S, et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol. 2013; 126:845–857. [PubMed: 24170096]

10. Gendron TF, 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; 126:829–844. [PubMed: 24129584]

11. Lagier-Tourenne C, et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc Natl Acad Sci U S A. 2013; 110:E4530–E4539. [PubMed: 24170860]

12. 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:E4968–E4977. [PubMed: 24248382]

13. Lee YB, et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. 2013; 5:1178–1186. [PubMed: 24290757]

14. Ash PE, et al. Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. Neuron. 2013; 77:639–646. [PubMed: 23415312]

15. Lashley T, Hardy J, Isaacs AM. RANTing about C9orf72. Neuron. 2013; 77:597–598. [PubMed: 23439112]

16. Mori K, et al. The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. Science. 2013; 339:1335–1338. [PubMed: 23393093]

17. Mori K, et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. 2013; 126:881–893. [PubMed: 24132570]

18. Zu T, et al. Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A. 2011; 108:260–265. [PubMed: 21173221]

19. Meyer DE, Chilkoti A. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. Biomacromolecules. 2002; 3:357–367. [PubMed: 11888323]

20. de Haro M, et al. MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. Hum Mol Genet. 2006; 15:2138–2145. [PubMed: 16723374]

21. Orengo JP, et al. Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. Proc Natl Acad Sci U S A. 2008; 105:2646–2651. [PubMed: 18272483]

22. Sakamoto N, et al. GGA*TCC-interrupted triplets in long GAA*TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. J Biol Chem. 2001; 276:27178–27187. [PubMed: 11325966]

23. Fratta P, et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci Rep. 2012; 2:1016. [PubMed: 23264878]

24. Duffy JB. GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis. 2002; 34:1–15. [PubMed: 12324939]

25. Mackenzie IR, et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. Acta Neuropathol. 2013; 126:859–879. [PubMed: 24096617]

26. Suzuki N, et al. The mouse C9ORF72 ortholog is enriched in neurons known to degenerate in ALS and FTD. Nat Neurosci. 2013; 16:1725–1727. [PubMed: 24185425]

27. Almeida S, et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. 2013; 126:385–399. [PubMed: 23836290]

28. Donnelly CJ, et al. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron. 2013; 80:415–428. [PubMed: 24139042]

29. Mori K, et al. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol. 2013; 125:413–423. [PubMed: 23381195]

30. Sareen D, et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci Transl Med. 2013; 5:208ra149.

31. Xu Z, et al. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc Natl Acad Sci U S A. 2013; 110:7778–7783. [PubMed: 23553836]
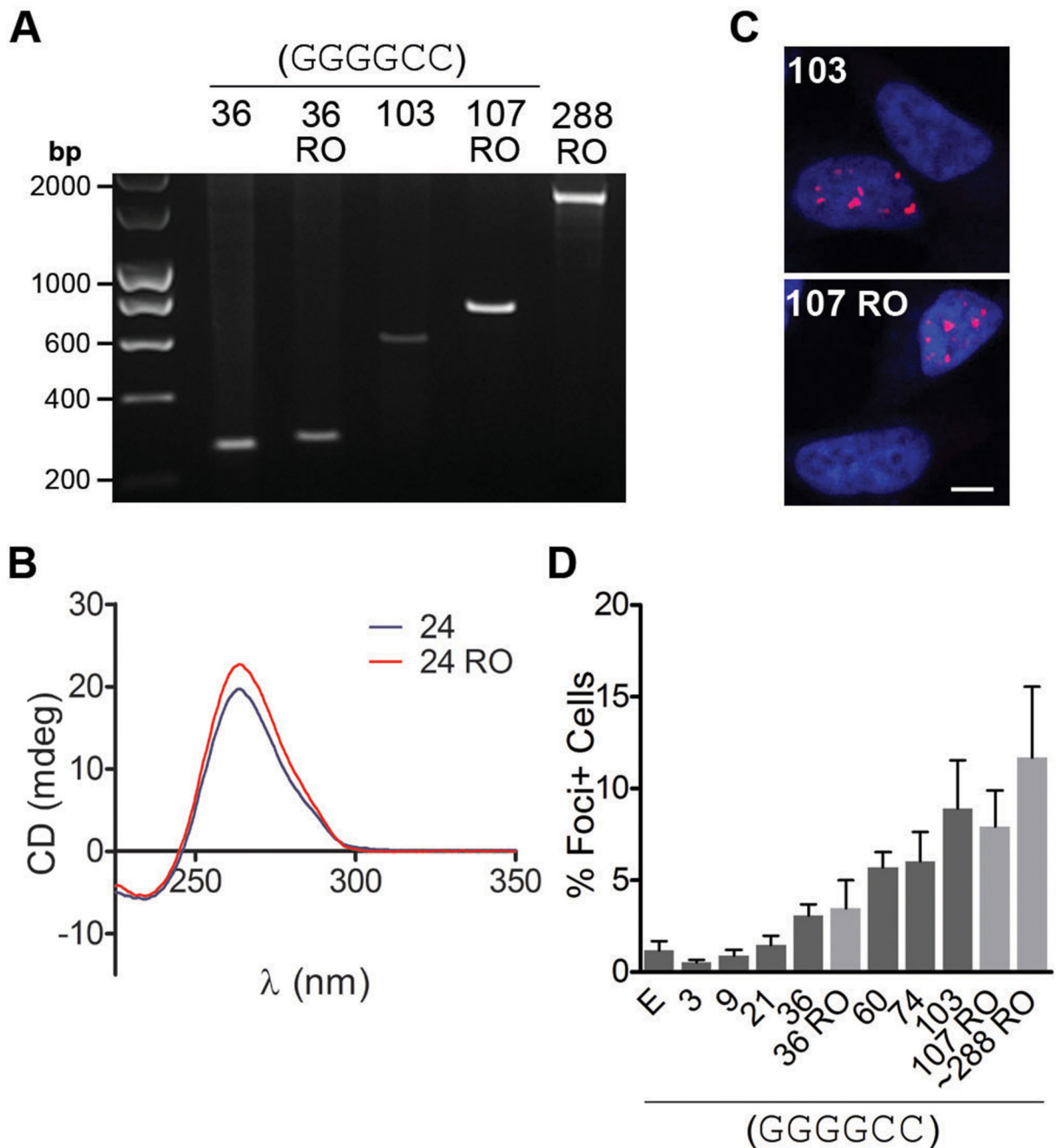
Fig. 1. Generation and characterisation of expanded pure and RNA-only GGGGCC repeats.
(**A**) Agarose gel showing pure GGGGCC repeats and stop codon-interrupted RNA-only
(RO) repeats. (**B**) Circular dichroism (CD) spectra of both 24 pure and 24 RO repeats
showed characteristic RNA G-quadruplex structure with minima and maxima at 237 and 262
nm respectively (23). (**C**) Confocal microscope images of nuclei (blue) in RNA FISH-
labelled SH-SY5Y cells showed that 103 pure and 107 RO repeats both produced nuclear
RNA foci (red). Scale bar represents 5 μm. (**D**) Quantification of the number of SH-SY5Y
cells containing RNA foci after transfection with pure and RO repeats of different lengths,

and empty vector (E). No difference was observed between equivalent length pure and RNA-only repeats (36 vs. 36 RO, 103 vs. 107 RO, 1 way ANOVA with Bonferroni test (selected pairs), n > 3, error bars represent SEM).
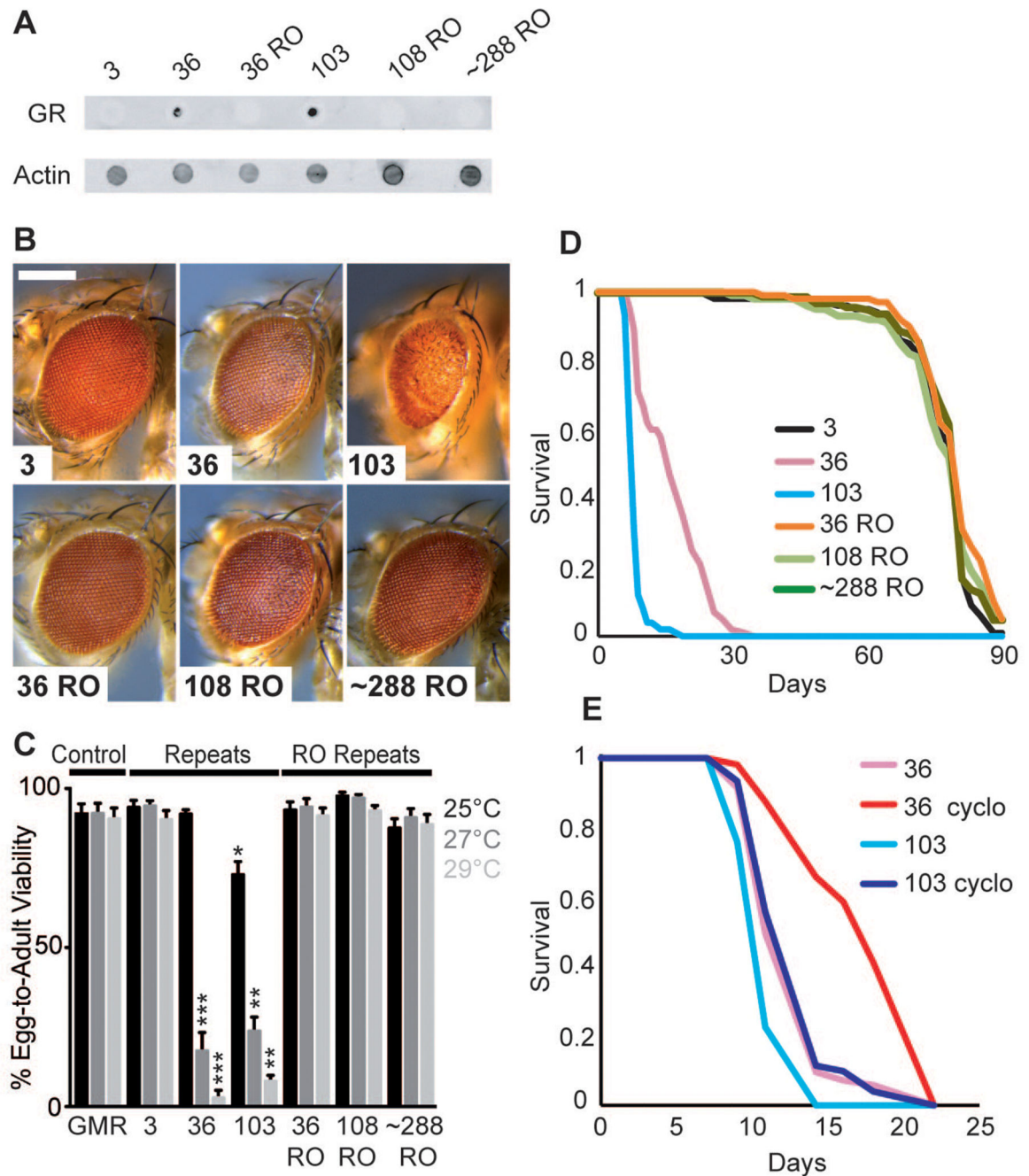
Fig. 2. Pure GGGGCC repeats caused toxicity via DPR proteins.
(**A**) Dot blot showing that 36 and 103 pure repeats generated poly-(GR) proteins while 3 pure repeats, and 36, 108 and ~288 RNA-only (RO) repeats did not. Genotypes were: *w; UAS-3/hsGal4, w; UAS-36/hsGal4, w; UAS-103/ hsGal4, w; UAS-36 RO/hsGal4, w; UAS-108 RO/ hsGal4, w; UAS-288 RO/hsGal4.* (**B**) Stereomicroscopy images of representative *Drosophila* eyes expressing pure or RO repeats using the GMR-GAL4 driver. 36 pure repeats were mildly toxic. 103 pure repeats showed more overt toxicity. 3 repeats and 36 and 108 RO repeats had no effect. Genotypes were: *w; GMR-Gal4/+, w; GMR-Gal4/*

UAS-3, w; GMR-Gal4/UAS-36, w; GMR-Gal4/UAS-103, w; GMR-Gal4/UAS-36RO, w; GMR-Gal4/UAS-108RO, w; GMR-Gal4/UAS-288RO. Scale bar represents 200 μm. **(C)** Quantification of egg-to-adult viability showed that 36 and 103 pure repeats were lethal at higher temperatures, whereas RO repeats had no effect (Kruskal Wallis test with Dunn's multiple comparison (selected pairs), ***p<0.001, **p<0.01, *p<0.05, error bars represent SEM). Genotypes were as in (B). **(D)** Survival of female flies expressing repeats in adult neurons using the elav-GeneSwitch (elavGS) driver. 36 and 103 pure repeats substantially decreased survival, while 36, 108 and 288 RO repeats had no effect (*p*<0.0001, log-rank test). Genotypes were: *w; UAS-3/+; elavGS/+, w; UAS-36/+; elavGS/+, w; UAS-103/+; elavGS/+, w; UAS-36 RO/+; elavGS/+, w; UAS-108 RO/+; elavGS/+, w; UAS-288 RO/+; elavGS/+.* **(E)** Flies expressing 36 and 103 pure repeats survived longer in the presence of cycloheximide than in its absence (*p*<0.001, log-rank test). Genotypes were: *w; UAS-36/+; elavGS/+, w; UAS-103/+; elavGS/+.*
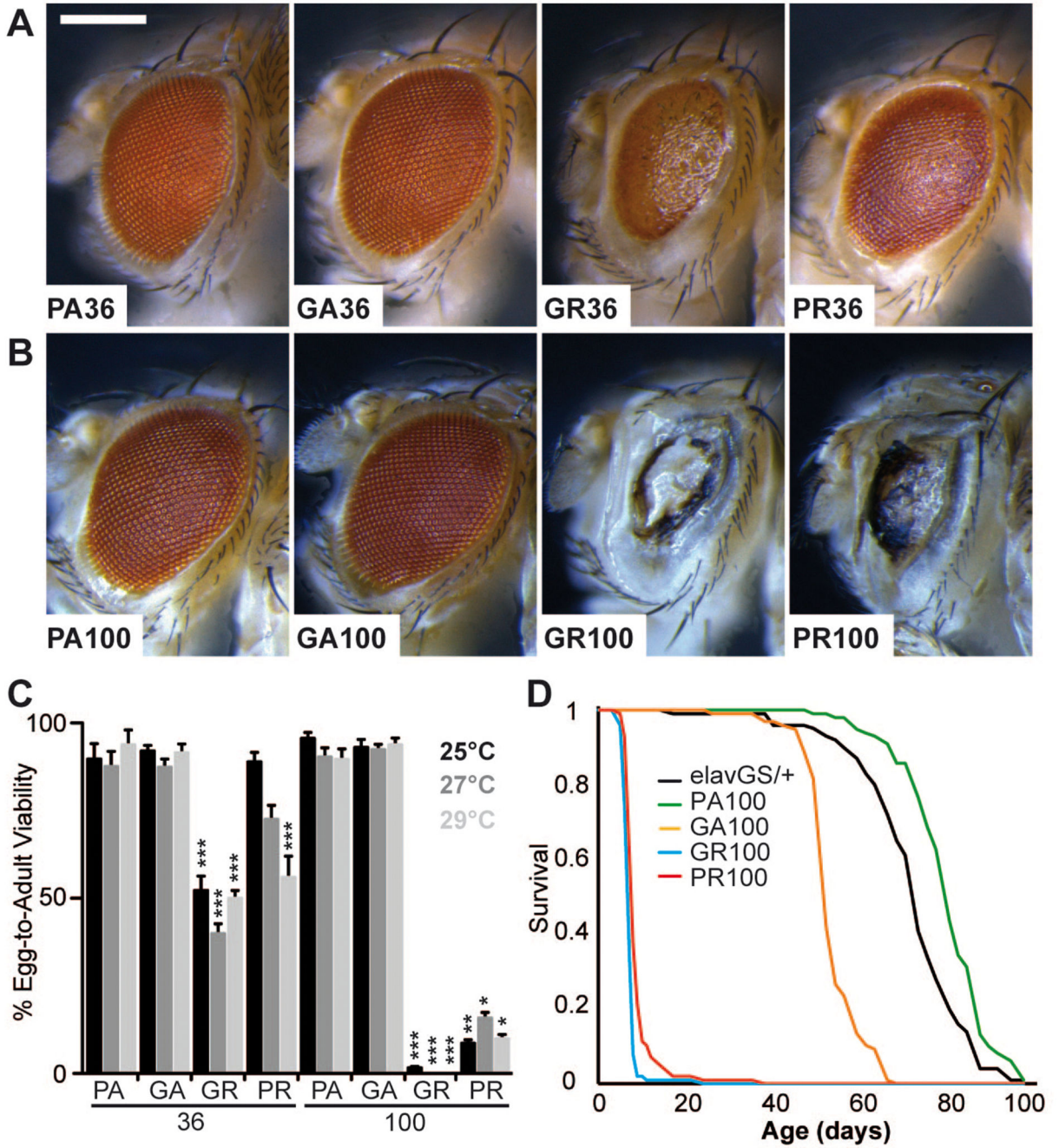
Fig. 3. DPR toxicity was caused by poly-GR and poly-PR proteins.
"Protein-only" constructs for individual DPR proteins were expressed in the *Drosophila* eye
(A-C), and the adult nervous system (D). **(A)** $(GR)_{36}$ and $(PR)_{36}$ caused eye degeneration,
while $(GA)_{36}$ and $(PA)_{36}$ had no effect. Genotypes were: *w; UAS-PA36/GMR-Gal4, w;
UAS-GA36/GMR-Gal4, w; UAS-GR36/GMR-Gal4, w; UAS-PR36/GMR-Gal4.* Scale bar
represents 200 μm. **(B)** $(GR)_{100}$ and $(PR)_{100}$ caused extensive eye degeneration, while
$(GA)_{100}$ and $(PA)_{100}$ had no effect. Genotypes were: *w; UAS-PA100/GMR-Gal4, w; UAS-
GA100/GMR-Gal4, w; UAS-GR100/GMR-Gal4, w; UAS-PR100/GMR-Gal4.* **(C)**

Quantification of egg-to-adult viability showed $(GR)_{100}$ and $(PR)_{100}$ caused a substantial reduction in survival, whereas $(GA)_{100}$ and $(PA)_{100}$ had no effect (Kruskal Wallis test with Dunn's multiple comparison, selected pairs, ***$p < 0.001$, **$p < 0.01$, error bars represent SEM). Genotypes were as in (A) and (B). **(D)** Expression of $(GR)_{100}$ and $(PR)_{100}$ in adult neurons using the elav-GeneSwitch (elavGS) driver caused a substantial decrease in viability ($p < 0.001$, log-rank test); $(GA)_{100}$ caused a late-onset decrease in survival, and $(PA)_{100}$ or elavGS driver alone had no effect. Genotypes were: *w; elavGS/+, w; UAS-PA100/+; elavGS/+, w; UAS-GA100/+; elavGS/+, w; UAS-GR100/+; elavGS/+*, w; *UAS-PR100/+; elavGS/+.*