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

Authors:
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†] Adrian M. Isaacs[1†]

Affiliations:
[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.

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

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.

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

References and Notes:

1. P. Lillo, J. R. Hodges, Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. *J Clin Neurosci* **16**, 1131-1135 (2009).

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

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

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

5. J. Beck *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* **92**, 345-353 (2013).

6. M. van Blitterswijk *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* **12**, 978-988 (2013).

7. P. Fratta *et al.*, Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathol* **126**, 401-409 (2013).

8. M. B. Harms *et al.*, Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. *Neurobiol Aging* **34**, 2234-2239 (2013).

9. S. Mizielinska *et al.*, C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol* **126**, 845-857 (2013).

10. T. F. Gendron *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* **126**, 829-844 (2013).

11. C. Lagier-Tourenne *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* **110**, E4530-E4539 (2013).

12. T. Zu *et al.*, RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-E4977 (2013).

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

14. P. E. Ash *et al.*, Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron* **77**, 639-646 (2013).

15. T. Lashley, J. Hardy, A. M. Isaacs, RANTing about C9orf72. *Neuron* **77**, 597-598 (2013).

16. K. Mori *et al.*, The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science* **339**, 1335-1338 (2013).

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

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

19. D. E. Meyer, A. Chilkoti, 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* **3**, 357-367 (2002).

20. M. de Haro *et al.*, MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. *Hum Mol Genet* **15**, 2138-2145 (2006).

21. J. P. Orengo *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* **105**, 2646-2651 (2008).

22. N. Sakamoto *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* **276**, 27178-27187 (2001).

23. P. Fratta *et al.*, C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2**, 1016 (2012).

24. J. B. Duffy, GAL4 system in Drosophila: a fly geneticist's Swiss army knife. *Genesis* **34**, 1-15 (2002).

25. I. R. Mackenzie *et al.*, Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* **126**, 859-879 (2013).

26. N. Suzuki *et al.*, The mouse C9ORF72 ortholog is enriched in neurons known to degenerate in ALS and FTD. *Nat Neurosci* **16**, 1725-1727 (2013).

27. S. Almeida *et al.*, Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol* **126**, 385-399 (2013).

28. C. J. Donnelly *et al.*, RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**, 415-428 (2013).

29. K. Mori *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* **125**, 413-423 (2013).

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

31. Z. Xu *et al.*, Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc Natl Acad Sci U S A* **110**, 7778-7783 (2013).

Acknowledgments:
Funding was provided by Alzheimer's Research UK (AMI), The Motor Neurone Disease Association (AMI, EF, PF), the MHMS General Charitable Trust (AMI), NIHR-UCLH-BRC (PF), NIHR ACF (IW), the UK Medical Research Council (AMI, EF, PF), the Wellcome Trust (LP), and the Max Plank Society (LP). We thank Dr J Wadsworth and Dr N Alic for helpful discussion. LP dedicates this work to the memory of Noreen Murray.

Supplementary Materials:
Materials and Methods
Figs. S1-S7
References (*32*-*35*)
Submitted 30 May 2014; accepted 28 July 2014

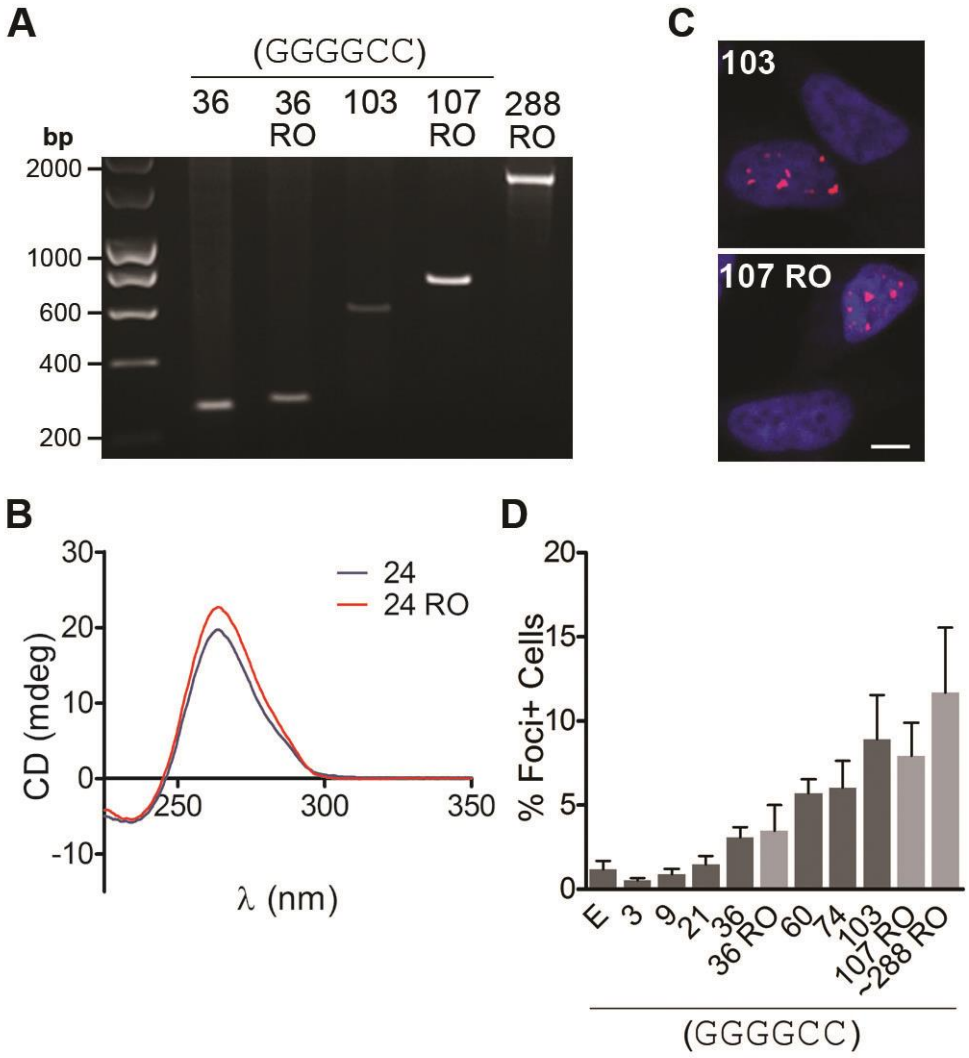

Fig. 1. **Generation and characterisation of expanded pure and RNA-only GGGGCC repeats. (A)** Agarose gel showing pure GGGGCC repeats and stop codon-interrupted RNA-only (RO) repeats. **(B)** Circular dichroism (CD) spectra of both 24 pure and 24 RO repeats showed characteristic RNA G-quadruplex structure 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).

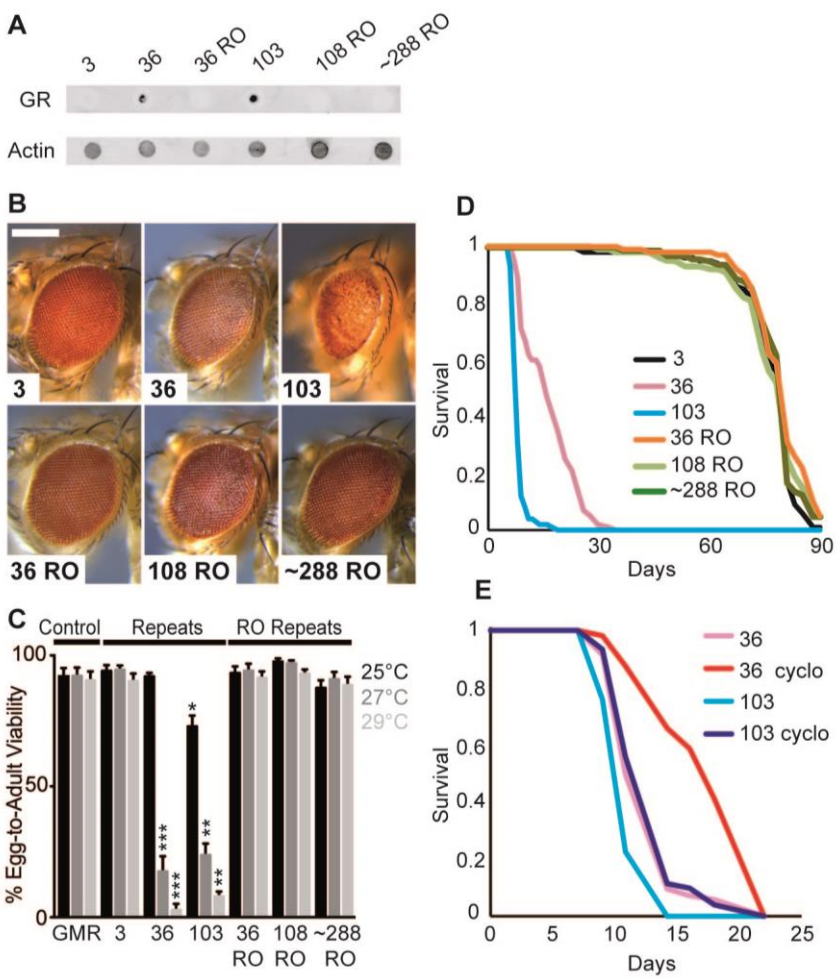

Fig. 2. Pure GGGGCC repeats caused toxicity via DPR proteins. (A) Dot blot showing that 36 and 103 pure repeats generated poly-(GR) proteins while 3 pure repeats, and 36, 108 and ~288 RNA-only (RO) repeats did not. Genotypes were: *w; UAS-3/hsGal4, w; UAS-36/hsGal4, w; UAS-103/ hsGal4, w; UAS-36 RO/hsGal4, 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/+*.

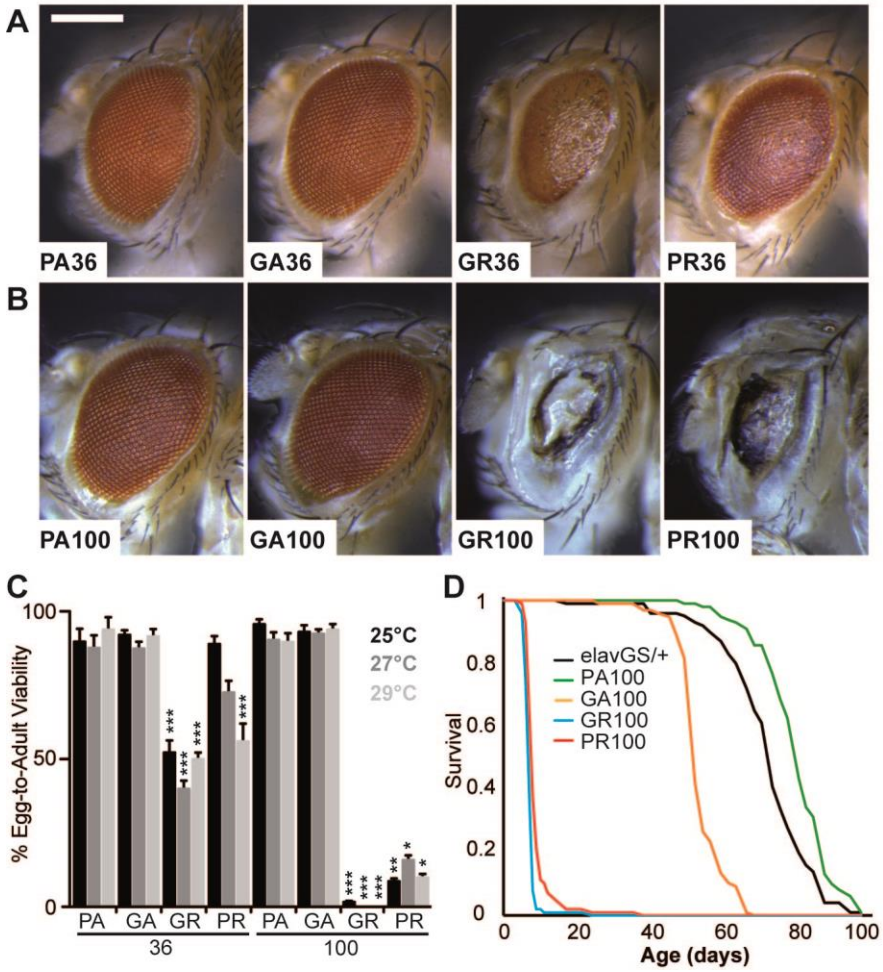

Fig. 3. DPR toxicity was caused by poly-GR and poly-PR proteins. "Protein-only" constructs for individual DPR proteins were expressed in the *Drosophila* eye (A-C), and the adult nervous system (D). **(A)** $(GR)_{36}$ and $(PR)_{36}$ caused eye degeneration, while $(GA)_{36}$ and $(PA)_{36}$ had no 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/+*.

Supplementary Materials for

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†] Adrian M. Isaacs[1†]

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

This PDF file includes:

Materials and Methods

Generation of repeat constructs

A range of increasing repeat sizes was generated using recursive directional ligation (RDL, (*19*), see fig. S1). Complementary DNA oligonucleotides (Sigma) containing 3 or 4 GGGGCC repeats, flanked 5' by BamHI and BspQI recognition sites and 3' by EcoO109I and NotI recognition sites, were annealed and ligated into the vector pBluescript using the BamHI and NotI sites. Oligomerization of repeat units was achieved by digestion of the vector with BspQI and EcoO109I (both NEB) to release an insert, followed by ligation into the vector, which had been linearized by digestion with only BspQI. Additional rounds of RDL proceeded identically, using the product from the previous round as the starting material. Repeat units were joined seamlessly, with no interruption of the GGGGCC repeat sequence, and loss of the BspQI and EcoO109I recognition sites at the site of ligation due to their non-palindromic overhangs. Repeat constructs were subcloned into pcDNA3.1(+) or pUAST attB for expression in neuroblastoma cells or *Drosophila*, respectively. Protein-only alternative codon constructs were synthesised (GeneArt, Life Technologies) and subcloned into pUAST attB as above (sequences available on request). Transformations were performed using recombination-deficient Stbl3 *E. coli* (Life Technologies) at 30°C to minimize retraction of repeats. DNA was extracted using QIAprep Spin Miniprep or Plasmid Maxi kits (QIAGEN), following the manufacturer's instructions. Constructs were screened using standard restriction enzyme digest and agarose gel electrophoresis, and were sequence-verified (Source Bioscience) using dGTP BigDye (Applied Biosystems) up to 108 repeats, to confirm repeat size and interruptions.

Circular Dichroism (CD)

An HPLC-purified RNA oligonucleotide of 24 pure or 24 RNA-only repeats (IDT) was annealed by heating the samples to 90°C and allowing them to cool overnight to 20°C. CD experiments were performed at 25°C in 40 mM KCl, 10 mM potassium phosphate, pH 7.0 buffer using a Jasco J715 spectropolarimeter (Jasco Hachioji, Tokyo, Japan). A CD spectrum of the buffer was recorded and subtracted from the spectrum obtained for the RNA-containing solution. Data were zero-corrected between 340–350 nm.

Cell culture and transfection

The human neuroblastoma cell line SH-SY5Y was grown using standard cell culture techniques in supplemented DMEM media (10% FCS, 5 mM L-glutamine, 1 mM sodium pyruvate (Life Technologies)) in a 37°C/5% CO_2 incubator. Cells were plated on glass coverslips in a 24-well dish at a density of $1.5x10^{-4}$ cells per well. The following day, cells were transfected using Lipofectamine2000 (Life Technologies), according to the manufacturer's instructions, using 0.4 ng of DNA, 0.04 ng of the co-transfection marker ZsGreen (pBI-CMV3, Clontech) and 1 µl Lipofectamine2000, each diluted in 30 µl of OptiMEM (Life Technologies) per well. Cells were incubated with transfection mix for 3 hours at 37°C/5% CO_2, washed once, and returned to fresh culture medium. Expression was analysed 24 hours later.

Generation and maintenance of transgenic fly lines

Transgenic fly lines were generated by phiC31-integrase-mediated, site-directed insertion into the attP40 locus (*32*). Constructs were injected into y1, M{vas-int.B}ZH-2A w*;

P{CaryP}attP40 embryos and the phiC31 integrase was removed by crossing transgenic males to *white* females for two successive generations. Integration into the attP40 landing site was verified by PCR using primers SOL734 (GGAACCTTACTTCTGTGGTGTGAC) and SOL459 (AATTCGGGGGCTTTCGGTGTT), which bind in the SV40 3´ UTR of the pUAST attB vector and in the genomic region flanking the attP40 insertion site, respectively.

All fly stocks were maintained at 25°C on a 12:12 hour light:dark cycle at constant humidity on a standard sugar-yeast (SY) medium (15 g/L agar, 50 g/L sugar, 100 g/L autolysed yeast, 30 ml/L nipagin (10% in ethanol) and 3 ml/L propionic acid). Elav-GeneSwitch (elavGS) (*33*) and daughterless-GeneSwitch (daGS) (*34*) flies were obtained as a generous gift from Dr H. Tricoire (CNRS, France). The GMR-Gal4, attP40 and vas phiC31 integrase lines were obtained from the Bloomington *Drosophila* Stock Centre.

Northern Blotting

To induce adult-specific, ubiquitous expression of GGGGCC repeats or DPR protein-only constructs, daughterless-GeneSwitch (da-GS) driver flies were crossed to the corresponding UAS lines. da-GS/UAS flies were fed for 5 days with 200 µM mifepristone (RU486) to induce expression. RNA was extracted using Trizol and the Qiagen RNeasy Mini Kit with on-column DNase-treatment. Northern blots were performed using the Northern Max Kit (Ambion) according to the manufacturer's instructions, using 20 µg total RNA per lane. A $(GGCCCC)_5$ oligonucleotide probe was used to detect sense repeats and a $(GGGGCC)_5$ oligonucleotide probe was used to detect antisense repeats. DPR protein-only constructs were detected using an oligonucleotide probe (CATGGTGGGATCCGAATTCCCAATTCCCT), targeted to a sequence in the 5´ UTR common to all DPR protein-only constructs. Oligonucleotides were end-labelled with T4 Polynucleotide Kinase and γ32-P-ATP. Hybridisation was performed using Ultrahyb-Oligo hybridisation buffer (Ambion) at 42°C overnight. Membranes were washed twice with 2xSSC at 42°C and then exposed to X-ray films. For normalization, blots were re-hybridised with a probe detecting ribosomal protein RpL32 transcripts. For control dot blots, $(GGCCCC)_5$ and $(GGGGCC)_5$ oligonucleotides were bound to a bright star plus nylon membrane (Ambion) and co-hybridised with the corresponding northern blots.

RNA fluorescent in situ hybridisation

Neuroblastoma cell lysates – 24 hours after transfection, cells were fixed in 4% paraformaldehyde (in PBS) for 10 minutes, washed briefly in 2xSCC, and then incubated in pre-hybridisation solution (40% formamide, 2xSSC) for 30 minutes at 37°C. Hybridisation using a Cy3-labelled $(GGCCCC)_4$ RNA probe (IDT) (0.4 ng/µl probe, 40% formamide, 2xSSC, 0.2% BSA, 1 mg/ml salmon sperm DNA, 1 mg/ml tRNA, 10% dextran sulphate, 2 mM vanadyl ribonucleoside) was performed at 37°C for 2 hours in a humidified chamber. This was followed by three 20-minute washes (40% formamide, 2xSCC) at 37°C with shaking, and three 10-minute washes (2xSCC) at room temperature. Samples were mounted using ProLong Gold antifade reagent, which includes DAPI (Life Technologies).

Drosophila – Salivary glands from fly larvae heat-shocked at 37°C for 1.5 hours and allowed to recover for 3 hours at 25°C were dissected in PBS and fixed in 4% paraformaldehyde (in PBS) for 20 minutes, washed 3 times in PBS-T (PBS containing 0.3% Tween), and then placed in methanol at -20°C for at least 12 hours. Samples were re-hydrated in a methanol gradient back to PBS, treated with 0.5 µg/ml proteinase K (in 50 mM Tris, 50 mM EDTA) for 1 hour at room temperature, then washed in PBS, followed by 2xSSC. Samples were then

incubated in pre-hybridisation solution (50% formamide, 2xSSC) for 30 minutes at 80°C. Hybridisation using an Alexa488-labelled $(GGCCCC)_4$ RNA probe (IDT) (0.2 ng/µl probe, 50% formamide, 2xSSC, 0.02% BSA, 1 mg/ml salmon sperm DNA, 1 mg/ml tRNA, 10% dextran sulphate, 1.6 mM vanadyl ribonucleoside) was performed at 80°C for 2.5 hours. This was followed by three 10-minute washes (50% formamide/0.2xSCC) at 80°C, three 10-minute washes (0.2xSCC) at room temperature, and then briefly in PBS. Samples were mounted using Vectashield (Vector Labs), which includes DAPI.

Imaging and quantification of RNA foci

Neuroblastoma cell lysates - Fifteen 40× z-stack images per coverslip were taken at high resolution (2048×2048 pixels) using a 1.4 NA objective on a confocal microscope (LSM710, Zeiss) to yield a minimum of 100 transfected cells for analysis. Quantification of foci was performed using Volocity image analysis software (Perkin Elmer) on maximum intensity projections of each z-stack. ZsGreen-positive cells, DAPI-stained nuclei and RNA foci were identified by fluorescence intensity, using the same threshold for all images; objects smaller than 0.05 μm^2 or larger than 6 μm^2 were excluded for RNA foci analysis. Presence of RNA foci in transfected cells was determined by co-localisation with DAPI and compartmentalisation into ZsGreen-positive cells. Data are presented as the percentage of transfected cells containing RNA foci.

Drosophila – One 40× z-stack image was taken per salivary gland at high resolution (2048×2048 pixels) using a 1.4 NA objective on a confocal microscope (LSM710, Zeiss). RNA foci-positive nuclei were counted and used to calculate the percentage of foci-positive cells per salivary gland.

Immunoblotting

DPR Antibodies – Anti-poly-(GP) antibody was generated by immunising rabbits with a $(GP)_7$-LPH (limulus polyphemus hemocyanin) fusion protein (Biogenes). Resulting antiserum was purified against the immunisation peptide (Biogenes) producing a monospecific antibody to poly-(GP). Anti-poly-(GR) antibody was rabbit polyclonal 23978-1-AP from Proteintech.

Drosophila – For western blotting, adult flies were fed for 4 days with 200 µM mifepristone (RU486) to induce expression. Fly heads were homogenised in 2x Laemmli sample buffer (4% SDS, 20% glycerol, 120 mM Tris-HCl pH6.8, 200 mM DTT with bromophenol blue) and heated to 95°C for 5 minutes. Samples were separated on 4-12% NuPAGE Bis-Tris gels (Invitrogen) and transferred to Immobilon membrane (Millipore). The membrane was blocked in 5% milk in TBS containing 0.05% Tween-20 (TBS-T) for 1 hour and then probed with rabbit anti-poly-(GP) antibody (1:1000 in TBS-T) at 4°C overnight, or with mouse anti-actin antibody (ab8224, Abcam) for 2 hours at room temperature. HRP-conjugated anti-mouse and anti-rabbit secondary antibodies (ab6789 and ab6721, Abcam) were used. Bands were visualized with Luminata Forte (Millipore) and imaged with ImageQuant LAS4000 (GE Healthcare Life Sciences).

To identify poly-(GR) DPR proteins, dot blots were used, due to the difficulty of analysing highly positively-charged proteins with SDS-PAGE. Heads and thoraxes were dissected from adult flies heat-shocked for 1.5 hours and allowed to recover for 24 hours, homogenised in RIPA buffer with protease inhibitors (cOmplete protease inhibitor cocktail without EDTA, Roche), and heated to 95°C for 5 minutes (5 µl of buffer per fly). An aliquot was taken at this point, added in equal amounts to 2x Laemmli sample buffer and heated to 95°C for 5 minutes. This was used to dot blot for the actin loading control. The rest of the sample was

treated with 10 µg/ml of proteinase K at 37°C for 30 minutes, added in equal amounts to 2x Laemmli sample buffer, and heated to 95°C for 5 minutes. Samples (2 µl) were blotted onto dry nitrocellulose membrane and allowed to dry for 15 minutes. Membranes were processed as above, probed with rabbit anti-poly-(GR) antibody (1:1000 in TBS-T).

Neuroblastoma cell lysates – 24 hours after transfection, cells were washed briefly in PBS, lysed in 2x Laemmli Sample Buffer (as above) and heated to 95°C for 10 minutes. Samples were separated on 4-12% NuPAGE Bis-Tris gels (Invitrogen) and transferred to Immobilon membrane (Millipore). The membrane was blocked in 5% milk in PBS containing 0.05% Tween-20 (PBS-T) for 1 hour and then probed with rabbit anti-poly-(GP) antibody (1:2000 in 1% BSA/PBS-T) or rabbit anti-poly-(GR) antibody (1:1000 in 1% BSA/PBS-T) for 2 hours at room temperature. HRP-conjugated anti-rabbit secondary antibody (1:5000, Jackson ImmnoResearch) was used and bands visualized with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific). For loading control, the membrane was stripped using Restore Western Blot Stripping Buffer (Thermo Scientific), blocked in 5% milk in PBS-T, re-probed with rabbit anti-GFP antibody (1:5000 in PBS-T, A-11122, Life Technologies), and continued as above.

Egg-to-adult viability assay and eye phenotypes

Three virgin GMR-Gal4 females were crossed with three UAS males, and flies were allowed to lay eggs for 12-16 hours at 25°C on SYA food. Eggs were counted, and then vials were incubated at 25°C, 27°C or 29°C to increase GAL4-mediated overexpression (*24*). Adult flies were counted, and viability was calculated by dividing the number of adult flies by the number of eggs. 10 replicates per genotype and temperature were used, n = >400 flies per condition.

Eye images of 1-day-old flies expressing repeats or DPR proteins under the control of the GMR-Gal4 driver were taken using a Leica M165 FC stereomicroscope equipped with a motorized stage and a multifocus tool (Leica application suite software). 20 images per fly eye were taken per genotype and treatment.

Lifespan assay

Flies were reared at a standard density on SY medium at 25°C. Two days after eclosion flies were allocated to experimental vials at a density of 10-15 flies per vial (total of 100-150 flies per condition) containing SY medium with or without 200 µM mifepristone (RU486) to induce expression. To reduce protein synthesis in flies, cycloheximide was added to the medium at a concentration of 28 mg/L, together with RU486 (200 µM). Flies were transferred to fresh vials 3 times per week and, at the same time, deaths in each vial were scored. Data are presented as survival curves and comparison between groups was performed using a log-rank test. For simplicity only induced flies are presented in the main text, uninduced controls are presented in fig. S7.

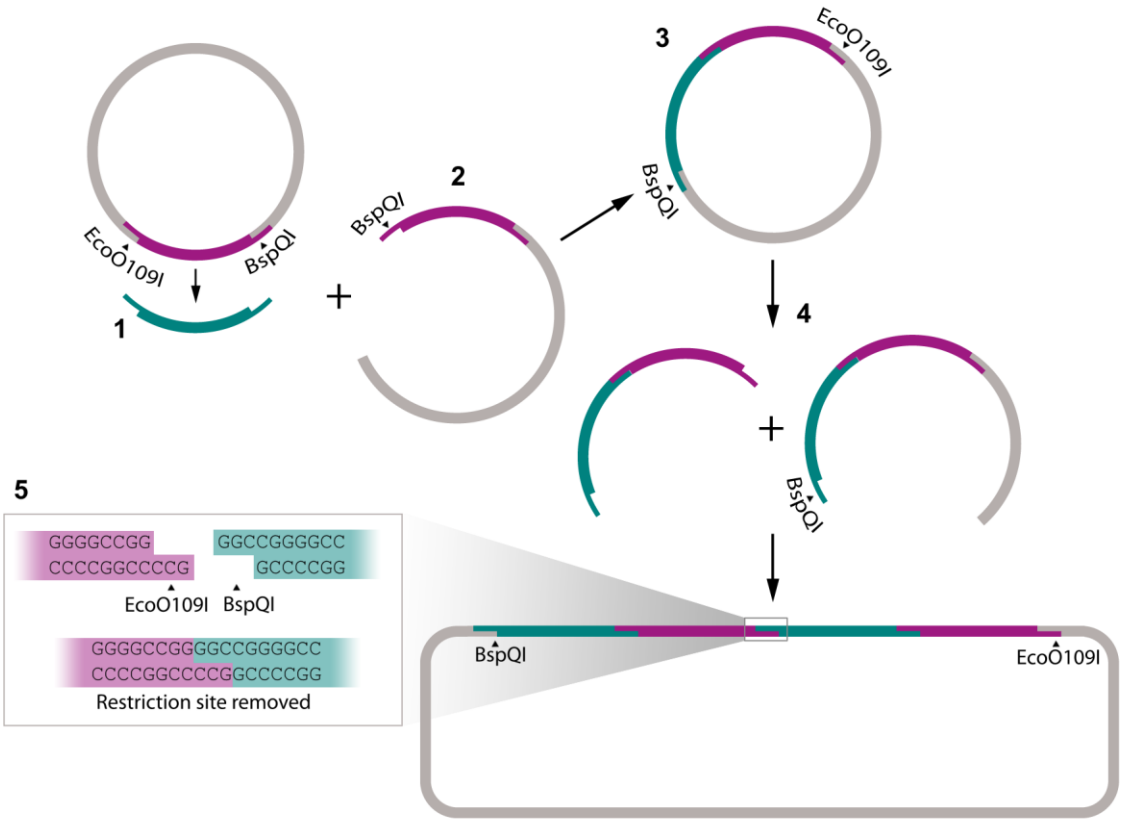

Fig. S1. **GGGGCC repeat cloning strategy.**
A range of GGGGCC repeat sizes was generated using recursive directional ligation (RDL) (*19*). Initial inserts containing 3 or 4 GGGGCC repeats were designed with recognition sites for BspQI (5') and EcoO109I (3'), which produce compatible overhangs when digested. Oligomerization of repeat units was achieved by digestion of the vector with both BspQI and EcoO109I to release an insert (**1**), followed by ligation into the vector, which had been linearized by digestion with only BspQI (**2**). The product contained one (or more) inserts of the original repeat unit, with the restriction recognition sites only maintained at the end of the full insert (**3**). Additional rounds of RDL proceeded identically, using the product from the previous round as the starting material (**4**). Repeat units were joined seamlessly, with no interruption of the GGGGCC repeat sequence, and loss of the BspQI and EcoO109I recognition sites at the site of ligation (**5**).

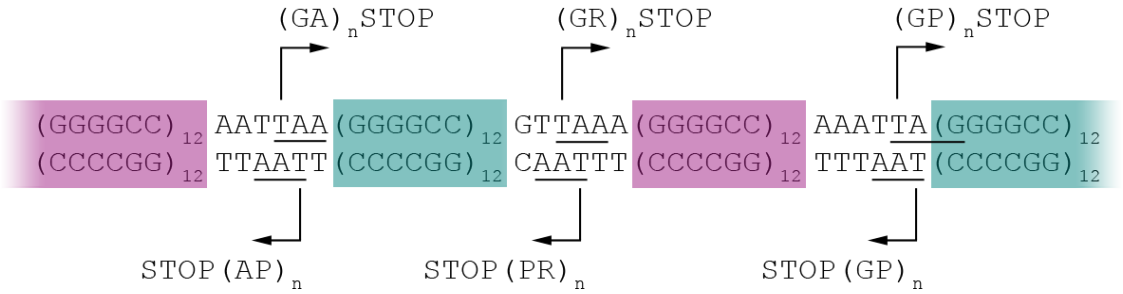

Fig. S2. **Schematic of RNA-only repeat construct design.**
To dissect repeat RNA and DPR protein toxicity, three 6-base-pair interruptions were designed, each containing one stop codon in the sense and one in the antisense direction, and inserted sequentially using recursive directional ligation, described in fig. S1. The resulting RNA-only repeat constructs contained 6-base-pair interruptions every 12 GGGGCC repeats with a stop codon in each sense and antisense frame every 36 repeats.

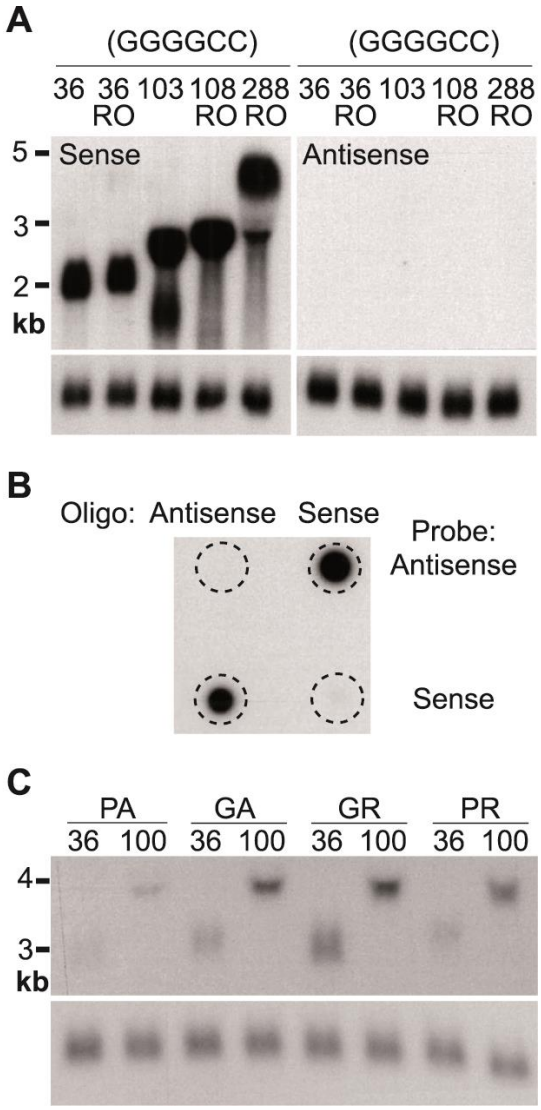

Fig. S3. GGGGCC repeat and DPR protein-only expression in *Drosophila*.
(A) Northern blotting using a $(GGCCCC)_5$ probe to detect sense repeats confirmed that pure and RNA-only (RO) repeat transcripts were expressed at similar levels and that their relative sizes were as expected. Hybridisation with a $(GGGGCC)_5$ probe did not detect antisense transcripts. A probe for RpL32 (lower panel) confirmed equal loading of the gel. All repeat-containing RNA ran at a higher molecular weight than predicted, likely due to the repeat sequence retarding progression through the gel. Genotypes were: *w; UAS-36/daGS, w; UAS-36 RO/daGS, w; UAS-103/daGS, w; UAS-108RO/daGS, w; UAS-288RO/daGS.* **(B)** Control dot blot for the $(GGCCCC)_5$ antisense probe and the $(GGGGCC)_5$ sense probe confirmed that both probes specifically detected the corresponding $(GGGGCC)_5$ sense and $(GGCCCC)_5$ antisense oligonucleotide, respectively. (C) Northern blotting with a probe that targets a common sequence in the 5´ UTR confirmed expression of 36 and 100 protein-only DPRs. GR and GA constructs were expressed at higher levels than PA and PR constructs. Equal RNA loading was controlled by an RpL32 probe (lower panel). Genotypes were: *w; UAS-PA(36)/daGS, w; UAS-PA(100)/daGS, w; UAS-GA(36)/daGS, w; UAS-GA(100)/daGS, w; UAS-GR(36)/daGS, w; UAS-GR(100)/daGS, w; UAS-PR(36)/daGS, w; UAS-PR(100)/daGS.*

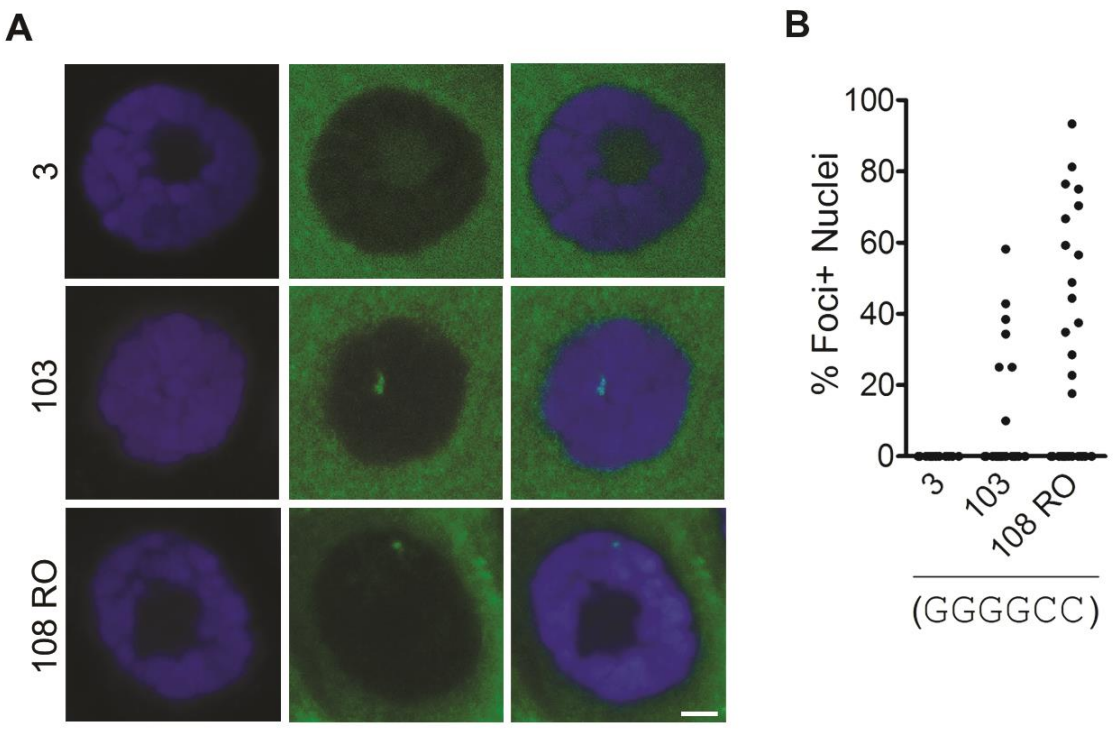

Fig. S4. Pure and RNA-only GGGGCC repeats generated RNA foci in *Drosophila*.
A heat-shock GAL4 (hsGal4) driver was used to express 3 and 103 pure repeats, and 108 RNA-only (RO) repeats in *Drosophila* salivary glands, chosen because of their large nuclei, which allowed visualisation of RNA foci. **(A)** RNA FISH detected RNA foci (green) within nuclei (blue) in flies expressing 103 pure repeats and 108 RO repeats, but not in flies expressing 3 repeats. Scale bar represents 5 μm. **(B)** Quantification of the percentage of salivary gland cells containing nuclear RNA foci. Each dot represents an individual salivary gland. Since the hsGal4 is a lethal insertion, heterozygous driver lines were crossed to homozygous UAS lines. The salivary glands of the progeny were scored without genotyping, therefore only 50% of samples are expected to express the repeat constructs and RNA foci are expected in a maximum of 50% of flies. N = >250 nuclei (12-27 salivary glands). The crosses were: *hsGal4/CyO* males crossed to females of the following genotypes: w; *UAS-3, w; UAS-100, w; UAS-108RO.*

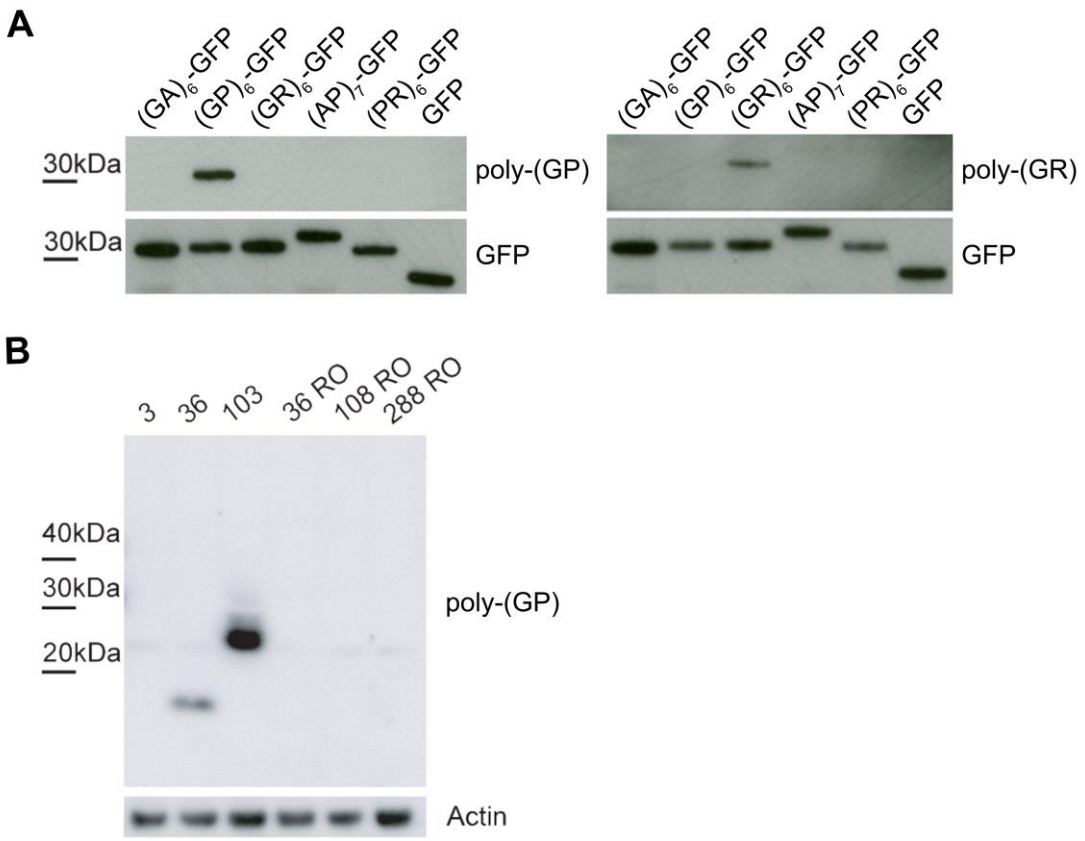

Fig. S5. Poly-(GP) proteins were generated by *Drosophila* expressing pure, but not RNA-only repeats.

(**A**) Characterisation of poly-(GP) (generated by Biogenes) and poly-(GR) (23978-1-AP from Proteintech) antibodies by immunoblotting of neuroblastoma cell line lysates expressing GFP-tagged alternative codon DPR proteins or GFP alone. Blotting showed specific binding of each antibody. Blotting with anti-GFP antibody confirmed expression of all constructs. **(B)** Pure or RNA-only (RO) repeats were expressed in *Drosophila* adult neurons using the elav-GeneSwitch (elavGS) driver. Immunoblots were performed on fly head protein extracts with anti-poly-(GP) antibody, confirming that 36 and 103 pure repeats generated GP DPR protein, and RNA-only repeats did not. Loading control was performed with anti-actin antibody. 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/+*.

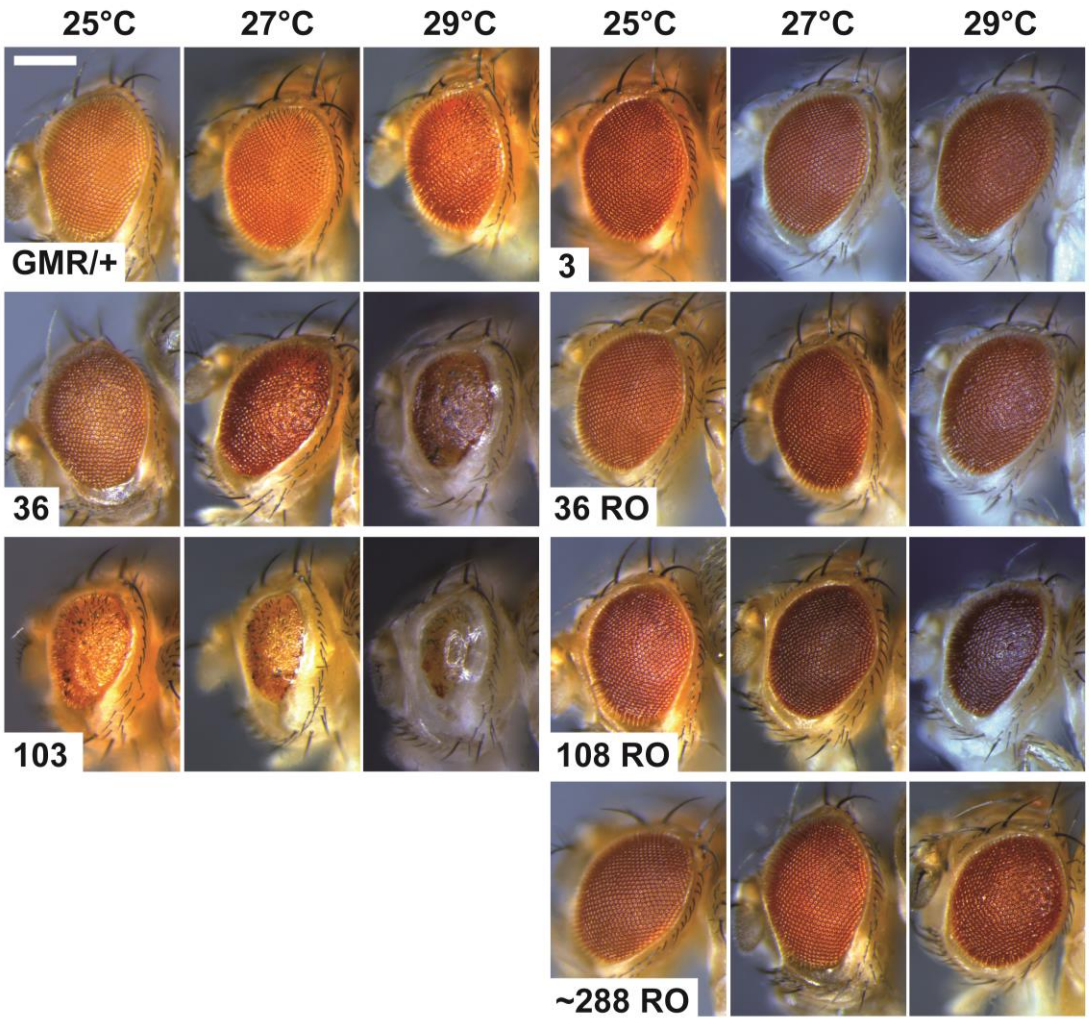

Fig. S6. Temperature-dependent increase in expression exacerbated toxicity of pure GGGGCC repeats.

Increasing the temperature to 29°C caused severe rough eye phenotypes when 36 and 103 pure repeats were expressed in the *Drosophila* eye using the GMR-Gal4 driver. Eye-specific expression of RNA-only (RO) repeats caused no phenotype even at 29°C, when compared to GMR/+ or 3 pure repeat controls. GMR driver alone caused mild toxicity at 29°C, as previously described (*35*). Genotypes were *w; GMR-Gal4/+*, *w; GMR-Gal4/UAS-3*, *w; GMR-Gal4/UAS-36*, *w; GMR-Gal4/UAS-36RO*, *w; GMR-Gal4/UAS-103*, *w; GMR-Gal4/UAS-108RO*, *w; GMR-Gal4/UAS-288RO*. Scale bar represents 200 μm.

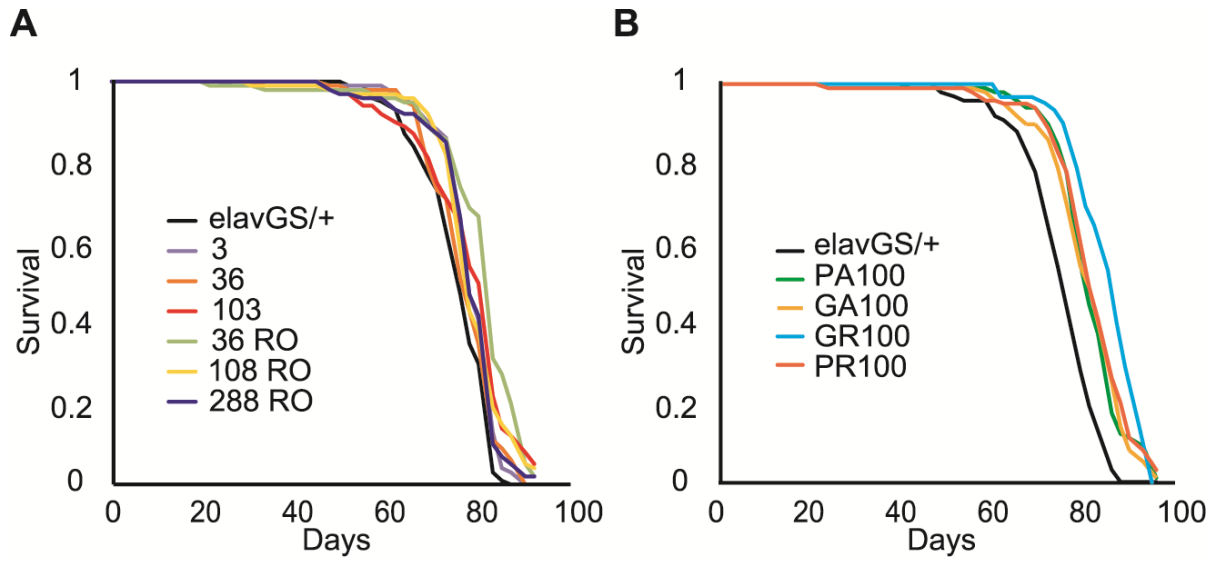

Fig. S7. Control survival curves for uninduced *Drosophila*.
Survival of uninduced female control flies for lifespans shown in Fig. 2D (**A**) and 3D (**B**). (**A**) While median lifespan was not significantly decreased in elav-GeneSwitch (elavGS) driver alone, 3 pure repeats, or 36, 108 and 288 RNA-only (RO) repeat flies compared to their respective uninduced control flies, 36 and 103 pure repeat-expressing flies showed a significantly decreased survival compared to their uninduced controls ($p<0.0001$, 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/+*. (**B**) There was no significant difference in survival between induced and uninduced elavGS driver only or PA100 flies. In contrast, survival was substantially decreased in GA100-, GR100- and PR100-expressing flies compared to their corresponding uninduced controls ($p<0.0001$, log-rank test). Genotypes were: *w; elavGS/+*, *w; UAS-PA100/+; elavGS/+*, *w; UAS-GA100/+; elavGS/+*, *w; UAS-GR100/+; elavGS/+*, w; *UAS-PR100/+; elavGS/+*.

References

1. P. Lillo, J. R. Hodges, Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. *J Clin Neurosci* **16**, 1131-1135 (2009).

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

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

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

5. J. Beck *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* **92**, 345-353 (2013).

6. M. van Blitterswijk *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* **12**, 978-988 (2013).

7. P. Fratta *et al.*, Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathol* **126**, 401-409 (2013).

8. M. B. Harms *et al.*, Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. *Neurobiol Aging* **34**, 2234-2239 (2013).

9. S. Mizielinska *et al.*, C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol* **126**, 845-857 (2013).

10. T. F. Gendron *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* **126**, 829-844 (2013).

11. C. Lagier-Tourenne *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* **110**, E4530-E4539 (2013).

12. T. Zu *et al.*, RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-E4977 (2013).

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

14. P. E. Ash *et al.*, Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron* **77**, 639-646 (2013).

15. T. Lashley, J. Hardy, A. M. Isaacs, RANTing about C9orf72. *Neuron* **77**, 597-598 (2013).

16. K. Mori *et al.*, The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science* **339**, 1335-1338 (2013).

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

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

19. D. E. Meyer, A. Chilkoti, 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* **3**, 357-367 (2002).

20. M. de Haro *et al.*, MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. *Hum Mol Genet* **15**, 2138-2145 (2006).

21. J. P. Orengo *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* **105**, 2646-2651 (2008).

22. N. Sakamoto *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* **276**, 27178-27187 (2001).

23. P. Fratta *et al.*, C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2**, 1016 (2012).

24. J. B. Duffy, GAL4 system in Drosophila: a fly geneticist's Swiss army knife. *Genesis* **34**, 1-15 (2002).

25. I. R. Mackenzie *et al.*, Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* **126**, 859-879 (2013).

26. N. Suzuki *et al.*, The mouse C9ORF72 ortholog is enriched in neurons known to degenerate in ALS and FTD. *Nat Neurosci* **16**, 1725-1727 (2013).

27. S. Almeida *et al.*, Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol* **126**, 385-399 (2013).

28. C. J. Donnelly *et al.*, RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**, 415-428 (2013).

29. K. Mori *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* **125**, 413-423 (2013).

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

31. Z. Xu *et al.*, Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc Natl Acad Sci U S A* **110**, 7778-7783 (2013).

32. M. Markstein, C. Pitsouli, C. Villalta, S. E. Celniker, N. Perrimon, Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* **40**, 476-483 (2008).

33. T. Osterwalder, K. S. Yoon, B. H. White, H. Keshishian, A conditional tissue-specific transgene expression system using inducible GAL4. *Proc Natl Acad Sci U S A* **98**, 12596-12601 (2001).

34. H. Tricoire *et al.*, The steroid hormone receptor EcR finely modulates Drosophila lifespan during adulthood in a sex-specific manner. *Mech. Ageing Dev* **130**, 547-552 (2009).

35. J. M. Kramer, B. E. Staveley, GAL4 causes developmental defects and apoptosis when expressed in the developing eye of Drosophila melanogaster. *Genet Mol Res* **2**, 43-47 (2003).