NEURODEGENERATION

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

Sarah Mizielinska,[1]* Sebastian Grönke,[2]* Teresa Niccoli,[2,3]* 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,[5] Jacqueline Dols,[2] Melissa Cabecinha,[3] Oliver Hendrich,[2] Pietro Fratta,[1,6] Elizabeth M. C. Fisher,[1,6] Linda Partridge,[2,3]† Adrian M. Isaacs[1]†

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, but not stop codon–interrupted "RNA-only" repeats in *Drosophila* caused adult-onset neurodegeneration. Thus, expanded repeats promoted neurodegeneration through dipeptide repeat proteins. Expression of individual dipeptide repeat proteins with a non-GGGGCC RNA sequence revealed that 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 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 two repeats being the most common, whereas 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 the brains of C9FTD/ALS patients (*9–13*). The GGGGCC repeat can be translated in all sense and antisense frames, two of which encode the same DPR; this results 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.
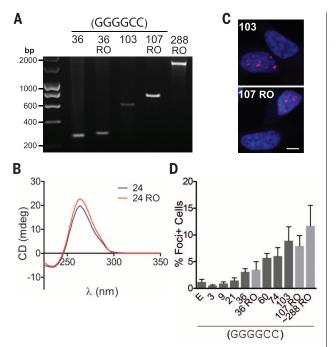
A major obstacle in the investigation of large expanded repeats is that they are inherently unstable. We used recombination-deficient *Escherichia coli* and a cloning strategy termed recursive directional ligation (*19*) to sequentially build seamless pure repeats from small GGGGCC-repeat units (fig. 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% 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 six (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), hence the interruptions did not affect the tertiary structure of the RNA. To investigate the formation of RNA foci, we expressed constructs in the human neuroblastoma cell line SH-SY5Y. RNA fluorescence in situ hybridization (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. 1, C and D).

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 with an antibody to poly-(GR) (Fig. 2A) or to poly-(GP) (fig. S5B) showed that, as expected, the pure repeats generated DPR proteins and the RNA-only repeats did not. Constitutive expression of both 36 and 103 pure repeats in the eye, using the GMR-Gal4 driver, 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 and 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 *Drosophila* 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 after eclosion. No effect was observed for 36, 108, and ~288 RNA-only repeats, which suggested that the neurotoxicity of the pure repeats was attributable to DPR protein production (Fig. 2D). To confirm this, we used a sublethal dose of cycloheximide to reduce protein synthesis in the flies expressing 36 and 103 pure repeats, which ameliorated the reduction in life span caused by the pure repeats (Fig. 2E). This result again showed that toxicity was attributable to DPR proteins.

To assess whether DPR protein expression alone was sufficient for toxicity, we generated "protein-only" constructs by using alternative codons to those found within the GGGGCC repeat. We compared the two arginine-containing DPR proteins, glycine-arginine (GR) and proline-

Fig. 1. Generation and characterization 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 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–labeled SH-SY5Y cells showed that 103 pure and 107 RO repeats both produced nuclear RNA foci (red). Scale bar, 5 μm. (**D**) Quantification of the number of SH-SY5Y cells containing RNA foci after transfection with pure and RO repeats of different lengths; E, empty vector. No difference was observed betwee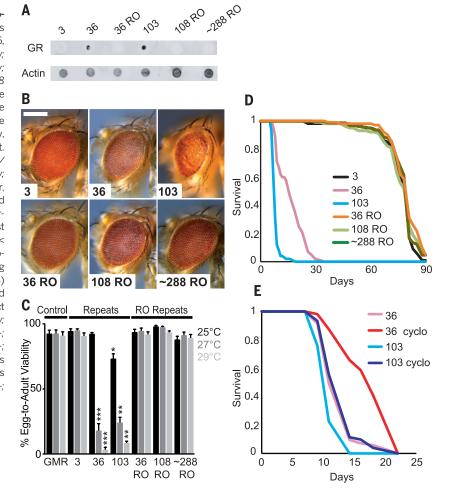n equivalent-length pure and RO repeats [36 pure versus 36 RO, 103 pure versus 107 RO, one-way analysis of variance with Bonferroni test (selected pairs), *n* > 3; error bars represent SEM].

arginine (PR), with two neutral DPR proteins, proline-alanine (PA) and glycine-alanine (GA). When constructs containing *36* DPRs (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, whereas GA and PA DPR proteins had no effect (Fig. 3, A and 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, whereas $(PA)_{100}$ and $(GA)_{100}$ had no effect (Fig. 3, B and 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, whereas $(PA)_{100}$ 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

Fig. 2. Pure GGGGCC repeats caused toxicity via DPR proteins. (**A**) Dot blot showing that 36 and 103 pure repeats generated poly-(GR) proteins, whereas 3 pure repeats and 36, 108, and ~288 RNA-only (RO) repeats did not. Genotypes: *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. We found that 36 pure repeats were mildly toxic, 103 pure repeats showed more overt toxicity, and 3 pure repeats and 36 and 108 RO repeats had no effect. Genotypes: *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, 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, whereas 36, 108, and 288 RO repeats had no effect (*P* < 0.0001, log-rank test). Genotypes: *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: *w;UAS-36/+; elavGS/+, w;UAS-103/+;elavGS/+*.
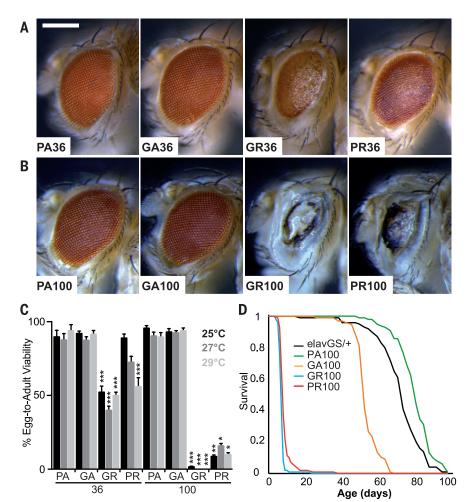
Fig. 3. DPR toxicity was caused by poly-GR and poly-PR proteins. "Protein-only" constructs for individual DPR proteins were expressed in the *Drosophila* eye [(A) to (C)] and the adult nervous system (D). (**A**) $(GR)_{36}$ and $(PR)_{36}$ caused eye degeneration, whereas $(GA)_{36}$ and $(PA)_{36}$ had no effect. Genotypes: *w;UAS-PA36/GMR-Gal4, w;UAS-GA36/GMR-Gal4, w;UAS-GR36/GMR-Gal4, w;UAS-PR36/GMR-Gal4*. Scale bar, 200 μm. (**B**) $(GR)_{100}$ and $(PR)_{100}$ caused extensive eye degeneration, whereas $(GA)_{100}$ and $(PA)_{100}$ had no effect. Genotypes: *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 that $(GR)_{100}$ 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, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; 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: *w;elavGS/+, w;UAS-PA100/+; elavGS/+, w;UAS-GA100/+;elavGS/+, w;UAS-GR100/+;elavGS/+, w;UAS-PR100/+;elavGS/+.*

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. Although 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 suggest that the major toxic species are the DPR proteins. However, the DPR protein toxicity that we observed from overexpression 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 brains of C9FTD patients, 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, mod-

eling 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, *J. Clin. Neurosci.* **16**, 1131–1135 (2009).
2. M. DeJesus-Hernandez et al., *Neuron* **72**, 245–256 (2011).
3. A. E. Renton et al., *Neuron* **72**, 257–268 (2011).
4. E. Majounie et al., *Lancet Neurol.* **11**, 323–330 (2012).
5. J. Beck et al., *Am. J. Hum. Genet.* **92**, 345–353 (2013).
6. M. van Blitterswijk et al., *Lancet Neurol.* **12**, 978–988 (2013).
7. P. Fratta et al., *Acta Neuropathol.* **126**, 401–409 (2013).
8. M. B. Harms et al., *Neurobiol. Aging* **34**, 2234.e13–2234.e19 (2013).
9. S. Mizielinska et al., *Acta Neuropathol.* **126**, 845–857 (2013).
10. T. F. Gendron et al., *Acta Neuropathol.* **126**, 829–844 (2013).
11. C. Lagier-Tourenne et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4530–E4539 (2013).
12. T. Zu et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4968–E4977 (2013).
13. Y. B. Lee et al., *Cell Rep.* **5**, 1178–1186 (2013).
14. P. E. Ash et al., *Neuron* **77**, 639–646 (2013).
15. T. Lashley, J. Hardy, A. M. Isaacs, *Neuron* **77**, 597–598 (2013).
16. K. Mori et al., *Science* **339**, 1335–1338 (2013).
17. K. Mori et al., *Acta Neuropathol.* **126**, 881–893 (2013).
18. T. Zu et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 260–265 (2011).
19. D. E. Meyer, A. Chilkoti, *Biomacromolecules* **3**, 357–367 (2002).
20. M. de Haro et al., *Hum. Mol. Genet.* **15**, 2138–2145 (2006).
21. J. P. Orengo et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2646–2651 (2008).
22. N. Sakamoto et al., *J. Biol. Chem.* **276**, 27178–27187 (2001).
23. P. Fratta et al., *Sci. Rep.* **2**, 1016 (2012).
24. J. B. Duffy, *Genesis* **34**, 1–15 (2002).
25. I. R. Mackenzie et al., *Acta Neuropathol.* **126**, 859–879 (2013).
26. N. Suzuki et al., *Nat. Neurosci.* **16**, 1725–1727 (2013).
27. S. Almeida et al., *Acta Neuropathol.* **126**, 385–399 (2013).
28. C. J. Donnelly et al., *Neuron* **80**, 415–428 (2013).
29. K. Mori et al., *Acta Neuropathol.* **125**, 413–423 (2013).
30. D. Sareen et al., *Sci. Transl. Med.* **5**, 208ra149 (2013).
31. Z. Xu et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7778–7783 (2013).

Science

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

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

Dipeptide repeat peptides on the attack

Certain neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS), are associated with expanded dipeptides translated from RNA transcripts of disease-associated genes (see the Perspective by West and Gitler). Kwon *et al.* show that the peptides encoded by the expanded repeats in the C9orf72 gene interfere with the way cells make RNA and kill cells. These effects may account for how this genetic form of ALS causes disease. Working in *Drosophila*, Mizielinska *et al.* aimed to distinguish between the effects of repeat-containing RNAs and the dipeptide repeat peptides that they encode. The findings provide evidence that dipeptide repeat proteins can cause toxicity directly.

Science, this issue p. 1139 and p. 1192; see also p. 1118

ARTICLE TOOLS	http://science.sciencemag.org/content/345/6201/1192
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2014/08/06/science.1256800.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/345/6201/1118.full http://science.sciencemag.org/content/sci/345/6201/1139.full http://stm.sciencemag.org/content/scitransmed/6/248/248ra104.full http://stm.sciencemag.org/content/scitransmed/5/208/208ra149.full http://stm.sciencemag.org/content/scitransmed/4/165/165ra164.full http://stm.sciencemag.org/content/scitransmed/4/145/145ra104.full
REFERENCES	This article cites 35 articles, 9 of which you can access for free http://science.sciencemag.org/content/345/6201/1192#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions