Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration

Zihui Xu[a,b,1], Mickael Poidevin[a,1], Xuekun Li[a], Yujing Li[a], Liqi Shu[a], David L. Nelson[c], He Li[b], Chadwick M. Hales[d], Marla Gearing[e], Thomas S. Wingo[f,a,d,2], and Peng Jin[a,2]

Departments of [a]Human Genetics, [d]Neurology, and [e]Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322; [b]Division of Histology and Embryology, Department of Anatomy, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, People's Republic of China; [c]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; and [f]Division of Neurology, Department of Verterans Affairs Medical Center, Atlanta, GA 30033

Edited* by Stephen T. Warren, Emory University School of Medicine, Atlanta, GA, and approved February 26, 2013 (received for review November 10, 2012)

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) share phenotypic and pathologic overlap. Recently, an expansion of GGGGCC repeats in the first intron of *C9orf72* was found to be a common cause of both illnesses; however, the molecular pathogenesis of this expanded repeat is unknown. Here we developed both *Drosophila* and mammalian models of this expanded hexanucleotide repeat and showed that expression of the expanded GGGGCC repeat RNA (rGGGGCC) is sufficient to cause neurodegeneration. We further identified Pur α as the RNA-binding protein of rGGGGCC repeats and discovered that Pur α and rGGGGCC repeats interact in vitro and in vivo in a sequence-specific fashion that is conserved between mammals and *Drosophila*. Furthermore, overexpression of Pur α in mouse neuronal cells and *Drosophila* mitigates rGGGGCC repeat-mediated neurodegeneration, and Pur α forms inclusions in the fly eye expressing expanded rGGGGCC repeats, as well as in cerebellum of human carriers of expanded GGGGCC repeats. These data suggest that expanded rGGGGCC repeats could sequester specific RNA-binding protein from their normal functions, ultimately leading to cell death. Taken together, these findings suggest that the expanded rGGGGCC repeats could cause neurodegeneration, and that Pur α may play a role in the pathogenesis of amyotrophic lateral sclerosis and frontotemporal dementia.

RNA-mediated neurodegeneration | fly model

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are heterogeneous illnesses that share a number of clinical, pathological, and genetic features (1). ALS is a fatal degenerative illness primarily affecting motor neurons. The cardinal feature is progressive weakness; however, cognitive impairment and behavioral changes, similar to those seen in FTD, are increasingly recognized symptoms (2). FTD is a progressive dementing illness affecting neurons primarily of the frontal, insular, and anterior temporal cortex, leading to profound changes in personality, behavior, and/or language abilities. A minority of individuals with FTD also develop ALS (3). Pathologically, ALS and FTD are most commonly characterized by abnormal accumulations of TAR DNA-binding protein (TDP-43) (4, 5).

Genetically, the clustering of families who segregate ALS or FTD, typically as an autosomal dominant trait, has supported the idea that each disease has a strong genetic basis. Linkage analysis and family-based genetic studies have found that mutations in several genes, notably TAR DNA-binding protein (*TARDBP*) and fused in sarcoma (*FUS*), are rare causes of both ALS and FTD (6–9). A number of families also have been found to cosegregate ALS, FTD, or both as an autosomal dominant trait and show evidence of linkage to a locus on 9p21 (10).

Recently, the cause of the linkage to the 9p21 locus was found to be an abnormally expanded GGGGCC repeat between exons 1a and 1b in *C9orf72* (11–13). The exact size of the expanded GGGGCC repeat necessary to cause disease is not precisely known. The typical detection method, repeat-primed PCR, is only semiquantitative; however, Southern blot analysis in one family indicated that carriers had 700–1,600 repeats (11). The actual size required to cause disease may be much smaller, and in healthy controls, two and eight repeats appear to be the most common sizes (11). The expansion of GGGGCC repeats has since been shown to be the most common genetic cause of ALS and FTD, with a prevalence of ∼5–7% in European and North American cohorts for each disease (10), and to account for ∼40% of familial ALS and 20% of familial FTD. The GGGGCC expansion has an age-dependent penetrance, with nearly full penetrance by age 80 y (14). Biochemically, *C9orf72* is largely uncharacterized, but appears to be widely transcribed in the brains of normal controls (11, 12).

The slowly progressive nature of ALS and FTD and the noncoding nature of the GGGGCC expansion bear striking similarity to other noncoding nucleotide repeat disorders, such as myotonic dystrophy, fragile X-associated tremor/ataxia syndrome (FXTAS), and spinocerebellar ataxias 8, 10, and 12. How noncoding repeats cause disease is not known. One possibility is that noncoding repeats exert their toxic effects through disruption of nuclear and cytoplasmic RNA processing (15). It was recently reported that noncoding repeats also can be translated into homopolymeric proteins that are presumably toxic (16). Further work is needed clarify the relative contribution of these processes to disease pathogenesis. One possible mechanism for RNA-mediated toxicity is through sequestration of normal RNA-binding proteins (RBPs) by transcribed expanded repeats, causing a depletion of RBPs available for normal RNA metabolism. The effects of depletion of these RBPs include loss of developmentally specific transcripts and frank splicing abnormalities.

Given the genetic evidence that some RBPs, namely *TARDBP* and *FUS*, can cause ALS or FTD, we hypothesized that an expanded riboGGGGCC (rGGGGCC) repeat causes neurodegeneration through alteration of RNA metabolism by sequestering specific RBPs that could bind to rGGGGCC repeats. To test this hypothesis, we developed both mammalian neuronal cells and *Drosophila* models that could express either 3 (normal) or 30 (expanded) rGGGGCC repeats. Repeat sizes were guided by the sizes seen in controls, most commonly two repeats (11, 17). We found that the expression of the expanded rGGGGCC repeats is sufficient to cause neurodegeneration. We next identified Pur α as the main

Author contributions: T.S.W. and P.J. designed research; Z.X., M.P., X.L., Y.L., L.S., C.M.H., M.G., T.S.W., and P.J. performed research; D.L.N., H.L., T.S.W., and P.J. contributed new reagents/analytic tools; Z.X., M.P., X.L., Y.L., D.L.N., H.L., C.M.H., M.G., T.S.W., and P.J. analyzed data; and Z.X., T.S.W., and P.J. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

See Commentary on page 7533.

[1]Z.X. and M.P. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: thomas.wingo@emory.edu or peng.jin@emory.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1219643110/-/DCSupplemental.

RBP that binds the rGGGGCC repeat in vitro, and found that Pur α and rGGGGCC repeats interact in vitro and in vivo in a sequence-specific manner in mammals and *Drosophila* models. We also found that overexpression of Pur α in mammalian neuronal cells and *Drosophila* could mitigate rGGGGCC-mediated neurodegeneration, suggesting that rGGGGCC repeat-binding proteins may be sequestered from their normal function by binding to rGGGGCC repeats. Finally, we found that Pur α could form inclusions in the fly eye expressing expanded rGGGGCC repeats and in the cerebellum of human carriers of expanded GGGGCC repeats in *C9orf72*. Together, these findings imply that the expanded rGGGGCC repeat is sufficient to cause neurodegeneration, and that Pur α could possibly play a role in the pathogenesis of ALS/FTD.

Results

Expression of Expanded rGGGGCC Repeat Causes Neuronal Toxicity in Mammalian Neuronal Cells and *Drosophila*. To determine whether the expanded rGGGGCC repeat could cause neuronal toxicity, we first cloned both normal ($n = 3$) GGGGCC repeats [$(GGGGCC)_3$] and expanded ($n = 30$) GGGGCC repeats [$(GGGGCC)_{30}$] into a mammalian expression vector containing the EFGP reporter gene (Fig. 1*A*). GGGGCC repeats were inserted into the 5′-UTR between the transcriptional start site and translational start site; there is no alternative ATG translational start site between the transcriptional start site and the GGGGCC repeat sequence. These constructs were transiently transfected into Neuro-2a cells, and cell viability determined. At 48 h post-transfection, cells transfected with the construct expressing $r(GGGGCC)_{30}$ displayed significantly reduced viability compared with those transfected with either EGFP alone or the construct expressing $r(GGGGCC)_3$ ($P < 0.0001$; Fig. 1*C*), suggesting that $r(GGGGCC)_{30}$ is sufficient to cause neuronal cell death.

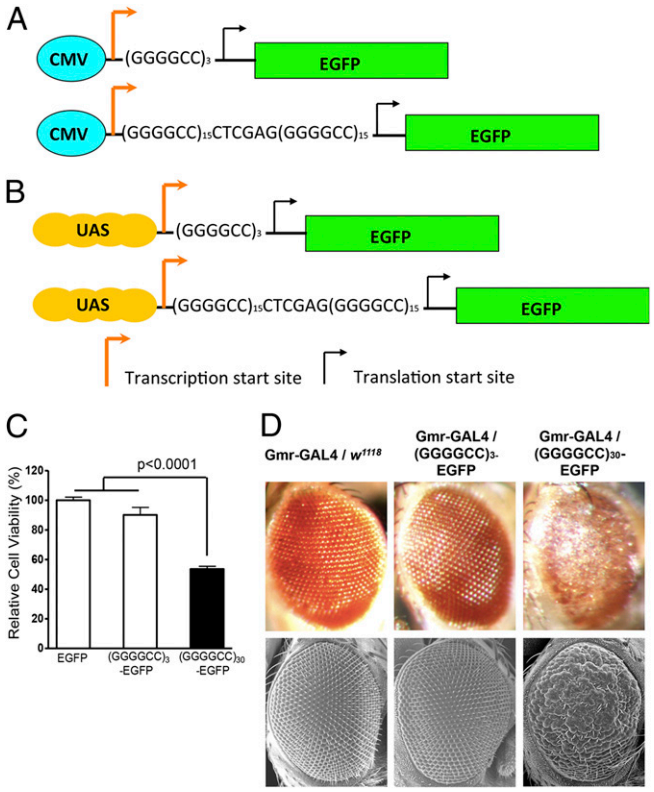

Fig. 1. Expression of expanded rGGGGCC repeat causes neuronal toxicity in mammalian neuronal cells and *Drosophila*. (*A*) Schematic representation of pCMV-$(GGGGCC)_n$-EGFP constructs. Either a 3 GGGGCC repeat or a 30 GGGGCC repeat was inserted upstream of the EGFP coding region, between the indicated transcription and translation start sites. (*B*) Schematic representation of pUAST-$(GGGGCC)_n$-EGFP constructs. Either a 3 GGGGCC repeat or a 30 GGGGCC repeat was inserted upstream of the EGFP coding region between the indicated transcription and translation start sites. (*C*) Results of a cell viability assay of Neuro-2a cells transfected with the pEGFP control vector, pCMV-$(GGGGCC)_3$-EGFP, or pCMV-$(GGGGCC)_{30}$-EGFP. Cells transfected with CMV-$(GGGGCC)_{30}$-EGFP showed significantly reduced survival ($P < 0.0001$, *t* test) compared with cells transfected with either pCMV-$(GGGGCC)_3$-EGFP or pEGFP control. (*D*) Expression of the rGGGGCC repeat disrupts the *Drosophila* eye morphology by light (*top*) and scanning electron (*bottom*) microscope. All flies are shown at 2 wk after eclosion. (*Left*) Flies expressing EGFP alone. (*Center*) Flies expressing $(GGGGCC)_3$-EGFP. (*Right*) Flies expressing $(GGGGCC)_{30}$-EGFP.

Expression of Expanded rGGGGCC Repeat Causes Neuronal Toxicity in Mammalian Neuronal Cells and *Drosophila*. To determine whether the expanded rGGGGCC repeat could cause neuronal toxicity, we first cloned both normal ($n = 3$) GGGGCC repeats [$(GGGGCC)_3$] and expanded ($n = 30$) GGGGCC repeats [$(GGGGCC)_{30}$] into a mammalian expression vector containing the EFGP reporter gene (Fig. 1*A*). GGGGCC repeats were inserted into the 5′-UTR between the transcriptional start site and translational start site; there is no alternative ATG translational start site between the transcriptional start site and the GGGGCC repeat sequence. These constructs were transiently transfected into Neuro-2a cells, and cell viability determined. At 48 h post-transfection, cells transfected with the construct expressing $r(GGGGCC)_{30}$ displayed significantly reduced viability compared with those transfected with either EGFP alone or the construct expressing $r(GGGGCC)_3$ ($P < 0.0001$; Fig. 1*C*), suggesting that $r(GGGGCC)_{30}$ is sufficient to cause neuronal cell death.

We also examined the expression levels of EGFP mRNA and protein and found reduced expression of EGFP (both mRNA and protein) in the cells expressing $r(GGGGCC)_{30}$ compared with those expressing $r(GGGGCC)_3$ or EGFP alone. This finding suggests that the expanded GGGGCC repeat could potentially interfere with gene transcription (Fig. S1).

To evaluate the effect of rGGGGCC expression in vivo, we expressed either $r(GGGGCC)_3$ or $r(GGGGCC)_{30}$ in *Drosophila melanogaster*. To do this, we created multiple transgenic lines using the *Drosophila* transformation vector pUAST-EGFP with normal ($n = 3$) and expanded ($n = 30$) GGGGCC repeats similar to the foregoing mammalian vectors (Fig. 1*B*). Control of transgene expression and tissue specificity was achieved with the upstream activator sequence (UAS)/GAL4 system (18). Transgenic flies produced using the pUAST-EGFP vector alone were used as controls. Transgene expression was directed to neuronal tissues by crossing the lines with different GAL4 drivers. Multiple transgenic lines expressing different levels of EGFP alone, $(GGGGCC)_3$-EGFP, or $(GGGGCC)_{30}$-EGFP were generated. In no case did the expression of EGFP or $(GGGGCC)_3$-EGFP exhibit a phenotypic effect. In contrast, the expression of $(GGGGCC)_{30}$-EGFP had deleterious consequences. When expression was targeted to all developing cells of the peripheral and central nervous systems using Elav-GAL4, the expression of $(GGGGCC)_{30}$-EGFP caused lethality in early development. Furthermore, the expression of $(GGGGCC)_{30}$-EGFP severely disrupted eye morphology when expression was directed to the retina using Gmr-GAL4 (Fig. 1*D*). This finding was replicated with multiple different $(GGGGCC)_{30}$-EGFP transgenic lines that showed different degrees of cell death, loss of pigmentation, and ommatidial disruption not seen with any EGFP- or $(GGGGCC)_3$-EGFP–expressing transgenic lines (Fig. 1*D*). Taken together, these data suggest that expanded rGGGGCC repeats could cause neuronal toxicity in vivo.

Expression of rGGGGCC Repeats Causes Progressive Neurodegeneration in Eye and Motor Neurons of *Drosophila*. Because both ALS and FTD are age-dependent illnesses, we examined the effect of expression of rGGGGCC repeats in the eyes of aged flies. We detected increased disruption of eye morphology in the aged transgenic flies expressing $r(GGGGCC)_{30}$ repeats using the Gmr-GAL4 driver (Fig. 2 *A* and *B*), but no disruption of eye morphology in the flies expressing EGFP or $r(GGGGCC)_3$ repeats using the Gmr-GAL4 driver up to age 28 d.

Given that the primary neuronal type affected in ALS is motor neurons, we used the UAS/GAL4 system to drive the expression of transgenes in motor neurons using the motor neuron-specific driver Ok371-GAL4. Ok371-GAL4 was crossed with UAS-EGFP alone and with the UAS-$(GGGGCC)_3$-EGFP and UAS-$(GGGGCC)_{30}$-EGFP lines that we used previously. To determine the impact of the rGGGGCC repeat on motor neurons, we examined the locomotor activity of these flies using a *Drosophila* activity monitoring (DAM) system. At day 7 after eclosion, we found no difference in either the $rGGGGCC_3$- or $rGGGGCC_{30}$-expressing flies compared with controls; however, at day 28 after eclosion, we observed a significant ($P = 0.0014$) reduction in locomotor activity in the flies expressing the $rGGGGCC_{30}$ repeat, but not in those expressing the $rGGGGCC_3$ repeat, compared with controls (Fig. 2*C*).

Identification of rGGGGCC Repeat RBPs. We posit that one mechanism for the expanded GGGGCC repeat-mediated neuronal

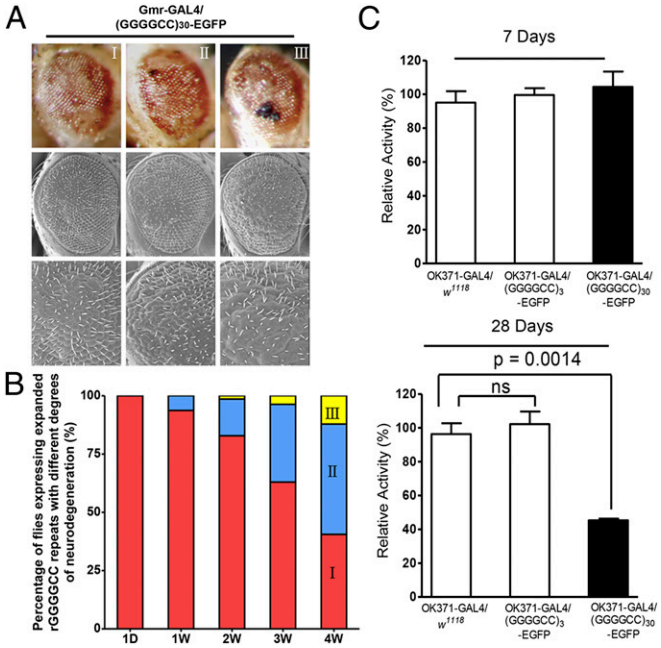

Fig. 2. Expression of rGGGGCC repeats cause progressive neurodegeneration in the eye and motor neurons of *Drosophila*. (*A*) Light microscopy of flies expressing $(GGGGCC)_{30}$-EGFP, with grade I eye disruption defined as <25% ommatidia loss (*left*), grade II eye disruption defined as 25–50% ommatidia loss with small areas of necrosis (*center*), and grade III eye disruption defined as >50% ommatidia loss with large regions of necrosis (*right*). SEM images of fly eyes are shown below. (*B*) Quantification of *Drosophila* eye disruption on day 1 and week 1, 2, 3 and 4. The eye defects are grouped into three categories: I, II, and III. (*C*) Effect of expression of rGGGGCC repeats in the fly motor neurons using the OK371-GAL4 driver shown at day 7 (*Upper*) and day 28 (*Lower*). Locomotion is given relative to the locomotion observed in the control flies at each time point. Thirty flies were tested in each group. No significant difference was observed at 7 d, but significantly decreased locomotion was observed at 28 d ($P = 0.0014$, *t* test) in the flies expressing $rGGGGCC_{30}$ repeats, but not in flies expressing the $rGGGGCC_3$ repeats.

toxicity is sequestration of normal RBPs. Thus, we would expect one or more RBPs to bind the expanded rGGGGCC repeat. To test this idea, we synthesized biotinylated $r(GGGGCC)_{10}$ repeat RNA and incubated it with the whole-cell lysate from mouse spinal cord. Streptavidin-coated magnetic beads were used to capture proteins binding the $r(GGGGCC)_{10}$, and eluted proteins were separated on 4–20% gradient SDS/PAGE gel. We observed two specific bands that intensified with increased amounts of spinal cord lysate, suggesting that one or more RBPs bound the rGGGGCC repeat (Fig. 3*A*).

To determine the specificity of the protein(s) for the rGGGGCC repeat, we performed competition assays using unlabeled $r(GGGGCC)_{10}$ that abolished our ability to capture these protein(s) (Fig. 3*B*). The addition of unlabeled r(CGG), which is very similar to the rGGGGCC repeat and has been implicated in FXTAS, partially reduced the specific RBPs (Fig. 3*C*) (19). This finding suggests that the RBPs have higher avidity for rGGGGCC repeats than for rCGG repeats; alternatively, binding to rGGGGCC and rCGG repeats could occur at different sites or could involve overlapping but distinct protein complexes.

To identify the specific RBPs binding to the rGGGGCC repeats, we excised the specific RBP bands and analyzed them by MS (Fig. 3*D*). The most abundant protein sequenced was Pur α, followed by Pur β and Pur γ (Table S1 and Dataset S1). The MS spectral quantification of the relative abundance of Pur α was corroborated by quantitative RT-PCR (qRT-PCR) of Pur α, Pur β, and Pur γ mRNA of the lysates used for the binding reaction (Fig. S2). Consequently, we focused our attention on Pur α.

Pur α Binds rGGGGCC Repeats in a Dose-Dependent Manner in Vitro and in Vivo. To assess the specificity of Pur α binding to the rGGGGCC repeat, we expressed recombinant GST-tagged mouse and *Drosophila* Pur α, GST-tagged Pur α, and mouse GST-hnRNP A2/B1. We also included hnRNP A2/B1 because it has been shown to bind to rCGG repeats and to modulate rCGG repeat-mediated neurodegeneration, and is predicted to bind to rGGGGCC (11, 19, 20). We performed a series of RNA-binding assays using various recombinant proteins and ^{32}P-labeled $r(GGGGCC)_{10}$. We found that both mouse and *Drosophila* Pur α could bind to rGGGGCC repeats in a dose-dependent fashion, but essentially no binding occurred with the GST control or hnRNP A2/B1 (Fig. 4*A*). For the mouse Pur α, we estimated the dissociation constant, K_d, as 14.8 nM (95% confidence interval, 11.5–18.1 nM) and B_{max} as 1.44 fmol rGGGGCC per μg protein. For *Drosophila*, we estimated the Pur α K_d as 5.1 nM (95% confidence interval, 1.95–8.25 nM) and the B_{max} as 4.67 fmol rGGGGCC per μg of protein. These findings suggest that the interaction between Pur α and rGGGGCC repeat is specific and relatively well conserved between flies and mammals.

To confirm the interaction between Pur α and rGGGGCC repeats in vivo, we performed an immunoprecipitation experiment using mouse Neuro-2a cells. In this experiment, we cotransfected the mammalian EGFP, $(GGGGCC)_3$-EGFP, and $(GGGGCC)_{30}$-EGFP constructs, along with a mammalian expression vector expressing a FLAG-tagged mouse Pur α, into Neuro-2a cells. Immunoprecipitation of cellular lysate was performed with the anti–FLAG-M2 antibody. Immunoprepcipitated proteins and RNA were measured by Western blot analysis using anti–FLAG-M2 antibody and by qRT-PCR using EGFP-specific primers, respectively (Fig. 4*B*). With similar amounts of Pur α immunoprecipitated, we found significant enrichment of EGFP mRNA containing expanded rGGGGCC repeats, but not of EGFP alone or $(GGGGCC)_3$-EGFP. These data demonstrate that Pur α is indeed associated with rGGGGCC repeat-containing mRNA in vivo.

To examine whether the interaction between Pur α and rGGGGCC repeats is conserved in humans, we performed a RBP pull-down experiment using the biotinylated rGGGGCC repeat oligo. In this experiment, we incubated the biotinylated $r(GGGGCC)_{10}$ with brain lysate from mouse and control human frontal cortex. Western blot analysis of eluted proteins using antibodies against Pur α and TDP-43 showed that Pur α, but not TDP-43, is associated with rGGGGCC repeats, suggesting a specific interaction between Pur α and rGGGGCC repeats that is conserved between mouse and human.

Overexpression of Pur α Rescues rGGGGCC Repeat-Induced Neurodegeneration in Mammalian and *Drosophila* Model Systems. If rGGGGCC repeats exert their toxicity, at least in part, by binding Pur α, then coexpression of Pur α with $r(GGGGCC)_{30}$ should ameliorate cell death. To test this idea, we transiently cotransfected the $r(GGGGCC)_{30}$-EGFP construct along with Pur α or hnRNP A2/B1 expression vectors into Neuro-2a cells. At 48 h after transfection, we found that Pur α was able to rescue cell viability but this was not seen with hnRNP A2/B1 (Fig. 5*A*). We also examined the genetic interaction between Pur α and rGGGGCC repeat-mediated neurodegeneration in *Drosophila*. We crossed $(GGGGCC)_{30}$ repeat transgenic fly with UAS-Pur α fly lines that we generated previously in the presence of the Gmr-GAL4 driver (19), and found that overexpression of Pur α could suppress rGGGGCC-mediated neurodegeneration (Fig. 5*B*).

Furthermore, considering that overexpression of Pur α reduced the cell death induced by the rGGGGCC repeat, we reasoned that the loss of Pur α per se should induce cell death. To test this idea, we transfected Neuro-2a cells with siRNA against Pur α, and found significantly reduced cell viability (Fig. 5*C*).

Finally, given that our MS analyses also identified Pur β as a potential rGGGGCC repeat-binding protein, albeit at a much lower abundance, and because Pur β has high homology to Pur α, we also performed cotransfection rescue and siRNA knockdown experiments using Neuro-2a cells to evaluate the role of Pur β in

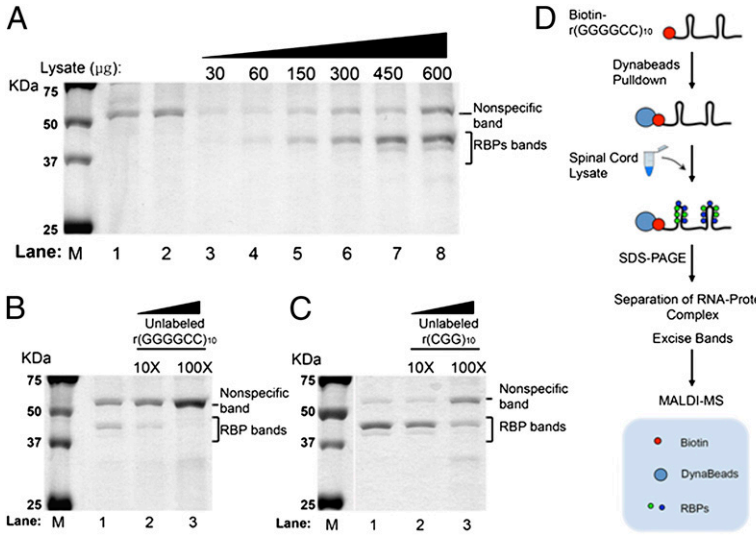

Fig. 3. Identification of rGGGGCC RBPs. (*A*) GGGGCC RNA-binding assays with mouse spinal cord lysates. Biotinylated $r(GGGGCC)_{10}$ repeat was incubated with increasing concentrations of mouse spinal cord lysates. Lane M refers to the molecular weight marker in all blots. Lane 1, 600 μg of spinal cord lysate only; lane 2, 300 μM biotin incubated with 600 μg of spinal cord lysate; lanes 3–8, 300 μM biotinylated $r(GGGGCC)_{10}$ repeat incubated with 30, 60, 150, 300, 450, and 600 μg of spinal cord lysate. (*B*) rGGGGCC repeat RBP competition assay with excess unlabeled $r(GGGGCC)_{10}$ repeat. Lane 1, 300 μM biotinylated $r(GGGGCC)_{10}$ repeat only; lane 2, 300 μM biotinylated $r(GGGGCC)_{10}$ repeat and 10× $r(GGGGCC)_{10}$ repeat; lane 3, 300 μM biotinylated $r(GGGGCC)_{10}$ repeat and 100× $r(GGGGCC)_{10}$ repeat. All lanes were incubated with 300 μg of spinal cord lysate. (*C*) rGGGGCC repeat RBP competition assay with excess unlabeled $r(CGG)_{10}$ repeat. Lane 1, 300 μM biotinylated $r(GGGGCC)_{10}$ repeat only; lane 2, 300 μM biotinylated $r(GGGGCC)_{10}$ repeat and 10× $r(CGG)_{10}$ repeat; lane 3, 300 μM biotinylated $r(GGGGCC)_{10}$ repeat and 100× $r(CGG)_{10}$ repeat. All lanes were incubated with 300 μg of spinal cord lysate. (*D*) Work flow schematic for identification of RBPs by MS.

rGGGGCC repeat-mediated neuronal toxicity. Intriguingly, Pur β behaved similarly to Pur α in these experiments (Fig. S3).

Pur α Forms Inclusions in Expanded GGGGCC Repeat *Drosophila* Model and Brain in Human Carriers. In addition to TDP-positive inclusions, human carriers of the expanded GGGGCC repeat in *C9orf72* were recently shown to have p62-positive TDP-negative inclusions in a number of brain regions (21). Interestingly, p62-positive TDP-negative inclusions appear fairly specific for expanded repeat carriers (22, 23), which led us to question whether Pur α forms inclusions in *Drosophila* expressing the expanded repeat or in humans who carry the expanded *C9orf72* hexamer repeat. In *Drosohpila*, Pur α inclusions colocalizing with ubiquitin were present in $rGGGGCC_{30}$-expressing flies, but not in $rGGGGCC_3$-expressing flies (Fig. 6*A*). In human cerebellum from individuals with neuropathologically defined frontotemoral lobar degeneration with TDP-43 inclusions (FTLD-TDP), Pur α formed inclusions (Fig. 6*B* and Fig. S4). Specifically, in the cerebellar molecular layer, we found intranuclear inclusions in three of four carriers of expanded GGGGCC repeats, in three of four FTLD-TDP noncarriers, and in one of six controls. The control with inclusions was the oldest individual, examined at age 94 y, and had subjectively fewer inclusions compared with cases. Additional staining results are provided in *SI Text*.

Discussion

Recently, an expanded GGGGCC repeat in *C9orf72* was found to be the most common genetic cause of both ALS and FTD (10). How the expanded GGGGCC repeats cause neurodegeneration is not known. Here we hypothesized that rGGGGCC repeats bind RBPs, promoting sequestration of RBPs away from their normal role in RNA metabolism, thereby leading to neurodegeneration. Here we report that expression of rGGGGCC repeats can cause neurodegeneration. We have identified Pur α as one of the major RBPs binding the expressed rGGGGCC repeats, and have presented both biochemical and genetic evidence indicating that rGGGGCC repeats interact with Pur α in a sequence-specific manner, implicating Pur α in ALS/FTD.

To test whether repeat RNA-mediated toxicity is a potential mechanism of expanded GGGGCC repeats in ALS/FTD, we expressed expanded and normal rGGGGCC repeats in mammalian neuronal cells, and found that only expression of the expanded rGGGGCC repeat induced cell death. We then used a *Drosophila* model to show that expression of expanded rGGGGCC repeats causes age-dependent disruption of the fly eye and reduction in locomotion, expression of rGGGGCC repeats were directed to the fly eye or to the motor neuron, respectively. These effects were specific to the expanded rGGGGCC repeat-expressing flies and were not found in files expressing EGFP or normal rGGGGCC, strongly suggesting that the observed cell death is a direct result of rGGGGCC RNA expression in these cells and not an off-target effect or related to relative expression levels in different lines.

After showing that the expanded rGGGGCC repeats induce neurodegeneration, we identified Pur α as the main RBP that binds the expanded rGGGGCC repeat in vitro using mouse spinal cord lysates, and then confirmed this interaction in vivo. We also found that Pur β bound the rGGGGCC repeat; however, we chose to focus on Pur α because it is more abundant in the brain. The interaction between the rGGGGCC repeat and RBPs in the mouse

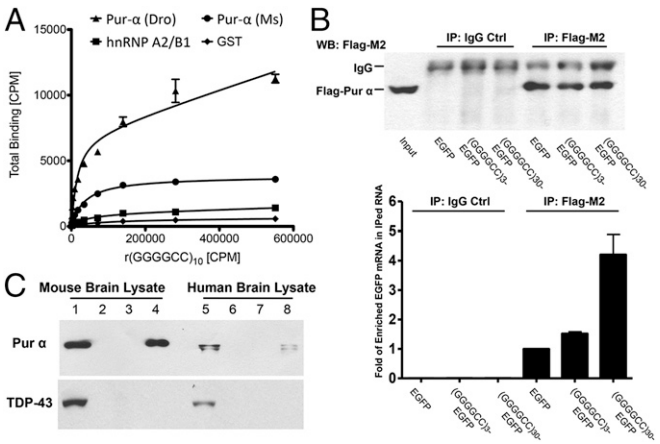

Fig. 4. Pur α binds rGGGGCC repeats in a dose-dependent manner in vitro and in vivo. (*A*) rGGGGCC binding assay with mouse and *Drosophila* Pur α and mouse hnRNP A2/B1. ^{32}P-labeled $r(GGGGCC)_{10}$ was incubated with increasing concentrations of recombinant Pur α (mouse), Pur α (*Drosophila*), hnRNP A2/B1, and GST alone. (*B*) In vivo interaction of rGGGGCC repeats and Pur α. Neuro-2a cells were cotransfected with the FLAG-tagged Pur α and either EGPF, $(GGGGCC)_3$-EGFP, or $(GGGGCC)_{30}$-EGFP constructs, and immunoprecipitation was performed with either mouse IgG or anti–FLAG-M2 antibodies. (*Upper*) Western blots of precipitate using IgG control and anti–FLAG-M2 antibodies. The input lane is a Neuro-2a cell expressing FLAG-tagged Pur α alone. (*Lower*) Results of qRT-PCR for EGFP RNA for each precipitate. (*C*) Pur α binds rGGGGCC from mouse or human brain lysate. Biotinylated $(GGGGCC)_{10}$ repeats were incubated with brain lysate. Lanes 1 and 5, 30 μg of brain lysate; lanes 2 and 6, 300 μg of brain lysate; lanes 3 and 7, 300 μg of brain lysate incubated with 300 μM biotin; lanes 4 and 8, 300 μg brain lysate incubated with 300 μM biotinylated $r(GGGGCC)_{10}$ repeat.

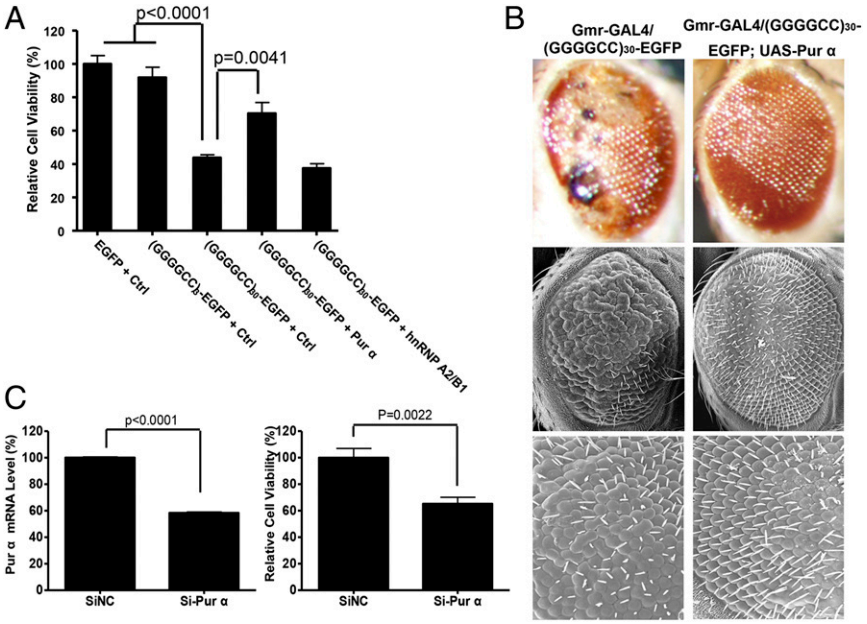

Fig. 5. Overexpression of Pur α suppresses rGGGGCC repeat-mediated neurodegeneration. (*A*) Pur α overexpression attenuates rGGGGCC repeat-mediated cell death in Neuro-2a cells. Neuro-2a cells were cotransfected using GGGGCC repeat-expressing constructs and Pur α or hnRNP A2/B1 expression vectors. A significant reduction in cell viability was seen for Neuro-2a cells transfected with the $(GGGGCC)_{30}$ construct and pcDNA control vector compared with cells transfected with the $(GGGGCC)_3$ construct or control EGFP expression vector ($P < 0.0001$, *t* test). Pur α overexpression in the presence of $(GGGGCC)_{30}$ construct resulted in significantly greater cell viability ($P = 0.0041$, *t* test). (*B*) Light (*top*) and electron (*middle* and *bottom*) microscopy of flies overexpressing Pur α and rGGGGCC repeats in the fly eye. (*Left*) Flies expressing $(GGGGCC)_{30}$-EGFP. (*Right*) Flies expressing $(GGGGCC)_{30}$-EGFP and Pur α. (*C*) Knockdown of Pur α induces cell death in Neuro-2a cells. Transient transfection of siRNA against Pur α significantly reduced Pur α mRNA levels (*Left*) ($P < 0.0001$, *t* test), as detected by qRT-PCR, and cell viability (*Right*) ($P = 0.0022$, *t* test).

brain lysates was specific and only partially reduced rCGG repeats. This finding suggests that Pur α binds rGGGGCC repeats with higher avidity than rCGG repeats, or possibly that there are separate binding sites for rGGGGCC and rCGG repeats that could allosterically modify one another. We also found no evidence of TDP-43 binding the rGGGGCC repeats using either mouse or human brain lysate. Interestingly, we found that hnRNP A2/B1 did not appreciably bind the rGGGGCC repeats either in vitro or in vivo, despite the fact that hnRNP A2/B1 is known to bind the rCGG repeat, similar to Pur α. This result may appear to be at odds with a recent report of hnRNP A2/B1 binding of rGGGGCC repeats in vitro (24), but this discrepancy is likely related to differences in binding conditions and the relative abundance of RBPs in input material between the two studies. These differences also may account for the hnRNP A3 binding of rGGGGCC repeats reported in the previous study (24) but not in the present study.

Finally, we found that overexpression of Pur α could mitigate the rGGGGCC-mediated neurodegeneration in the *Drosophila* and mammalian cells. We also found that Pur β, but not Pur γ, acted similarly to Pur α by inducing cell death after being knocked down and by attenuating rGGGGCC-mediated cell death. Importantly, we detected Pur α inclusions in $rGGGGCC_{30}$-expressing flies. In human subjects, our most intriguing finding was the different frequency of intranuclear inclusions in the cerebellar molecular layer of FTLD-TDP cases compared with controls. The significance of this finding is unclear, given that FTLD-TDP cases with and without expanded GGGGCC repeats show these inclusions, and follow-up in a larger case series is needed. Overall, our data support the hypothesis that the expanded GGGGCC repeat can act by causing a toxic gain of function via sequestration of RBPs from their normal cellular role, and suggest a potential role for Pur α in disease.

In ALS/FTD, there is compelling genetic and pathological evidence that RBPs are involved in disease development. Genetically, mutations in two RBPs, namely TDP-43 and *FUS*, can cause either disease. Pathologically, both illnesses are most commonly characterized by TDP-43–positive protein inclusion. We found evidence that Pur α, as well as Pur β, may be involved in ALS/FTD. An evolutionarily conserved RBP with several purported functions, including modulation of gene transcription and translation, Pur α plays a prominent role in controlling the cell cycle and differentiation (25). It is also a component of the RNA-transport granule and interacts with numerous proteins, including Pur β (26). If GGGGCC repeat-associated ALS/FTD were due to RNA toxicity, then we would expect neurons expressing the *C9orf72* gene to gradually accumulate rGGGGCC repeats that are bound by Pur α, which would eventually lead to the depletion of Pur α. This gradual loss of Pur α could be particularly important given its role in RNA-transport granules, which could lead to a loss of multiple mRNAs transported within the neuron. Indeed, axonal transport defects have long been recognized as a common problem in neurodegenerative diseases (27).

Recent evidence has emerged that potentially toxic homopolymeric proteins are translated from noncoding repeats (16), through a process called repeat-associated non-ATG (RAN) translation,

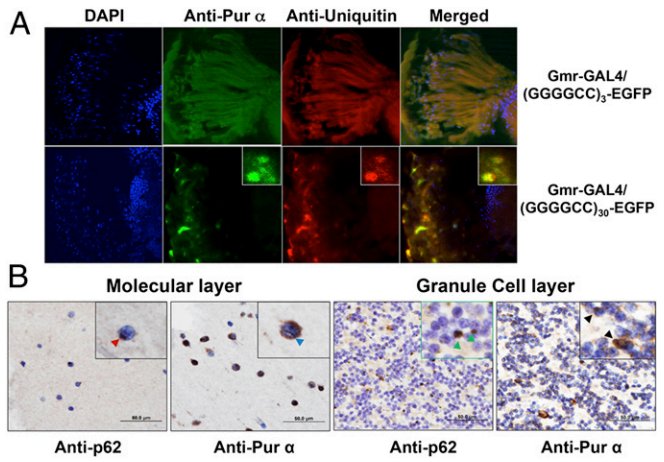

Fig. 6. Pur α inclusions in rGGGGCC *Drosophila* and human GGGGCC expansion repeat carriers. (*A*) *Drosophila* Pur α and ubiquitin colocalize in rGGGGCC-induced inclusions in flies expressing expanded rGGGGCC repeats. Confocal images are shown of eye transverse sections from 14-d-old flies with either $rGGGGCC_3$ (control) or $GGGGCC_{30}$ *in trans* to gmr-GAL4, stained with antibodies against ubiquitin (red) and Pur α (green). The nuclei were stained with DAPI (blue). (*B*) Pur α forms inclusions in the cerebellum of humans with FTLD-TDP. Molecular and granule cell layers of the cerebellum in individuals with FTLD-TDP with expanded rGGGGCC repeats are shown stained for p62 and Pur α. In the molecular layer, p62-positive TDP-negative intranuclear inclusions are shown as red arrowheads, and Pur α-staining intranuclear inclusions are shown as blue arrowheads. In the granule cell layer, p62-positive TDP-negative inclusions are shown as green arrowheads, and Pur α inclusions are shown as black arrowheads.

and evidence of RAN translation has been reported in *C9orf72* expanded repeat carriers (28, 29). Whether these homopolymeric proteins contribute to disease pathogenesis remains to be determined, however. Both RAN translation and RNA-mediated toxicity may contribute to disease additively or synergistically. Indeed, it is plausible that RAN translation could be blocked by the addition of Pur α and other RBPs that specifically bind rGGGGCC repeats either through direct interaction with rGGGGCC repeats or through recruitment of other proteins.

All models have limitations, and our use of the 30 GGGGCC repeats as a surrogate for expanded repeats potentially may limit the generalizability of our work. Available data indicate that 90% of human alleles are at or below eight repeats, with two repeats accounting for >50% of alleles (11, 17). On the other hand, most affected carriers appear to have hundreds to thousands of repeats, not 30 repeats; however, there is evidence of somatic instability of the repeat, and the minimal repeat size necessary to cause disease is unknown (11). In an overexpression system such as the one we used in the present study, the actual copy number of repeats is likely far greater that seen under physiological conditions. In addition, we found that Pur α from different species has varying avidity for the rGGGGCC repeats, which is highly likely to influence the threshold for detectable cell loss. Our model represents the largest repeat that we could stably introduce; we deemed it infeasible to use GGGGCC repeats in the context of the *C9orf72* gene itself. Indeed, although it is possible that reduced *C9orf72* gene expression may contribute to human disease; it is clear that the specific gene is neither necessary nor sufficient for GGGGCC repeats to exert their toxic effects.

In conclusion, our findings indicate that the expression of rGGGGCC in both mammalian and *Drosophila* systems can cause neurodegeneration. Pur α is a major protein binding rGGGGCC, and the interaction between Pur α and rGGGGCC is sequence-specific and dose-dependent. We found that overexpression of Pur α in either mammalian cells or *Drosophila* can suppress rGGGGCC-mediated neurodegeneration, and we also found evidence of accumulation of Pur-α in *Drosophila* expressing $rGGGGCC_{30}$. Staining for Pur α in human cases reveals inclusions in *C9orf72* carriers and noncarriers with FTLD-TDP. Taken together, these findings imply that the newly identified GGGGCC expansion in ALS/FTD acts in part via RNA-mediated toxicity, and point to Pur α as an RBP with a possible role in ALS/FTD disease pathogenesis.

Materials and Methods

Plasmid Construction. Synthetic oligonucleotides were generated with the following sequence and restriction enzyme sites: EcoR1-$(GGGGCC)_3$-Kpn1, EcoR1-$(GGGGCC)_{15}$-Xho1, and Xho1-$(GGGGCC)_{15}$-Kpn1 (Integrated DNA Technologies). DNA oligonucleotides were digested and cloned between the transcription site and the translation site of a mammalian expression vector (pEGFP-N3). Further details are provided in *SI Text*.

***Drosophila* Genetics.** The pUAST constructs with 3 and 30 GGGGCC repeats were injected in a w^{1118} strain using standard methods. Transgenic flies were generated by standard P-element transgene injection (BestGene). All flies were maintained at 25 °C. UAS-Pur α transgenic flies were generated as described previously (30). The Gmr-GAL4 fly line was obtained from the Bloomington Stock Center (no. 8605). Further details are provided in *SI Text*.

Immunohistochemistry. Immunohisotchemistry analyses of *Drosophila* and human cases were performed following standard procedures (31); further details are provided in *SI Text*.

ACKNOWLEDGMENTS. We thank the Emory Integrated Microscopy and Microanalytical Facility, members of the P.J. laboratory, and Deborah Cooper for technical assistance and C. Strauss for a critical reading of the manuscript. D.L.N. and P.J. are supported by National Institutes of Health (NIH) Grants R01 NS051630 and R21 NS067461. P.J. is also supported by NIH Grant P50AG025688. T.S.W. is supported by the Department of Veterans Affairs Medical Center, Atlanta, GA. This work does not represent the views of the Department of Veterans Affairs or the United States Government. This work was supported in part by the Proteomics and Neuropathology Cores of the Emory Neuroscience National Institute of Neurological Disorders and Stroke Core Facilities (Grant P30 NS055077).

1. Fecto F, Siddique T (2011) Making connections: Pathology and genetics link amyotrophic lateral sclerosis with frontotemporal lobe dementia. *J Mol Neurosci* 45(3):663–675.
2. Ringholz GM, et al. (2005) Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology* 65(4):586–590.
3. Lomen-Hoerth C, Anderson T, Miller B (2002) The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. *Neurology* 59(7):1077–1079.
4. Neumann M, et al. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314(5796):130–133.
5. Mackenzie IR, et al. (2007) Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Ann Neurol* 61(5):427–434.
6. Daoud H, et al. (2009) Contribution of TARDBP mutations to sporadic amyotrophic lateral sclerosis. *J Med Genet* 46(2):112–114.
7. Vance C, et al. (2009) Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* 323(5918):1208–1211.
8. Kwiatkowski TJ, Jr., et al. (2009) Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* 323(5918):1205–1208.
9. Kabashi E, et al. (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet* 40(5):572–574.
10. Rademakers R, Neumann M, Mackenzie IR (2012) Advances in understanding the molecular basis of frontotemporal dementia. *Nat Rev Neurol* 8:423–434.
11. DeJesus-Hernandez M, et al. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72(2):245–256.
12. Renton AE, et al.; ITALSGEN Consortium (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72(2): 257–268.
13. Gijselinck I, et al. (2012) A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: A gene identification study. *Lancet Neurol* 11(1):54–65.
14. Majounie E, et al.; Chromosome 9-ALS/FTD Consortium; French Research Network on FTLD/FTLD/ALS; ITALSGEN Consortium (2012) Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *Lancet Neurol* 11(4):323–330.
15. Ranum LP, Cooper TA (2006) RNA-mediated neuromuscular disorders. *Annu Rev Neurosci* 29:259–277.
16. Zu T, et al. (2011) Non–ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci USA* 108(1):260–265.
17. Rutherford NJ, et al. (2012) Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype. *Neurobiol Aging* 33(12):2950.e5–2950.e7.
18. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118(2):401–415.
19. Jin P, et al. (2007) Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. *Neuron* 55(4):556–564.
20. Sofola OA, et al. (2007) RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron* 55(4):565–571.
21. Al-Sarraj S, et al. (2011) p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. *Acta Neuropathol* 122(6):691–702.
22. Snowden JS, et al. (2012) Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. *Brain* 135(Pt 3):693–708.
23. Simón-Sánchez J, et al. (2012) The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. *Brain* 135(Pt 3):723–735.
24. Mori K, et al. (2013) 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(3):413–23.
25. White MK, Johnson EM, Khalili K (2009) Multiple roles for Pur-α in cellular and viral regulation. *Cell Cycle* 8(3):414–420.
26. Kanai Y, Dohmae N, Hirokawa N (2004) Kinesin transports RNA: Isolation and characterization of an RNA-transporting granule. *Neuron* 43(4):513–525.
27. Roy S, Zhang B, Lee VMY, Trojanowski JQ (2005) Axonal transport defects: A common theme in neurodegenerative diseases. *Acta Neuropathol* 109(1):5–13.
28. Mori K, et al. (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* 339(6125):1335–8.
29. Ash PEA, et al. (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 77(4):639–646.
30. Jin P, et al. (2003) RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron* 39(5):739–747.
31. Seyfried NT, et al. (2012) Quantitative analysis of the detergent-insoluble brain proteome in frontotemporal lobar degeneration using SILAC internal standards. *J Proteome Res* 11(5):2721–2738.