

Perspective

Cell Biology

Clogging information flow in ALS

Joseph W. Paul III^{1,2} and Aaron D. Gitler^{1,3}

1. Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305

2. Department of Biology, East Carolina University, Greenville, NC 27858.

3. Correspondence should be addressed to: A.D.G.

Aaron D. Gitler

300 Pasteur Drive

M322 Alway Building

Stanford, CA 94305

agitler@stanford.edu

NOTICE

This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in *Science* 5 September 2014: Volume 345, number 6201, pages 1118-1119, doi:10.1126/science.1259461.

Full text available with subscription or payment at:

<http://www.sciencemag.org/content/345/6201/1118.long>

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, is a devastating neurodegenerative disorder that causes a progressive loss of motor neurons, leading to paralysis and death typically within 2 to 5 years of onset. There are no cures and few treatments. ALS shares some genetic and pathological overlap with another neurodegenerative disease, frontotemporal dementia (FTD), which causes changes to personality and language. Mutations in the gene called *chromosome 9 open reading frame 72 (C9orf72)* are the most common genetic cause of both ALS and FTD. On pages [1139](#) and [1192](#) of this issue, Kwon *et al.* (1) and Mizielińska *et al.* (2), respectively, describe how *C9orf72* mutations might cause neurodegeneration.

Mutations in the *C9orf72* gene were identified as the most common cause of ALS and FTD (3, 4). This discovery completely revolutionized research on both conditions, with intense efforts now focused on understanding the mechanisms by which *C9orf72* mutations cause disease and developing therapeutic strategies to treat patients harboring these mutations. The mutational mechanism is a massive expansion of a hexanucleotide repeat, GGGGCC (G is guanine and C is cytosine) located in a noncoding region of the *C9orf72* gene. Normally, the number of repeat copies ranges from 2 to 23, whereas hundreds or even thousands of copies constitute the disease-causing range. How do these extra GGGGCC repeats cause disease? There are three main hypotheses. One proposal is that expansion of repeats may interfere with the expression of *C9orf72*, thereby resulting in a loss of gene function that is detrimental (5). Another possibility is that the GGGGCC repeat transcript (as well as the antisense CCCC GG repeat transcript) may accumulate in RNA foci (6) and sequester RNA binding proteins, which could disrupt RNA metabolism. There is also the idea based on the perplexing finding that the long repeat is translated into protein in all six reading frames of the RNA (sense and antisense directions) in a manner that does not depend on the presence of the start codon ATG (A, adenine; T, thymine; G, guanine). This repeat-associated non-ATG translation (RAN translation) thus produces the dipeptide repeat proteins glycine-alanine (GA), glycine-proline (GP), proline-alanine (PA), glycine-

arginine (GR), and proline-arginine (PR). These dipeptide repeat proteins are themselves aggregation-prone and accumulate in affected brain regions (7). Of course, these three mechanisms are not mutually exclusive, but determining the contribution of each aberration will be critical for the development of effective therapeutic interventions.

To dissect potential contributions from the repeat RNA and those of the dipeptide translation products to disease, Mizielinska *et al.* generated repeat RNA constructs harboring interruptions that preclude their translation. The authors engineered a series of stop codons in both the sense and antisense RNAs every 12 GGGGCC repeats. This way, the RNA could be expressed but the dipeptide translation products would not. They validated that the pure repeats and interrupted repeats formed the G-quadruplex structures characteristic of GGGGCC repeats (5), indicating that the interruption strategy did not grossly affect the repeat structure. Both pure and interrupted constructs also formed RNA foci when expressed in cultured neuronal cells. The authors then generated transgenic fruit flies expressing either the pure or interrupted GGGGCC repeats. Remarkably, expression of pure repeats (able to form dipeptides) in the fly eye caused degeneration, whereas the interrupted constructs (producing RNA only) did not. Likewise, expressing pure repeats in the nervous system of adult flies led to toxicity and early lethality, whereas the interrupted repeats had no effect. Blocking translation by cycloheximide treatment of flies partially suppressed the phenotype, indicating that toxicity was attributable to translation. Thus, Mizielinska *et al.* demonstrate that the GGGGCC repeats can cause toxicity through the production of aberrant translation products and not from the RNA alone.

But which of the five possible dipeptides (GA, GP, PA, GR, PR) translated from the C9orf72 repeat drives neurodegeneration? Mizielinska *et al.* generated transgenic flies expressing each dipeptide independently, but not in the context of a repeat sequence—that is, rather than using GGGGCC to produce the dipeptides, because the genetic code is degenerate (more than one codon codes for the same amino acid), the authors could use different codons to make the dipeptides of

interest (e.g., GR) without having a repetitive sequence. Expression of two of the dipeptides, GR and PR, was sufficient to cause toxicity and early lethality when directed to the eye or nervous system. The others were less toxic, but the GA dipeptide did show some toxicity later in life. These results focus attention on the arginine-containing dipeptides (GR and PR), and, to a lesser extent, GA, on driving neurodegeneration caused by *C9orf72* mutations in FTD and ALS.

Kwon *et al.* also identify the arginine-containing *C9orf72* dipeptides GR and PR as drivers of toxicity, but present a new hypothesis to explain how they might contribute to FTD and ALS pathogenesis. In a broader sense, their study connects two of probably the most exciting new concepts in ALS and FTD research: *C9orf72* nucleotide repeat expansion and the remarkable property of some RNA binding proteins (e.g., FUS, EWSR1, TAF15, hnRNPA2) harboring low-complexity sequences (also known as prion-like domains) to form polymeric assemblies, which incorporate into hydrogels *in vitro* (8). Such assemblies are akin to RNA granules (e.g., P-bodies and stress granules) that play key regulatory roles in gene expression (9). Many of the RNA binding proteins that can associate with hydrogels have serine-arginine (SR) repeat domains. These SR domain-containing proteins are regulated by phosphorylation (on the serine residues of the SR domains) by kinases. Kwon *et al.* show that this phosphorylation can control the association of SR proteins with hydrogels made up of RNA binding proteins such as hnRNPA2. Phosphorylation releases the SR domain protein from the gel. Mutating most of the serines in the SR domain protein to glycine (e.g., to convert these domains to GR instead of SR) blocks the effects of the kinases, and the mutant SR domain proteins are no longer able to be released from the hydrogel.

Kwon *et al.* recognized that some of the dipeptide RAN translation products produced from the *C9orf72* repeat transcripts resembled the mutated SR domains (e.g., GR and PR). The authors thus proposed an intriguing hypothesis: These RAN translation products associate with hydrogels (or similar types of assemblies *in vivo*, such as RNA-containing granules) but are impervious to the

regulatory action of kinases, and consequently, might clog up the trafficking of SR domain-containing RNA binding proteins that are moving in and out of these granules to transfer information throughout the cell. This block in information flow (e.g., alterations in RNA processing) could lead to cellular dysfunction and might underpin neurodegeneration in ALS and FTD. Indeed, recombinant RAN translation products comprising either polymers of GR or PR bound hnRNPA2 hydrogels and were resistant to kinase treatment (just like the mutated SR domains). When Kwon *et al.* added these recombinant GR and PR proteins to cultured cells, they entered the cells, got into the nucleus, and then localized to the nucleolus, where they caused cell death (see the figure).

To determine why GR and PR proteins are toxic and to define the potential consequences of their accumulation in nucleoli, Kwon *et al.* added one of the dipeptides (PR) to cultured astrocytes (a key cell type relevant for ALS) and performed RNA sequencing to assess potential RNA processing alterations. They identified a number of splicing changes as well as changes in the amounts of different mRNAs and major impairments in synthesis of ribosomal RNA (rRNA). These results suggest that by binding to nucleoli and clogging multiple steps along the pathway of RNA biogenesis, dipeptide products of the C9orf72 repeat expansion could cause disease. Kwon *et al.* hypothesize that the arginine-rich stretches of the GR and PR peptides make them rogue mimics of nuclear localization signals, which may facilitate their entering the nucleus and wreaking havoc on multiple steps of RNA biogenesis. The derangements in RNA processing pathways, perhaps initially mild but accumulating over decades, may eventually lead to neurodegenerative disease pathogenesis.

The studies of Mizielinska *et al.* and Kwon *et al.* strongly implicate dipeptide toxicity as a central component of neurodegeneration. A major challenge is to decipher the mechanism by which RAN translation products contribute to neurotoxicity, and to reexamine dipeptide pathology in human disease with a focus on distinguishing GR and PR from the other ones (e.g., GP, GA, PA). Another recent study by Su *et al.* (10) presents a potentially game-changing biomarker assay to detect and

quantify one of the RAN translation products from patient cerebrospinal fluid. Assays for the abundance of the disease-associated dipeptides could be used in a clinical trial setting to monitor the efficacy of potential disease-modifying therapies that target the repeat RNA expansion. These approaches could include small molecules, as shown by Su *et al.* (10), and antisense oligonucleotides (6, 11, 12). More focus on understanding the molecular mechanisms of RAN translation could reveal new and unexpected therapeutic strategies. Perhaps there are ways to specifically prevent the production of the more toxic dipeptides in exchange for expressing more of some of the benign ones. Unbiased screens using model organisms (yeast, flies, worms, etc.) to identify suppressors of dipeptide toxicity, combined with validation in human patient samples and cell lines, may also reveal mechanisms underpinning dipeptide toxicity and may suggest approaches to mitigate this toxicity.

References

1. Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., et al. (2014). Polydipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science*, 345(6201), 1139–1145. doi:10.1126/science.12549172.
2. Mizielińska, S., Gronke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science*, 345(6201), 1192–1194. doi:10.1126/science.1256800
3. M. DeJesus-Hernandez *et al.*, Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245 (2011).
4. A. E. Renton *et al.*, A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257 (2011).
5. A. R. Haeusler *et al.*, C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507, 195 (2014).

6. C. J. Donnelly *et al.*, RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**, 415 (2013).
7. S. C. Ling, M. Polymenidou, D. W. Cleveland, Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416 (2013).
8. M. Kato *et al.*, Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753 (2012).
9. T. W. Han *et al.*, Cell-free Formation of RNA Granules: Bound RNAs Identify Features and Components of Cellular Assemblies. *Cell* **149**, 768 (2012).
10. Y. R. Li, O. D. King, J. Shorter, A. D. Gitler, Stress granules as crucibles of ALS pathogenesis. *The Journal of cell biology* **201**, 361 (2013).
11. I. Kwon *et al.*, Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049 (2013).
12. Su, Z., Zhang, Y., Gendron, T. F., Bauer, P. O., Chew, J., Yang, W.-Y., et al. (2014). Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron*, 83(5), 1043–1050. doi:10.1016/j.neuron.2014.07.041