# A yeast functional screen predicts new candidate ALS disease genes

Julien Couthouis[a,1], Michael P. Hart[a,1], James Shorter[b,1], Mariely DeJesus-Hernandez[c], Renske Erion[a], Rachel Oristano[a], Annie X. Liu[a], Daniel Ramos[a,d], Niti Jethava[a,d], Divya Hosangadi[a,e], James Epstein[a,f], Ashley Chiang[a,g], Zamia Diaz[b], Tadashi Nakaya[h], Fadia Ibrahim[h], Hyung-Jun Kim[i], Jennifer A. Solski[j], Kelly L. Williams[j,k], Jelena Mojsilovic-Petrovic[l], Caroline Ingre[m], Kevin Boylan[n], Neill R. Graff-Radford[n], Dennis W. Dickson[c], Dana Clay-Falcone[h,o], Lauren Elman[p], Leo McCluskey[p], Robert Greene[h,o], Robert G. Kalb[l], Virginia M.-Y. Lee[h,o], John Q. Trojanowski[h,o], Albert Ludolph[q], Wim Robberecht[r], Peter M. Andersen[m], Garth A. Nicholson[j,k], Ian P. Blair[j,k], Oliver D. King[s], Nancy M. Bonini[i], Vivianna Van Deerlin[h,o], Rosa Rademakers[c], Zissimos Mourelatos[h], and Aaron D. Gitler[a,2]

Departments of [a]Cell and Developmental Biology, [b]Biochemistry and Biophysics, [h]Pathology and and Laboratory Medicine, and [p]Neurology, and [o]Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, PA 19104; Departments of [c]Neuroscience and [n]Neurology, Mayo Clinic, Jacksonville, FL 32224; [d]J. R. Masterman Laboratory and Preparation School, Philadelphia, PA 19130; [e]Central Bucks High School East, Doylestown, PA 18902; [f]Radnor High School, Radnor, PA 19087; [g]Conestoga High School, Berwyn, PA 19312; [i]Department of Biology and the Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, Philadelphia, PA 19104; [j]Northcott Neuroscience Laboratory, ANZAC Research Institute, Sydney 2139, NSW, Australia; [k]Sydney Medical School, University of Sydney, Sydney 2006, NSW, Australia; [l]Department of Pediatrics, Division of Neurology, Abramson Research Center, Children's Hospital of Philadelphia, Philadelphia, PA 19104; [m]Department of Neurology, Umeå University, SE-901 87 Umeå, Sweden; [q]Department of Neurology, University of Ulm, 89081 Ulm, Germany; [r]VIB Vesalius Research Center and Laboratory of Neurobiology, Department of Neurology, Katholieke Universiteit Leuven, B - 3000 Leuven, Belgium; and [s]Boston Biomedical Research Institute, Watertown, MA 02472

This Feature Article is part of a series identified by the Editorial Board as reporting findings of exceptional significance.

Edited by Gregory A. Petsko, Brandeis University, Waltham, MA, and approved October 13, 2011 (received for review June 13, 2011)

**Amyotrophic lateral sclerosis (ALS) is a devastating and universally fatal neurodegenerative disease. Mutations in two related RNA-binding proteins, TDP-43 and FUS, that harbor prion-like domains, cause some forms of ALS. There are at least 213 human proteins harboring RNA recognition motifs, including FUS and TDP-43, raising the possibility that additional RNA-binding proteins might contribute to ALS pathogenesis. We performed a systematic survey of these proteins to find additional candidates similar to TDP-43 and FUS, followed by bioinformatics to predict prion-like domains in a subset of them. We sequenced one of these genes, *TAF15*, in patients with ALS and identified missense variants, which were absent in a large number of healthy controls. These disease-associated variants of TAF15 caused formation of cytoplasmic foci when expressed in primary cultures of spinal cord neurons. Very similar to TDP-43 and FUS, TAF15 aggregated in vitro and conferred neurodegeneration in *Drosophila*, with the ALS-linked variants having a more severe effect than wild type. Immunohistochemistry of postmortem spinal cord tissue revealed mislocalization of TAF15 in motor neurons of patients with ALS. We propose that aggregation-prone RNA-binding proteins might contribute very broadly to ALS pathogenesis and the genes identified in our yeast functional screen, coupled with prion-like domain prediction analysis, now provide a powerful resource to facilitate ALS disease gene discovery.**

In the future, personalized genome sequencing will become routine, empowering us to define the genetic basis of many human diseases. Currently, however, complete genome sequencing for individuals to discover rare pathogenic mutations is still too costly and time consuming. Thus, more creative approaches are needed to accelerate the discovery of disease genes. Moreover, even once genes are revealed, the need for innovative approaches to elucidate causality remains critical.

ALS, also known as Lou Gehrig's disease, is a devastating adult-onset neurodegenerative disease that attacks upper and lower motor neurons (1). A progressive and ultimately fatal muscle paralysis ensues, usually causing death within 2–5 y of disease onset. ALS is mostly sporadic, but ~10% of cases are familial. Pathogenic mutations in several genes have been linked to familial and sporadic ALS, including *SOD1*, *TARDBP*, *FUS/TLS*, *VAPB*, *OPTN*, *VCP*, and others (2). Two of these genes, *TARDBP* (TDP-43) and *FUS/TLS* (FUS) are notable because they encode related RNA-binding proteins that harbor a prion-like domain (3–6). Moreover, both of these proteins have been identified as components of pathological inclusions in neurons of patients with ALS (7–9). Indeed, an emerging concept suggested by the association of FUS and TDP-43 to ALS is that defects in RNA metabolism might contribute to disease pathogenesis. These observations suggested an intriguing possibility: Could TDP-43 and FUS be just the tip of an iceberg? In other words, could other human RNA-binding proteins with properties similar to those of TDP-43 and FUS also contribute to ALS?

Here we report a simple yeast functional screen, followed by bioinformatics to predict prion-like domains, to identify human proteins with similar properties to TDP-43 and FUS. We then identify mutations in human patients with ALS in one gene from this screen, *TAF15*, which ranks with the highest prion-like domain score after FUS. Importantly, we show that TAF15 has similar in vitro and in vivo properties to TDP-43 and FUS. Moreover, the ALS-associated TAF15 mutations are more aggregation prone in vitro, have a more severe effect on lifespan than WT when expressed in *Drosophila*, and increase cytoplasmic mislocalization in mammalian spinal cord neurons. The identification of mutations in an additional RNA-binding protein harboring a prion-like domain further underscores a key role for RNA metabolism defects in ALS and suggests that this class of aggregation-prone RNA-binding proteins might contribute very broadly to ALS and perhaps other related neurodegenerative disorders.

## Results

**Yeast Screen to Identify RNA-Binding Proteins with Properties Similar to TDP-43 and FUS.** TDP-43 and FUS are both RNA recognition motif (RRM)-containing proteins (RRM proteins) (3) and both

Author contributions: J.C., M.P.H., J.S., M.D.-H., T.N., F.I., H.-J.K., G.A.N., I.P.B., N.M.B., V.V.D., R.R., Z.M., and A.D.G. designed research; J.C., M.P.H., J.S., M.D.-H., R.E., R.O., X.A.L., D.R., N.J., D.H., J.E., A.C., Z.D., T.N., F.I., H.-J.K., J.A.S., K.L.W., D.C.-F., L.E., L.M., R.G., N.M.B., V.V.D., R.R., Z.M., and A.D.G. performed research; J.M.-P., C.I., K.B., N.R.G.-R., D.W.D., R.G.K., V.M.-Y.L., J.Q.T., A.L., W.R., and P.M.A. contributed new reagents/analytic tools; J.C., M.P.H., J.S., M.D.-H., R.E., Z.D., T.N., F.I., H.-J.K., I.P.B., O.D.K., N.M.B., V.V.D., R.R., Z.M., and A.D.G. analyzed data; and A.D.G. wrote the paper.

Conflict of interest statement: A.D.G. is an inventor on patents and patent applications that have been licensed to FoldRx Pharmaceuticals.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]J.C., M.P.H., and J.S. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: gitler@mail.med.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1109434108/-/DCSupplemental.

form cytoplasmic inclusions and are toxic when expressed in yeast (Fig. 1*A*) (10–13). Including FUS and TDP-43, there are at least 213 RRM proteins (Protein Families database ID PF00076) present in the human proteome (Dataset S1). This result raised the question of whether other human RRM proteins would show properties like TDP-43 and FUS (e.g., aggregation prone and toxic in yeast). If so, these properties might make these genes top candidates for unique ALS disease genes. We therefore designed a simple yeast functional screen to identify such genes (Fig. 1*B*). Of the 213 human RRM proteins, we were able to obtain a gene clone for 133. We cloned these 133 different human RRM-containing ORFs into yeast expression vectors as YFP fusion proteins, under the control of a galactose-inducible promoter, and individually transformed them into yeast cells. We have previously found that the addition of the YFP tag to TDP-43 and FUS does not affect the aggregation or toxicity properties (10–12). Fluorescence microscopy was used to determine the localization of each protein (nuclear, cytoplasmic, diffuse, foci; Fig. 1*C* and Dataset S1) and spotting assays were used to assess toxicity (Fig. 1*D* and Dataset S1). Some proteins localized to the nucleus (38/133) whereas others were diffusely localized in the cytoplasm (26/133). Interestingly, several others formed multiple foci in the cytoplasm in a pattern strikingly similar to that of FUS and TDP-43 (54/133). Of the proteins that accumulated in the cytoplasm, 38 were also toxic, including FUS and TDP-43 (Table 1). Thus, 38 of 133 human RRM proteins behave like FUS and TDP-43 in yeast cells.

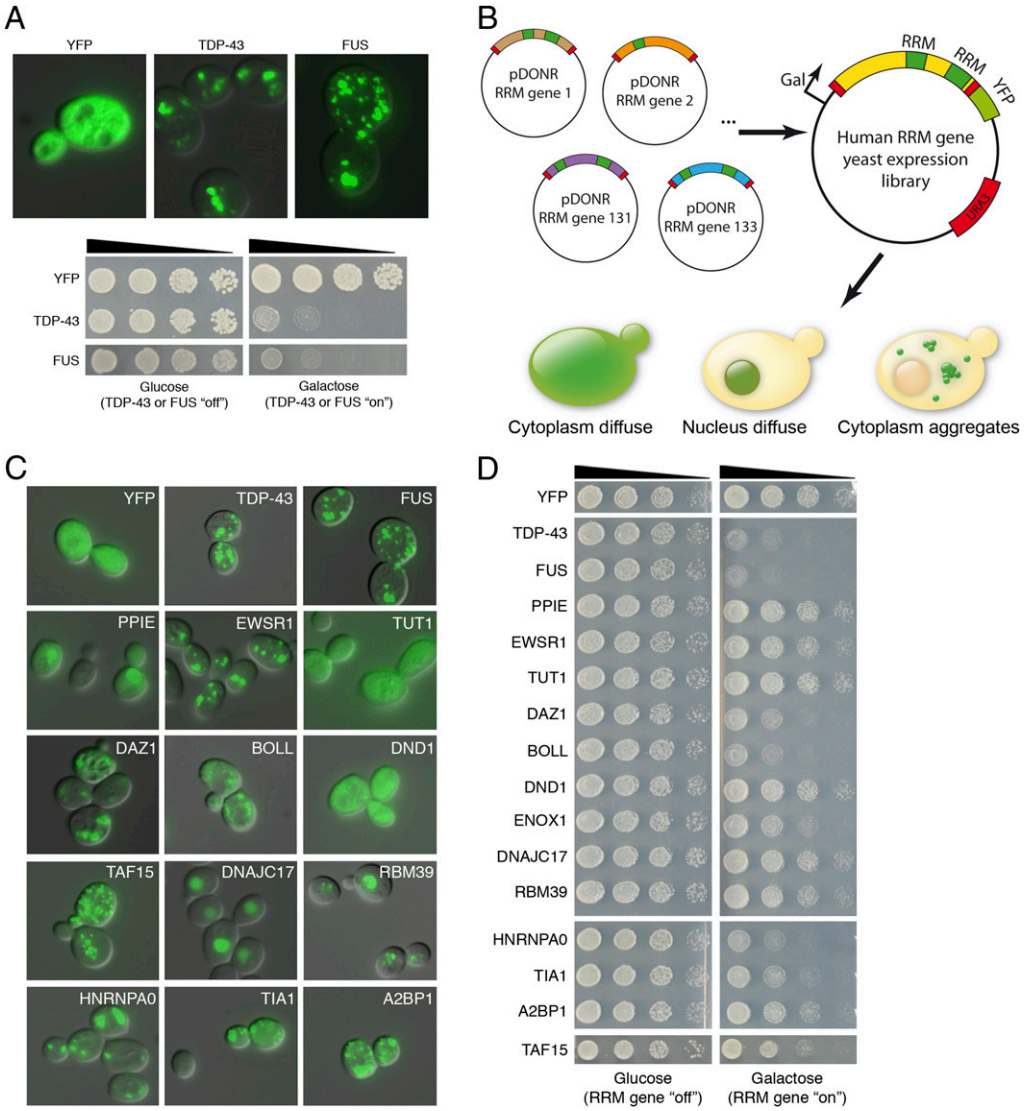


**Fig. 1.** A yeast functional screen identifies human RRM proteins with properties similar to FUS and TDP-43. (*A*) When expressed in yeast, TDP-43 and FUS form multiple cytoplasmic aggregates (*Upper*) and are toxic (*Lower*). (*B*) We designed a yeast functional screen to identify additional human RRM proteins that aggregate and are toxic in yeast. A library of 133 different human ORFs encoding the proteins was cloned into yeast expression vectors as galactose-inducible YFP fusions. We individually transformed each of these plasmids into yeast cells and assessed their effect on aggregation and toxicity by fluorescence microscopy and spotting assays, respectively. (*C*) Examples of various localization patterns in yeast cells of human RRM proteins. Some proteins were localized diffusely throughout the cytoplasm (TUT1 and DND1) and others were localized diffusely in the nucleus (DNAJC17). Some formed multiple foci in the nucleus (RBM39) and several others resembled FUS and TDP-43, which formed multiple cytoplasmic foci (EWSR1, TAF15, HNRNPA0, and DAZ1). (*D*) Spotting assays to assess the toxicity of human RRM proteins. Transformants were grown on synthetic media containing either glucose (control, RRM gene ''off'') or galactose (to induce expression of candidate ORFs, RRM gene ''on''). Some proteins were very toxic when overexpressed (DAZ1, HNRNPA0, FUS, and TDP-43) whereas others were moderately toxic (EWSR1 and TAF15) and others were not toxic (PPIE and DNAJC17). See Table 1 and Dataset S1 for toxicity and aggregation scores.

**Prion-Like Domains in TDP-43, FUS, and Several Other RRM Proteins.** To focus this list further, we used a bioinformatics approach. In addition to the RRM domain, FUS and TDP-43 share a glycine-rich domain and a bioinformatics-predicted prion-like domain (4). Like prion domains found in fungal prion proteins (e.g., Sup35, Ure2, and Rnq1), these domains are enriched in uncharged polar amino acids (such as asparagine, glutamine, and tyrosine) and glycine (14). In TDP-43, the predicted prion-like domain overlaps the glycine-rich domain; in FUS, a QGSY-rich region defines the prion-like domain, although there is some overlap with the glycine-rich domain. The prion-like domain is a shared feature that may be important, given the "prionoid" aggregation propensity of many proteins associated with human neurodegenerative disease (15). Moreover, the prion-like domains of TDP-43 and FUS are critical for the aggregation of these proteins (10–13, 16). Remarkably, using an algorithm to score 21,873

**Table 1. Human RRM proteins with similar properties to FUS and TDP-43 when expressed in yeast**

| Name | Description | Toxicity score (1–4) | Prion domain score | Prion domain rank |
|---|---|---|---|---|
| BOLL | Boule-like (*Drosophila*) | 2 | | |
| CPSF6 | Cleavage and polyadenylation-specific factor | 2.5 | | |
| DAZ1 | Deleted in azoospermia 1 | 2.5 | 14 | 143 |
| DAZ2 | Deleted in azoospermia 2 | 3 | 14 | 143 |
| DAZ3 | Deleted in azoospermia 3 | 3.5 | 15 | 136 |
| DAZAP1 | DAZ-associated protein 1 | 2 | 12 | 198 |
| ELAVL1 | Embryonic lethal, abnormal vision-like 1 | 1 | | |
| ELAVL2 | Embryonic lethal, abnormal vision-like 2 | 1 | | |
| ELAVL3 | Embryonic lethal, abnormal vision-like 3 | 2.5 | | |
| ELAVL4 | Embryonic lethal, abnormal vision-like 4 | 1 | | |
| ENOX1 | Ecto-NOX disulfide-thiol exchanger 1 | 2.5 | | |
| EWSR1 | Ewing sarcoma breakpoint region 1 | 3.5 | 32 | 25 |
| FUS | Fusion (involved in malignant liposarcoma) | 1.5 | 38 | 13 |
| G3BP1 | Ras-GTPase–activating protein | 2 | | |
| HNRNPA0 | Heterogeneous nuclear ribonucleoprotein | 1 | 21 | 81 |
| HNRNPA1 | Heterogeneous nuclear ribonucleoprotein | 1 | 28 | 38 |
| HNRNPA2B1 | Heterogeneous nuclear ribonucleoprotein | 1 | 30 | 32 |
| HNRNPM | Heterogeneous nuclear ribonucleoprotein | 3 | | |
| IGF2BP2 | IGF-II mRNA-binding protein 2 | 2.5 | | |
| IGF2BP3 | IGF-II mRNA-binding protein 3 | 2.5 | | |
| MSI2 | Musashi homolog 2 | 2 | | |
| RALYL | RNA-binding protein-like | 2.5 | | |
| RBM12B | RNA-binding motif protein | 3.5 | | |
| RBM14 | RNA-binding motif protein | 2 | 16 | 117 |
| RBM4 | RNA-binding motif protein | 3 | | |
| RBM41 | RNA-binding motif protein | 2.5 | | |
| RBM4B | RNA-binding motif protein | 2.5 | | |
| RBM5 | RNA-binding motif protein | 3 | | |
| RBM9 | RNA-binding motif protein | 3.5 | | |
| RBMS1 | RNA-binding motif, single-stranded interacting protein | 2 | | |
| RBMS2 | RNA-binding motif, single-stranded interacting protein | 2 | | |
| RBPMS | RNA-binding motif, single-stranded interacting protein | 3 | | |
| ROD1 | Regulator of differentiation | 1 | | |
| SNRPA | Small nuclear ribonucleoprotein polypeptide | 2 | | |
| SNRPB2 | Small nuclear ribonucleoprotein polypeptide | 2 | | |
| TAF15 | TATA box binding protein (TBP)-associated factor | 3.5 | 33 | 22 |
| TARDBP | TAR DNA-binding protein (TDP-43) | 1.5 | 27 | 43 |
| TIA1 | Cytotoxic granule-associated RNA binding protein | 2 | 23 | 55 |

Thirty-eight human RRM proteins that formed cytoplasmic aggregates and were toxic when expressed in yeast are shown. Toxicity was scored from 1 (most toxic) to 4 (not toxic). Prion domain score, based on ref. 14, indicates the maximum log-likelihood for prion-like amino acid composition vs. non-prion–like amino acid composition in any 60 consecutive aa window contained in a region parsed as prion-like by the hidden Markov model. No prion score indicates that no region of length ≥60 was parsed as prion-like. Prion domain rank is from 21,873 human proteins.

human proteins for likelihood of harboring a prion-like domain, FUS and TDP-43 ranked 13th and 43rd, respectively. We therefore interrogated the list of human RRM proteins to identify whether others ranked highly, using the prion domain prediction algorithm (14). Interestingly, 31 of the 213 human RRM proteins ranked in the top 250 (Dataset S1). Among these, FUS and TDP-43 ranked first and 10th, respectively. Of the 38 proteins that were toxic and formed cytoplasmic inclusions in yeast, 13, including FUS and TDP-43, scored highly for a prion-like domain (Table 1). Thus, using the combined yeast screen and prion-like domain analysis, we narrowed the list of RRM proteins by 10-fold (133 human RRM proteins → 38 that aggregate and are toxic in yeast → 13 that also contain a prion domain). We therefore focused on these proteins because they shared similar functional and structural features with FUS and TDP-43: (*i*) they formed cytoplasmic accumulations, (*ii*) they were toxic in yeast, and (*iii*) they contained a predicted prion domain.

**Identification of *TAF15* Sequence Variants in Patients with ALS.** With this list of 13 "FUS- and TDP-43–like" proteins in hand, we sought to test the hypothesis that these additional RRM proteins might contribute to ALS. We gave top priority to TAF15 [RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68 kDa] because it ranked second of 213 human RRM proteins (and 22nd of all 21,873 human proteins) on the basis of the prion domain prediction algorithm (Table 1). Moreover, TAF15 belongs to the same protein family as FUS and is remarkably similar, especially within the RRM, the glycine-rich domain, and the C-terminal RGG domain- and PY-motif–containing region (Fig. 2*A*). Spotting assays showed that TAF15 expression was also toxic, albeit not as toxic as TDP-43 and FUS (Fig. 1*D*). Given these commonalities, we proceeded to sequence exons 13–16 of *TAF15* (NM_139215), a region analogous to where many *FUS* mutations are located (3). These exons comprise the RGG- and PY-motif–containing C-terminal domain, which is important for nuclear localization of FUS (17–19). We performed complete sequencing of these exons in 735 individuals diagnosed with ALS and in 1,328 geographically matched healthy population control individuals (see *SI Materials and Methods* for patient and control demographic information). We found three patient-specific nonsynonymous missense variants (Fig. 2 *B–D* and Dataset S2), c.1258G > A, p.G391E; c.1308C > T, p.R408C; and c.1504G > A, p.G473E. These variants were found in individuals with ages of onset of sporadic ALS of 67 y, 47 y, and 68 y, respectively, and were all located in highly conserved regions of TAF15 (Fig. 2*F*). In addition, we identified one nonsynonymous missense variant that was present in both patients and controls (*TAF15* c.1249G > A, p.R388H). The presence of this variant in control individuals suggests that it likely represents a benign variant, although functional studies are required to distinguish

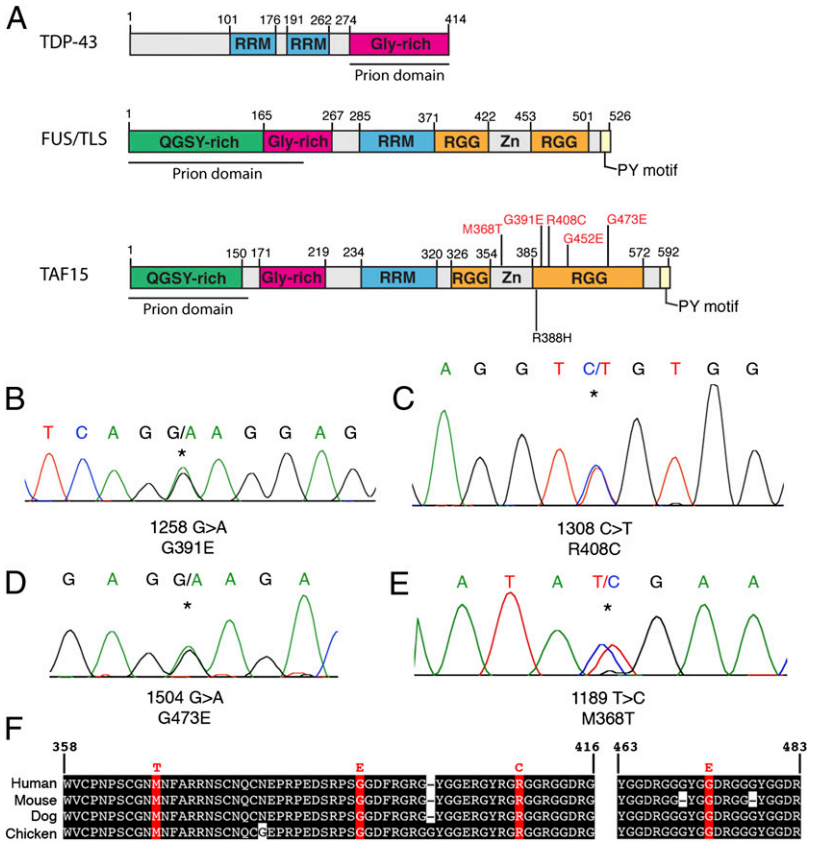


**Fig. 2.** Missense mutations in *TAF15* in patients with ALS. (*A*) Comparison of FUS and TAF15 demonstrates similar domain architecture. Both proteins contain a single RRM, a glycine-rich domain, a predicted prion domain, RGG domains, and a C-terminal PY motif, which can function as an NLS (18). Similar domains are also present in TDP-43. The TAF15 variants identified in exons 13–16 in ALS cases are highlighted in red and the variant, R388H, identified in both cases and controls is highlighted in black. (*B–E*) DNA sequence analysis of *TAF15* in North American Caucasian patients with ALS identified three missense variants. (*B*) A *TAF15* mutation in an ALS case: c.1258 G > A, predicted to lead to p.G391E. (*C*) A *TAF15* mutation in an ALS case: c.1308 C > T, predicted to lead to p. R408C. (*D*) An additional *TAF15* variant c.1504G > A, predicted to lead to p.G473E, identified in the ALS cohort from Mayo Clinic. (*E*) An additional *TAF15* variant c.1189T > C, predicted to lead to p.M368T, identified in an ALS cohort from Sweden. (*F*) Sequence alignment of amino acids 358–416 and 463–483 of TAF15 from diverse vertebrate species indicates that the mutated residues in TAF15 are highly conserved. Identical amino acids have a black background and mutation sites are red.

potentially damaging variants from benign variants (see below). Thus, nonsynonymous missense variations in *TAF15* were detected in 4 of 735 North American ALS cases and 1 of 1,328 North American controls. We also identified a fifth variant (M368T; Fig. 2*E*) in an independent cohort of 351 Swedish patients with ALS and a sixth variant (G452E; Fig. 2*A*) in a cohort of 176 Australian patients with ALS. Because the *TAF15* variants were identified in sporadic ALS cases, familial evidence for segregation with disease was not possible. Notably, however, *TARDBP* and *FUS* mutations have also been confirmed in apparent sporadic ALS cases (20). In addition, the parents of the affected individuals were not available to determine whether the mutations occurred de novo or were inherited.

**ALS-Specific TAF15 Variants Form Cytoplasmic Foci in Spinal Cord Neurons.** To provide evidence that the specific sequence variants that we found in ALS cases might be pathogenic, we performed an unbiased assessment of all nonsynonymous variants identified in cases and controls. We designed an assay that potentially discriminates functionally deleterious variants from benign variants. ALS-linked mutations in TDP-43 and FUS have been shown to disrupt protein localization, leading to enhanced accumulation of ALS-linked variants in cytoplasmic structures (8, 9, 18, 21–23). Given this common feature, we assessed the effects of the TAF15 variants (Fig. 2*A*) on subcellular localization. We transfected myc-tagged WT or the mutant forms of TAF15 into spinal cord neurons cultured from rat embryos (Fig. 3*A*). The transfection efficiency in these primary neuronal cultures was too low to quantify protein expression levels; however, immunoblotting of HEK293T cells transfected with these same constructs confirmed that the WT and variant constructs are expressed at similar levels (Fig. 3*C*). WT or mutant TAF15 expression under these conditions did not result in toxicity; however, we did observe significant effects of the variants on localization.

The transfected WT and mutant TAF15 proteins were restricted to the nucleus in ~40% of cells and present in both nucleus and cytoplasm in ~60% of cells. The TAF15 mutations did not affect nuclear localization per se, but rather promoted the formation of cytoplasmic foci. WT TAF15 and the R388H variant, which was present in both patients and controls, localized to the nucleus and cytoplasm, in a diffuse pattern (Fig. 3 *A* and *B*). In contrast, all three of the ALS-linked mutant forms of TAF15 (G391E, R408C, and G473E) found in the North American patients with ALS showed a striking accumulation of TAF15 in cytoplasmic foci in dendrites and axons (Fig. 3 *A* and *B*). Thus, all of the TAF15 variants that we found in patients with ALS were deleterious; that is, they caused mislocalization of the protein to cytoplasmic foci within neurons, whereas the one variant found in controls was benign and behaved like WT. This unbiased assessment of all variants provides strong support for the variants that we found in ALS cases as being risk factors for ALS (3 damaging TAF15 variants of 735 ALS cases vs. 0 damaging TAF15 variants of 1,328 controls, $P = 0.0451$, Fisher's exact test). The fourth ALS-specific variant (M368T; Fig. 2*E*), found in an independent cohort of 351 Swedish patients with ALS, also affected localization in this assay (Fig. 3 *A* and *B*), but we do not include this variant in the above statistical assessment because we did not sequence an equivalent number of matched controls. The fifth ALS-specific variant (G452E; Fig. 2*A*), in 1 Australian patient with ALS of a cohort of 176 patients with ALS, was not present in 72 sequenced Australian controls or in our collection of 1,328 sequenced North American controls. We have not yet tested the effects of this variant on TAF15 localization in our functional assay. Unlike several of the ALS-linked FUS mutations, which are located in the conserved PY motif at the extreme C terminus of the

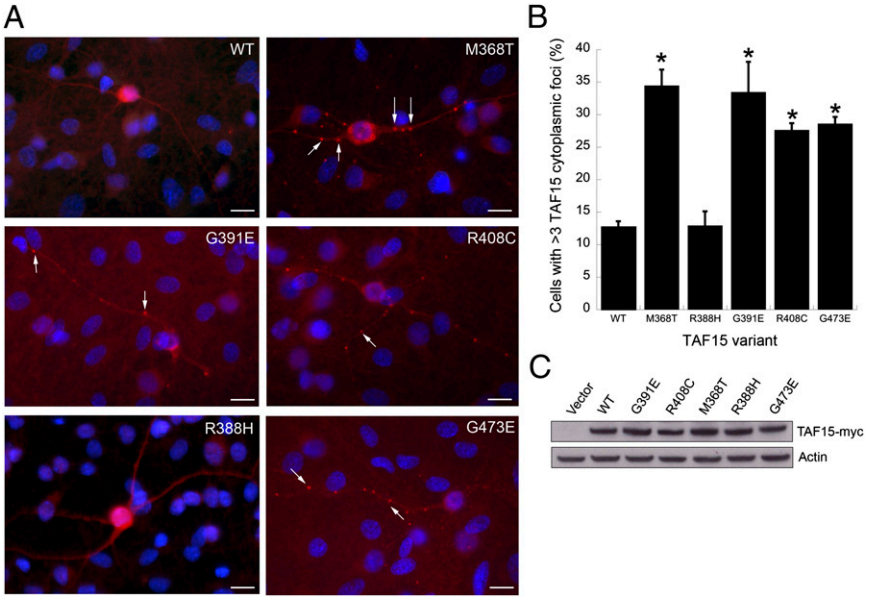


**Fig. 3.** A functional assay to distinguish potentially damaging TAF15 variants from benign variants. We performed an unbiased assessment of all TAF15 missense variants identified from sequencing of ALS cases (M368T, G391E, R408C, and G473E) and controls (R388H). (*A*) Primary rat embryonic neuron cultures were transfected with myc-tagged WT or mutant TAF15 and stained with α-myc (red). Transfection of WT TAF15 or the R388H variant found in controls results in localization within the nucleus and cytoplasm of neurons in a diffuse pattern. In contrast, the ALS-linked mutant forms of TAF15 showed a striking accumulation of cytoplasmic foci (arrows) in dendrites and axons. (Scale bar, 20 μm.) (*B*) Quantitation of transfected WT or mutant TAF15 that accumulates in cytoplasmic puncta. Four of four TAF15 variants found in ALS cases (M368T, G391E, R408C, and G473E) showed cytoplasmic puncta formation whereas the one variant found in both cases and controls (R388H) behaved like WT. *$P < 0.01$ (cytoplasmic puncta formation of TAF15 variants compared with WT, Student's *t* test). Error bars show mean ± SEM. (*C*) Because the transfection efficiency in the primary rat spinal cord neuron cultures was not high enough to detect TAF15 overexpression by immunoblot, we transfected the same constructs of WT and TAF15 mutants into HEK293T cells. Immunoblotting with α-myc antibodies was used to detect myc-tagged TAF15 and α-actin antibody was used as a loading control. In contrast to the primary neuronal cultures, we did not observe a significant difference in aggregation between WT TAF15 and the variants in HEK293T cells.

protein and have strong effects on nuclear localization (18), the TAF15 variants that we have identified are not located in TAF15's PY motif (Fig. 2*A*), possibly explaining why they do not affect nuclear localization per se but seem to promote the accumulation of TAF15 in cytoplasmic foci. As additional variants in TAF15 are identified, this functional assay will be a powerful tool for assessing their potential pathogenicity.

**TAF15 Is Inherently Aggregation Prone.** These genetic and functional studies highlight a potential role for TAF15 in ALS pathogenesis. We next sought further functional evidence that TAF15 has properties similar to TDP-43 and FUS. We asked, does TAF15 spontaneously aggregate in vitro like TDP-43 and FUS (11, 12) and does TAF15 confer neurodegeneration when expressed in the nervous system, like TDP-43 and FUS (24–31)?

We purified bacterially expressed GST-tagged TAF15 as a soluble protein under native conditions, as previously described for TDP-43 and FUS (11, 12). TAF15, TDP-43, and FUS were all competent in RNA binding, suggesting that the RRM domains were correctly folded (11, 12). Upon addition of tobacco etch virus (TEV) protease to specifically remove the N-terminal GST tag and SDS-PAGE to assess purity and expected molecular

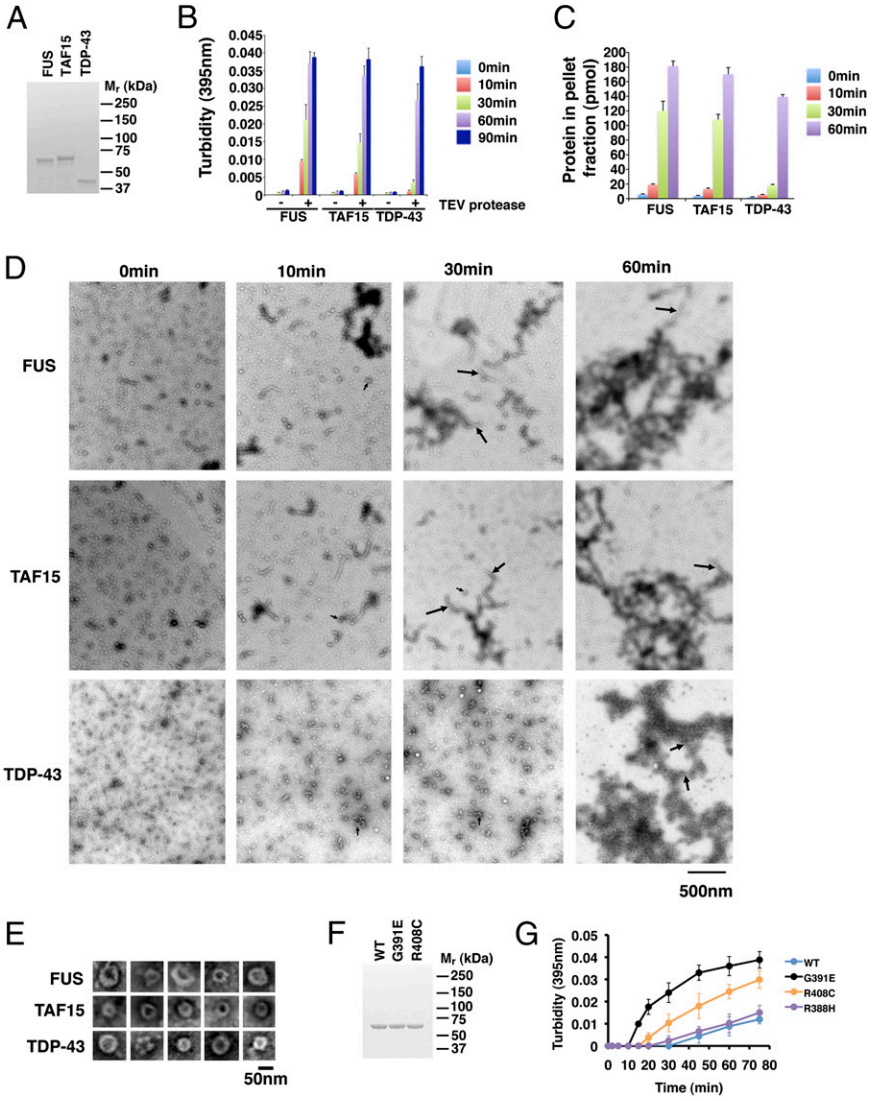


**Fig. 4.** TAF15 is an aggregation-prone protein like TDP-43 and FUS. (*A*) Following TEV protease cleavage to remove the N-terminal GST tag, FUS, TAF15, and TDP-43 proteins were processed for SDS-PAGE and Coomassie stained to confirm purity and expected molecular weight. (*B*) GST-TDP-43, GST-FUS, or GST-TAF15 (3 μM) were incubated in the presence or absence of TEV protease at 25 °C for 0–90 min with agitation. Note that very little aggregation occurs in the absence of TEV protease. The extent of aggregation was determined by turbidity. Values represent means ± SEM ($n$ = 3). (*C*) GST-TDP-43, GST-FUS, or GST-TAF15 (3 μM) were incubated in the presence of TEV protease at 25 °C for 0–60 min. At the indicated times, reactions were processed for sedimentation analysis. Pellet and supernatant fractions were resolved by SDS-PAGE and stained with Coomassie Brilliant Blue. The amount of protein in the pellet fraction was determined by densitometry in comparison with known quantities of the appropriate protein. Values represent means ± SEM ($n$ = 3). A human RRM protein, DND1, which did not aggregate and was not toxic in yeast (Fig. 1 *C* and *D*), was also soluble and did not form aggregates in this assay. (*D*) GST-TDP-43, GST-FUS, or GST-TAF15 (3 μM) were incubated in the presence of TEV protease at 25 °C for 0–60 min. At various times, reactions were processed for EM. Small arrows denote small pore-shaped oligomers and large arrows denote linear polymers. (Scale bar, 500 nm.) (*E*) Gallery of TDP-43, FUS, and TAF15 oligomers formed during aggregation reactions. (Scale bar, 50 nm.) (*F*) Following TEV protease cleavage to remove the N-terminal GST tag, TAF15 wild-type (WT), G391E, and R408C proteins were processed for SDS-PAGE and Coomassie stained to confirm purity and expected molecular weight. (*G*) ALS-linked TAF15 variants G391E and R408C displayed accelerated aggregation kinetics whereas the R388H variant found in both cases and controls aggregated with similar kinetics to WT TAF15.

weight (Fig. 4*A*), we found that TAF15 rapidly aggregated at 25 °C with gentle agitation (Fig. 4 *B* and *C*). Turbidity and sedimentation analysis revealed that TAF15 aggregated with kinetics similar to FUS and slightly more rapidly than TDP-43, assessed by turbidity (Fig. 4*B*). If TEV protease was omitted, then little aggregation occurred (Fig. 4*B*). Electron microscopy studies established that TAF15 rapidly accessed oligomeric forms (Fig. 4*D*), which would frequently adopt a pore-like conformation [Fig. 4*D*

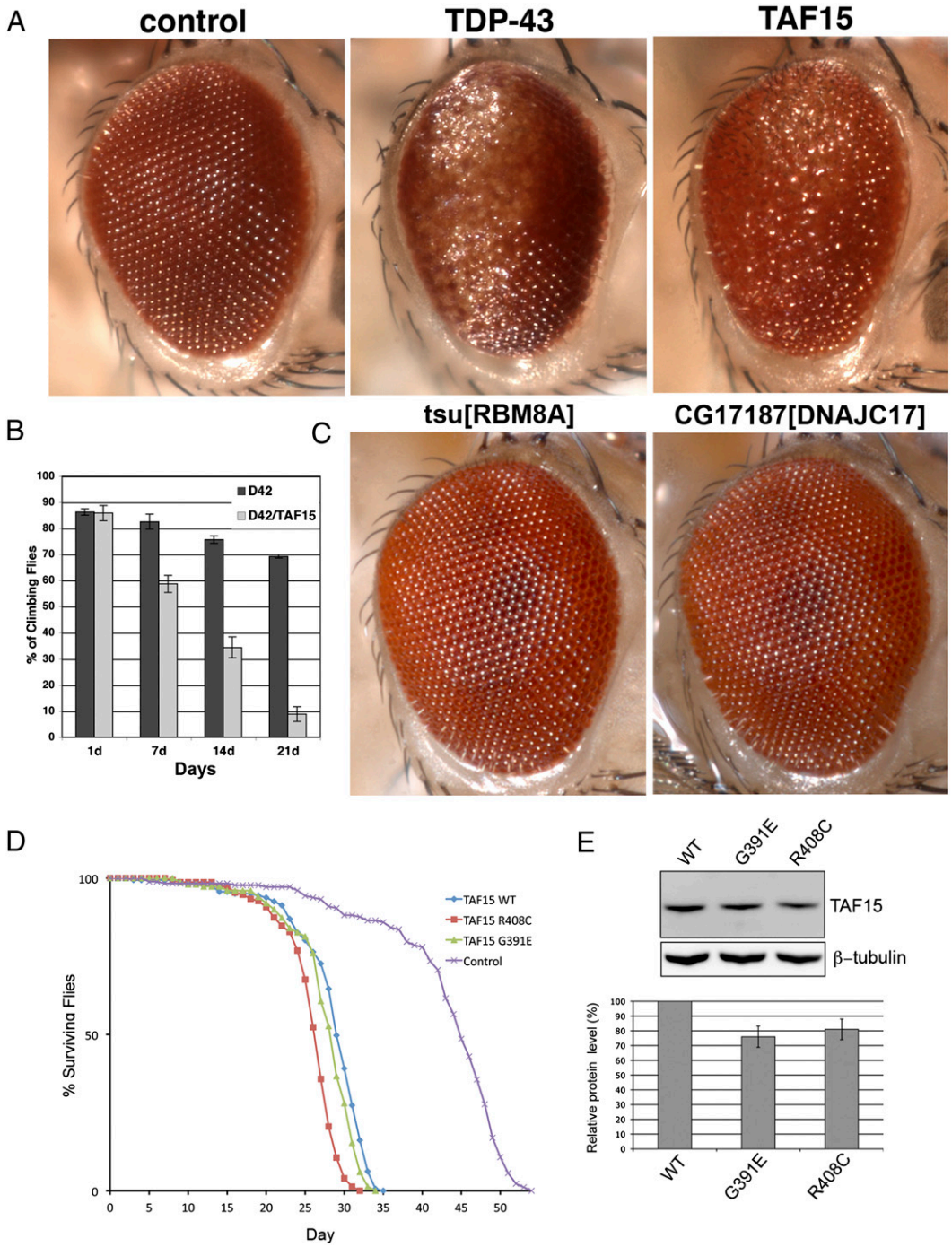


**Fig. 5.** TAF15 confers neurodegeneration and dysfunction in *Drosophila*. (*A*) Toxicity of RRM proteins in the eye. TAF15 causes degeneration and disruption of the retinal structure, akin to TDP-43 (also see ref. 24). Control is driver line alone, gmr-GAL4/+. TDP-43 is gmr-GAL4/UAS-TDP-43-YFP. TAF15 is gmr-GAL4/UAS-TAF15 (grown at 29 °C). (*B*) Progressive loss of climbing behavior upon expression of TAF15 selectively in motor neurons using the motor neuron-specific D42-GAL4 driver. (*C*) Up-regulation of two other RRM proteins does not cause neurodegeneration in *Drosophila*. As a specificity control for the neurodegenerative phenotype conferred by up-regulation of TDP-43 and TAF15 in *Drosophila*, we tested the effects of up-regulating the fly homologs of two other human RRM proteins in the eye. Up-regulation of tsu and CG17187, which are *Drosophila* homologs of human genes *RBM8A* and *DNAJC17*, respectively, did not cause neurodegeneration. (*D*) Expression of TAF15 in the nervous system reduces lifespan (blue, compared with normal in purple). Up-regulation of TAF15 variants G391E and R408C causes more rapid death (red and green, compared with WT TAF15 in blue). (*E*) Immunoblot showing TAF15 WT and mutant expression levels in transgenic flies. β-Tubulin levels were used as loading control.

(small arrows) and *E*], similar to those formed by TDP-43 and FUS (Fig. 4 *D* and *E*) (11, 12), which are reminiscent of pathological oligomers formed by α-synuclein and amyloid-β (32). Furthermore, TAF15 also assembled into linear polymers with a cross-sectional diameter of ∼15–20 nm (Fig. 4*D*, large arrows) that increased in length over time and would often become tangled into large masses by 60 min (Fig. 4*D*). In general, the morphology of TAF15 aggregates was more similar to FUS (12) than to TDP-43 (11), which over this time frame formed shorter polymers that would clump together to form large masses (Fig. 4*D*) (11). Importantly, a human RRM protein, DND1, which did not aggregate and was not toxic in yeast (Fig. 1 *C* and *D*), remained soluble and did not aggregate in this in vitro assay, providing evidence that in vitro aggregation is not a property shared by all RRM proteins. Thus, similar to TDP-43 and FUS, and concordant with the yeast data, TAF15 is an inherently aggregation-prone protein.

It is likely that sophisticated cellular proteostasis mechanisms, not recapitulated with this in vitro assay, mitigate rapid TAF15 aggregation in vivo. However, age-associated decline in these quality control measures, along with environmental triggers (for example, injury or exposure to toxins), might allow TAF15 to aggregate in disease. In vitro aggregation assays similar to these have been tremendously powerful tools in defining basic mechanisms underpinning TDP-43 and FUS aggregation (11, 12) as well as the aggregation events in Parkinson's disease and Alzheimer's disease (33–35).

**ALS-Specific Variants Accelerate TAF15 Aggregation in Vitro.** Next, we tested two of the ALS-linked mutant forms of TAF15 (G391E and R408C) as well as the variant that we also found in controls (R388H) in this aggregation assay (Fig. 4*F*). We noticed that if we omitted agitation during the in vitro aggregation reaction, we were able to slow down the aggregation process, allowing us to detect potential differences in aggregation between WT and mutant TAF15 proteins. R388H aggregated with similar kinetics to WT TAF15 (Fig. 4*G*). By contrast, the two ALS-linked mutants, G391E and R408C, aggregated with more rapid kinetics (Fig. 4*G*). This increased aggregation propensity might help explain why ALS-linked mutant TAF15 is more prone to accumulate in cytoplasmic inclusions in spinal cord neurons (Fig. 3) and supports the hypothesis that these are likely pathogenic variants.

**TAF15 Up-Regulation Causes Neurodegeneration in *Drosophila* and ALS-Linked Mutants Have More Severe Effect on Lifespan.** To extend these findings from in vitro to in vivo, and analyze the effects of TAF15 in the nervous system, we used *Drosophila*. We and others have previously shown that directing TDP-43 or FUS expression to the fly nervous system causes neurodegeneration (24–28, 30, 31). A series of transgenic lines were generated that expressed WT and mutant human TAF15. Directing expression of TAF15 to the eye of the fly caused degeneration of the structure (Fig. 5*A*) and led to progressive loss of motility when directed to motor neurons (Fig. 5*B*). As for the in vitro aggregation assay (Fig. 4), the effect was specific because two other unrelated *Drosophila* RRM proteins, the human counterparts of which did not aggregate and were not toxic in yeast, did not confer neurodegeneration when up-regulated in *Drosophila* (Fig. 5*C*). Thus, TAF15 possesses activity sufficient to confer neurodegeneration in the nervous system, in a manner similar to that of TDP-43 and FUS. TAF15 also induced a markedly shortened lifespan when expressed in the nervous system (Fig. 5*D*). Notably, expression of ALS-linked TAF15 variants G391E and R408C, inserted at the same genomic location to ensure similar protein expression (Fig. 5*E*), had a more severe effect, resulting in an even shorter lifespan than that of flies expressing WT TAF15 (Fig. 5*D*), supporting the notion that these missense variants in TAF15 are potentially deleterious.

**TAF15 Is Mislocalized in Motor Neurons of Patients with ALS.** We next performed immunohistochemistry on human postmortem spinal cord tissue to determine whether TAF15 was present in motor neurons and whether its localization was affected in ALS, as for TDP-43 and FUS. TAF15 was robustly expressed in spinal cord motor neurons and localized to the nucleus, as for TDP-43 and FUS (Fig. 6*A*). We next examined TAF15 localization in autopsy tissue from three sporadic ALS cases (without mutations in *TAF15*), all with confirmed TDP-43 pathology. Immunostaining of spinal cord sections from these patients with ALS revealed significant cytoplasmic staining in a punctate granular pattern (Fig. 6 *B–H*), which did not colocalize with TDP-43 inclusions. In contrast to TDP-43 (7), and similar to FUS (8, 9), we did not observe significant nuclear clearing of TAF15. The strong punctate granular localization for TAF15 was not seen in any of the control cases that we examined ($n = 3$). Because these sporadic ALS cases that we examined also contained prominent TDP-43 cytoplasmic inclusions, which did not colocalize with TAF15, future studies will be required to extend these initial findings and to determine whether TAF15 mislocalization in disease is an early initiator or rather a later consequence of motor neuron degeneration. Thus, TAF15 is expressed in a disease-relevant cell type and can be mislocalized to the cytoplasm in ALS, further supporting the notion that TAF15 can contribute to pathogenesis. Notably, a recent study has also observed TAF15 mislocalization in neurons and glial cells of patients with frontotemporal lobar degeneration (FTLD)-FUS (36).

## Discussion

In an effort to streamline the identification of new ALS genetic risk factors, we devised a simple yeast functional screen to define additional RNA-binding proteins with properties shared by the known ALS disease genes FUS and TDP-43. This screen resulted in the enrichment of 38 proteins that behave like FUS and TDP-43 in yeast (cytoplasmic inclusions and toxicity), 13 of which contain a predicted prion-like domain (Table 1). Indeed, the combination of yeast screen and prion prediction algorithm enabled us to significantly focus our list of candidate genes ∼10-fold. As evidence of the usefulness of this approach to define genes with a role in ALS, we identified patient-specific missense variants in one of these genes, *TAF15*, in five unrelated patients with ALS (three variants from our initial cohort of North

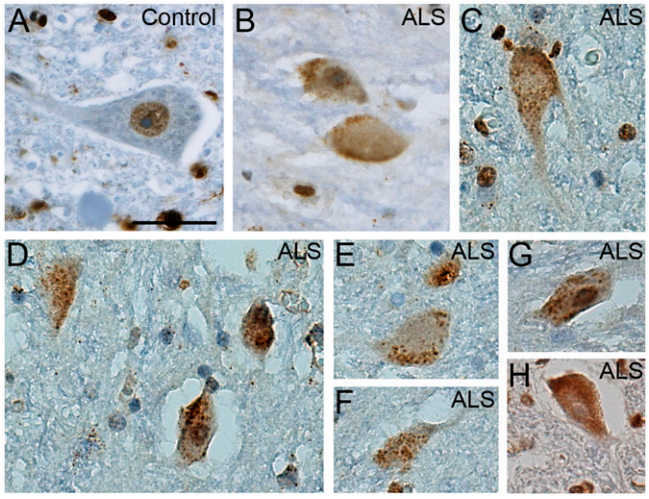


**Fig. 6.** TAF15 accumulates in the cytoplasm in some sporadic spinal cord neurons of patients with ALS. (*A*) Immunohistochemistry for TAF15 in a postmortem control spinal cord neuron indicates nuclear localization, whereas, in addition to nuclear localization, TAF15 accumulated in cytoplasmic puncta of spinal cord neurons of patients with ALS (*B–H*). (Scale bar, 25 μm.)

American Caucasian patients with ALS, a fourth variant from a cohort of Swedish patients with ALS, and a fifth variant from a cohort of Australian patients with ALS). Further, we provide in vitro and in vivo evidence that TAF15 has functional properties similar to those of TDP-43 and FUS: It is intrinsically aggregation prone and can confer neurodegeneration in *Drosophila* and the ALS-linked variants can increase aggregation in vitro, decrease lifespan in *Drosophila*, and alter protein subcellular localization in spinal cord neurons. Although familial segregation could not be assessed, the absence of the variants in a large number of healthy controls, the shared structural evidence with known ALS genes, and functional in vitro and in vivo data strongly support the notion that these variants in *TAF15* represent pathogenic disease mutations for ALS.

Future studies will be required to determine the relative contribution of TAF15 variants to ALS risk compared with known genetic risk factors such as TDP-43, FUS, SOD1, and others. Our initial analyses with TAF15 in patients with ALS and control populations, as well as recent studies by Ticozzi and colleagues with TAF15 and the related gene EWSR1 (37), suggest that if indeed TAF15 mutations contribute to ALS, they will likely be rarer than FUS and TDP-43 mutations. However, as for all complicated human diseases there will very likely be common genetic contributors as well as rare genetic risk factors. For ALS, we propose that there may be a delicate balance in RNA processing within motor neurons such that slight perturbations from any one of several different aggregation-prone RNA-binding proteins could lead to neurodegeneration. An interesting additional concept that emerges from our findings is that perhaps variants in multiple RNA-binding proteins could synergize with each other to contribute to ALS. There are likely to be some variants that are extremely damaging and thus fully penetrant and aggressive on their own. Case in point: P525L and R495X mutations in FUS lead to relatively severe ALS clinical phenotypes and very early age of disease onset (18, 23), whereas other FUS mutations are less severe (e.g., R521G) (18). This result seems to be due to the effect of the mutations on FUS nuclear localization, with variants having the strongest effect on nuclear localization resulting in the earliest age of onset of ALS (18). Those aggressive FUS variants might sit at one end of a spectrum, with weaker variants at the other. Perhaps then the accumulation of multiple weaker variants in two, three, or more different aggregation-prone RNA-binding proteins (e.g., the top candidates listed in Table 1) might be required to tip the balance in RNA metabolism toward ALS. Future studies will be required to test this hypothesis and to better resolve the complexities of the ALS genetic landscape.

These findings predict that additional aggregation-prone RRM or other RNA-binding proteins, like TAF15, FUS, and TDP-43, could contribute to ALS. Notably, the prion-like domain algorithm ranked FUS and TAF15 first and second of 213 RRM proteins, respectively, and ranked TDP-43 10th. We suggest that genes ranked third through ninth should now be given top priority for genetic analysis in populations of patients with ALS, especially EWSR1, which ranked third and is a close relative of both FUS and TAF15 (38). In addition to ALS, these candidates should also be examined in related clinico-pathological disorders including FTLD and inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia (IBMPFD). For example, mutations in ALS genes *TARDBP* and *FUS* have been identified in patients with FTLD and mutations in an IBMPFD gene, *VCP*, have been identified in patients with ALS (39–41).

Next-generation sequencing and exon capture approaches will eventually become routine in personalized medicine (42–44) and promise to identify all genetic contributors to ALS. Meanwhile, the list of ALS candidate genes that we provide here (Table 1 and Dataset S1), generated by the combination of the yeast functional screen and prion-like domain prediction, will be a powerful resource to jumpstart efforts to identify new genetic risk factors for ALS and spur innovative new diagnostic and therapeutic approaches.

## Materials and Methods

**Plasmids, Yeast Strains, and Media.** Strain and plasmid construction is detailed in *SI Materials and Methods*.

**Human RNA-Binding Protein Plasmid Library.** We constructed a library of yeast expression plasmids containing 133 unique human RRM-containing ORFs, as detailed in *SI Materials and Methods*.

**Yeast Transformation and Spotting Assays.** Yeast transformation and spotting assays were performed as described in ref. 10.

**Microscopy.** Fluorescence microscopy was performed as described in *SI Materials and Methods*.

**Prion-Prediction Algorithm.** Putative prion domains were predicted using a hidden Markov model as described in *SI Materials and Methods*.

**Sequencing *TAF15* in Patients with ALS and Controls.** ALS patient and control population demographics and *TAF15* mutational analysis procedures are described in *SI Materials and Methods*.

**TAF15 Protein Purification and in Vitro Aggregation Assays.** TDP-43 and FUS were purified as described (11, 12). WT and mutant TAF15 proteins were expressed and purified from *Escherichia coli* as GST-tagged proteins and in vitro aggregation assays were performed as described in *SI Materials and Methods*.

***Drosophila* Experiments.** Transgenic flies expressing human TAF15 were generated by standard techniques using the pUAST vector and analyzed as described in *SI Materials and Methods*.

**Rat Primary Neuron Transfection and Immunofluorescence.** Primary neurons were isolated from rat embryos, cultured, and transfected as described in *SI Materials and Methods*.

**Statistical Analysis.** Two-tailed Fisher's exact tests were used to evaluate genetic association between *TAF15* sequence variants and ALS. We tested all TAF15 nonsynonymous missense variants detected in cases and controls in an assay that potentially discriminates functionally deleterious variants from benign variants (e.g., the subcellular localization assay reported in Fig. 3), then classified the variants as deleterious or benign on the basis of their properties in this assay, and then performed a statistical comparison of the frequency of functionally deleterious variants in cases vs. controls.

**Immunohistochemistry.** Immunohistochemistry to detect TAF15 localization in postmortem spinal cord tissue was performed as described in *SI Materials and Methods*.

**ACKNOWLEDGMENTS.** We thank the patients and their families for their dedication and for their invaluable contributions to this research. We thank the Packard Center for ALS Research at Johns Hopkins for their generosity and collaborative spirit. We thank contributors, including the Alzheimer's Disease Centers who collected samples used in this study. We thank C. Cecere for assistance in sample collection. We thank Rose Li, Phoebe Leboy, and Marisa Bartolomei for helpful suggestions on the manuscript. This work was supported by National Institutes of Health Director's New Innovator Awards 1DP2OD004417 (to A.D.G.) and 1DP2OD002177-01 (to J.S.); National Institutes of Health Grants 1R01NS065317 (to A.D.G.), 5R21NS067354-02 (to J.S.), AG17586 (to V.V.D., J.Q.T., and R.G.), AG10124 (to V.V.D. and J.Q.T.), P01-AG-09215 (to N.M.B.), NS056070 and NS072561 (to Z.M.), T32-AG00255 [to F.I. and V.M.-Y.L. (program director)], R01 AG26251-03A1 (to R.R.), R01 NS065782 (to R.R.), and P50 AG16574 (to R.R.); the University of Pennsylvania Institute on Aging and Alzheimer's Disease Core Center Pilot Grant Program (AG10124) (to V.V.D.); the ALS Association (R.R); a grant from the Packard Center for ALS Research at Johns Hopkins (to A.D.G. and J.S.); and an Ellison Medical Foundation New Scholar in Aging Award (to J.S.). A.D.G. is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts, and a Rita Allen Scholar, supported by the Rita Allen Foundation. J.Q.T. is the William Maul Measey–Truman G. Schnabel, Jr., Professor of Geriatric Medicine and Gerontology. N.M.B. is an Investigator of the Howard Hughes Medical Institute. Samples from the National Cell

Repository for Alzheimer's Disease, which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging, were used in this study. In Australia, the work was supported by the National Health and Medical Research Council of Australia (1004670 and 511941) and a Peter Stearne grant from the Motor Neurone Disease Research Institute of Australia. This research was conducted while J.C. was an Ellison Medical Foundation/AFAR Post-doctoral Fellow.

1. Cleveland DW, Rothstein JD (2001) From Charcot to Lou Gehrig: Deciphering selective motor neuron death in ALS. *Nat Rev Neurosci* 2:806–819.
2. Van Damme P, Robberecht W (2009) Recent advances in motor neuron disease. *Curr Opin Neurol* 22:486–492.
3. Lagier-Tourenne C, Cleveland DW (2009) Rethinking ALS: The FUS about TDP-43. *Cell* 136:1001–1004.
4. Cushman M, Johnson BS, King OD, Gitler AD, Shorter J (2010) Prion-like disorders: Blurring the divide between transmissibility and infectivity. *J Cell Sci* 123:1191–1201.
5. Udan M, Baloh RH (2011) Implications of the prion-related Q/N domains in TDP-43 and FUS. *Prion* 5:1–5.
6. Gitler AD, Shorter J (2011) RNA-binding proteins with prion-like domains in ALS and FTLD-U. *Prion* 5:179–187.
7. Neumann M, et al. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314:130–133.
8. Kwiatkowski TJ, Jr., et al. (2009) Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* 323:1205–1208.
9. Vance C, et al. (2009) Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* 323:1208–1211.
10. Johnson BS, McCaffery JM, Lindquist S, Gitler AD (2008) A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proc Natl Acad Sci USA* 105:6439–6444.
11. Johnson BS, et al. (2009) TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J Biol Chem* 284:20329–20339.
12. Sun Z, et al. (2011) Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol* 9:e1000614.
13. Ju S, et al. (2011) A yeast model of FUS/TLS-dependent cytotoxicity. *PLoS Biol* 9: e1001052.
14. Alberti S, Halfmann R, King O, Kapila A, Lindquist S (2009) A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137: 146–158.
15. Aguzzi A, Rajendran L (2009) The transcellular spread of cytosolic amyloids, prions, and prionoids. *Neuron* 64:783–790.
16. Furukawa Y, Kaneko K, Watanabe S, Yamanaka K, Nukina N (2011) A seeding reaction recapitulates intracellular formation of Sarkosyl-insoluble transactivation response element (TAR) DNA-binding protein-43 inclusions. *J Biol Chem* 286:18664–18672.
17. Araya N, et al. (2003) Cooperative interaction of EWS with CREB-binding protein selectively activates hepatocyte nuclear factor 4-mediated transcription. *J Biol Chem* 278:5427–5432.
18. Dormann D, et al. (2010) ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *EMBO J* 29:2841–2857.
19. Lee BJ, et al. (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* 126:543–558.
20. Lagier-Tourenne C, Polymenidou M, Cleveland DW (2010) TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration. *Hum Mol Genet* 19(R1): R46–R64.
21. Barmada SJ, et al. (2010) Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J Neurosci* 30:639–649.
22. Kabashi E, et al. (2010) Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Hum Mol Genet* 19:671–683.
23. Bosco DA, et al. (2010) Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. *Hum Mol Genet* 19:4160–4175.
24. Elden AC, et al. (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 466:1069–1075.
25. Hanson KA, Kim SH, Wassarman DA, Tibbetts RS (2010) Ubiquilin modifies TDP-43 toxicity in a Drosophila model of amyotrophic lateral sclerosis (ALS). *J Biol Chem* 285: 11068–11072.
26. Li Y, et al. (2010) A Drosophila model for TDP-43 proteinopathy. *Proc Natl Acad Sci USA* 107:3169–3174.
27. Lu Y, Ferris J, Gao FB (2009) Frontotemporal dementia and amyotrophic lateral sclerosis-associated disease protein TDP-43 promotes dendritic branching. *Mol Brain* 2:30.
28. Ritson GP, et al. (2010) TDP-43 mediates degeneration in a novel Drosophila model of disease caused by mutations in VCP/p97. *J Neurosci* 30:7729–7739.
29. Estes PS, et al. (2011) Wild-type and A315T mutant TDP-43 exert differential neurotoxicity in a Drosophila model of ALS. *Hum Mol Genet* 20:2308–2321.
30. Lanson NA, Jr., et al. (2011) A Drosophila model of FUS-related neurodegeneration reveals genetic interaction between FUS and TDP-43. *Hum Mol Genet* 20:2510–2523.
31. Wang JW, Brent JR, Tomlinson A, Shneider NA, McCabe BD (2011) The ALS-associated proteins FUS and TDP-43 function together to affect Drosophila locomotion and life span. *J Clin Invest* 121:4118–4126.
32. Lashuel HA, Hartley D, Petre BM, Walz T, Lansbury PT, Jr. (2002) Neurodegenerative disease: Amyloid pores from pathogenic mutations. *Nature* 418:291.
33. Conway KA, Harper JD, Lansbury PT (1998) Accelerated in vitro fibril formation by a mutant alpha-synuclein linked to early-onset Parkinson disease. *Nat Med* 4: 1318–1320.
34. Conway KA, et al. (2000) Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *Proc Natl Acad Sci USA* 97: 571–576.
35. Lashuel HA, et al. (2003) Mixtures of wild-type and a pathogenic (E22G) form of Abeta40 in vitro accumulate protofibrils, including amyloid pores. *J Mol Biol* 332: 795–808.
36. Neumann M, et al. (2011) FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain* 134:2595–2609.
37. Ticozzi N, et al. (2011) Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. *Am J Med Genet B Neuropsychiatr Genet* 156B:285–290.
38. Tan AY, Manley JL (2009) The TET family of proteins: Functions and roles in disease. *J Mol Cell Biol* 1:82–92.
39. Broustal O, et al.; French clinical and genetic research network on FTD/FTD-MND (2010) FUS mutations in frontotemporal lobar degeneration with amyotrophic lateral sclerosis. *J Alzheimers Dis* 22:765–769.
40. Gitcho MA, et al. (2009) TARDBP 3′-UTR variant in autopsy-confirmed frontotemporal lobar degeneration with TDP-43 proteinopathy. *Acta Neuropathol* 118:633–645.
41. Johnson JO, et al.; ITALSGEN Consortium (2010) Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* 68:857–864.
42. Biesecker LG (2010) Exome sequencing makes medical genomics a reality. *Nat Genet* 42:13–14.
43. Ng SB, et al. (2010) Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 42:30–35.
44. Ng SB, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461:272–276.

# Supporting Information

## Couthouis et al. 10.1073/pnas.1109434108

### SI Materials and Methods

**Plasmids, Yeast Strains, and Media.** The yeast strain used in the human RRM screen and follow-up analyses was BY4741 [genotype *Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*]. Strains were manipulated and media prepared using standard techniques. TAF15 expression constructs were generated by Gateway cloning (Invitrogen), starting with entry clones in pDONR221 and shuttled from entry clones into a modified pGW vector (for motor neuron transfection experiments), created by incorporating the Gateway B cassette into the SmaI site of the pGW vector using the Gateway conversion kit (Invitrogen). ALS and control patient mutations in TAF15 were introduced by site-directed mutagenesis using the QuikChange Site-Directed Mutagenesis kit (Stratagene).

**Human RNA-Binding Protein Plasmid Library.** We constructed a library of yeast expression plasmids containing 133 unique human RRM-containing ORFs. The ORFs were obtained from the human ORFeome collection (Open Biosystems) as Gateway entry clones in plasmid pDONR223. We selected 133 unique clones contained within the library predicted to encode RRM-domain proteins (Protein Families database ID PF00076). ORFs from the entry clones were shuttled into the 2-μm galactose-inducible yeast expression plasmid pAG426Gal-ccdB-EYFP by Gateway LR cloning reaction (1) to generate C-terminally tagged RRM-protein–YFP fusions. Restriction digest and DNA sequencing were used to confirm the integrity of each expression construct.

**Yeast Transformation and Spotting Assays.** The PEG/lithium acetate method was used to transform yeast with each plasmid DNA from the RRM ORF library (2). For spotting assays, yeast cells were grown overnight at 30 °C in liquid media containing raffinose (SRaf/−Ura) until log or midlog phase. Cultures were then normalized for $OD_{600}$, serially diluted, and spotted onto synthetic solid media containing glucose or galactose lacking uracil and were grown at 30 °C for 2–3 d.

**Microscopy.** For fluorescence microscopy experiments, single-colony isolates of the yeast strains were grown to midlog phase in SRaf/−Ura media at 30 °C. Cultures were spun down and resuspended in the same volume of SGal/−Ura to induce expression of the TDP-43 constructs. Cultures were induced with galactose for 4–6 h and processed for microscopy. Images were obtained using an Olympus IX70 inverted microscope and a Photometrics CoolSnap HQ 12-bit CCD camera.

**Prion-Prediction Algorithm.** Proteins were parsed into prion-like and non-prion–like regions using a hidden Markov model developed to identify regions that have the unusual amino acid composition characteristic of yeast prions (3, 4). Prion-like regions of length ≥60 were given a prion-domain score, defined as the maximum log-likelihood for the prion-like state vs. the non-prion–like state over any 60 consecutive amino acids within the regions (3). Among the 21,873 human genes analyzed (Ensembl GrCh37.59), 246 had prion-like regions of length ≥60 and were ranked by prion-domain score. For genes with multiple transcripts, the longest one was used, with the one with lowest Ensembl Transcript ID used in case of ties.

**Sequencing *TAF15* in Patients with ALS and Controls.** Genomic DNA from 735 non-Latino Caucasian individuals with ALS was obtained from a combination of Coriell Institute for Medical Research ($n$ = 406), the University of Pennsylvania Center for Neurodegenerative Disease Research ($n$ = 204), and the Mayo Clinic ($n$ = 125). Coriell ALS samples were distributed in 96-well plates NDPT025, NDPT026, NDPT030, NDPT100, NDPT103, and NDPT106. The University of Pennsylvania genomic DNA samples were from patients verified to meet El Escorial criteria for definite or probable ALS by a neurologist or with neuropathologic findings consistent with ALS. All subjects were collected with University of Pennsylvania Institutional Review Board approval. The racial background of the University of Pennsylvania subjects was 90% non-Latino Caucasian, 5% Black, and 5% mixed or other. The University of Pennsylvania subjects were 57% male and had an average age of onset of 57 y (8–89) and an average duration of disease of 4 y (range 1–46). A family history of ALS (FALS) was present in 29 of 204 patients (11.6%) for which family history was available. Mutations in *SOD1* and *FUS/TLS* were excluded in all of the familial ALS cases and *TARDBP* mutations excluded in all University of Pennsylvania cases. All cases with potentially pathogenic variants in *TAF15* were also sequenced for *TARDBP*, *FUS*, and *SOD1*. The Mayo Clinic ALS samples consisted of 105 unrelated non-Latino Caucasian patients with ALS (56 males, 49 females) from a consecutive clinical case series seen at Mayo Clinic Florida by the ALS Center in the period 2008–2010 and 20 pathologically confirmed patients with ALS selected from the Mayo Clinic Florida Brain Bank (7 males, 13 females). All patients agreed to be in the study and biological samples were obtained after informed consent. Mutations in *SOD1*, *FUS/TLS*, or *TARDBP* were previously excluded in all patients included in this patient series (5). The average age of onset in the clinical patient population was 57.2 ± 10.6 y (range 17–78 y), and the average age at death in pathologically confirmed samples was 68.9 ± 12.1 y (range 46–83 y). Sixteen percent of patients showed a positive family history of ALS defined as having at least one affected relative within three generations.

We also sequenced the same TAF15 C-terminal exons in 351 Swedish patients with ALS and identified one missense variant, which was damaging in our functional assay (Fig. 3) but not included in our statistical analysis because we did not sequence Swedish controls for direct comparison.

We sequenced all 16 exons of *TAF15* in 176 Australian patients with ALS (135 familial and 41 sporadic ALS cases) and identified one missense variant, G449E, which was not present in 72 sequenced Australian controls or in our cohort of 1,328 sequenced controls. We did not test this variant in our functional assay (Fig. 3).

A total of 713 neurologically normal control samples from Coriell were distributed in 96-well plates NDPT084, NDPT090, NDPT093, NDPT094, NDPT095, NDPT096, NDPT098, and NDPT099. An additional 90 neurologically normal control samples were obtained from the Children's Hospital of Philadelphia. A total of 179 DNA samples from cognitively normal individuals >60 y of age were obtained from the National Cell Repository for Alzheimer's Disease (Indianapolis, IN). DNA samples of an additional 346 healthy control individuals (range 61–90 y) were also ascertained at the Department of Neurology at Mayo Clinic Florida. The racial background of all controls was non-Latino Caucasian.

We sequenced exons 13–16 of *TAF15*, which encode the C-terminal domain of TAF15. *TAF15* was sequenced in 735 ALS cases ($n$ = 406 Coriell, $n$ = 204 University of Pennsylvania, $n$ = 125 Mayo Clinic) and 1,328 controls ($n$ = 713 Coriell, $n$ = 269

University of Pennsylvania, $n = 346$ Mayo Clinic). Bidirectional sequencing was performed by separately amplifying *TAF15* exons 13–16 from samples using PCR. PCR primers and cycling conditions used for amplification and sequencing are available upon request. Amplicons were purified, processed, and sequenced using Big-Dye Terminator v3.1 sequencing on an ABI3730 Genetic Analyzer (Applied Biosystems). All variants identified were confirmed by repeat sequencing. Sequence analysis was performed using Sequencher DNA Software.

**TAF15 Protein Purification.** TDP-43 and FUS were purified as described (6, 7). WT and mutant TAF15 proteins were expressed and purified from *Escherichia coli* as GST-tagged proteins. TAF15 was cloned into GV13 to yield GST-TEV-TAF15 and overexpressed in *E. coli* BL21 Star (Invitrogen). Protein was purified over a glutathione-Sepharose column (GE) according to manufacturer's instructions. GST-TAF15 was eluted from the glutathione Sepharose with 50 mM Tris·HCl, pH 7.4, 100 mM potassium acetate, 200 mM trehalose, 0.5 mM EDTA, and 20 mM glutathione. After purification, proteins were concentrated to ≥10 μM using Amicon Ultra-4 centrifugal filter units (10-kDa molecular mass cutoff; Millipore). Protein was then filtered through a 0.22-μm filter to remove any aggregated material. After filtration, the protein concentration was determined by Bradford assay (Bio-Rad) and the proteins were used immediately for aggregation reactions.

**TAF15 in Vitro Aggregation Assays.** Filtered, purified GST-TAF15 proteins were used immediately for aggregation assays. Aggregation was initiated by the addition of TEV protease (Invitrogen) to TAF15 (3 μM) in assembly buffer (AB): 50 mM Tris·HCl, pH 7.4, 100 mM potassium acetate, 200 mM trehalose, 0.5 mM EDTA, and 20 mM glutathione. Aggregation reactions were incubated at 25 °C for 0–90 min with agitation at 700 rpm in an Eppendorf Thermomixer. No aggregation occurred unless TEV protease was added to separate GST from TAF15. Turbidity was used to assess aggregation by measuring absorbance at 395 nm. For sedimentation analysis, reactions were centrifuged at $16{,}100 \times g$ for 20 min at 25 °C. Supernatant and pellet fractions were then resolved by SDS/PAGE and stained with Coomassie Brilliant Blue, and the amount in either fraction was determined by densitometry in comparison with known quantities of TAF15. For electron microscopy (EM) of in vitro aggregation reactions, protein samples (20 μL of a 3-μM solution) were adsorbed onto a glow-discharged 300-mesh Formvar/carbon-coated copper grid (Electron Microscopy Sciences) and stained with 2% (wt/vol) aqueous uranyl acetate. Excess liquid was removed, and grids were allowed to air dry. Samples were viewed using a JEOL 1010 transmission electron microscope.

***Drosophila* Experiments.** Transgenic flies expressing human TAF15 were generated by standard techniques using the pUAST vector. TDP-43 transgenic flies are described in ref. 8. To direct transgene expression to the eye, gmr-GAL4 driver was used. To direct expression to motor neurons, D42-GAL4 driver was used. Locomotor activity was assessed using a climbing assay as described in ref. 8. For lifespan analysis, flies were raised at 25 °C, and lifespan was assessed at 29 °C. For specificity controls, we up-regulated other RNA-binding proteins, which did not aggregate and were not toxic in our yeast screen (Dataset S1). We used transgenic EP lines to up-regulate two endogenous fly genes, tsu and CG17187, which are the homologs of human genes *RBM8A* and *DNAJC17*, respectively.

**Rat Primary Neuron Transfection and Immunofluorescence.** Embryonic Sprague–Dawley rat spinal cord neurons were grown on previously established cortical astrocyte monolayers as described previously (9). Briefly, the cerebral cortex of postnatal day 1–3 (P1–P3) rat pups was dissociated, plated on laminin/polylysine-coated glass coverslips, and maintained in minimal essential media (10% horse serum, 10% FBS) and 10 μM AraC was added for 1 d to arrest additional proliferation when confluent. Under these conditions, no cortical neurons are present in these cultures. Spinal cord from embryonic day 15 (E15) rats was dissociated and plated on the astrocytes at a density of one spinal cord per 4 mL of media conditioned previously over astrocytes with 10 ng/mL of the following trophic factors added: ciliary neurotrophic factor, cardiotrophin-1 (CT1), and glial-derived neurotrophic factor (all from Alomone Labs).

Primary neuron cultures were transfected with myc-tagged TAF15 constructs after 5 d in vitro, using Lipofectamine LTX with PLUS reagent (Invitrogen) according to the manufacturer's protocol in media lacking antibiotics. Media were replaced 12 h following transfection with media containing antibiotics. Cells were harvested for immunofluorescence 168 h after transfection. Briefly, cultures were washed in PBS, fixed in 4% paraformaldehyde for 15 min, and then washed four times in 1× PBS. Cells were blocked for 1 h in blocking solution (2% FBS, 0.02% Triton X-100, 1× PBS) and then incubated for 1 h in primary antibody at room temperature. Cells were then washed three times in 1× PBS and then incubated with secondary antibody for 1 h at room temperature. Cells were then washed with blocking solution and mounted in Vectashield mounting media with DAPI (Vector). Antibodies used were α-myc rabbit antibody (Sigma), 1:1,000; α-Doublecortin goat antibody (Santa Cruz), 1:500; Cy-3–conjugated α-mouse IgG (Jackson Immunoresearch), 1:250; Cy-3–conjugated α-rabbit IgG (Jackson Immunoresearch), 1:250; and Cy-2–conjugated α-goat IgG (Jackson Immunoresearch), 1:250. Cells were visualized by fluorescence microscopy. An investigator blinded to the identity of the transfected construct quantified localization of transfected WT or mutant TAF15. The number of neurons containing accumulations of TAF15 in processes was divided by the total number of transfected neurons (TAF15-myc positive) counted to yield the percentage of neurons with TAF15 accumulations in processes. More than 50 neurons were analyzed for each condition, and transfection and quantification was repeated four or more independent times per construct. Neurons were identified by morphology as well as Doublecortin staining.

**Immunohistochemistry.** Formalin-fixed, paraffin-embedded human spinal cord sections were deparaffinized before pretreatment, using heat antigen retrieval with Bull's Eye Decloaker (BioCare Medical). Endogenous peroxidase was then blocked with 3% hydrogen peroxide in PBS for 10 min. After washing with 0.1% PBST, blocking was performed with 10% goat serum, 0.5% PBST for 30–60 min at 25 °C. Sections were incubated with rabbit anti-TAF15 (1:250; Bethyl Laboratories) in 0.1% PBST overnight at 4 °C. After washing with 0.1% PBST, sections were incubated with biotinylated goat anti-mouse or rabbit IgG (1:200; Vector Laboratories) for 1 h at 25 °C. After washing with 0.1% PBST, sections were then incubated with Vectastain ABC (Vector Laboratories) for 45 min. After washing with 0.1% PBST followed by 0.1 M Tris (pH 7.5) and 0.3 M NaCl, peroxidase activity was then detected with DAB (Sigma). Detailed immunohistochemistry protocols are available at http://www.med.upenn.edu/mcrc/histology_core/.


1. Alberti S, Gitler AD, Lindquist S (2007) A suite of Gateway cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. *Yeast* 24:913–919.
2. Ito H, Fukuda Y, Murata K, Kimura A (1983) Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* 153:163–168.
3. Alberti S, Halfmann R, King O, Kapila A, Lindquist S (2009) A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137:146–158.
4. Cushman M, Johnson BS, King OD, Gitler AD, Shorter J (2010) Prion-like disorders: Blurring the divide between transmissibility and infectivity. *J Cell Sci* 123:1191–1201.

5. Dejesus-Hernandez M, et al. (2010) De novo truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis. *Hum Mutat* 31):E1377–E1389.
6. Johnson BS, et al. (2009) TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J Biol Chem* 284:20329–20339.
7. Sun Z, et al. (2011) Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol* 9:e1000614.
8. Elden AC, et al. (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 466:1069–1075.
9. Hu P, Kalb RG (2003) BDNF heightens the sensitivity of motor neurons to excitotoxic insults through activation of TrkB. *J Neurochem* 84:1421–1430.

## Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)