

Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS

Hong Joo Kim^{1*}, Nam Chul Kim^{1*}, Yong-Dong Wang^{2*}, Emily A. Scarborough^{3*}, Jennifer Moore^{1*}, Zamia Diaz^{3*}, Kyle S. MacLea⁴, Brian Freibaum¹, Songqing Li¹, Amandine Molliex¹, Anderson P. Kanagaraj¹, Robert Carter⁵, Kevin B. Boylan⁶, Aleksandra M. Wojtas⁶, Rosa Rademakers⁶, Jack L. Pinkus⁷, Steven A. Greenberg⁷, John Q. Trojanowski⁸, Bryan J. Traynor⁹, Bradley N. Smith⁸, Simon Topp¹⁰, Athina-Soragia Gkazi¹⁰, Jack Miller¹⁰, Christopher E. Shaw¹⁰, Michael Kottlors¹¹, Janbernd Kirschner¹¹, Alan Pestronk¹², Yun R. Li¹³, Alice Flynn Ford³, Aaron D. Gitler¹⁴, Michael Benatar¹⁵, Oliver D. King¹⁶, Virginia E. Kimonis¹⁷, Eric D. Ross⁴, Conrad C. Weihl¹², James Shorter³ & J. Paul Taylor¹

Algorithms designed to identify canonical yeast prions predict that around 250 human proteins, including several RNA-binding proteins associated with neurodegenerative disease, harbour a distinctive prion-like domain (PrLD) enriched in uncharged polar amino acids and glycine. PrLDs in RNA-binding proteins are essential for the assembly of ribonucleoprotein granules. However, the interplay between human PrLD function and disease is not understood. Here we define pathogenic mutations in PrLDs of heterogeneous nuclear ribonucleoproteins (hnRNPs) A2B1 and A1 in families with inherited degeneration affecting muscle, brain, motor neuron and bone, and in one case of familial amyotrophic lateral sclerosis. Wild-type hnRNPA2 (the most abundant isoform of hnRNPA2B1) and hnRNPA1 show an intrinsic tendency to assemble into self-seeding fibrils, which is exacerbated by the disease mutations. Indeed, the pathogenic mutations strengthen a 'steric zipper' motif in the PrLD, which accelerates the formation of self-seeding fibrils that cross-seed polymerization of wild-type hnRNP. Notably, the disease mutations promote excess incorporation of hnRNPA2 and hnRNPA1 into stress granules and drive the formation of cytoplasmic inclusions in animal models that recapitulate the human pathology. Thus, dysregulated polymerization caused by a potent mutant steric zipper motif in a PrLD can initiate degenerative disease. Related proteins with PrLDs should therefore be considered candidates for initiating and perhaps propagating proteinopathies of muscle, brain, motor neuron and bone.

Elucidating the genetic basis of rare, inherited diseases can provide valuable insights to the molecular pathogenesis of common diseases. Inclusion body myopathy (IBM) associated with Paget's disease of the bone (PDB), fronto-temporal dementia (FTD) and amyotrophic lateral sclerosis (ALS)—sometimes called IBMPFD/ALS—is a rare disorder characterized by progressive degeneration of muscle, brain, motor neurons and bone accompanied by prominent TDP-43 (also called TARDBP) pathology¹. Patients with this rare, inherited syndrome experience features of IBM, FTD, ALS or PDB indistinguishable from familial and sporadic cases of these disorders, and the disease may manifest in several tissues in the same patient^{1,2}. Recently the name multisystem proteinopathy (MSP) has been adopted to reflect the expanding phenotype and prominent proteinaceous pathology that characterizes this syndrome. Some, but not all, cases of MSP are caused by mutations in the *VCP* gene³, which encodes the AAA+ ATPase VCP, a ubiquitin-dependent segregase.

The discovery that *VCP* mutations cause MSP led to the subsequent discovery of pathogenic *VCP* mutations in more common diseases such as sporadic or familial forms of ALS², FTD⁴, IBM⁵ and PDB⁶. These rare MSP families represent a unique opportunity to identify fundamental molecular defects shared among age-related diseases; thus it is highly desirable to identify additional genetic mutations responsible for this syndrome.

A pathogenic mutation in hnRNPA2B1 in MSP

We identified a family (family 1) with dominantly inherited degeneration affecting muscle, bone, brain and motor neurons that was clinically indistinguishable from previous families we have seen with *VCP*-related MSP (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1). Sequencing the entire *VCP* gene (including introns and exons) in affected patients revealed no synonymous or nonsynonymous variants. Genetic analysis of this family by exome sequencing and linkage analysis

¹Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38120, USA. ²Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, Tennessee 38120, USA. ³Department of Biochemistry and Biophysics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523, USA. ⁵Department of Computational Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38120, USA. ⁶Department of Neuroscience, Mayo Clinic, Jacksonville, Florida 32224, USA. ⁷Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Department of Pathology and Laboratory Medicine, Institute on Aging and Center for Neurodegenerative Disease Research, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁹Neuromuscular Diseases Research Group, Laboratory of Neurogenetics, Porter Neuroscience Building, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁰King's College London Centre for Neurodegeneration Research, Department of Clinical Neuroscience, Institute of Psychiatry, London SE5 8AF, UK. ¹¹Division of Neuropediatrics and Muscle Disorders, University Children's Hospital Freiburg, 79106 Freiburg, Germany. ¹²Department of Neurology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. ¹³Medical Scientist Training Program, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹⁴Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. ¹⁵Neurology Department, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. ¹⁶Boston Biomedical Research Institute, Watertown, Massachusetts 02472, USA. ¹⁷Department of Pediatrics, Division of Genetics and Metabolism, University of California-Irvine, 2501 Hewitt Hall, Irvine, California 92696, USA.

*These authors contributed equally to this work.

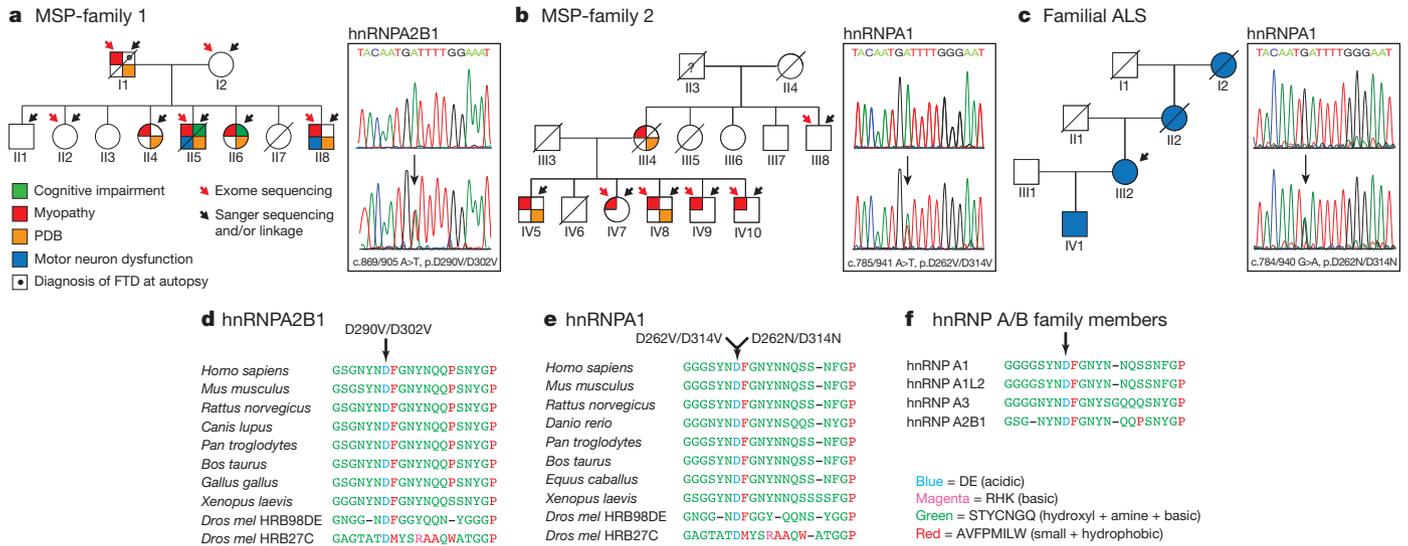


Figure 1 | Identification of previously known disease mutations in MSP and ALS. **a**, Family 1 pedigree indicating individuals affected by dementia, myopathy, PDB and ALS. The causative mutation was p.D290V/D302V in hnRNA2B1. Roman numerals denote generation and Arabic numbers denote family member within a generation. **b**, Family 2 pedigree indicating individuals affected by myopathy and PDB. The causative mutation was p.D262V/D314V in hnRNA1. **c**, The pedigree of a family with ALS. The causative mutation was p.D262N/D314N in hnRNA1. **d, e**, Sequence alignment of hnRNA2B1 (**d**) and hnRNA1 (**e**) orthologues showing evolutionary conservation of the

in parallel (Supplementary Fig. 1) identified a single, previously unknown variant (c.869/905A>T, p.D290V/D302V) that co-segregates with disease and alters the gene encoding hnRNA2B1, a ubiquitously expressed RNA-binding protein (Fig. 1a). hnRNA2B1 is expressed as two alternatively spliced isoforms: A2 and B1. The shorter hnRNA2, which lacks 12 amino acids in the amino-terminal region, is the main isoform, accounting for ~90% of the protein in most tissues. The identified mutation substitutes a valine residue in the place of an evolutionarily conserved aspartate residue (Fig. 1d) that is centred in a motif conserved in multiple human paralogs of the hnRNP A/B family (Fig. 1f and Supplementary Fig. 3).

Pathogenic mutations in hnRNA1 in MSP and ALS

Additional validation of the pathogenicity of the hnRNA2B1 mutation came from the analysis of family 2. The clinical features of this family with VCP-negative MSP have been previously described⁷. To identify the culprit mutation we took the same strategy as for family 1 (Supplementary Fig. 2 and Supplementary Table 2). This analysis identified five previously unknown single nucleotide variants and one indel that co-segregated with the disease. Of these, a mutation in hnRNA1 (c.785/941A>T, p.D262V/D314V) stood out because it was the only variant associated with a conserved residue, it was predicted to be deleterious, and because hnRNA1 is highly expressed in affected tissues (Fig. 1b and Supplementary Fig. 2). Moreover, this mutation in hnRNA1 was identical to that found in hnRNA2B1 in family 1, creating a substitution of a valine residue in the place of a highly conserved aspartate residue that is centred in a motif conserved in multiple human paralogs of the hnRNP A/B family (Fig. 1e, f and Supplementary Fig. 3). Subsequent to the identification of these MSP-causing mutations we screened the exomes of 212 familial ALS cases for sequence variants affecting hnRNA2B1 or hnRNA1. In one dominantly inherited case in which known ALS genes were formally excluded, we identified a mutation (c.784/940G>A; p.D262N/D314N) affecting the identical, conserved aspartate residue in hnRNA1 (Fig. 1c, e, f).

The identification of disease mutations in hnRNA2B1 and hnRNA1 was intriguing for three reasons. First, hnRNA2B1 and hnRNA1 interact directly with TDP-43 and function cooperatively

mutated aspartate and surrounding residues. Gaps (black lines) are inserted between residues so that identical or similar amino acids are aligned in successive columns. **f**, Sequence alignment of four human paralogs of the hnRNP A/B family in which the disease-affected residue and surrounding residues are highly conserved. Acidic residues (Asp, Glu) are in blue, basic residues (Arg, His, Lys) are in magenta, uncharged polar amino acids (Ser, Thr, Tyr, Asn, Gln) and Gly and Cys are in green, and nonpolar amino acids aside from Gly and Cys (Ala, Val, Phe, Pro, Met, Ile, Leu, Trp) are in red.

to regulate RNA metabolism⁸. Second, an unbiased genetic screen previously identified fly homologues of TDP-43, hnRNA2B1 and hnRNA1 as suppressors of VCP-related degeneration in a *Drosophila* model of MSP⁹ (Supplementary Fig. 3). Last, hnRNA2B1 has been previously implicated in neurodegenerative disease. Specifically, hnRNA2B1 is sequestered in RNA foci in the fragile-X-associated tremor ataxia syndrome (FXTAS)¹⁰, binds the expanded riboCGG repeats that underlie this disease^{11,12}, and is a genetic modifier in a *Drosophila* model of FXTAS^{11,12}.

hnRNA2B1 and hnRNA1 pathology in MSP

Muscle biopsies from patients II5 (family 1) and IV9 (family 2) showed atrophic fibres, central nuclei and rimmed vacuoles characteristic of IBM (Supplementary Fig. 4a–c). Whereas in normal muscle hnRNA2B1 and hnRNA1 are exclusively nuclear (Fig. 2a, e and Supplementary Fig. 4d), analysis of muscle tissue from patient II5 showed that hnRNA2B1 cleared from many nuclei and accumulated in cytoplasmic inclusions in ~10% of fibres (Fig. 2b and Supplementary Fig. 4e). Muscle from this patient also showed TDP-43 pathology consisting of nuclear clearance and cytoplasmic inclusions, consistent with previous observations in VCP-related and sporadic IBM (Fig. 2j and Supplementary Fig. 4p)¹³. Interestingly, hnRNA2B1 pathology was also observed in VCP-related and sporadic IBM (Fig. 2c, d and Supplementary Fig. 4h, i).

Analysis of muscle from patient IV9 also revealed nuclear clearance and cytoplasmic inclusions of hnRNA1 in ~10% of fibres (Fig. 2f and Supplementary Fig. 4j). hnRNA2B1 and TDP-43 pathology were also observed in the muscle biopsy from patient IV9 (Fig. 2h, k and Supplementary Fig. 4f, g). Moreover, hnRNA1 pathology was also observed in VCP-related and sporadic IBM (Fig. 2g and data not shown). Therefore, regardless of aetiology we have found hnRNA2B1, hnRNA1 and TDP-43 pathology in sporadic and familial IBM. Pathological redistribution of hnRNA2B1, hnRNA1 and TDP-43 was not observed in other types of muscle disease (Supplementary Fig. 4q). FUS (also known as TLS)-associated pathology was not observed in these cases, and thus pathological redistribution

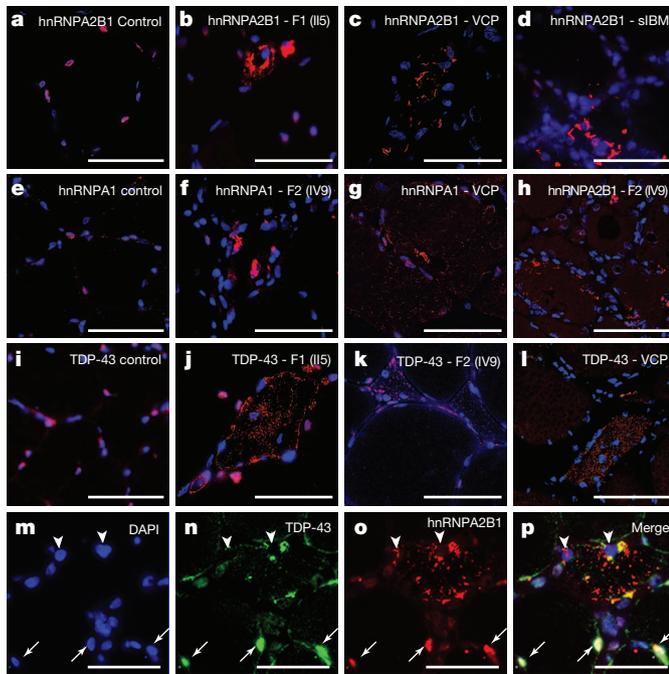


Figure 2 | Cytoplasmic pathology of hnRNPA2B1 and hnRNPA1. **a–d**, Immunohistochemical analysis of hnRNPA2B1 (red) in normal muscle (control, **a**), muscle biopsy from patient II5 from family (F) 1 (**b**), a patient with MSP caused by a VCP mutation (R155H) (**c**) and a patient with sporadic IBM (sIBM, **d**). hnRNPA2B1 (red) was cleared from 4',6-diamidino-2-phenylindole (DAPI)-stained nuclei (blue) and accumulated in cytoplasmic inclusions (**b–d**). **e–g**, Immunohistochemical analysis of hnRNPA1 (red) in a normal muscle (**e**), a muscle biopsy from patient IV9 from family 2 (**f**) and a patient with MSP caused by a VCP mutation (R155H) (**g**). hnRNPA1 was cleared from nuclei and accumulated in cytoplasmic inclusions (**f, g**). **h**, In muscle tissue from patient IV9 (family 2), hnRNPA2B1 was cleared from nuclei and accumulated in cytoplasmic inclusions. (**i–l**) Immunohistochemical analysis of TDP-43 (red) in a normal muscle (**i**), muscle biopsy from patient II5 from family 1 (**j**), muscle biopsy from patient IV9 from family 2 (**k**) and a patient with MSP caused by a VCP mutation (R155H) (**l**). TDP-43 (red) was cleared from DAPI-stained nuclei (blue) and accumulated in cytoplasmic inclusions (**j–l**). **m–p**, Immunohistochemical analysis of hnRNPA2B1 and TDP-43 co-localization in a patient with sporadic IBM. DAPI (**m**), TDP-43 (**n**), hnRNPA2B1 (**o**) and merged images (**p**) are shown. hnRNPA2B1 (red) and TDP-43 (green) were cleared from nuclei (arrowheads) in an atrophic muscle fibre, but found in nuclei of neighbouring unaffected fibres (arrows). Scale bars, 50 μ m.

does not affect all hnRNPs. Double staining revealed that fibres with TDP-43 pathology typically also showed hnRNPA2B1 and hnRNPA1 pathology (Fig. 2m–p and Supplementary Fig. 4g, k). In these instances there was often partial co-localization of TDP-43, hnRNPA2B1 and hnRNPA1. We also observed limited immunopositivity for ubiquitin and p62 (Supplementary Fig. 4l–o).

Disease mutations affect hnRNPA2B1 and A1 PrLDs

hnRNPA2B1 and hnRNPA1 each have carboxy-terminal, glycine-rich domains that are essential for activity and mediate interaction with TDP-43 (ref. 8). FoldIndex predicts these domains to be intrinsically unfolded (Fig. 3a, b). Interestingly, these low-complexity domains have an amino-acid composition similar to that of yeast prion domains. Indeed, a previously devised algorithm¹⁴ predicted that the C-terminal regions of hnRNPA2 (residues 185–341) and hnRNPA1 (residues 186–320) are prion like (Fig. 3a–d), whereas these domains narrowly missed the cutoff for prion propensity according to a different algorithm¹⁵. Interestingly, similar PrLDs are present in many hnRNPs, including TDP-43 and FUS (Supplementary Fig. 5 and Supplementary Table 3). The disease-causing mutations fall at the centre of the predicted PrLD,

and, according to both algorithms, are expected to enhance prion-like behaviour (Fig. 3a, b).

We also examined hnRNPA2B1 and hnRNPA1 with ZipperDB, a structure-based threading algorithm, which scores six-amino-acid segments or their propensity to form two self-complementary beta strands termed ‘steric zippers’ that form the spine of amyloid fibrils¹⁶. Hexapeptides with Rosetta energy lower than $-23 \text{ kcal mol}^{-1}$ are predicted to form steric zippers, with lower energy predicting higher amyloidogenicity. ZipperDB predicted that disease mutations increase the potency of steric zipper motifs present in the PrLDs of hnRNPA2B1 and hnRNPA1 (Fig. 3c–f). This analysis is discussed in greater detail in Supplementary Fig. 6. Taken together, multiple algorithms predict that the PrLDs of hnRNPA2B1 and hnRNPA1 are intrinsically disordered, but poised to access higher order self-templating structures. Notably, the disease mutations are centred within these PrLDs and are predicted to promote transition to an ordered structure.

Disease mutations accelerate hnRNP fibrillization

We directly tested the predictions that (1) hnRNPA2 and hnRNPA1 are prone to fibrillization and (2) that this property is enhanced by disease-causing mutations. First, we experimentally assessed the ZipperDB prediction for the impact of disease mutations on steric zipper motifs found in hnRNPA2B1 and hnRNPA1. Notably, the synthetic mutant hexapeptides of hnRNPA2B1(D290V) (Asn-Tyr-Asn-Val-Phe-Gly) and hnRNPA1(D262V) (Ser-Tyr-Asn-Val-Phe-Gly) rapidly assembled into amyloid fibrils, as shown by thioflavin T fluorescence and electron microscopy, whereas the corresponding wild-type peptides did not, even after several weeks (Fig. 4a–d). Thus, the disease-causing mutations in hnRNPA2B1 and hnRNPA1 generate highly amyloidogenic hexapeptides precisely as predicted. The more potent steric zippers that result from disease mutations are likely to be important for two reasons. First, introduction of similarly potent steric zippers is sufficient to force fibril formation, even in model proteins that do not ordinarily fibrillize¹⁷. Second, in hnRNPA1 and hnRNPA2 these potent steric zippers are centred in the intrinsically disordered PrLD and are thus available to make intermolecular contacts and drive fibril formation.

We next assessed the fibrillization propensity of full-length proteins. We purified bacterially expressed, glutathione S-transferase-tagged wild-type and mutant hnRNPA2 and hnRNPA1 as soluble proteins under native conditions and assessed fibrillization by sedimentation analysis and electron microscopy. Wild-type hnRNPA2 and hnRNPA1 were both intrinsically aggregation prone with a lag phase of ~ 4 h (Fig. 4e, g). Electron microscopy revealed that wild-type hnRNPA2 and hnRNPA1 form fibrils (Fig. 4f, h). The pronounced lag suggests that a critical rate-limiting step in hnRNPA2 and hnRNPA1 fibrillization is nucleation. Indeed, a small quantity (5% wt/wt) of preformed hnRNPA2 or hnRNPA1 fibrils greatly accelerated the assembly of hnRNPA2 or hnRNPA1, respectively (Supplementary Fig. 7a, e). By contrast, preformed hnRNPA2 fibrils did not seed assembly of hnRNPA1 (Supplementary Fig. 7e). Likewise, pre-formed hnRNPA1 fibrils did not seed assembly of hnRNPA2 (Supplementary Fig. 7a). Thus, wild-type hnRNPA1 and hnRNPA2 spontaneously form self-seeding fibrils.

Importantly, the disease mutations greatly accelerated hnRNPA2 and hnRNPA1 fibrillization (Fig. 4e–h). In all cases, the lag phase was curtailed and fibrillization was well advanced while the wild-type protein remained in lag phase. Thus, the disease-causing mutations directly promote nucleation of hnRNPA2 and hnRNPA1 into fibrils. Moreover, fibrils formed by hnRNPA2(D290V), hnRNPA1(D262V) and hnRNPA1(D262N) not only seeded their own assembly (Supplementary Fig. 7b, f, g), but also promoted fibrillization of their respective wild-type counterparts (Supplementary Fig. 7d, i). Neither hnRNPA2(D290V) fibrils nor hnRNPA1(D262V) fibrils seeded the assembly of TDP-43, another RNA-binding protein with a PrLD.

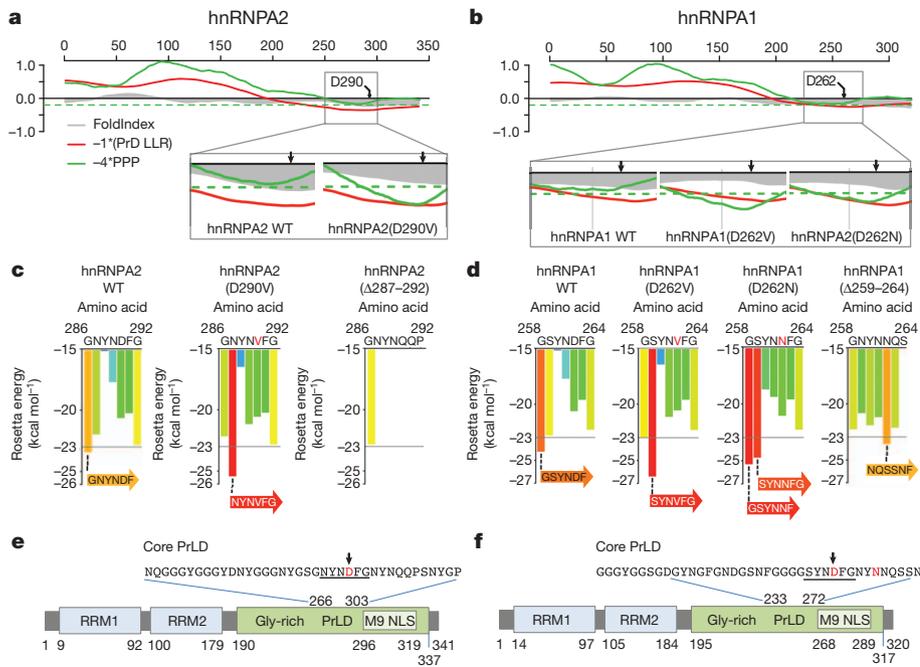


Figure 3 | The disease mutations affect a PrLD in hnRNPA2B1 and hnRNPA1. **a, b,** FoldIndex predicts an extended intrinsically unfolded region (grey curve less than zero) in the C-terminal region of hnRNPA2 (**a**) and hnRNPA1 (**b**). These regions were also predicted to be prion-like according to their log-likelihood ratio scores (PrD LLR) from the algorithm described in ref. 14 (red curve less than zero), and narrowly missed the cutoff for the predicted prion propensity (PPP) log-odds ratio scores from the algorithm described in ref. 15 (green curve below the dashed green line). All curves represent averages of 41 consecutive windows of 41 amino acids, corresponding to the criteria described in ref. 15. The disease mutations were predicted to make these domains more prionogenic (insets). **c, d,** ZipperDB-detected six-amino-acid stretches (underlined in **e, f**) within the core PrLDs for which the disease mutations increased the predicted amyloid-fibril-forming potential beyond the Rosetta threshold. **e, f,** Domain architecture of hnRNPA2 (**e**) and hnRNPA1 (**f**) shows the RNA-recognition motifs 1 and 2 (RRM1 and RRM2), the C-terminal glycine-rich domain and an M9 nuclear-localization signal. The PrLDs are centred in the C-terminal glycine-rich domain. Highly similar predictions were made for the minor isoforms of hnRNPA2B1 (hnRNPA2B1) and hnRNPA1 (hnRNPA1 isoform b). WT, wild type.

Finally, we assessed the importance of the steric zipper motifs in fibrillization. We deleted the steric zipper residues 287–292 from hnRNPA2 and 259–264 from hnRNPA1 and assessed fibrillization *in vitro*. Importantly, neither hnRNPA2(Δ 287–292) nor hnRNPA1(Δ 259–264) formed fibrils (Fig. 4e, g), even after 24 h of agitation. Moreover, hnRNPA2(Δ 287–292) did not fibrillize when seeded by fibrils of wild-type or mutant hnRNPA2 (Supplementary Fig. 7c), nor did hnRNPA1(Δ 259–264) fibrillize when seeded by fibrils of wild-type or mutant hnRNPA1 (Supplementary Fig. 7h). Thus, residues 287–292 of hnRNPA2 and 259–264 of hnRNPA1 are critical for spontaneous and seeded fibrillization of the full-length protein. Collectively, these findings indicate that steric zipper motifs centred in the PrLDs of wild-type hnRNPA2 and hnRNPA1 are critical to their intrinsic tendency to fibrillize and that disease mutations introduce more-potent steric zippers, which accelerate nucleation and polymerization (Supplementary Fig. 6).

Mutant hnRNPA2B1 PrLD supports yeast prion activity

Yeast prion proteins are generally modular, meaning that prion domains from one protein can be transferred to another, while retaining prion activity¹⁸. To determine whether the PrLD of hnRNPA2B1 could support prion activity in yeast, we replaced the Sup35 nucleation domain (residues 3–40) with the core PrLD (residues 261–303) from either wild-type or mutant hnRNPA2B1 and expressed these fusion proteins as the sole copies of Sup35 in the cell. Prion formation was detected by monitoring nonsense suppression of the *ade2-1* allele and verified by monitoring a series of prion-defining criteria, including curability with guanidine, transferability with cytoduction and sensitivity to expression levels (detailed in Supplementary Fig. 8). This analysis showed that the hnRNPA2B1 core PrLD can substitute for the Sup35 nucleation domain in supporting prion formation, and that the D290V mutation specifically promotes the nucleation activity. Moreover, full-length hnRNPA1 and hnRNPA2 form cytoplasmic aggregates and are toxic in yeast (Supplementary Fig. 9).

Mutations enhance hnRNP recruitment to RNA granules

Remarkably, the PrLDs in hnRNPA2B1 and hnRNPA1 correspond to the ‘low-complexity sequences’ found in various hnRNPs (including TDP-43 and FUS) that are essential determinants of RNA granule

assembly¹⁹. Stress granules are cytoplasmic ribonucleoprotein granules composed of repressed translation complexes^{20,21}. TDP-43 and FUS are recruited to stress granules and this is enhanced by disease mutations²⁰. Thus, we proposed that hnRNPA2B1 and hnRNPA1 would be recruited to stress granules and that this would be enhanced by disease mutations. To test this, we first examined the subcellular localization of endogenous hnRNPA2B1 in HeLa cells before and after stress granule induction with arsenite. At baseline, endogenous hnRNPA2B1 was localized exclusively to nuclei (Supplementary Fig. 10). After arsenite treatment hnRNPA2B1 re-localized to eIF4G-positive cytoplasmic puncta, showing the recruitment of endogenous hnRNPA2B1 to stress granules (Supplementary Fig. 10).

To examine the impact of disease mutations, we expressed Flag-tagged versions of wild-type and mutant hnRNPA2. There was significantly greater incorporation of mutant hnRNPA2 into constitutive stress granules than wild-type hnRNPA2 (Fig. 5a). Moreover, after arsenite treatment mutant hnRNPA2 was incorporated into stress granules more rapidly than wild-type hnRNPA2, although by 30 min their degree of incorporation was similar (Fig. 5b). Both wild-type and mutant hnRNPA2 formed high-molecular-weight species that accumulated in the detergent-insoluble fraction (Fig. 5c). Similar to the kinetics of stress granule incorporation, mutant hnRNPA2 accumulated as high-molecular-weight species more rapidly than wild-type hnRNPA2, although by 30 min their accumulation was comparable (Fig. 5c). Similar results were obtained upon examination of endogenous hnRNPA1 as well as exogenous wild-type and mutant hnRNPA1 (Supplementary Fig. 11). Interestingly, hnRNPA2-positive and hnRNPA1-positive stress granules were also immunopositive for TDP-43 (Supplementary Fig. 12). We also generated fibroblast cell lines from patients II4 and II6 (family 1). In these cells endogenous mutant hnRNPA2 accumulated in constitutive stress granules that were immunopositive for eIF4G in addition to TDP-43 and VCP (Supplementary Fig. 13).

Mutant hnRNPs form pathological inclusions *in vivo*

We generated transgenic *Drosophila* expressing wild-type or mutant forms of human hnRNPA2 and hnRNPA1, as well as wild-type or mutant forms of the fly homologue HRB98DE (Supplementary Figs 14 and 15). Expression of wild-type hnRNPA2 in *Drosophila*

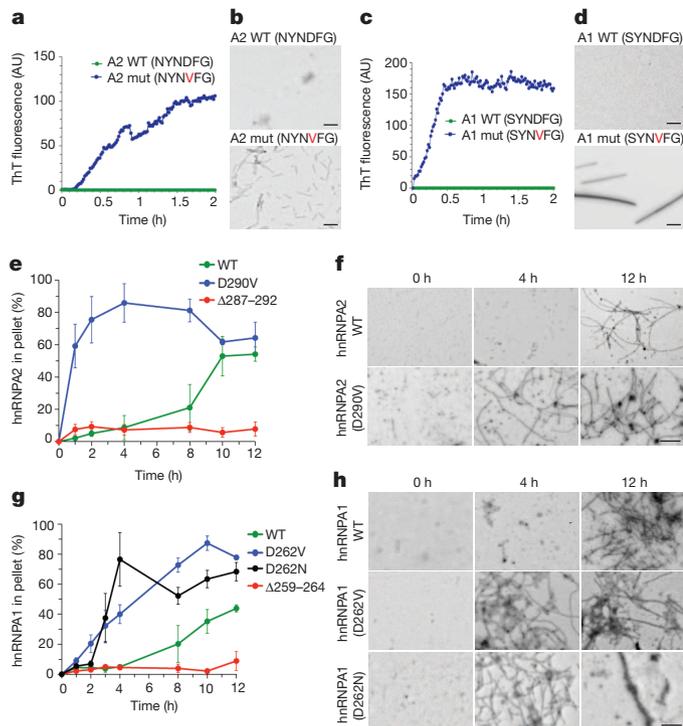


Figure 4 | Disease mutations accelerate hnRNP A2 and hnRNP A1 fibrillization. **a**, Synthetic hexapeptides A2 wild-type (Asn-Tyr-Asn-Asp-Phe-Gly) or mutant (Asn-Tyr-Asn-Val-Phe-Gly) were incubated at 25 °C for 2 h. Fibrillization was monitored by thioflavin T (ThT) fluorescence. **b**, Electron micrograph of A2 wild-type or mutant hexapeptides after 10 min at 25 °C. Scale bars, 0.1 μ m. **c**, **d**, Fibrillization analysis of A1 wild-type (Ser-Tyr-Asn-Asp-Phe-Gly) or mutant (Ser-Tyr-Asn-Val-Phe-Gly) as in **a**, **b**. Scale bars, 0.5 μ m. **e**, Full-length hnRNP A2 wild-type, hnRNP A2(D290V) or hnRNP A2(Δ 287–292) was incubated at 25 °C with agitation for 0–12 h. At various times, the amount of aggregated hnRNP A2 was determined. Values represent means \pm s.e.m. ($n = 3$). **f**, Electron micrograph of hnRNP A2 fibrillization reactions after 0, 4 and 12 h at 25 °C. Note the absence of fibres after 4 h for wild-type hnRNP A2. Scale bar, 0.5 μ m. **g**, **h**, Fibrillization of full-length hnRNP A1 wild-type, hnRNP A1(D262V), hnRNP A1(D262N) or hnRNP A1(Δ 259–264), monitored as in **e**, **f**. Scale bar, 0.5 μ m.

indirect flight muscle led to mild degeneration affecting the rostral portion of several muscles, whereas expression of hnRNP A2(D290V) caused severe degeneration that affected all muscles (Fig. 6a). Importantly, flies overexpressing hnRNP A2(Δ 287–292), in which the hexapeptide (Asn-Tyr-Asn-Asp/Val-Phe-Gly) in the PrLD was deleted, had virtually normal muscles (Fig. 6a). This result indicates that the severe toxicity found in flies overexpressing mutant hnRNP A2 requires the presence of this potent steric zipper motif.

Immunohistochemical analysis showed that wild-type hnRNP A2 localized appropriately to nuclei, as did hnRNP A2(Δ 287–292), whereas hnRNP A2(D290V) largely accumulated in cytoplasmic inclusions (Fig. 6b). We observed a correlation between cytoplasmic inclusion formation and solubility of hnRNP A2. Specifically, mutant hnRNP A2(D290V) protein was largely recovered from the detergent-insoluble fraction, whereas wild-type hnRNP A2 and hnRNP A2(Δ 287–292) proteins were mostly found in the detergent-soluble fraction (Fig. 6c, d). Thus, the degree of muscle degeneration in flies expressing hnRNP A2 correlates with the extent of cytoplasmic inclusions and with hnRNP A2 solubility. Entirely consistent results were obtained upon expression of wild-type and mutant versions of hnRNP A1 and HRB98DE (Supplementary Fig. 15).

Last, we evaluated the impact of the disease mutation on hnRNP A2 cellular localization in mammalian muscle. We electroporated mouse tibialis anterior with plasmid expressing Flag-tagged versions of wild-type or mutant hnRNP A2. Wild-type hnRNP A2 was localized

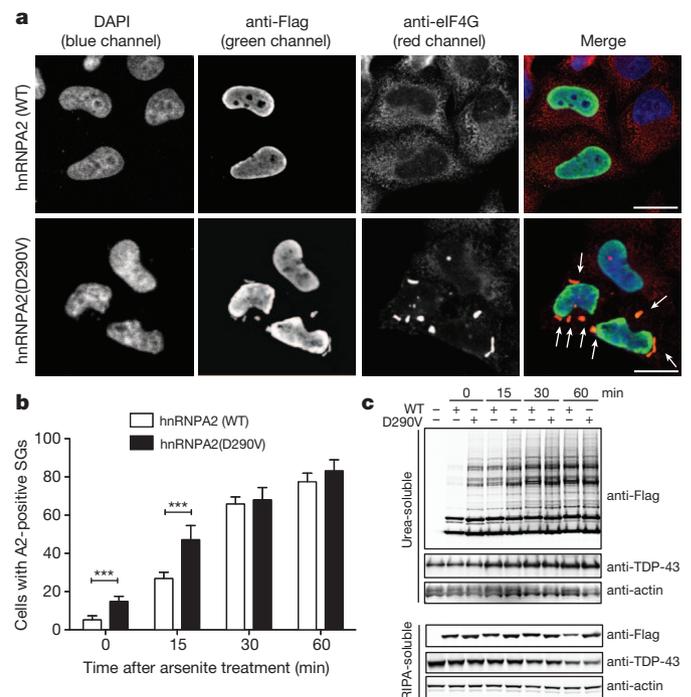


Figure 5 | hnRNP A2 recruitment to stress granules is accelerated by disease mutation. **a**, HeLa cells were transfected with Flag-tagged wild-type or mutant hnRNP A2 and stained with anti-Flag (green), anti-eIF4G (red) and DAPI (blue). Arrows indicate hnRNP A2- and eIF4G-positive stress granules. Scale bar, 20 μ m. **b**, HeLa cells were transfected as in **a**, treated with 0.5 mM sodium arsenite for the indicated time, and immunostained as in **a**. The percentage of cells displaying hnRNP A2-positive stress granules (SGs) at indicated time points after treatment with arsenite cells are plotted. Data represent mean \pm s.e.m. $n = 3$; *** $P < 0.001$. **c**, HeLa cells were transfected and stimulated as in **b**, and sequentially extracted with RIPA (radio immunoprecipitation assay) buffer and urea buffer. Immunoblotting was conducted with anti-Flag, anti-TDP-43 or anti-actin (loading control) antibodies.

appropriately within nuclei, whereas mutant hnRNP A2(D290V) was excluded from nuclei and accumulated in cytoplasmic inclusions (Supplementary Fig. 16), highly reminiscent of the pathology seen in MSP patients.

Discussion

Polymerization of PrLDs underlies the orderly, regulated phase transition that drives assembly of non-membrane-bound organelles including RNA granules, which serve as ‘reaction crucibles’ in which many aspects of RNA metabolism occur²². Disease mutations introduce more potent steric zippers into the PrLDs of hnRNP A2 and hnRNP A1, dysregulating and accelerating nucleation and polymerization, altering the dynamics of RNA granule assembly, which likely has adverse consequences for RNA metabolism. In addition to hnRNP A2 and hnRNP A1, at least four other RNA-binding proteins that harbour PrLDs (TDP-43, FUS, EWSR1 and TAF15) accumulate in disease pathology²³. By analogy to hnRNP A2B1 and hnRNP A1, we speculate that these and perhaps other PrLD-containing proteins contribute to the initiation or propagation of disease by a similar mechanism. Indeed, ~250 human proteins are predicted to harbour PrLDs, many of which are RNA-binding proteins (Supplementary Fig. 5 and Supplementary Table 3), and should be considered candidates for contributing to degenerative disease.

Thus, disease could ensue from unregulated polymerization initiated spontaneously by PrLDs upon environmental stress, or by a mutation affecting either the PrLD or another factor (for example, VCP) that regulates the ribonucleoprotein granule assembly with altered metabolism of RNA as a consequence.

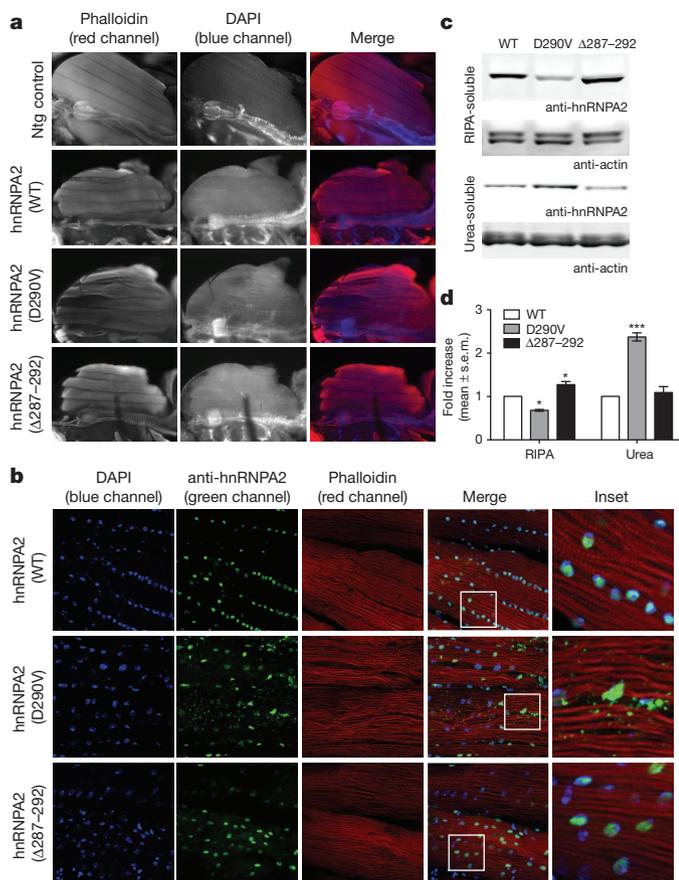


Figure 6 | Mutant hnRNPA2 forms cytoplasmic inclusions in *Drosophila*. **a**, Adult flies were dissected to expose the dorsal longitudinal indirect flight muscle and stained with Texas Red-phalloidin (red) and DAPI (blue). Flies expressing human wild-type hnRNPA2 under control of the *Mhc*-Gal4 driver showed mild degeneration, whereas flies expressing mutant human hnRNPA2 show severe degeneration affecting all muscles. Flies expressing hnRNPA2(Δ 287-292) show muscle histology similar to flies expressing wild-type hnRNPA2. Ntg indicates non-transgenic control. **b**, Wild-type hnRNPA2 localizes exclusively to nuclei, whereas hnRNPA2(D290V) also accumulates extensively in cytoplasmic inclusions. hnRNPA2(Δ 287-292) localizes exclusively to nuclei. **c**, Thoraces of adult flies were dissected and sequential extractions were performed to examine the solubility profile of hnRNPA2. **d**, Quantification of the blot shown in **c**. Data represent mean \pm s.e.m. $n = 3$; * $P < 0.05$, *** $P < 0.001$; two-way analysis of variance with Bonferroni's post hoc test.

Notably, diseases associated with pathological inclusions of PrLD-containing proteins frequently show 'spreading' pathology, in which cellular degeneration with intracellular inclusions starts in one epicentre and subsequently spreads to neighbouring tissue²⁴. Although not directly addressed here, this study suggests that cell-to-cell transmission of a self-templating conformer could contribute to the spreading pathology that is characteristic of these diseases.

The discovery of *VCP* mutations in MSP¹ led to the subsequent discovery of pathogenic *VCP* mutations in sporadic or familial forms of ALS², FTD⁴, IBM⁵ and PDB⁶. Therefore, subsequent to the discovery of mutations in hnRNPA2B1 and hnRNPA1 underlying MSP, we screened 212 familial ALS cases and identified a pathogenic hnRNPA1 mutation in one family, as described earlier. As this manuscript was being assembled for publication we completed screening of 305 sporadic ALS cases and identified a non-synonymous variant in hnRNPA1 (c.800/956A>G, p.N267S/N319S) in one classic, late-onset case in which mutations in known ALS genes had been excluded (Supplementary Fig. 17). Although not definitive, this variant is probably pathogenic because it is centred in the core PrLD and introduces a potent steric zipper similar to that introduced by D262V/N mutations. The frequency of hnRNPA2B1 and hnRNPA1 mutations in

ALS, and the possibility that mutations in these genes underlie some sporadic and familial forms of FTD, IBM and PDB, will be important to address in the future.

METHODS SUMMARY

Exome enrichment of 3 μ g genomic DNA was performed by using the Agilent SureSelect Human All Exon Kit capture library (G3362 for five DNA samples from family 1 and G3370 for five DNA samples from family 2) according to the manufacturer's protocol (SureSelect Human All Exon Illumina Paired-End protocol version 1.0.1). Captured DNAs were sequenced on an Illumina GAIIX sequencer (Illumina) with 76-base-pair (family 1) or 100-base pair (family 2) paired-end reads. At least two lanes of sequencing data were collected for each sample to generate sufficient coverage. Image analyses and base calling were performed by using the Illumina Genome Analyzer Pipeline software (GAPipeline version 1.5 or higher) with default parameters. Reads were aligned to a human reference sequence (University of California, Santa Cruz (UCSC) assembly hg19, National Center for Biotechnology Information (NCBI) build 37) and genotypes were called at all positions at which there were high-quality sequence bases (Phred-like score Q25 or greater) at minimum coverage of 8, using CLC Genomics Workbench v4.5.1 (CLC Bio). For each sample, more than 50% of at least 100 million reads were uniquely mapped to the targeted human exon regions to give mean depth of coverage of 123. For linkage analysis samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. Single nucleotide polymorphism genotype calls were performed according to the standard Affymetrix protocols. A subset of single nucleotide polymorphisms chosen on the basis of heterozygosity was used for linkage analysis, using dChip software (<http://www.hsph.harvard.edu/cli/complab/dchip/>) under parametric linkage analysis and a dominant model with 99% penetrance in heterozygotes and a disease allele frequency of 0.001.

Full Methods and any associated references are available in the online version of the paper.

Received 5 January 2012; accepted 17 January 2013.

Published online 3 March 2013.

- Nalbandian, A. *et al.* The multiple faces of valosin-containing protein-associated diseases: inclusion body myopathy with Paget's disease of bone, frontotemporal dementia, and amyotrophic lateral sclerosis. *J. Mol. Neurosci.* **45**, 522-531 (2011).
- Johnson, J. O. *et al.* Exome sequencing reveals *VCP* mutations as a cause of familial ALS. *Neuron* **68**, 857-864 (2010).
- Watts, G. D. *et al.* Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nature Genet.* **36**, 377-381 (2004).
- Neumann, M., Tolnay, M. & Mackenzie, I. R. The molecular basis of frontotemporal dementia. *Expert Rev. Mol. Med.* **11**, e23 (2009).
- Shi, Z. *et al.* Characterization of the Asian myopathy patients with *VCP* mutations. *Eur. J. Neurol.* **19**, 501-509 (2012).
- Chung, P. Y. *et al.* Indications for a genetic association of a *VCP* polymorphism with the pathogenesis of sporadic Paget's disease of bone, but not for *TNFSF11* (RANKL) and *IL-6* polymorphisms. *Mol. Genet. Metab.* **103**, 287-292 (2011).
- Kottlors, M. *et al.* Late-onset autosomal dominant limb girdle muscular dystrophy and Paget's disease of bone unlinked to the *VCP* gene locus. *J. Neurol. Sci.* **291**, 79-85 (2010).
- Buratti, E. *et al.* TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J. Biol. Chem.* **280**, 37572-37584 (2005).
- Ritson, G. P. *et al.* TDP-43 mediates degeneration in a novel *Drosophila* model of disease caused by mutations in *VCP/p97*. *J. Neurosci.* **30**, 7729-7739 (2010).
- Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256-271 (2006).
- Sofola, O. A. *et al.* RNA-binding proteins hnRNPA2/B1 and CUGBP1 suppress fragile X CCG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS. *Neuron* **55**, 565-571 (2007).
- Jin, P. *et al.* Pur α binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a *Drosophila* model of fragile X tremor/ataxia syndrome. *Neuron* **55**, 556-564 (2007).
- Salajegheh, M. *et al.* Sarcoplasmic redistribution of nuclear TDP-43 in inclusion body myositis. *Muscle Nerve* **40**, 19-31 (2009).
- Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146-158 (2009).
- Toombs, J. A., McCarty, B. R. & Ross, E. D. Compositional determinants of prion formation in yeast. *Mol. Cell. Biol.* **30**, 319-332 (2010).
- Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **107**, 3487-3492 (2010).
- Teng, P. K. & Eisenberg, D. Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng. Des. Sel.* **22**, 531-536 (2009).

18. Li, L. & Lindquist, S. Creating a protein-based element of inheritance. *Science* **287**, 661–664 (2000).
19. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
20. Wolozin, B. Regulated protein aggregation: stress granules and neurodegeneration. *Mol. Neurodegener.* **7**, 56 (2012).
21. Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **36**, 932–941 (2009).
22. Weber, S. C. & Brangwynne, C. P. Getting RNA and protein in phase. *Cell* **149**, 1188–1191 (2012).
23. Neumann, M. *et al.* FET proteins TAF15 and EWS are selective markers that distinguish FTL with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain* **134**, 2595–2609 (2011).
24. King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* **1462**, 61–80 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the patients whose participation made this work possible. We thank the St Jude Pediatric Cancer Genome Project and J. Zhang in particular for providing access to control sequencing data. We thank C. Gellera,

B. Baloh, M. Harms, S. Krause, G. Dreyfuss and T. Cundy for sharing reagents. We thank S. Donkervoort and S. Mumm for coordinating samples, and A. Taylor for editorial assistance. J.P.T. was supported by ALSAC, the Packard Foundation and the National Institutes of Health (NIH) (NS053825); J.P.T. and M.B. were supported by the ALS Association; J.Q.T. was supported by the NIH (AG032953); J.S. was supported by the NIH (DP20D002177 and NS067354) and the Ellison Medical Foundation; E.D.R. was supported by the National Science Foundation (MCB-1023771). C.C.W. was supported by the NIH (AG031867).

Author Contributions H.J.K., N.C.K., E.D.R., C.C.W., J.S. and J.P.T. designed experiments. H.J.K., N.C.K., E.A.S., J.Moore, Z.D., K.S.M., B.F., S.L., A.M., A.P.K., Y.R.L. and A.F.F. performed the experiments. K.B.B., A.M.W., R.R., J.L.P., S.A.G., J.Q.T., B.N.S., S.T., A.-S.G., J.Miller, C.E.S., M.K., J.K., A.P., M.B. and V.E.K. provided patient clinical material, clinical evaluation, or evaluation of patient clinical material. H.J.K., N.C.K., Y.-D.W., R.C., B.J.T., A.D.G., O.D.K., E.D.R., J.S. and J.P.T. contributed to data analysis. E.D.R., O.D.K. and C.C.W. contributed to manuscript preparation. H.J.K., J.S. and J.P.T. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P.T. (JP.Taylor@STJUDE.ORG) or J.S. (jshorter@mail.med.upenn.edu).

METHODS

Patients. Patients were examined by M.B. and J.P.T. except where indicated. The Institutional Review Board of the University of Miami approved the study protocol and all participants provided written informed consent.

Exome capture and next-generation sequencing. Exome enrichment of 3 µg genomic DNA was performed by using the Agilent SureSelect Human All Exon Kit capture library (G3362 for five DNA samples from family 1 and G3370 for five DNA samples from family 2) according to the manufacturer's protocol (SureSelect Human All Exon Illumina Paired-End protocol version 1.0.1). The capture library, containing regions totalling approximately 38 megabases (G3362) or 50 megabases (G3370), is designed to target all human exons in the NCBI consensus coding sequence (CCDS) database. Captured DNAs were sequenced on an Illumina GAIIX sequencer (Illumina) with 76-base-pair (family 1) or 100-base-pair (family 2) paired-end reads. At least two lanes of sequencing data were collected for each sample to generate sufficient coverage. Image analyses and base calling were performed by using the Illumina Genome Analyzer Pipeline software (GAPipeline version 1.5 or higher) with default parameters. Reads were aligned to a human reference sequence (UCSC assembly hg19, NCBI build 37) and genotypes were called at all positions at which there were high-quality sequence bases (Phred-like score Q25 or greater) at minimum coverage of 8, using CLC Genomics Workbench v4.5.1 (CLC Bio). For each sample, more than 50% of at least 100 million reads were uniquely mapped to the targeted human exon regions to give a mean depth of coverage of 123. Under such coverage, approximately 94% of targeted regions were covered by 5 reads or more, and more than 85% were covered by more than 20 reads (Supplementary Fig. 1). To identify the pathogenic mutations, we performed a series of filtering steps (Supplementary Fig. 1). We first discarded the variants that did not change the amino-acid sequence. We then generated a list of variants that are present in affected individuals but absent in nonaffected individuals. Variants previously reported in the Single Nucleotide Polymorphism Database build 132 (dbSNP132) hosted by the National Center for Biotechnology Information were excluded. The variants were further filtered by using the in-house database (a collection of variants from whole-exome sequencing of more than 625 individuals) by a similar methodology. Finally Sanger sequencing was performed on DNA of three additional members (both affected and nonaffected) from each family to exclude the nonpathogenic mutations. Presence of mutations was confirmed in all candidate disease-causing variants by Sanger sequencing.

Linkage analysis. Eight DNA samples (parents and six siblings) from family 1 were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. SNP genotype calls were performed according to the standard Affymetrix protocols. A subset of SNPs chosen on the basis of heterozygosity was used for linkage analysis, using dChip software (<http://www.hsph.harvard.edu/cli/complab/dchip>) under parametric linkage analysis and a dominant model with 99% penetrance in heterozygotes and a disease allele frequency of 0.001. Five genomic segments with a size greater than 100 kilobases, at least 20 SNP markers, and a log odds ratio (lod) score greater than 0 are shown in Supplementary Fig. 1.

Generation of phylogram for hnRNP A/B gene family. TBPH of *D. melanogaster* was used to BLAST (basic local alignment search tool) against the RefSeq protein database including species of *M. musculus*, *D. melanogaster*, *A. gambiae*, *H. sapiens*, *D. rerio* (zebrafish), *C. elegans* and *G. gallus* (jungle fowl). The top 100 sequences (according to BLAST *E* values) were obtained. Sequences sharing more than 98% similarity were filtered by UCLUST²⁵ because they likely represent alternatively spliced variants of the same genes. Eighty-one sequences remaining after this process were aligned using the MAFFT multiple sequence alignment program²⁶. The DNA-coding sequences of all proteins were extracted and threaded onto the protein sequences to form an in-frame codon alignment. Squint²⁷ was used to extract the DNA-coding alignment of the two RRM domains present in each protein using the protein alignment as a guide. Phylogenetic trees of the codon alignment including 21 genes from only human and *Drosophila* were generated using the genetic algorithm-based GARLI²⁸ method.

Constructs. Complementary DNAs encoding human hnRNP2 (accession number NM_002137.2) or hnRNP1 (accession number NM_002136.1) in the plasmid pCMV6-XL5 were obtained from OriGene. A Flag epitope tag was added to the 5' end of hnRNP2 or hnRNP1 by PCR. The cDNAs were then subcloned into a pcDNA 3.1 plasmid, using restriction sites EcoRI and XhoI, creating pcDNA 3.1-Flag-hnRNP2 or -Flag-hnRNP1. cDNAs encoding *Drosophila* HRB98DE (accession number NM_170373.1) were obtained from the *Drosophila* Genomics Research Center. Missense mutations of hnRNP2 (hnRNP2 (D290V)), hnRNP1 (hnRNP1 (D262V)) and *Drosophila* HRB98DE (HRB98DE (D302V)) were generated by site-directed mutagenesis. Sequence analysis was conducted on all plasmid constructs.

Antibodies. The following commercial antibodies were used in this study: mouse monoclonal anti-hnRNP2B1 antibodies EF-67 and DP3B3 (Santa

Cruz Biotechnology), rabbit polyclonal anti-TDP-43 antibody (Protein Tech Group), anti-eIF4G antibody (Santa Cruz Biotechnology), anti-hnRNP1 antibody (Cell Signaling Technology), anti-VCP antibody (Epitomics) and anti-Flag M2 antibody (Sigma). Anti-hnRNP1 antibodies (4B10 and 9H10) were generous gifts from G. Dreyfuss. Validation of these antibodies is shown in Supplementary Fig. 18.

Cell culture and transfection. HEK293 cells and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 1% L-glutamate. Human fibroblasts were grown in DMEM nutrient mixture F-12 (DMEM/F12) supplemented with 20% FBS, 1% penicillin/streptomycin and 1% L-glutamate. Cells were transfected using Lipofectamine LTX with Plus Reagent (Invitrogen), according to the manufacturer's instructions.

RNA interference. Short interfering RNAs (siRNAs) against hnRNP2B1 were ordered from Santa Cruz Biotechnology (sc-43841, target sequences: 5'-CAGUCCGUAAGCUCUUUAtt, GGAUGGCUAUAUUGGGUAUtt-3' and 5'-GGAUCAUGGUGUAAUAAGAtt-3') and Dharmacon (ON-TARGETplus SMARTpool, human HNRPA2B1 (3181), target sequences: 5'-CGGUGGA AAUUUCGGACCA-3', 5'-GCUGUUUGUUGCGGAAUU-3', 5'-GGAGAG UAGUUGAGCCAAA-3' and 5'-GAGGAGGAUCUGAUGGAUA-3'). Lower-case letters indicate non-targeting 3'-overhang included for stability. Control siRNAs were ordered from Dharmacon (ON-TARGETplus Non-targeting siRNA no. 1). HeLa cells were transfected in six-well plates with Lipofectamine RNAiMax (Invitrogen) according to the manufacturer's protocol and collected after 48 and 72 h.

Immunofluorescence studies. Cells were fixed in 4% paraformaldehyde in PBS buffer, permeabilized with 0.5% Triton X-100 in PBS for 10 min, blocked with 5% goat serum in PBS for 30 min and incubated with primary antibody for 2 h at room temperature (20–25 °C) or overnight at 4 °C. Primary antibodies were visualized with secondary antibodies conjugated with Alexa Fluor 488 and Alexa Fluor 594 (Molecular Probes, Invitrogen) and nuclei were detected using DAPI. Stained cells were examined using a confocal microscope (Zeiss LSM 510 NLO Meta) with Zeiss ZEN software.

Solubility and biochemical analyses. Sequential extractions were performed to examine the solubility profile of hnRNP2B1. Cells were washed twice with PBS, lysed in cold RIPA buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS and 1 mM EDTA) and sonicated. Cell lysates were cleared by centrifugation at 100,000g for 30 min at 4 °C to generate RIPA-soluble samples. To prevent carry-overs, the resulting pellets were washed twice with PBS (that is, re-sonicated and re-centrifuged). Only supernatants from the first centrifugation were analysed. RIPA-insoluble pellets were then extracted with urea buffer (7 M urea, 2 M thiourea, 4% CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulphonate), 30 mM Tris, pH 8.5), sonicated and centrifuged at 100,000g for 30 min at 22 °C. Protease inhibitors were added to all buffers before use. Protein concentration was determined by the bicinchoninic acid method (Pierce) and proteins were resolved by NuPAGE Novex 4–12% Bis-Tris Gel (Invitrogen).

In vivo electroporation of mice. All animal work was performed with the approval of the St Jude Children's Research Hospital Committee on Animal Care. Mice were anesthetized and the skin overlying the tibialis anterior muscle was shaved, and the animals were injected with endotoxin-free Flag-tagged wild-type hnRNP2 or Flag-tagged hnRNP2(D290V) mutant diluted in sterile PBS to a volume of 50 µl by using a 0.5-ml syringe fitted with a 29-gauge needle. Two-needle array electrodes (BTX Technologies) were inserted into the muscle immediately after DNA delivery for electroporation. The distance between the electrodes was 5 mm, and the array was inserted longitudinally relative to the muscle fibres. *In vivo* electroporation parameters were as follows: voltage, 75 V; pulse length, 50 ms; number of pulses, 6; pulse interval, 200 ms; desired field strength, 200 V cm⁻¹, given by a BTX ECM830 Electro Square Porator. After 7 days of recovery, mice were processed for immunofluorescence analysis.

Muscle histochemistry and immunohistochemistry. Human samples were processed for routine histochemical analysis as previously described¹³. For skeletal muscle immunohistochemistry, isolated skeletal muscle was mounted by using tragacanth gum and quick-frozen in liquid-nitrogen-cooled 2-methylbutane. Sections (8 µm) were placed on slides and fixed in 4% paraformaldehyde for 5 min and then ice-cold acetone for 5 min. Sections were incubated with PNB blocking reagent (Perkin Elmer) for 1 h and then incubated overnight in primary antibody diluted in PNB. After serial washes in 1× PBS, slides were incubated in secondary antibody diluted in PNB for 2 h. Sections were mounted with Mowiol containing DAPI. Specimens were examined using a fluorescent microscope (80i upright; Nikon) and charge-coupled device camera (EZ monochrome; Roper Industries) with deconvolution software analysis (NIS Elements; Nikon). Nonfluorescent images were taken with a 5-megapixel colour charge-coupled

device (Nikon). Image processing and analysis were performed with NIS Elements 4.0 software and Photoshop CS3 (Adobe).

Yeast strains, media and plasmids. Yeast cells were grown in rich media (YPD) or in synthetic media lacking uracil and containing 2% glucose (SD/–Ura), raffinose (SRaf/–Ura) or galactose (SGal/–Ura). The TDP-43 and FUS yeast expression constructs have been previously described^{29–31}. Wild-type and IBMPFD-associated mutant hnRNPA2B1 and hnRNPA1 Gateway entry clones were generated by PCR, incorporating the flanking Gateway attB1 and attB2 sites along with a Kozak consensus sequence. Resulting PCR products were shuttled into pDONR221, using a Gateway BP Clonase reaction. Two versions of each entry clone (with or without stop codon) were then used in LR reactions with pAG416Gal-ccdB, pAG426Gal-ccdB, pAG416Gal-ccdB-GFP or pAG426Gal-ccdB-GFP³² to generate the CEN (centromeric) and 2- μ m untagged or YFP-fusion constructs. Primer sequences are available upon request. To prevent unwanted recombination events owing to unstable repetitive sequences in the hnRNPA2B1 and hnRNPA1 DNA sequence, we propagated plasmids in SURE-2 or STBL-3 *Escherichia coli* at 30 °C. For testing the phenotype of wild-type and IBMPFD-associated mutant hnRNPA2 as Sup35-fusion proteins, PCR was used to amplify amino acids 261–303 of hnRNPA2 and insert it into Sup35 in the place of amino acids 3–40. The resulting PCR products were co-transformed with BamHI/HindIII-cut pJ526 (from D. Masison, NIH; ref. 33) into strain 780-1D (*MAT α kar1-1 SUQ5 ade2-1 his3 Δ 202 leu2 Δ 1 trp1 Δ 63 ura3-52 sup35::KanMX*) containing the SUP35 maintainer plasmid pJ533 (from D. Masison; ref. 34). Transformants were selected on medium lacking leucine, and then transferred to 5-fluoroorotic-acid-containing medium to select for loss of pJ533. Sequences were confirmed by DNA sequencing. Overexpression plasmids were generated by PCR amplification, restriction digest and cloning into compatible sites in pKT24 (from K. Taylor, NABI), which contains a *GAL1* promoter, as previously described³⁵.

Yeast transformation and spotting assays. Yeast procedures were performed according to standard protocols³⁶. The polyethylene glycol/lithium acetate method was used to transform yeast with plasmid DNA³⁷. For spotting assays, yeast cells were grown overnight at 30 °C in SRaf/–Ura until they reached log or mid-log phase. Cultures were then normalized for $D_{600\text{ nm}}$, serially diluted, spotted onto synthetic solid media containing SD/–Ura or SGal/–Ura, and grown at 30 °C for 2–3 days. hnRNPA2B1–Sup35 fusions were tested for prion formation and for stability and curability of the Ade⁺ phenotype as previously described³⁵. Cytoconductions were performed as Ade⁺ previously described³⁸. The recipient strain was YER746 (*MAT α kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 arg1::HIS3 sup35::KanMX*), carrying plasmid pER697 (URA3), which expresses the hnRNPA2(D290V)–Sup35 fusion from the SUP35 promoter.

Fluorescence microscopy for yeast. For fluorescence microscopy experiments, single-colony isolates of yeast strains were grown to mid-log phase in SRaf/–Ura media at 30 °C. Cultures were spun down and re-suspended in the same volume of SGal/–Ura to induce expression of the hnRNPA2B1 or hnRNPA1 constructs. Cultures were induced with galactose for 4–6 h and processed for microscopy. Images were obtained by an Olympus IX70 inverted microscope and a Photometrics CoolSnap HQ 12-bit CCD camera. Z stacks of several fields were collected for each strain. The images were de-blurred using a nearest neighbour algorithm in the DeltaVision softWoRx software and representative cells were chosen.

Hexapeptide assembly. The wild-type hnRNPA2 (Asn-Tyr-Asn-Asp-Phe-Gly), mutant hnRNPA2(D290V) (Asn-Tyr-Asn-Val-Phe-Gly), wild-type hnRNPA1 (Ser-Tyr-Asn-Asp-Phe-Gly) and mutant hnRNPA1(D262V) (Ser-Tyr-Asn-Val-Phe-Gly) hexapeptides were synthesized at the Keck Biotechnology Resource Laboratory at Yale University School of Medicine. A1 peptides were dissolved at 1 mM and A2 peptides at 5 mM in 150 mM KCl, 40 mM HEPES-KOH, pH 7.4 and 1 mM dithiothreitol and used immediately for assembly reactions. Fibre-assembly reactions were monitored at $\lambda_{\text{ex/em}}$ 440/482 nm, in the presence of 25 μ M thioflavin T, at room temperature on a Tecan Safire² or Tecan Infinite M1000 plate reader. Alternatively, reactions were processed for electron microscopy as for the full-length hnRNPs (see below).

hnRNP purification. Wild-type and mutant hnRNPA2 or hnRNPA1 were expressed and purified from *E. coli* as GST-tagged proteins. Expression constructs were generated in pDuet to contain a TEV-cleavable site, resulting in a GST-TEV-hnRNP construct^{30,31}. GST-TEV-hnRNP was overexpressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells (Agilent) and purified under native conditions using a glutathione-sepharose column (GE) according to the manufacturer's instructions. Proteins were eluted from the glutathione sepharose with assembly buffer (AB; 40 mM HEPES-KOH, 150 mM KCl, 5% glycerol, 20 mM glutathione, pH 7.4). Protein was centrifuged for 10 min at 16,100g, and supernatant was separated from pellet to remove any protein aggregates. Protein concentration was determined by Bradford assay (Bio-Rad) in comparison to BSA standards.

In vitro fibril-formation assay. For GST-tagged protein, aggregation was initiated by adding TEV protease (Invitrogen) to GST-TEV-hnRNPA2 (3 μ M) or GST-TEV-hnRNPA1 (5 μ M) in AB. Aggregation reactions were incubated at 25 °C for 0–12 h with or without agitation at 1,200 r.p.m. in an Eppendorf Thermomixer. For self-seeded and cross-seeded reactions, hnRNPA2, hnRNPA2(D290V), hnRNPA1 or hnRNPA1(D262V) fibrils were assembled at 5 μ M for 24 h with agitation at 1,400 r.p.m. in AB. These preformed fibrils then used to seed (5% wt/wt) the assembly of hnRNPA2 (2.5 μ M), hnRNPA2(D290V) (2.5 μ M), hnRNPA1 (5 μ M), hnRNPA1(D262V) (5 μ M) or hnRNPA1(D262N) (5 μ M) in AB as indicated. Here, assembly reactions were agitated at 1,200 r.p.m. at 25 °C. For hnRNPA2 cross-seeding reactions the amount of preformed fibrils added was increased to 10% wt/wt and reactions were not agitated. In hnRNPA1 cross-seeding reactions, preformed fibrils were briefly sonicated to fragment fibrils before addition to the assembly reaction. For hnRNPA1 cross-seeding reactions the amount of preformed fibrils added was 5% wt/wt and reactions were not agitated. For sedimentation analysis, samples were centrifuged at 16,100g for 10 min at 25 °C. Pellet fractions were resolved by SDS-PAGE and stained with Coomassie blue. The amount of protein in the pellet fraction was determined by densitometry in comparison to known amounts of hnRNP. For electron microscopy of *in vitro* aggregation reactions, samples (10 μ l) were adsorbed onto glow-discharged 300-mesh Formvar/carboncoated copper grid (Electron Microscopy Sciences) and stained with 2% (w/v) aqueous uranyl acetate. Excess liquid was removed, and grids were allowed to air dry. Samples were viewed by a JEOL 1010 transmission electron microscope.

Fly stocks and culture. The wild-type hnRNPA2 and mutant hnRNPA2(D290V) cDNAs were subcloned into the pUASTattB plasmid, using restriction sites EcoRI and XhoI, creating pUASTattB-wild type hnRNPA2 or pUASTattB-mutant hnRNPA2(D290V). Wild-type HRB98DE and mutant HRB98DE(D302V) were subcloned into pUASTattB plasmid using NotI and XhoI sites. Flies carrying pUASTattB transgenes were generated by a standard injection and ϕ C31 integrase-mediated transgenesis technique. To express a transgene in muscles, *Mhc-Gal4* was used (from G. Marqués). All *Drosophila* stocks were maintained in a 25 °C incubator with a 12-h day/night cycle.

Adult fly muscle preparation and immunohistochemistry. Adult flies were embedded in a drop of OCT compound (Sakura Finetek) on a slide glass, frozen with liquid nitrogen and bisected sagittally by a razor blade. After fixing with 4% paraformaldehyde in PBS, hemithoraces were stained by Texas Red-X phalloidin (Invitrogen) and DAPI according to manufacturer's instructions. Stained hemithoraces were mounted in 80% glycerol, and the musculature was examined by DMIRE2 (Leica, 10 \times). For hnRNPA2 staining, hemithoraces were permeabilized with PBS containing 0.2% Triton X-100 and stained with anti-hnRNPA2B1 (EF-67) antibody (Santa Cruz Biotechnology) and Alexa-488-conjugated secondary antibody (Invitrogen). Stained muscle fibres were dissected and mounted in Fluoromount-G (Southern Biotech) and imaged with a Marianas confocal microscope (Zeiss, \times 63).

Fly western blotting. Thoraces of adult flies were prepared and ground in PBS containing 0.2% Triton X-100. After adding SDS sample buffer, samples were boiled for 5 min and analysed by the standard western blotting method provided by Odyssey system (LI-COR) with 4–12% NuPAGE Bis-Tris Gel (Invitrogen) and anti-hnRNPA2B1 antibody (Santa Cruz, 1:1,000).

- Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
- Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
- Goode, M. G. & Rodrigo, A. G. SQUINT: a multiple alignment program and editor. *Bioinformatics* **23**, 1553–1555 (2007).
- Zwickl, D. J. *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion*. PhD thesis, Univ. Texas at Austin (2006).
- Johnson, B. S., McCaffery, J. M., Lindquist, S. & Gitler, A. D. 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 (2008).
- Johnson, B. S. *et al.* TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J. Biol. Chem.* **284**, 20329–20339 (2009).
- Sun, Z. *et al.* Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLN. *PLoS Biol.* **9**, e1000614 (2011).
- Alberti, S., Gitler, A. D. & Lindquist, S. A suite of Gateway@ cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**, 913–919 (2007).
- Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. Primary sequence independence for prion formation. *Proc. Natl Acad. Sci. USA* **102**, 12825–12830 (2005).
- Song, Y. *et al.* Role of Hsp70 chaperone in *Saccharomyces cerevisiae* prion seed replication. *Eukaryot. Cell* **4**, 289–297 (2005).

35. Ross, C. D., McCarty, B. M., Hamilton, M., Ben-Hur, A. & Ross, E. D. A promiscuous prion: efficient induction of [URE3] prion formation by heterologous prion domains. *Genetics* **183**, 929–940 (2009).
36. Guthrie, C. & Fink, G. R. *Methods in Ezymology: Guide to Yeast Genetics and Molecular and Cell Biology* 169 (Academic, 2002).
37. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168 (1983).
38. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. Primary sequence independence for prion formation. *Proc. Natl Acad. Sci. USA* **102**, 12825–12830 (2005).

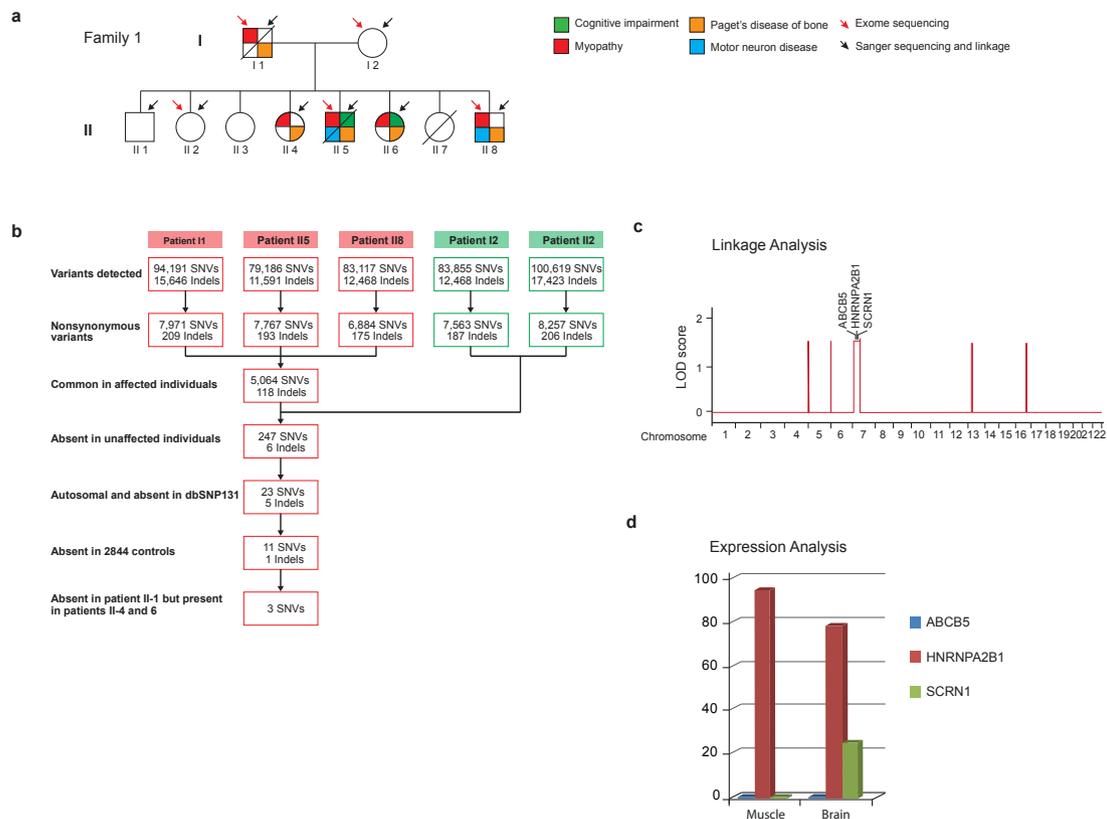


Figure 1. Identification of the disease-causing mutation in family 1. **a.** The pedigree of family 1 indicating individuals affected by cognitive impairment, myopathy, PDB, and motor neuron disease¹. Patient I1 had behavioral changes, weakness, muscle atrophy, and progressive skeletal abnormalities. At autopsy, he was diagnosed with FTD, IBM, and PDB. The affected children had onset of weakness and skeletal abnormalities in their 20s, which was slowly progressive. Two of the children (patients II5 and II6) showed evidence of impaired executive function upon cognitive exam. Muscle biopsy (patients II4, II5, and II6) showed groups of small atrophic fibers and rimmed vacuoles characteristic of IBM in addition to regions of grouped atrophy. Radiological and serum evaluation (patients II4, II5, II6, and II8) confirmed diagnoses of PDB. Electrophysiological examination (II5 and II8) showed motor neuron dysfunction in addition to myopathic features. Patient II8 subsequently succumbed to rapidly progressive weakness. **b.** The filtering paradigm used to detect novel sequence variants that co-segregated with disease is shown. Affected patients represented by red and unaffected patients by green. **c.** The results of linkage analysis in family 1 is shown and the location of the 3 SNVs that co-segregated with disease is indicated. **d.** The relative expression levels of hnRNPA2/B1, SCR1, and ABCB5 in muscle and brain are indicated as determined by RNAseq. Not only is there limited to no expression of SCR1 and ABCB5 in affected tissues, the variants we identified in these genes were subsequently identified by the NHLBI Exome Sequencing Project as polymorphisms present in the normal population (see Table 1).

Gene	Location	Chromosome	Allele Variant (Ref/patient)	Variation Type	Amino Acid Change	Residue Conserved?	Expressed in Adult Muscle and CNS?	Polyphen Prediction	Absent polymorphism in NHLBI Exome Sequencing Project
HNRNPA2B1	26232966	7	A/T	SNV	D290V (isoform A2) D302V (isoform B1)	Yes	Yes	Damaging	Yes
ABCB5	20784996	7	T/G	SNV	L677V	No	No	Benign	No
SCRN1	30008638	7	G/A	SNV	R16C (isoform a) R36C (isoform b)	No	Detectable	Benign	No

Table 1. Summary table of data relating to the 3 SNVs that co-segregate with disease in family 1.

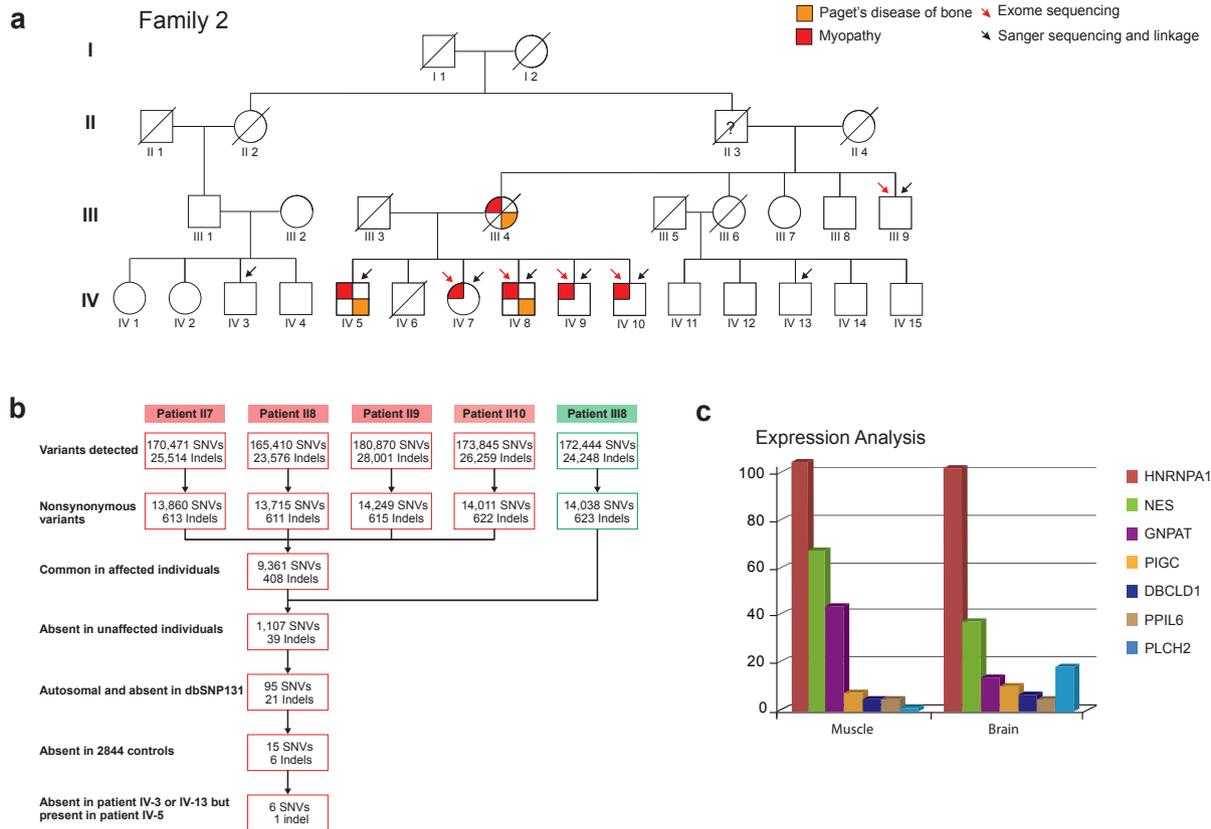


Figure 2. Identification of the disease-causing mutation in family 2. **a.** The full pedigree of family 2 indicating individuals affected by myopathy and PDB. **b.** The filtering paradigm used to detect novel sequence variants that co-segregated with disease is shown. Affected patients represented by red and unaffected patients by green. **c.** The relative expression levels of candidate genes in muscle and brain are indicated as determined by RNAseq. Subsequently, the PPIL6 variant was identified as a polymorphism in the normal population by the NHLBI Exome Sequencing Project (see Table 2).

Gene	Location	Chromosome	Allele Variant (Ref/patient)	Variant Type	Amino Acid Change	Residue Conserved?	Expressed in Adult Muscle and CNS?	Polyphen Prediction	Absent polymorphism in NHLBI Exome Sequencing Project
HNRNPA1	54677629	12	A/T	SNV	D262V (isoform a) D314V (isoform b)	Yes	Yes	Damaging	Yes
NES	156641885	1	G/A	SNV	P699S	No	Yes	Benign	Yes
DCBLD1	117825031	6	G/A	SNV	I72V	No	Detectable	Possibly damaging	Yes
PPIL6	109752445	6	A/G	SNV	V80A (isoform 1) V112A (isoform 2)	No	Detectable	Benign	No
PIGC	172411650	1	C/T	SNV	R38Q	No	Detectable	Benign	Yes
GNPAT	231386768	1	A/C	SNV	D47A	No	Yes	Possibly damaging	Yes
PLCH2	2435831	1	Tcc/---	Deletion	S1144del	N/A	Detectable	N/A	Yes

Table 2. Summary table of data relating to the 6 SNVs and 1 indel that co-segregate with disease in family 2.

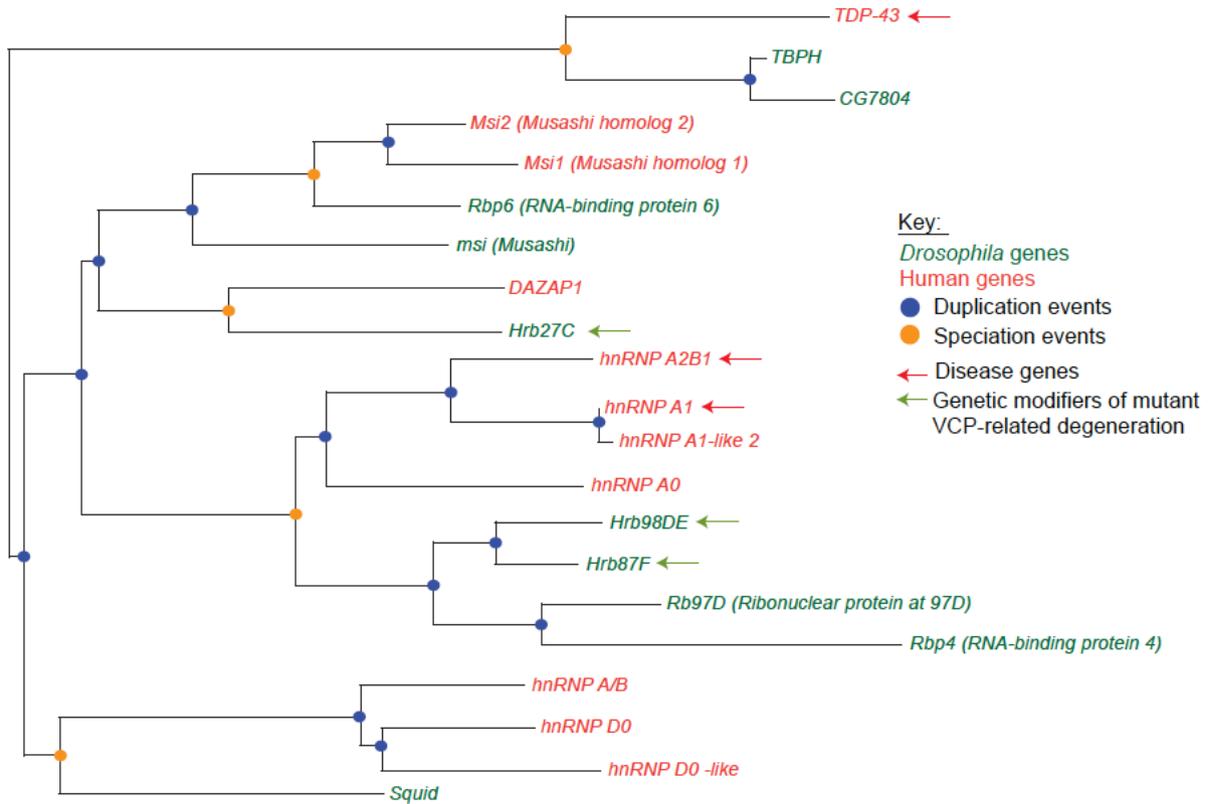


Figure 3. Orthology/paralogy assignment of hnRNP A/B proteins in *Drosophila melanogaster* and *Homo sapiens*. Phylogenetic analysis was used to resolve the orthological/paralogical relations among human and drosophila proteins that show homology to TBPH. Twenty-one genes were shown in the phylogram, in which drosophila and human genes are in red and green colors, respectively. The node colors denote duplication events (red nodes) and speciation events (green nodes). Disease genes indicated with red arrow and genetic modifiers of VCP-related degeneration indicated with green arrow².

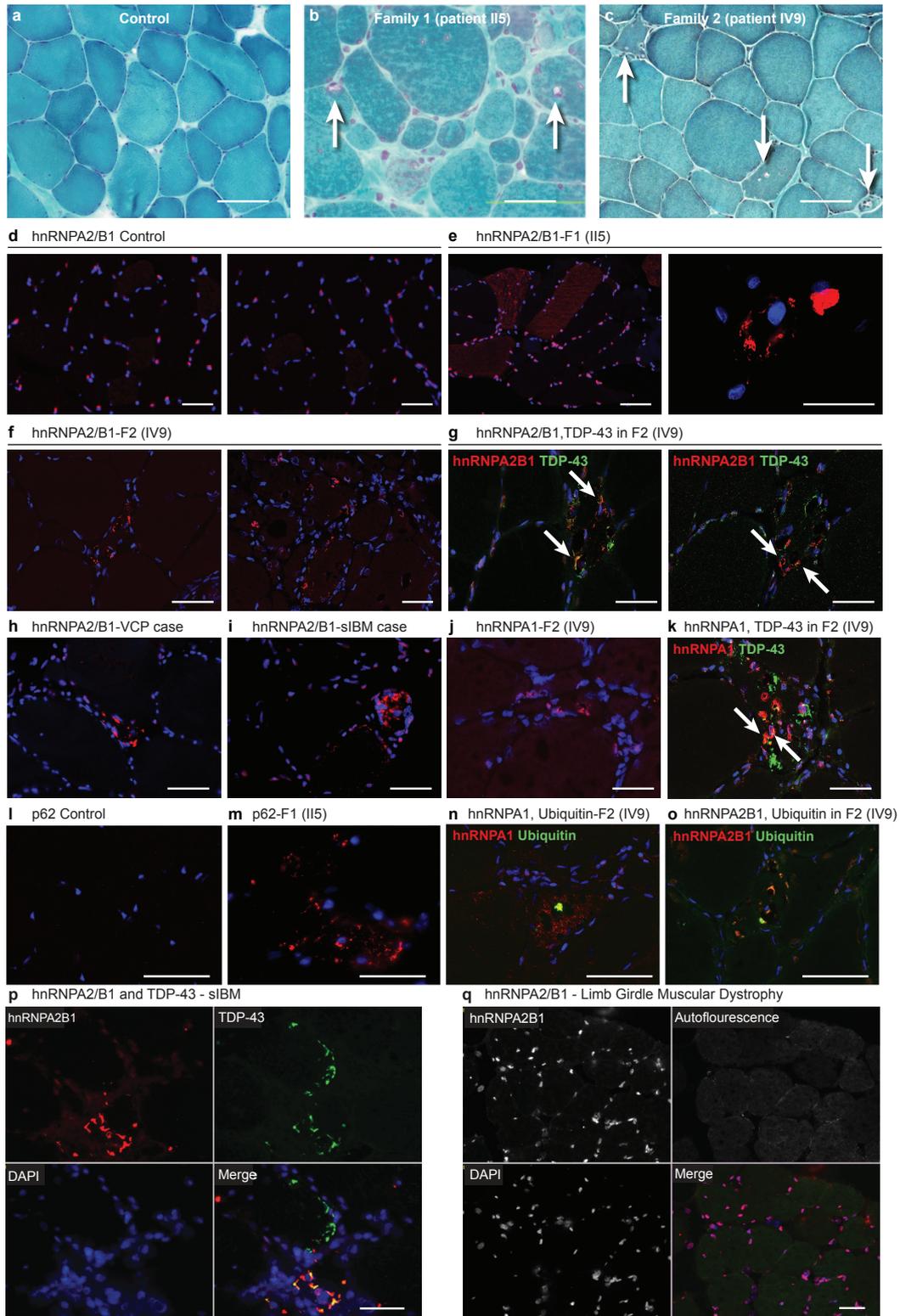


Figure 4. Nuclear clearance and accumulation of pathological sarcoplasmic inclusions of hnRNPA2/B1 and hnRNPA1 in patients. **a-c.** Gomori trichrome staining in a normal muscle sample (**a**), a muscle biopsy from patient II5 from family 1 (**b**) and patient IV9 from family 2 (**c**). Arrows indicate rimmed vacuoles. **d-f.** Immunohistochemical analysis of hnRNPA2/B1 (red) in a normal muscle sample (**d**), a

muscle biopsy from patient II5 from family 1 (**e**), a muscle biopsy from patient IV9 from family 2 (**f**). hnRNPA2/B1 (red) was cleared from DAPI-stained myonuclei (blue) and accumulated in large sarcoplasmic inclusions in patients (**e**, **f**). **g**. Immunohistochemical analysis of hnRNPA2/B1 and TDP-43 colocalization in a muscle biopsy from patient IV9 from family 2. DAPI (blue), TDP-43 (green), and hnRNPA2/B1 (red) are shown. hnRNPA2/B1 (red) and TDP-43 (green) were partially colocalizing (arrows) in an atrophic muscle fiber. **h-i**. Immunohistochemical analysis of hnRNPA2/B1 in a patient with MSP caused by VCP mutation (R155H) (**h**), and a patient with sporadic IBM (**i**). **j**. Immunohistochemical analysis of hnRNPA1 in a muscle biopsy from patient IV9 from family 2. **k**. Immunohistochemical analysis of hnRNPA1 and TDP-43 colocalization in a muscle biopsy from patient IV9 from family 2. DAPI (blue), TDP-43 (green), and hnRNPA1 (red) are shown. hnRNPA1 (red) and TDP-43 (green) were partially colocalizing (arrows) in an atrophic muscle fiber. **l-m**. Immunohistochemical analysis of p62 in a normal muscle sample (**l**) and a muscle biopsy from patient II5 from family 1 (**m**). p62 was accumulated in large sarcoplasmic inclusions in patient (**m**). **n-o**. Immunohistochemical analysis of ubiquitin and hnRNPA1 (**n**) and hnRNPA2/B1 (**o**) in a muscle biopsy from patient IV9 from family 2. In both cases, limited immunopositivity of ubiquitin was observed in a large sarcoplasmic inclusions (arrows). **p**. Immunohistochemical analysis of hnRNPA2/B1 and TDP-43 colocalization in a patient with sporadic IBM. hnRNPA2/B1 (red) and TDP-43 (green) were partially colocalizing. **q**. No hnRNPA2/B1 pathology was observed in a patient with limb-girdle muscular dystrophy.

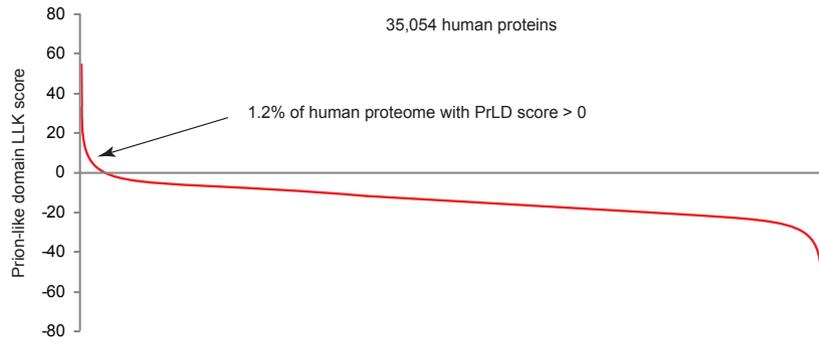


Figure 5. Examination of the human proteome (35,054 RefSeq proteins, including variant isoforms) identified possible PrLDs in approximately 1.2 %, according to the algorithm of Alberti et al. The list of possible PrLD-containing proteins was rich in proteins that regulate RNA metabolism.

Gene Name	PrLD Rank Among All Proteins	PrLD Score	Comment
FUS	13	37.6	ALS, FTD
TAF15	22	33.2	ALS, FTD
EWSR1	25	32.4	ALS, FTD
HNRPDL	28	31.5	
HNRNPD	30	30.6	
HNRNPA2B1	32	29.9	Family 1
HNRNPA1	38	28.2	Family 2
HNRNPAB	39	27.3	
HNRNPA3	41	27.2	
TARDBP	43	26.5	ALS, FTD, IBM
TIA1	55	23.2	
HNRNPA1L2	57	22.8	
HNRNPH1	63	22.3	
AC021224.2	68	21.6	
SFPQ	79	20.8	
HNRNPA0	81	20.6	
HNRNPH2	98	17.5	
RBM14	117	16.1	
CSTF2	122	15.7	
AC021593.2	124	15.4	
DAZ3	136	14.6	
DAZ1	143	14.1	
DAZ2	143	14.1	
DAZ4	143	14.1	
CSTF2T	148	14	
HNRNPH3	147	14	
TIAL1	158	13.5	
RBM33	172	12.9	
CELF4	176	12.8	
DAZAP1	198	11.7	
PSPC1	226	10	

Table 3. Predicted PrLD containing proteins that have an RRM are shown. In addition to hnRNPA2B1 and hnRNPA1, this list included several proteins previously implicated in human degenerative proteinopathies, as indicated. This observation was also made by Couthouis et al.³

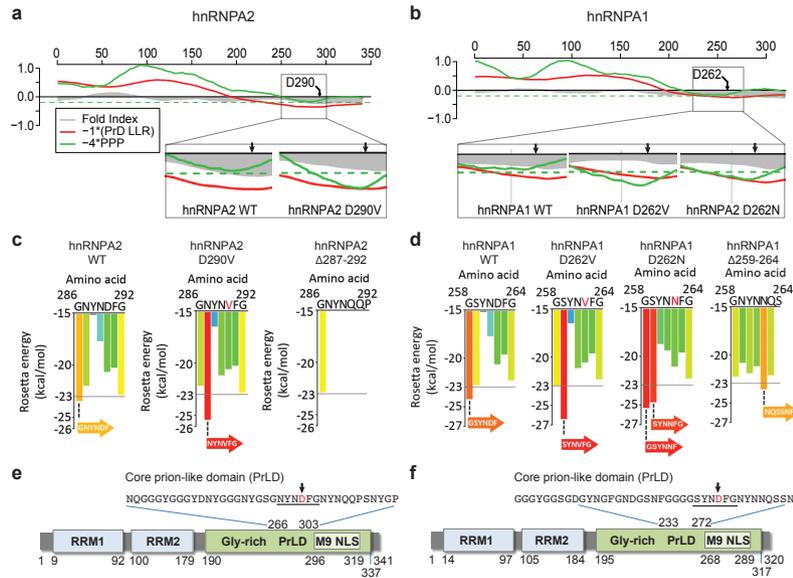


Figure 6. We examined hnRNPA2 and hnRNPA1 with ZipperDB, a structure-based threading algorithm, which scores 6-amino acid segments for their propensity to form 2 self-complementary beta strands termed “steric zippers” that form the spine of amyloid fibrils⁴. Hexapeptides with a Rosetta energy lower than -23 kcal/mol are predicted to form steric zippers, with lower energy predicting higher amyloidogenicity. ZipperDB predicted that the D290V mutation in hnRNPA2 causes a mutant hexapeptide (residues 287-292: NYNVFG) to become highly amyloidogenic (Rosetta energy = -25.5 kcal/mol), whereas the wild-type peptide (NYNDFG) is not (Rosetta energy = -21.9 kcal/mol). Similarly, ZipperDB predicted that the D262V mutation in hnRNPA1 causes a mutant hexapeptide (residues 259-264: SYNDFG) to become highly amyloidogenic (Rosetta energy = -26.4 kcal/mol), whereas the wild-type peptide (GSYNDF) is not (Rosetta energy = -22.8 kcal/mol).

Intriguingly, ZipperDB predicts that the ALS-connected D262N mutation in hnRNPA1 causes two hexapeptides to become more amyloidogenic. First, the hexapeptide comprising residues 258-263 (GSYNNF) becomes more amyloidogenic (Rosetta energy = -25.3 kcal/mol) compared to the wild-type counterpart (GSYNDF), which interestingly also breaches the critical threshold (Rosetta energy = -24.2 kcal/mol). Second, the hexapeptide comprising residues amino acids 259-264 (SYNNFG) also becomes more amyloidogenic (Rosetta energy = -24.7 kcal/mol) compared to the wild-type version (SYNNFG, Rosetta energy = -22.8 kcal/mol). Thus, in all cases the disease-associated mutations introduce potent steric zipper motifs into the PrLD. Importantly, in both hnRNPA1 D262V and hnRNPA2 D290V a mutant hexapeptide is predicted by ZipperDB to have the highest fibrillization propensity in the entire PrLD. For hnRNPA1 D262N, the mutant steric zipper motifs rank second and fourth for fibrillization propensity in the PrLD. The introduction of additional potent steric zipper motifs by the disease-causing mutations in a PrLD is likely to be significant for two reasons. First, introduction of similarly potent steric zipper motifs is sufficient to force fibril formation in model proteins, such as RNase A that would not ordinarily fibrillize⁵. Second, although many (if not all) proteins harbor steric zipper motifs they are usually buried or contorted in protein structures such that they are unable to make the intermolecular contacts necessary for fibril

formation⁴. This is not the case in hnRNPA1 and hnRNPA2 as the PrLD is likely to be intrinsically disordered. Thus, the potent steric zipper motifs of mutant hnRNPA1 and hnRNPA2 are available to make intermolecular contacts and drive fibril formation. The hexapeptide 287-292 in hnRNPA2 is critical for spontaneous and seeded fibrillization of full-length WT hnRNPA2 and D290V (Fig. 4e). Likewise, residues 259-264 are essential for spontaneous and seeded fibrillization of WT hnRNPA1, D262V, and D262N (Fig. 4g). The WT versions of these hexapeptides do not fibrillize (Fig. 4a-d). However, the adjacent hexapeptides shifted by one residue towards the N-terminus in WT hnRNPA2 (amino acids 286-291: GNYNDF; Rosetta energy=-23.4kcal/mol) or in WT hnRNPA1 (amino acids 258-263: GSYNDF; Rosetta energy=-24.2kcal/mol), are also predicted to be steric zipper motifs (although not as strongly as the disease-associated mutant hexapeptides). These steric zippers likely drive assembly of WT hnRNPA2 and A1 (Fig. 3c, d). Indeed, deletion of amino acids 259-264 in hnRNPA1 and 287-292 in hnRNPA2 also eliminates these WT steric zipper motifs (Fig. 3c, d) and fibrillization (Fig. 4e, g), which suggests that no other steric zipper motifs in the WT PrLD of hnRNPA2 or A1 are capable of driving fibrillization.

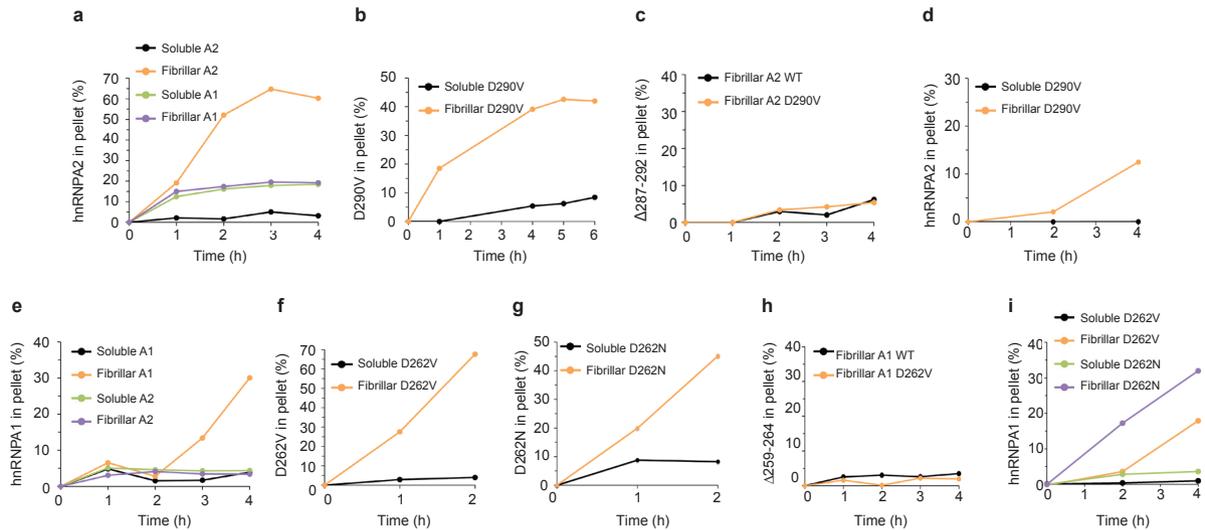


Figure 7. Preformed fibrillar form of mutant proteins can seed their wild type counterpart. **a.** hnRNPA2 (2.5 μ M) was incubated for 0-4h at 25°C with agitation (1200 rpm) in the presence of the indicated seed (5% wt/wt) comprised of either soluble hnRNPA2 (black), fibrillar hnRNPA2 (orange), soluble hnRNPA1 (green), or fibrillar hnRNPA1 (purple). At the indicated times, the amount of aggregated hnRNPA2 was determined by sedimentation analysis. A representative data set of several replicates is shown. **b.** hnRNPA2-D290V (2.5 μ M) was incubated for 0-6h at 25°C with agitation (1200 rpm) in the presence of the indicated seed (5% wt/wt) comprised of either soluble hnRNPA2-D290V (black) or fibrillar hnRNPA2 D290V (orange). At the indicated times, the amount of aggregated hnRNPA2-D290V was determined by sedimentation analysis. A representative data set of several replicates is shown. **c.** hnRNPA2 $\Delta^{287-292}$ (5 μ M) was incubated for 0-4h at 25°C with agitation (1200 rpm) in the presence of the indicated seed (5% wt/wt) comprised of either fibrillar hnRNPA2 (black) or fibrillar hnRNPA2-D290V (orange). At the indicated times, the amount of aggregated hnRNPA2 $\Delta^{287-292}$ was determined by sedimentation analysis. A representative data set of several replicates is shown. **d.** hnRNPA2 (5 μ M) was incubated for 0-4h at 25°C without agitation in the presence of the indicated seed (10% wt/wt) comprised of either soluble hnRNPA2-D290V (black) or fibrillar hnRNPA2-D290V (orange). At the indicated times, the amount of aggregated hnRNPA2 was determined by sedimentation analysis. A representative data set of several replicates is shown. **e.** hnRNPA1 was incubated for 0-4h at 25°C with agitation in the presence of the indicated soluble (black) or fibrillar (orange) hnRNPA1 seed (5% wt/wt), or soluble (green) or fibrillar (purple) hnRNPA2 seed (5% wt/wt). At various times, the amount of aggregated hnRNPA1 was determined. A representative dataset is shown. **f.** hnRNPA1-D262V was incubated for 0-2h at 25°C with agitation in the presence of the indicated soluble (black) or fibrillar (orange) hnRNPA1-D262V seed (5% wt/wt). At various times, the amount of aggregated hnRNPA1 D262V was determined. A representative dataset is shown. **g.** hnRNPA1-D262N was incubated for 0-2h at 25°C with agitation in the presence of the indicated soluble (black) or fibrillar (orange) hnRNPA1-D262N seed (5% wt/wt). At various times, the amount of aggregated hnRNPA1-D262N was determined. A representative dataset is shown. **h.** hnRNPA1 $\Delta^{259-264}$ was incubated for 0-4h at 25°C with agitation in the presence of the indicated fibrillar hnRNPA1 (black) or fibrillar hnRNPA1-D262V (orange) seed (5% wt/wt). At various times, the amount of

aggregated hnRNPA1^{Δ259-264} was determined. A representative dataset is shown. i. hnRNPA1 was incubated for 0-4h at 25°C without agitation in the presence of soluble hnRNPA1 D262V (black), soluble hnRNPA1-D262N (green), sonicated fibrillar hnRNPA1 D262V (orange), or sonicated fibrillar hnRNPA1-D262N (purple) seed (5% wt/wt). At various times, the amount of aggregated hnRNPA1 was determined. A representative dataset is shown.

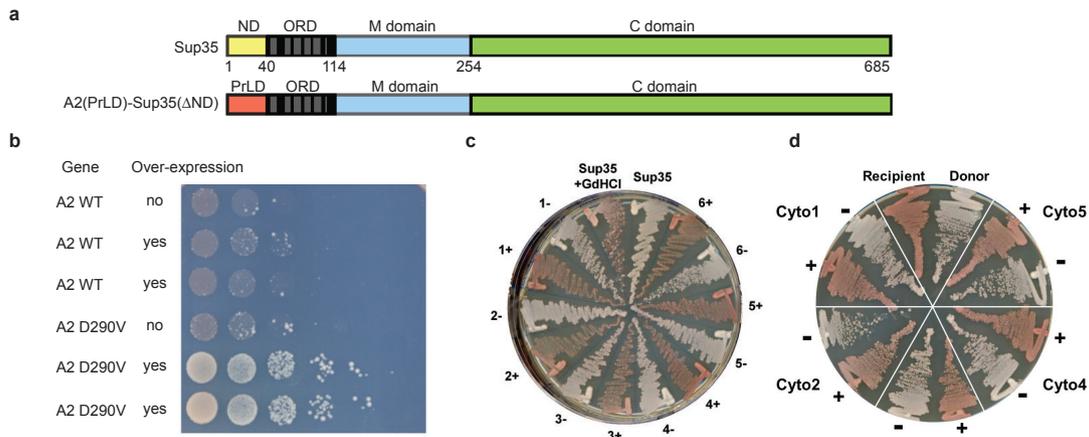


Figure 8. To determine whether the PrLD of hnRNPA2/B1 could support prion nucleation activity in yeast, we replaced the Sup35 nucleation domain (residues 3-40) with the wild-type or mutant hnRNPA2 core PrLD (residues 261-303), and expressed these fusions (referred to as A2-Sup35) as the sole copy of Sup35 in the cell (**a**). Prion formation was detected by monitoring nonsense suppression of the *ade2-1* allele. *ade2-1* mutants are unable to grow in the absence of adenine and grow red on limiting adenine. $[PSI^+]$ increases stop codon read-through, allowing *ade2-1* mutants to grow in the absence of adenine and grow white on limiting adenine. Strains expressing these fusions as the sole copy of Sup35 were transformed with either an empty vector (pKT24) or pKT24 modified to express the prion domain from the matching hnRNPA2/B1-Sup35 fusion from the GAL1 promoter. Cells were grown in galactose/raffinose dropout medium, and plated onto medium lacking adenine to select for prion formation. When cells expressing the wild-type and D290V A2-Sup35 fusions were plated on medium lacking adenine, rare colonies were formed, consistent with prion formation. One hallmark of prions is that the frequency of prion formation should be dependent on the expression level of the prion protein. Transient over-expression of the wild-type A2-Sup35 chimeric prion domain did not substantially increase Ade⁺ colony formation (**b**). By contrast over-expression of the D290V A2-Sup35 prion domain substantially increased Ade⁺ colony formation (**b**), suggesting that the D290V fusion is forming prions. To further demonstrate that the Ade⁺ phenotype observed for the D290V fusion was due to a prion formation, Ade⁺ isolates were tested for curability and stability of the Ade⁺ phenotype. Six independent Ade⁺ isolates formed by strains expressing the D290V-Sup35 fusion were grown on YPD (-) and YPD with 4 mM guanidine HCl (+), and then re-streaked on YPD to test for loss of the Ade⁺ phenotype. An identically treated $[PSI^+]$ strain expressing wild type Sup35 is included as a control. Almost all Ade⁺ isolates stably maintained the Ade⁺ phenotype when grown on non-selective medium, but lost the Ade⁺ phenotype when grown on medium containing low concentrations of guanidine HCl (**c**). Guanidine HCl cures almost all yeast prions by inhibiting Hsp104⁶⁻⁸, a chaperone required for prion propagation⁹. By contrast, none of the Ade⁺ isolates from the wild-type fusion were both stable and curable (**c**). Furthermore, the Ade⁺ phenotype in the mutants was transferable by cytoduction (**d**), a method that transfers cytoplasmic elements, but not nuclear genes¹⁰. Ade⁺ isolates from cells expressing the D290V A2-Sup35 fusion were used as donors for cytoduction reactions. Recipient cells also expressed the D290V A2-Sup35 fusion as

the sole copy of Sup35. Donor and recipient cells were streak onto YPD, along with individual cytoductants before (-) and after (+) treatment with guanidine HCl. These results demonstrate that the Ade⁺ phenotype is the result of prion formation. Collectively, these experiments show that the core D290V PrLD can substitute for the Sup35 nucleation domain in supporting prion formation, and that the D290V mutation specifically promotes the nucleation activity.

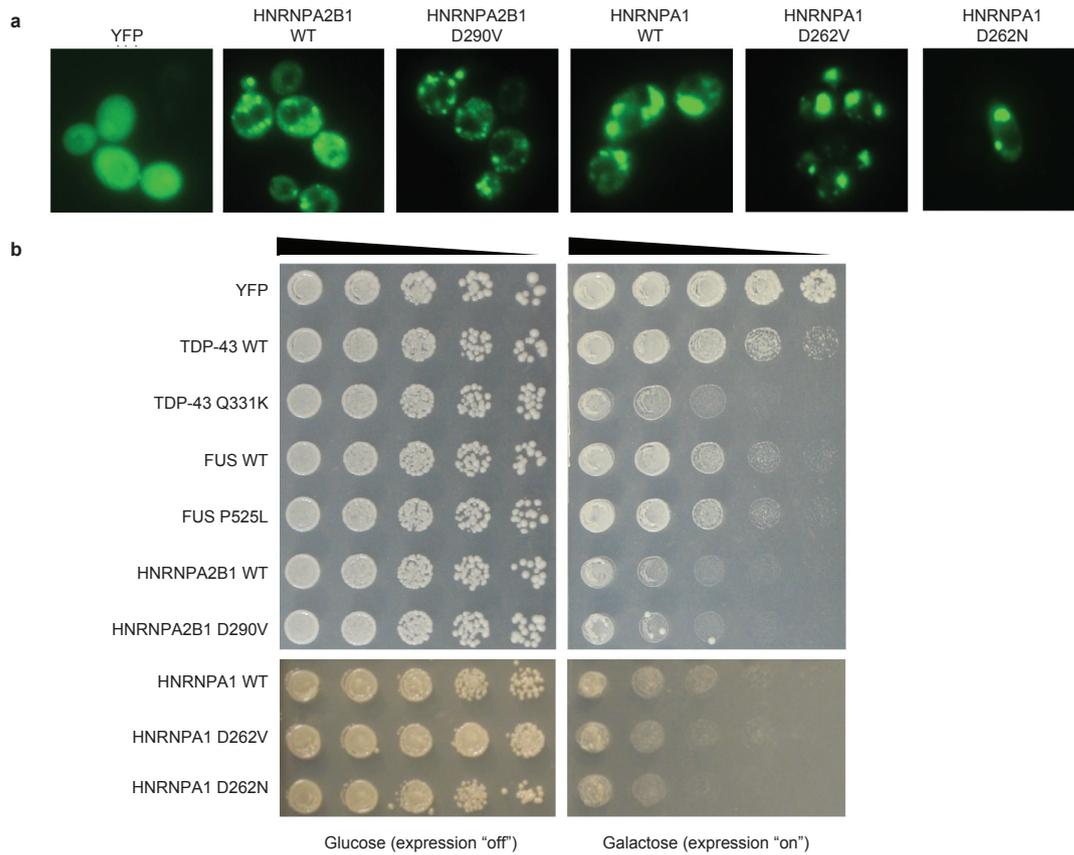


Figure 9. HnRNPA2 and hnRNPA1 aggregate in yeast and are toxic. **a.** Yeast cells expressing YFP alone or YFP-tagged WT and mutant human RBPs HNRNPA2B1 and HNRNPA1. YFP was diffusely localized throughout the cytosol and nucleus, whereas wild-type and mutant hnRNPA2 and hnRNPA1 formed multiple cytoplasmic foci when expressed in yeast. The mutant forms of hnRNPA2 and hnRNPA1 formed similar numbers of cytoplasmic foci as WT. **b.** HNRNPA2 and HNRNPA1 are toxic when expressed in yeast. Yeast spotting assays comparing toxicity of TDP-43, FUS, hnRNPA2, and hnRNPA1. WT TDP-43 is toxic when expressed in yeast and an ALS-linked mutation (Q331K) increases toxicity. WT FUS is also toxic when expressed in yeast and ALS-linked mutations (e.g. P525L) do not increase toxicity. WT HNRNPA2 and HNRNPA1 are very toxic when expressed in yeast and the MSP-associated mutant forms do not enhance toxicity in this assay. Because of the galactose-inducible promoter, expression of these RBPs is repressed when cells are grown in the presence of glucose (left panel, expression "off") and induced when grown in the presence of galactose (right panel, FUS expression "on").

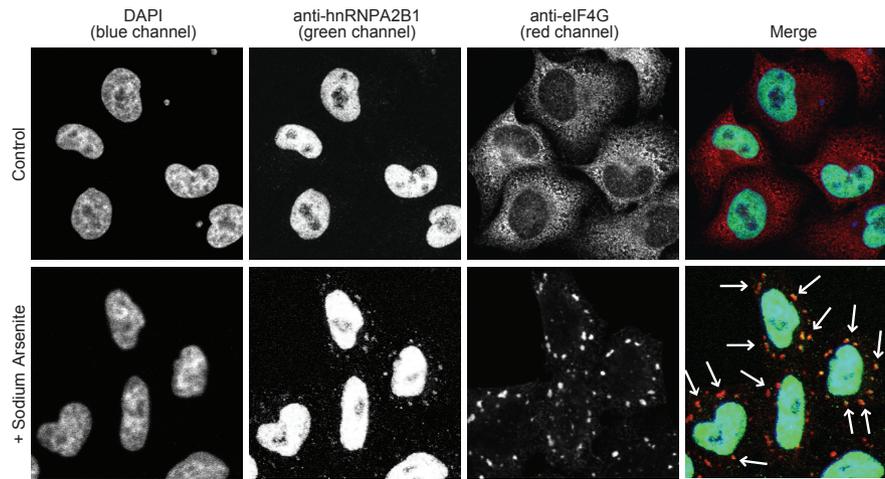


Figure 10. Endogenous hnRNPA2B1 is recruited to SGs following treatment with arsenite. HeLa cells were treated with 0.5 mM sodium arsenite for 1h and stained with anti-Flag (green), anti-eIF4G (red), and DAPI (blue).

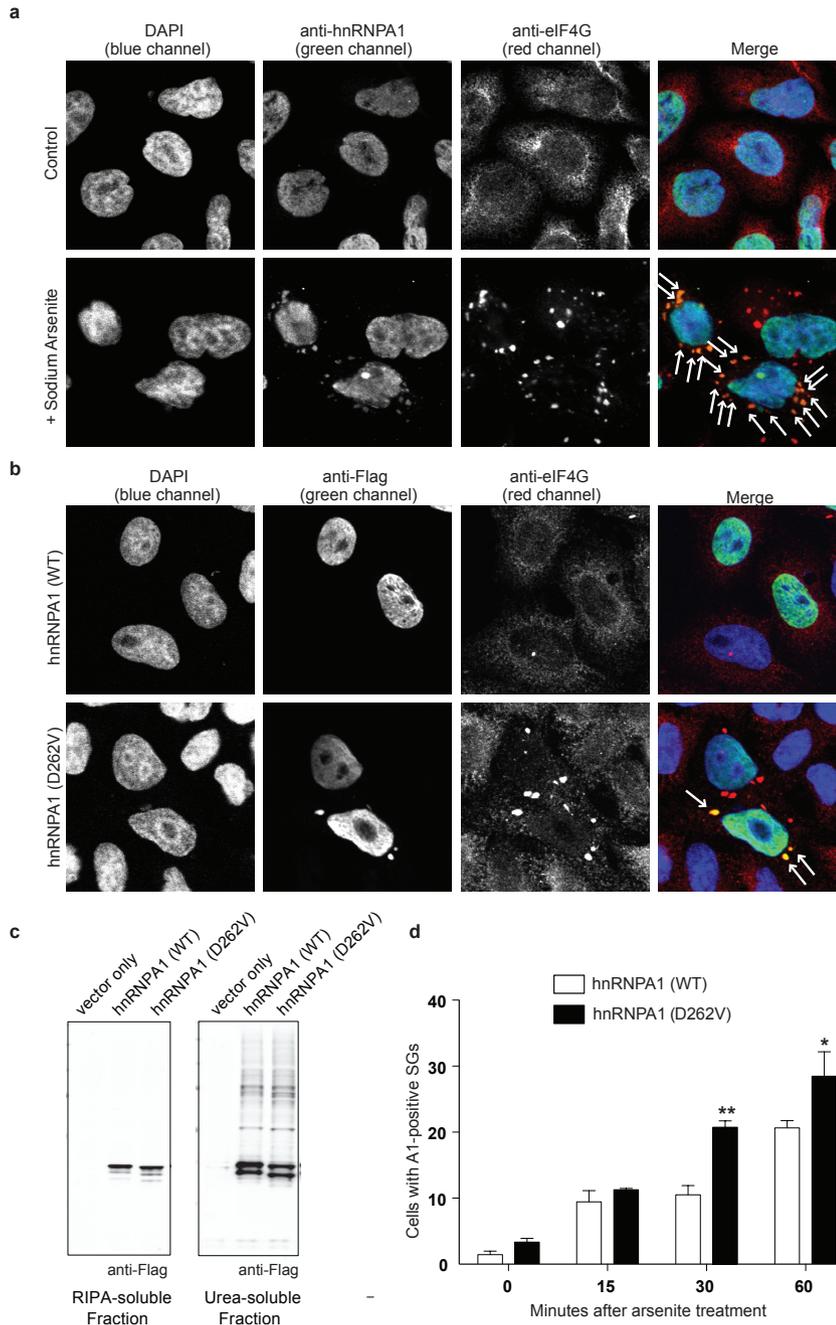


Figure 11. hnRNPA1 is recruited to SGs and this is augmented by disease-causing mutation. **a.** Endogenous hnRNPA1 was recruited to SGs following treatment with arsenite. **b.** HeLa cells were transfected with either the Flag-wild-type or Flag-D262V mutant form of hnRNPA1 and stained with anti-Flag (green), anti-eIF4G (red), and DAPI (blue). Arrows indicate hnRNPA1- and eIF4G-positive SGs. **c.** Immunoblot of hnRNPA1 in the RIPA-soluble and RIPA-insoluble (Urea-soluble) fractions from transfected HeLa cells. **d.** Quantification of the number of cells displaying hnRNPA1-positive SGs. The percentage of cells displaying hnRNPA1-positive SGs at indicated time points following treatment with arsenite cells are plotted. Data represent mean \pm SEM of at least 3 independent experiments (* $P < 0.05$, ** $P < 0.01$). P values are compared with wild type control.

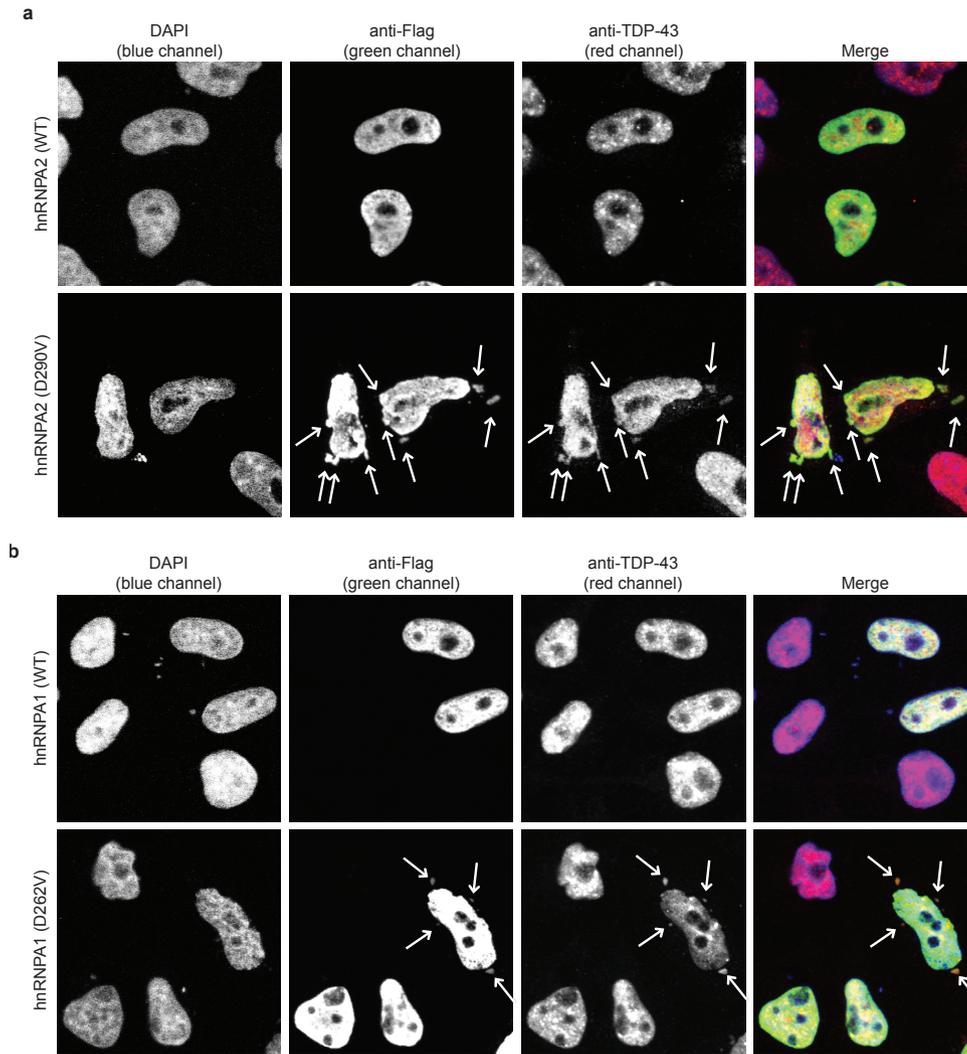


Figure 12. Colocalization of hnRNPA2 and hnRNPA1 with TDP-43 in SGs. **a.** HeLa cells were transfected with either Flag-wild-type or Flag-D290V mutant form of hnRNPA2 and stained with anti-Flag (green), anti-TDP-43 (red), and DAPI (blue). Arrows indicate hnRNPA2 and TDP-43 positive SGs. **b.** HeLa cells were transfected with either Flag-wild-type or Flag-D262V mutant form of hnRNPA1 and experiments are performed as in **a.**

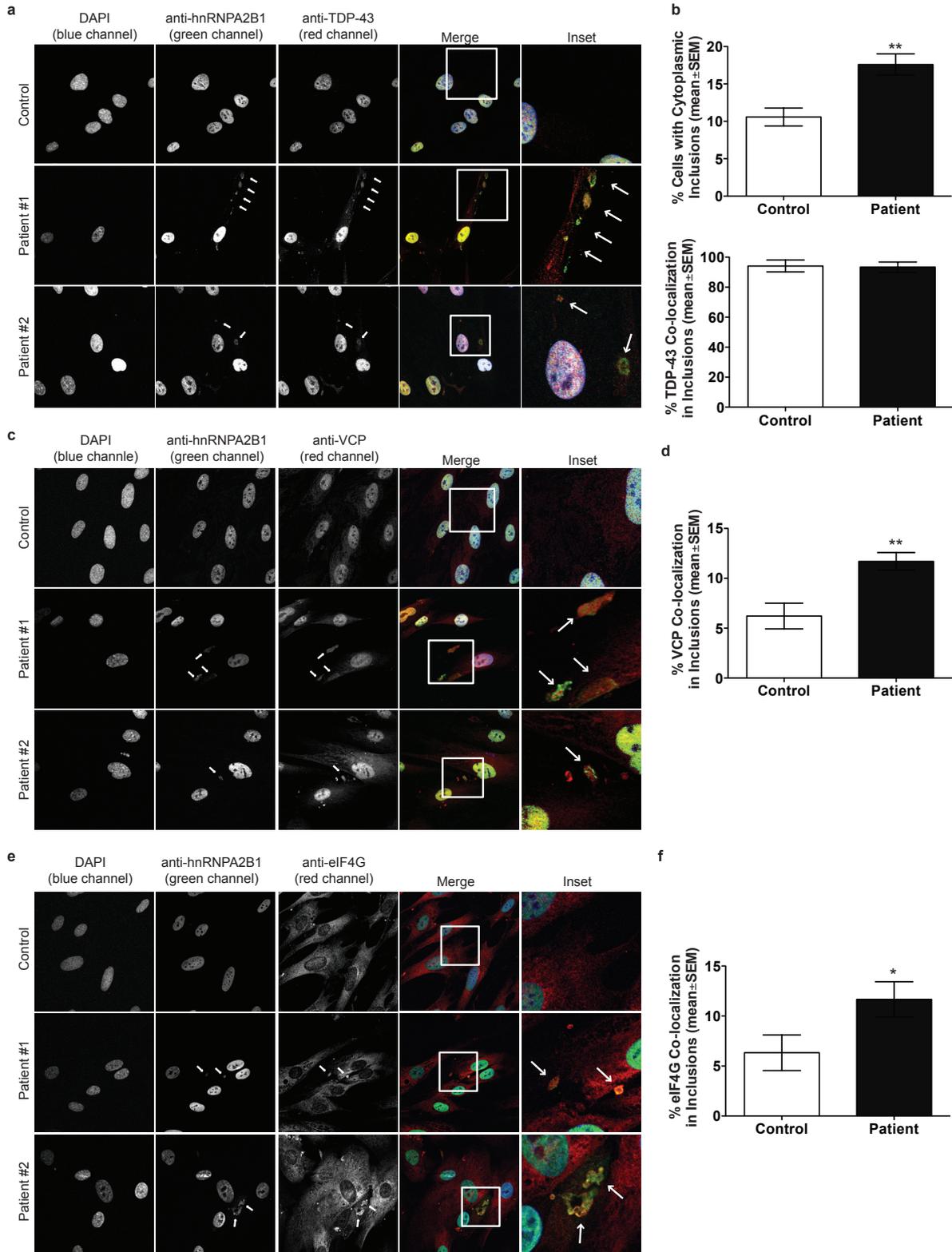


Figure 13. hnRNPA2B1 localization in control and patients fibroblasts. **a.** Control and patients fibroblasts were stained with anti-hnRNPA2B1 (green), anti-TDP-43 (red), and DAPI (blue). Arrows indicate cytoplasmic granules displaying colocalization of hnRNPA2B1 with TDP-43. **b.** Percentage of cells

displaying hnRNPA2B1-positive cytoplasmic inclusions (upper panel), and percentage of cells displaying hnRNPA2B1 and TDP-43 colocalization in inclusions (number of cells with hnRNPA2B1+, TDP-43+ in inclusions / number of cells with hnRNPA2/B1+ in inclusions, lower panel) in control and patients fibroblasts. **c.** Control and patient fibroblasts were stained with anti-hnRNPA2B1 (green), anti-VCP (red), and DAPI (blue). Arrows indicate cytoplasmic granules displaying colocalization of hnRNPA2 with VCP. **d.** Percentage of cells displaying hnRNPA2B1 and VCP colocalization in inclusions (number of cells with hnRNPA2B1+, VCP+ in inclusions / total number of cells) in control and patients fibroblasts. **e.** Control and patients fibroblasts were stained with anti-hnRNPA2B1 (green), anti-eIF4G (red), and DAPI (blue). Arrows indicate cytoplasmic granules displaying colocalization of hnRNPA2 with eIF4G. **f.** Percentage of cells displaying hnRNPA2B1 and eIF4G colocalization in inclusions (number of cells with hnRNPA2B1+, eIF4G+ in inclusions / total number of cells) in control and patients fibroblasts. Data represent mean \pm SEM of at least 3 independent experiments (* $P < 0.05$, ** $P < 0.01$). P values are compared with wild type control.

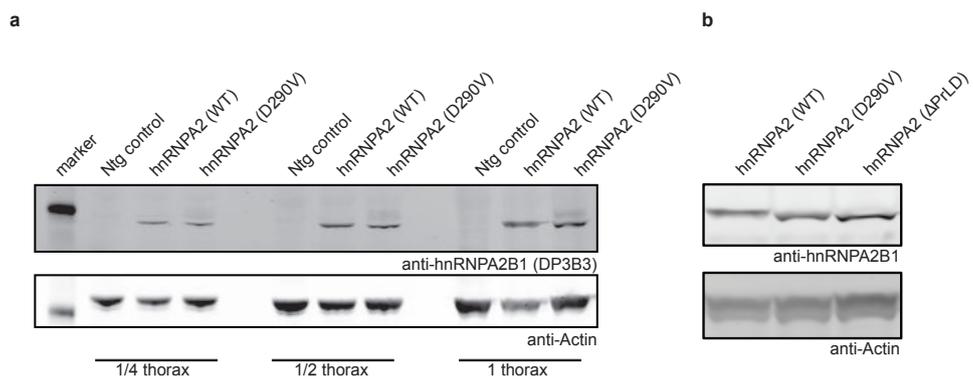


Figure 14. Comparable levels of hnRNPA2 expression transgenic flies. **a-b.** Thoraces of adult flies were processed for Western blot analysis with an antibody against hnRNPA2/B1. Actin was blotted as a loading control.

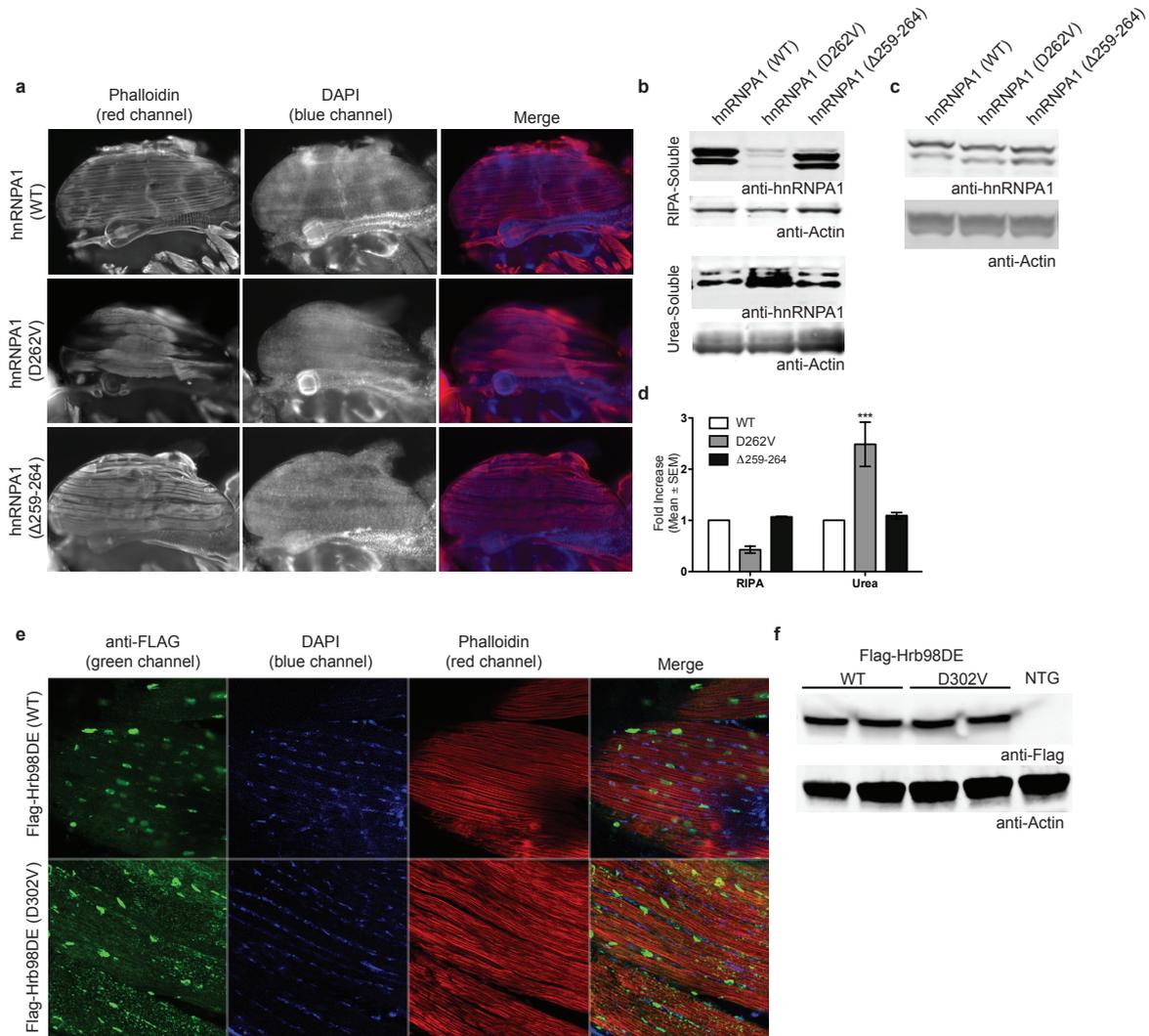


Figure 15. Mutant hnRNPA1 and Hrb98DE accumulate in pathological inclusions in *Drosophila*. **a.** Adult flies were dissected to expose the dorsal longitudinal indirect fly muscle (DLM) and stained with Texas Red-phalloidin (red), and DAPI (blue). **b.** Thoraces of adult flies were dissected. Sequential extractions were performed to examine the solubility profile of hnRNPA1. **c.** Thoraces of adult flies were processed for Western blot analysis with an antibody against hnRNPA1. **d.** Quantification of the blot shown in **b.** *** $p < 0.001$. P values are compared with wild type (WT). **e.** DLM of Adult flies were dissected and stained with anti-Flag (green), Texas Red-phalloidin (red), and DAPI (blue). **f.** Thoraces of adult flies were processed for Western blot analysis with an antibody against Flag.

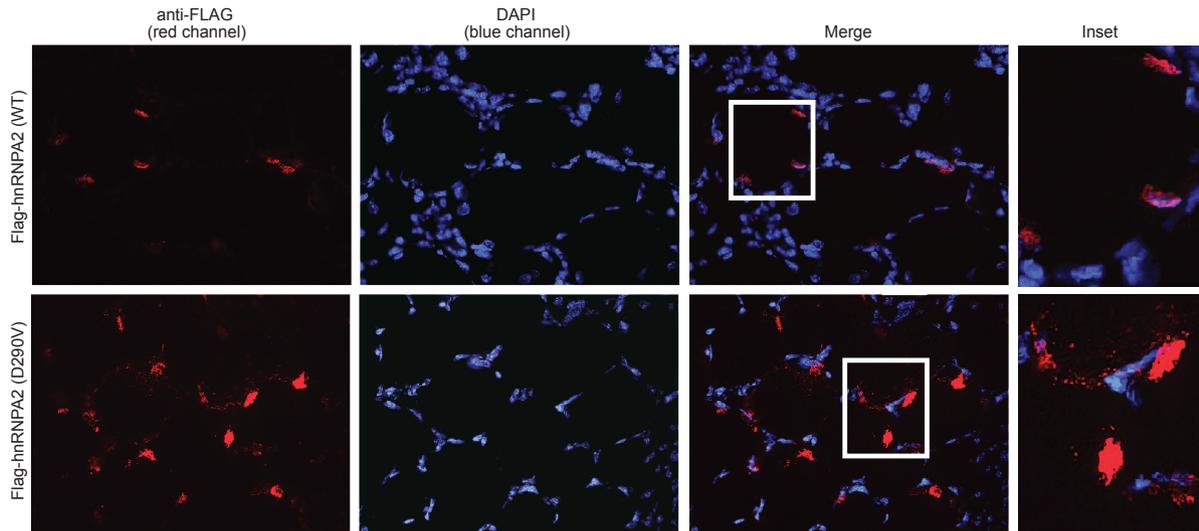


Figure 16. Mutant hnRNPA2 accumulates in pathological inclusions in mouse muscle. Mouse tibialis anterior was electroporated with plasmid expressing either wild-type Flag-hnRNPA2 or mutant Flag-hnRNP (D290V). After a 7-day recovery, wild-type hnRNPA2 was found to localize exclusively to myonuclei whereas mutant hnRNPA2 (D290V) also accumulated extensively in sarcoplasmic aggregates.

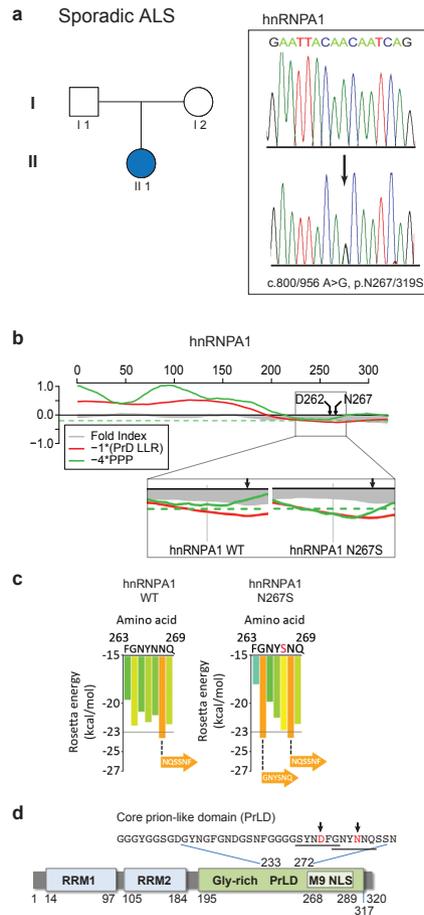


Figure 17. The disease-causing mutation impact a PrLD in hnRNPA1 in sporadic ALS. **a.** We screened 305 sporadic ALS cases and identified a nonsynonymous variant in hnRNPA1 (c.800/956A>G, p.N267/319S) in one classic, late-onset case in which mutations in known ALS genes had been excluded (c9orf72, SOD1, TARDBP, FUS and PNF1). **b.** PrLD prediction for hnRNPA1. The panel shows the FoldIndex prediction of an extended intrinsically unfolded region (grey curve less than zero) in the C-terminus of hnRNPA1. Part of this region was also predicted to be prion-like according to the algorithm of Alberti et al. (red curve less than zero), and narrowly missed the cutoff for the prion propensity algorithm of Toombs et al. (green curve below the dashed green line). All curves represent averages of 41 consecutive windows of 41 amino acids, corresponding to the criteria of Toombs et al. The N267S mutation in hnRNPA1 was predicted to make this region more prionogenic, with a core PrLD that now satisfied all 3 conditions. **c.** ZipperDB detected a 6-amino-acid stretch within the core PrLD for which the N267S mutation increased the predicted amyloid fibril-forming potential beyond the Rosetta threshold. **d.** Domain architecture of hnRNPA1 shows two RNA-recognition motifs, RRM1 and 2, a C-terminal glycine-rich domain, and an M9 nuclear localization signal. The predicted prion-like domain is within the C-terminal glycine-rich domain.

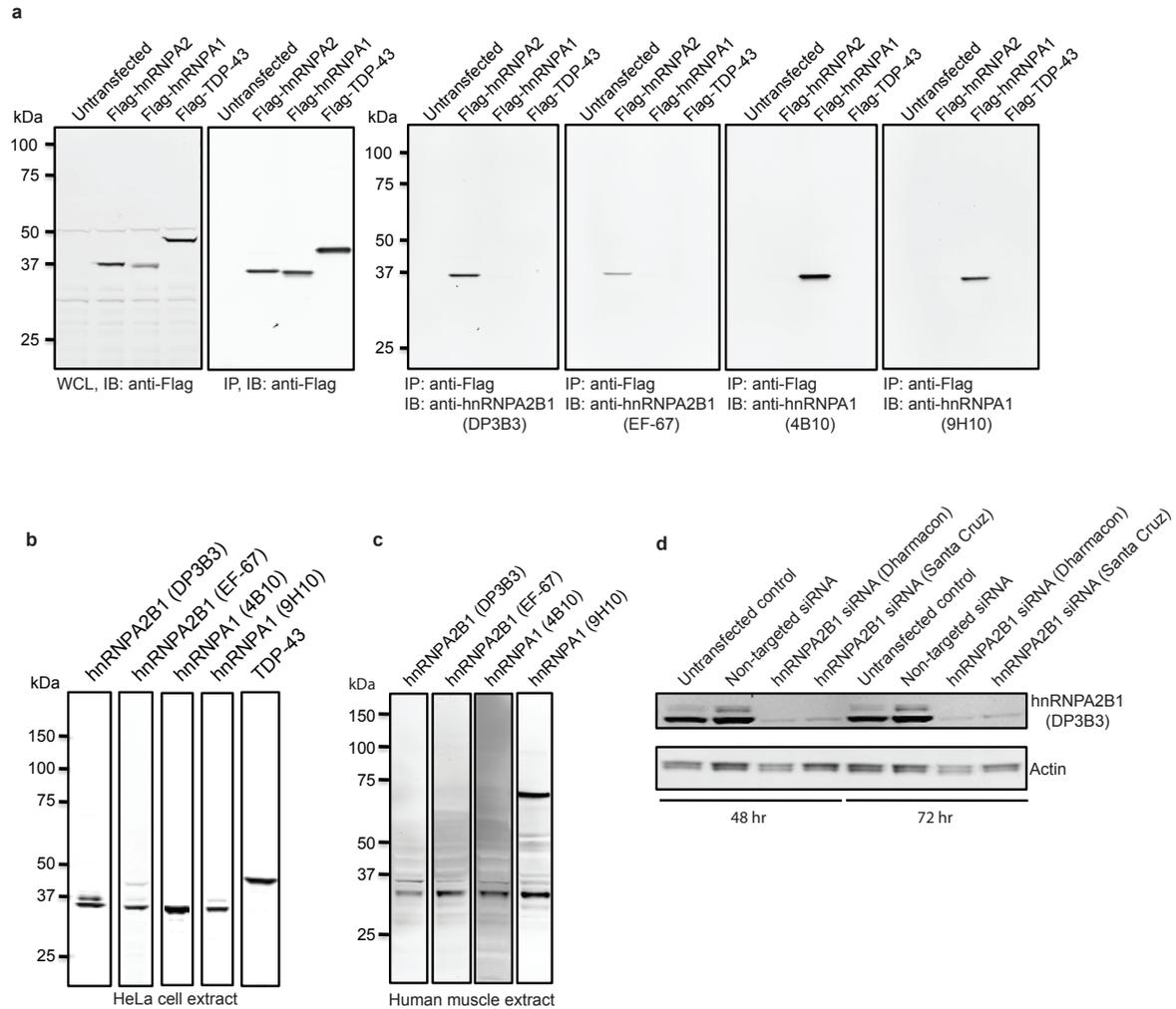


Figure 18. Validation of antibodies used in the experiments. **a.** HeLa cells were transfected with Flag-hnRNP2, Flag-hnRNP1 or Flag-TDP-43 as indicated. Immunoprecipitation (IP) was performed with anti-Flag antibody and immunoblotting (IB) was performed with antibodies indicated. **b.** HeLa cell lysates were immunoblotted with antibodies indicated. **c.** Human muscle extract was immunoblotted with antibodies indicated. **d.** HeLa cells were transfected with hnRNP2B1 specific siRNA. Two and three days post transfection, cell lysates were immunoblotted with hnRNP2B1 (DP3B3) antibody.

References

- 1 Waggoner, B. *et al.* Heterogeneity in familial dominant Paget disease of bone and muscular dystrophy. *Am J Med Genet* 108, 187-191, (2002).
- 2 Ritson, G. P. *et al.* TDP-43 mediates degeneration in a novel *Drosophila* model of disease caused by mutations in VCP/p97. *J Neurosci* 30, 7729-7739, (2010).
- 3 Couthouis, J. *et al.* A yeast functional screen predicts new candidate ALS disease genes. *Proc Natl Acad Sci U S A* 108, 20881-20890, (2011).
- 4 Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci U S A* 107, 3487-3492, (2010).
- 5 Teng, P. K. & Eisenberg, D. Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng Des Sel* 22, 531-536, (2009).
- 6 Ferreira, P. C., Ness, F., Edwards, S. R., Cox, B. S. & Tuite, M. F. The elimination of the yeast [PSI⁺] prion by guanidine hydrochloride is the result of Hsp104 inactivation. *Mol Microbiol* 40, 1357-1369, (2001).
- 7 Jung, G. & Masison, D. C. Guanidine hydrochloride inhibits Hsp104 activity in vivo: a possible explanation for its effect in curing yeast prions. *Curr Microbiol* 43, 7-10, (2001).
- 8 Tuite, M. F., Mundy, C. R. & Cox, B. S. Agents that cause a high frequency of genetic change from [psi⁺] to [psi⁻] in *Saccharomyces cerevisiae*. *Genetics* 98, 691-711, (1981).
- 9 Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi⁺]. *Science* 268, 880-884, (1995).
- 10 Conde, J. & Fink, G. R. A mutant of *Saccharomyces cerevisiae* defective for nuclear fusion. *Proc Natl Acad Sci U S A* 73, 3651-3655, (1976).