

RESEARCH ARTICLE

Open Access



Natural and pathogenic protein sequence variation affecting prion-like domains within and across human proteomes

Sean M. Cascarina and Eric D. Ross*

Abstract

Background: Impaired proteostatic regulation of proteins with prion-like domains (PrLDs) is associated with a variety of human diseases including neurodegenerative disorders, myopathies, and certain forms of cancer. For many of these disorders, current models suggest a prion-like molecular mechanism of disease, whereby proteins aggregate and spread to neighboring cells in an infectious manner. The development of prion prediction algorithms has facilitated the large-scale identification of PrLDs among “reference” proteomes for various organisms. However, the degree to which intraspecies protein sequence diversity influences predicted prion propensity has not been systematically examined.

Results: Here, we explore protein sequence variation introduced at genetic, post-transcriptional, and post-translational levels, and its influence on predicted aggregation propensity for human PrLDs. We find that sequence variation is relatively common among PrLDs and in some cases can result in relatively large differences in predicted prion propensity. Sequence variation introduced at the post-transcriptional level (via alternative splicing) also commonly affects predicted aggregation propensity, often by direct inclusion or exclusion of a PrLD. Finally, analysis of a database of sequence variants associated with human disease reveals a number of mutations within PrLDs that are predicted to increase prion propensity.

Conclusions: Our analyses expand the list of candidate human PrLDs, quantitatively estimate the effects of sequence variation on the aggregation propensity of PrLDs, and suggest the involvement of prion-like mechanisms in additional human diseases.

Keywords: Prion-like domains, Sequence variation, Protein aggregation, Prion, Prion prediction, Neurodegenerative disease

Background

Prions are infectious proteinaceous elements, most often resulting from the formation of self-replicating protein aggregates. A key component of protein aggregate self-replication is the acquired ability of aggregates to catalyze the conversion of identical proteins to the non-native, aggregated form. Although prion phenomena may occur in a variety of organisms, budding yeast has been used extensively as a model organism to study the relationship between protein sequence and prion activity [1–4]. Prion domains from yeast prion proteins tend to

share a number of unusual compositional features, including high glutamine/asparagine (Q/N) content and few charged and hydrophobic residues [2, 3]. Furthermore, the amino acid composition of these domains (rather than primary sequence) is the predominant feature conferring prion activity [5, 6]. This observation has contributed to the development of a variety of composition-centric prion prediction algorithms designed to identify and score proteins based on sequence information alone [7–13].

Many of these prion prediction algorithms were extensively tested and validated in yeast as well. For example, multiple yeast proteins with experimentally-demonstrated prion activity were first identified as high-scoring prion candidates by early prion prediction algorithms [9–11]. Synthetic prion domains, designed in silico using the

* Correspondence: Eric.Ross@colostate.edu
Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523, USA



Prion Aggregation Prediction Algorithm (PAPA), exhibited bona fide prion activity in yeast [14]. Additionally, application of these algorithms to proteome sequences for a variety of organisms has led to a number of important discoveries. The first native bacterial PrLDs with demonstrated prion activity in bacteria (albeit in an unrelated bacterial model organism) were also initially identified using leading prion prediction algorithms [15, 16]. A prion prediction algorithm was used in the initial identification of a PrLD from the model plant organism *Arabidopsis thaliana* [17], and this PrLD was shown to aggregate and propagate as a prion in yeast (though it is currently unclear whether it would also have prion activity in its native host). Similarly, multiple prion prediction algorithms applied to the *Drosophila* proteome identified a prion-like domain with bona fide prion activity in yeast [18]. A variety of PrLD candidates have been identified in eukaryotic virus proteomes using prion prediction algorithms [19], and one viral protein was recently reported to behave like a prion in eukaryotic cells [20]. These examples represent vital advances in our understanding of protein features conferring prion activity, and illustrate the broad utility of prion prediction algorithms.

Some prion prediction algorithms may even have complementary strengths: identification of PrLD candidates with the first generation of the Prion-Like Amino Acid Composition (PLAAC) algorithm led to the discovery of new prions [11], while application of PAPA to this set of candidate PrLDs markedly improved the discrimination between domains with and without prion activity in vivo [7, 14]. Similarly, PLAAC identifies a number of PrLDs within the human proteome, and aggregation of these proteins is associated with an assortment of muscular and neurological disorders [21–34]. In some cases, increases in aggregation propensity due to single amino acid substitutions are accurately predicted by multiple aggregation prediction algorithms, including PAPA [33, 35]. Furthermore, the effects of a broad range of mutations within PrLDs expressed in yeast can also be accurately predicted by PAPA and other prion prediction algorithms, and these predictions generally extend to multicellular eukaryotes, albeit with some exceptions [36, 37]. The complementary strengths of PLAAC and PAPA are likely derived from their methods of development. The PLAAC algorithm identifies PrLD candidates by compositional similarity to domains with known prion activity, but penalizes all deviations in composition (compared to the training set) regardless of whether these deviations enhance or diminish prion activity. PAPA was developed by randomly mutagenizing a canonical Q/N-rich yeast prion protein (Sup35) and directly assaying the frequency of prion formation, which was used to quantitatively estimate of the prion propensity of each of the 20 canonical

amino acids. Therefore, PLAAC seems to be effective at successfully identifying PrLD candidates, while PAPA is ideally-suited to predict which PrLD candidates are most likely to have true prion activity, and how changes in PrLD sequence might affect prion activity.

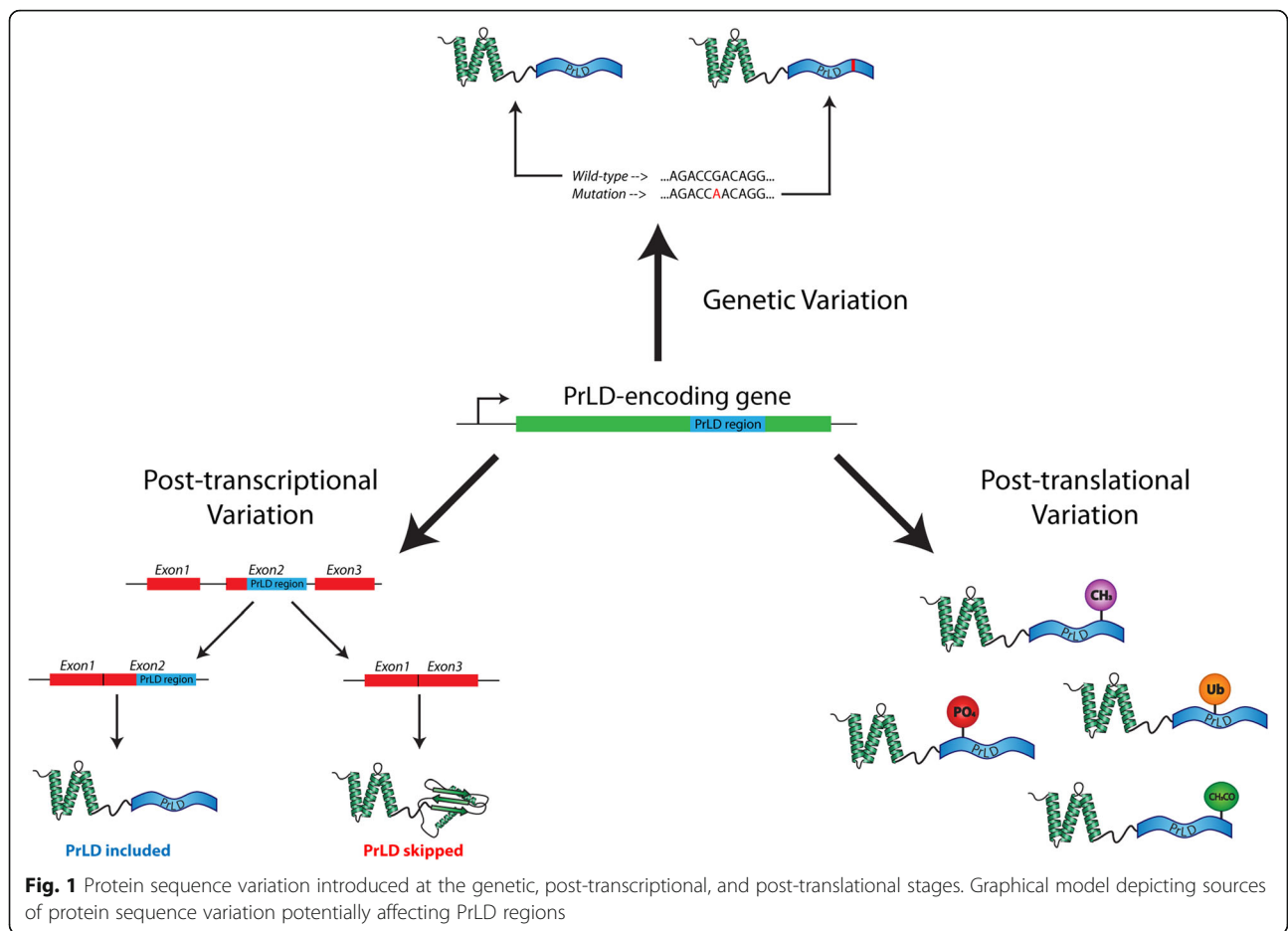
To date, most proteome-scale efforts of prion prediction algorithms have focused on the identification of PrLDs within reference proteomes (i.e. a representative set of protein sequences for each organism). However, reference proteomes do not capture the depth and richness of protein sequence variation that may affect PrLDs *within* a species. Here, we explore the depth of intraspecies protein sequence variation affecting human PrLDs at the genetic, post-transcriptional, and post-translational stages (Fig. 1). We estimate the range of aggregation propensity scores resulting from known protein sequence variation, for all high scoring PrLDs. To our surprise, aggregation propensity ranges are remarkably large, suggesting that natural sequence variation could potentially result in large inter-individual differences in aggregation propensity for certain proteins. Furthermore, we define a number of proteins whose aggregation propensities are affected by alternative splicing or pathogenic mutation. In addition to proteins previously linked to prion-like disorders, we identify a number of high-scoring PrLD candidates whose predicted aggregation propensity increases for certain isoforms or upon mutation, and some of these candidates are associated with prion-like behavior in vivo yet are not currently classified as “prion-like”. Finally, we provide comprehensive maps of PTMs within human PrLDs derived from a recently-collated PTM database.

Results

Sequence variation in human PrLDs leads to wide ranges in estimated aggregation propensity

Multiple prion prediction algorithms have been applied to specific reference proteomes to identify human PrLDs [8, 13, 38–41]. While these predictions provide important baseline maps of PrLDs in human proteins, they do not account for the considerable diversity in protein sequences across individuals. In addition to the ~ 42 k unique protein isoforms (spanning ~ 20 k protein-encoding genes) represented in standard human reference proteomes, the human proteome provided by the neXtProt database includes >6 million annotated single amino acid variants [42]. Importantly, these variants reflect the diversity of human proteins, and allow for the exploration of additional sequence space accessible to human proteins.

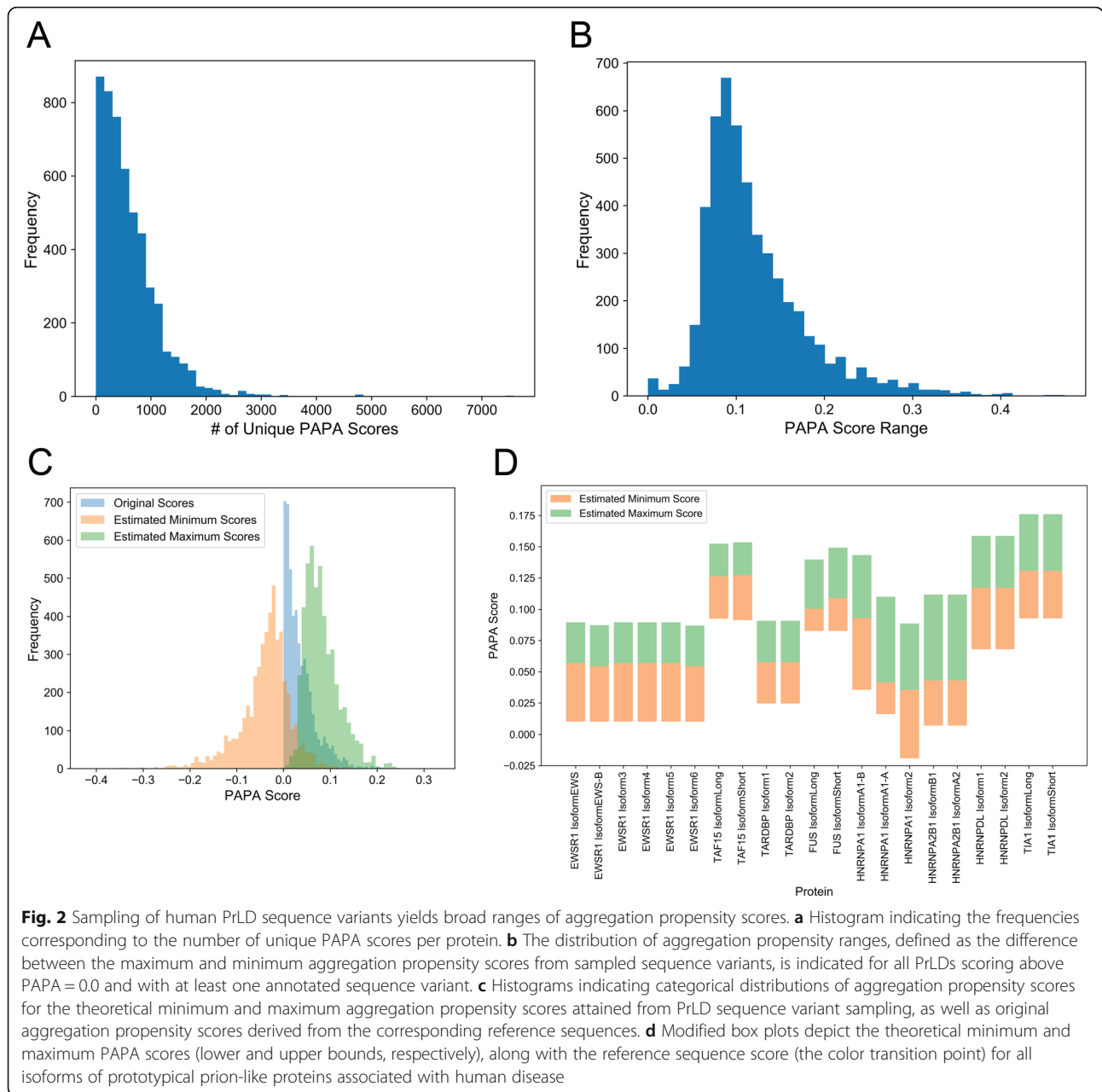
The majority of known variants in human coding sequences are rare, occurring only once in a dataset of ~ 60,700 human exomes [43]. However, the frequency of multiple-variant co-occurrence for each possible variant combination in a single individual has not been



quantified on a large scale. Theoretically, the frequency of rare variants would result in each pairwise combination of rare variants occurring in a single individual only a few times in the current human population. We emphasize that this is only a rough estimate, as it assumes independence in the frequency of each variant, and that the observed frequency of rare variants corresponds to the actual population frequency.

With these caveats in mind, we applied a modified version of our Prion Aggregation Prediction Algorithm (PAPA; see Methods for modifications and rationale) to the human proteome reference sequences to obtain baseline aggregation propensity scores and to identify relatively high-scoring PrLD candidates. Since sequence variants could increase predicted aggregation propensity, we employed a conservative aggregation propensity threshold (PAPA score ≥ 0.0) to define high-scoring PrLD candidates ($n = 5173$ unique isoforms). Nearly all PrLD candidates ($n = 5065$; 97.9%) have at least one amino acid variant within the PrLD region that influenced the PAPA score. Protein sequences for all pairwise combinations of known protein sequence variants were computationally generated for all proteins with moderately

high-scoring PrLDs (>20million variant sequences, derived from the 5173 protein isoforms with PAPA score ≥ 0.0). While most proteins had relatively few variants that influenced predicted aggregation propensity scores, a number of proteins had > 1000 unique PAPA scores, indicating that PrLDs can be remarkably diverse (Fig. 2a). To estimate the overall magnitude of the effects of PrLD sequence variation, the PAPA score range was calculated for each set of variants (i.e. for all variants corresponding to a single protein). PAPA score ranges adopt a right-skewed distribution, with a median PAPA score range of 0.10 (Fig. 2b, c; Additional file 1). Importantly, the estimated PAPA score range for a number of proteins exceeds 0.2, indicating that sequence variation can have a dramatic effect on predicted aggregation propensity (by comparison, the PAPA score range = 0.92 for the entire human proteome). Additionally, we examined the aggregation propensity ranges of prototypical prion-like proteins associated with human disease [21–25, 27–34], which are identified as high-scoring candidates by both PAPA and PLAAC. In most cases, the lowest aggregation propensity estimate derived from sequence variant sampling scored well-below the classical aggregation threshold



(PAPA score = 0.05), and the highest aggregation propensity estimate scored well-above the aggregation threshold (Fig. 2d). Furthermore, for a subset of prion-like proteins (FUS and hnRNPA1), aggregation propensity scores derived from the initial reference sequences differed considerably for alternative isoforms of the same protein, suggesting that alternative splicing may also influence aggregation propensity. It is possible that natural genetic variation between individuals may substantially influence the prion-like behavior of human proteins.

Alternative splicing introduces sequence variation that affects human PrLDs

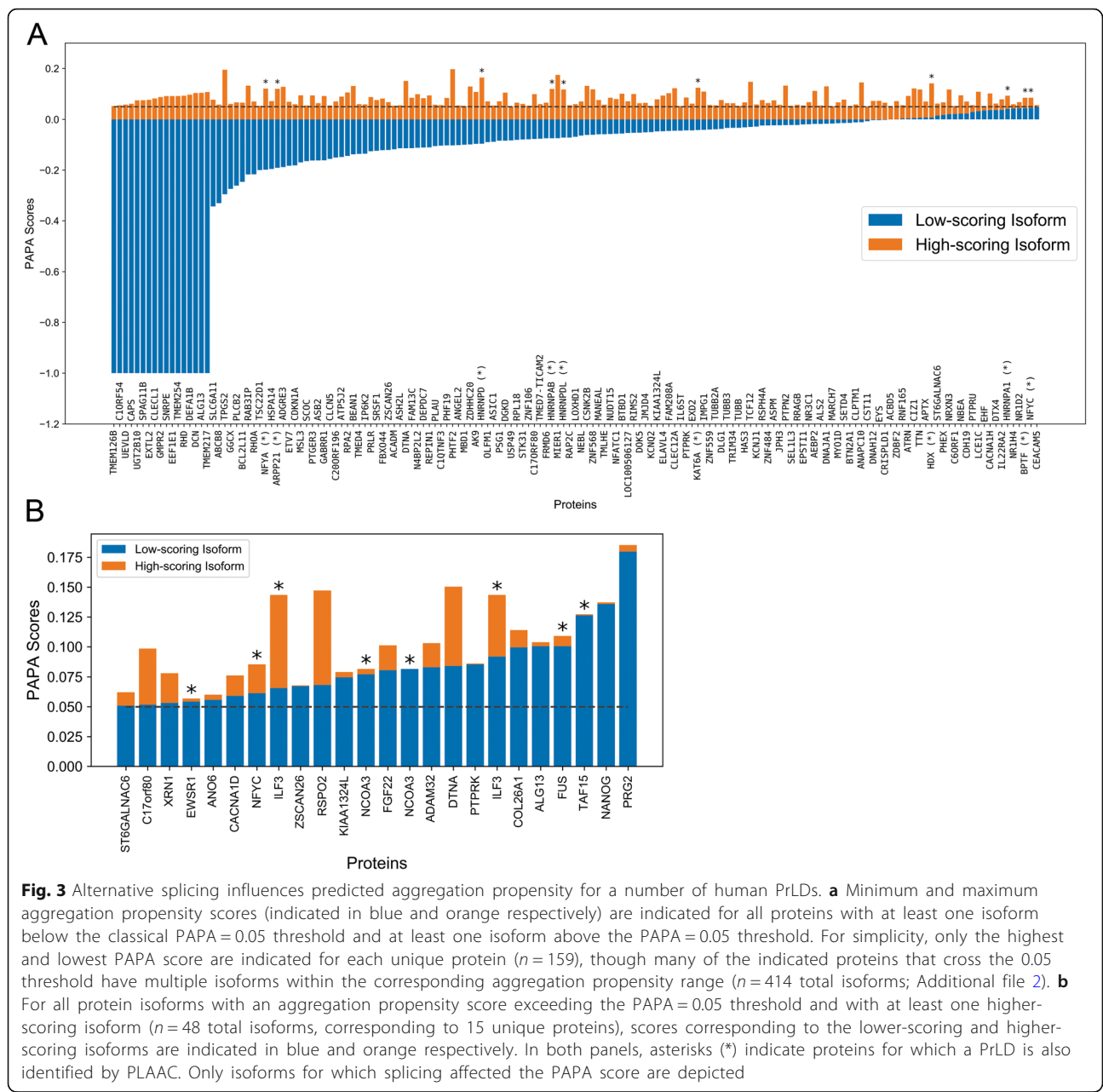
As observed in Fig. 2d, protein isoforms derived from the same gene can correspond to markedly different aggregation propensity scores. Alternative splicing essentially represents a form of post-transcriptional sequence variation *within* each individual. Alternative splicing could affect aggregation propensity in two main ways. First, alternative splicing could lead to the inclusion or exclusion of an entire PrLD, which could modulate prion-like activity in a tissue-specific manner, or in

response to stimuli affecting the regulation of splicing. Second, splice junctions that bridge short, high-scoring regions could generate a complete PrLD, even if the short regions in isolation are not sufficiently prion-like.

The ActiveDriver database [44] is a centralized resource containing downloadable and computationally accessible information regarding “high-confidence” protein isoforms, post-translational modification sites, and disease associated mutations in human proteins. We first examined whether alternative splicing would affect predicted aggregation propensity for isoforms that map to a common gene. In total, of the 39,532 high-confidence isoform sequences, 8018 isoforms differ from the highest-scoring isoform mapping to the same gene (Additional file 2). Most proteins maintain a low aggregation propensity score even for the highest-scoring isoform. However, we found 159 unique proteins for which both low-scoring and high-scoring isoforms exist (Fig. 3a; 414 total isoforms that differ from the highest-scoring isoform), suggesting that alternative splicing could affect prion-like activity. Furthermore, it is possible that known, high-scoring prion-like proteins are also affected by alternative splicing. Indeed, 15 unique proteins had at least one isoform that exceeded the PAPA threshold, and at least one isoform that scored even higher (Fig. 3b). Therefore, alternative splicing may affect aggregation propensity for proteins that are already considered high-scoring PrLD candidates.

Strikingly, many of the prototypical disease-associated prion-like proteins were among the high-scoring proteins affected by splicing. Consistent with previous analyses [45], PrLDs from multiple members of the hnRNP family of RNA binding proteins are affected by alternative splicing. For example, hnRNPD, which is linked to limb girdle muscular dystrophy type1G, has one isoform scoring far below the 0.05 PAPA threshold and another scoring far above the 0.05 threshold. hnRNPA1, which is linked to a rare form of myopathy and to amyotrophic lateral sclerosis (ALS), also has one isoform scoring below the 0.05 PAPA threshold and one isoform scoring above the threshold. Additionally, multiple proteins linked to ALS, including EWSR1, FUS, and TAF15 all score above the 0.05 PAPA threshold and have at least one isoform that scores even higher. Mutations in these proteins are associated with neurological disorders involving protein aggregation or prion-like activity. Therefore, in addition to well-characterized mutations affecting aggregation propensity of these proteins, alternative splicing may play an important and pervasive role in disease pathology, either by disrupting the intracellular balance between aggregation-prone and non-aggregation-prone variants, or by acting synergistically with mutations to further enhance aggregation propensity.

The fact that numerous proteins already linked to prion-like disorders have PAPA scores affected by alternative splicing raises the intriguing possibility that additional candidate proteins identified here may be involved in prion-like aggregation under certain conditions or when splicing is disrupted. For example, the RNA-binding protein XRN1 is a component of processing-bodies (or “P-bodies”), and can also form distinct synaptic protein aggregates known as “XRN1 bodies”. Prion-like domains have recently been linked to the formation of membraneless organelles, including stress granules and P-bodies [46]. Furthermore, dysregulation of RNA metabolism, mRNA splicing, and the formation and dynamics of membraneless organelles are prominent features of prion-like disorders [46]. However, XRN1 possesses multiple low-complexity domains that are predicted to be disordered, so it will be important to determine which (if any) of these domains are involved in prion-like activity. Interestingly, multiple β -tubulin proteins (TUBB, TUBB2A, and TUBB3) are among proteins with both low-scoring and high-scoring isoforms. Expression of certain β -tubulins is misregulated in some forms of ALS [47, 48], β -tubulins aggregate in mouse models of ALS [49], mutations in α -tubulin subunits can directly cause ALS [50], and microtubule dynamics are globally disrupted in the majority of ALS patients [51]. The nuclear transcription factor Y subunits NFYA and NFYC, which both contain high-scoring PrLDs affected by splicing, are sequestered in Htt aggregates in patients with Huntington’s disease [52]. NFYA has also been observed in aggregates formed by the TATA-box binding protein, which contains a polyglutamine expansion in patients with spinocerebellar ataxia 17 [53]. BPTF (also referred to as FAC1 or FALZ, for Fetal Alzheimer Antigen) is normally expressed in neurons in developing fetal tissue but largely suppressed in mature adults. However, FAC1 is upregulated in neurons in both Alzheimer’s and ALS, and is a characterized epitope of antibodies that biochemically distinguish diseased from non-diseased brain tissue in Alzheimer’s disease [54–56]. HNRNP A/B constitutes a specific member of the hnRNP A/B family, and encodes both a low-scoring and a high-scoring isoform. The high-scoring isoforms resembles prototypical prion-like proteins, containing two RNA-recognition motifs (RRMs) and a C-terminal PrLD (which is absent in the low-scoring isoform, and hnRNP A/B proteins were shown to co-aggregate with PABPN1 in a mammalian cell model of oculopharyngeal muscular dystrophy [57]. Alternative splicing of ILF3 mRNA leads to the direct inclusion or exclusion of a PrLD in the resulting protein isoforms NFAR2 and NFAR1, respectively [58, 59]. NFAR2 (but not NFAR1) is recruited to stress granules, its recruitment is dependent upon its PrLD, and recruitment of NFAR2 leads to stress granule enlargement [60]. A short “amyloid core” from the high-scoring NFAR2



PrLD forms amyloid fibers in vitro [40]. ILF3 proteins co-aggregate with mutant p53 (another PrLD-containing protein) in models of ovarian cancer [61]. ILF3 proteins are also involved in the inhibition of viral replication upon infection by dsRNA viruses, re-localize to the cytoplasm in response to dsRNA transfection (simulating dsRNA viral infection), and appear to form cytoplasmic inclusions [62]. Similarly, another RNA-binding protein, ARPP21, is expressed in two isoforms: a short isoform containing two RNA-binding motifs (but lacking a PrLD), and a longer isoform containing both RNA-binding motifs as well as a PrLD. The longer isoform (but not the short isoform)

is recruited to stress granules, suggesting that the recruitment is largely dependent on the C-terminal PrLD [63]. Furthermore, most of the proteins highlighted above have PrLDs that are detected by both PAPA and PLAAC (Additional file 2), indicating that these results are not unique to PAPA.

Collectively, these observations suggest that alternative splicing may play an important and pervasive role in regulating the aggregation propensity of certain proteins, and that misregulation of splicing could lead to an improper intracellular balance of a variety of aggregation-prone isoforms.

Disease-associated mutations influence predicted aggregation propensity for a variety of human PrLDs

Single-amino acid substitutions in prion-like proteins have already been associated with a variety of neurological disorders [46]. However, the role of prion-like aggregation/progression in many disorders is a relatively recent discovery, and additional prion-like proteins continue to emerge as key players in disease pathology. Therefore, the list of known prion-like proteins associated with disease is likely incomplete, and raises the possibility that PrLD-driven aggregation influences additional diseases in currently undiscovered or underappreciated ways.

We leveraged the ClinVar database of annotated disease-associated mutations in humans to examine the extent to which clinically-relevant mutations influence predicted aggregation propensity within PrLDs. For simplicity, we focused on single-amino acid substitutions that influenced aggregation propensity scores. Of the 33,059 single-amino acid substitutions (excluding mutation to a stop codon), 2385 mutations increased predicted aggregation propensity (Additional file 3). Of these proteins, 27 unique proteins scored above the 0.05 PAPA threshold and had mutations that increased predicted aggregation propensity (83 total mutants), suggesting that these mutations lie within prion-prone domains and are suspected to enhance protein aggregation (Fig. 4a). Additionally, 24 unique proteins (37 total mutants) scored below the 0.05 PAPA threshold but crossed the threshold upon mutation (Fig. 4b).

As observed for protein isoforms affecting predicted aggregation propensity, a number of mutations affecting prion-like domains with established roles in protein aggregation associated with human disease [21–25, 27–34, 64] were among these small subsets of proteins, including TDP43, hnRNPA1, hnRNPD, hnRNPA2B1, and p53. However, a number of mutations were also associated with disease phenotypes that have not currently linked to prion-like aggregation. For example, in addition to hnRNPA1 mutations linked to prion-like disorders (which are also detected in our analysis; Fig. 3, and Additional file 3), K277N, P275S, and P299L mutations in the hnRNPA1 PrLD increase its predicted aggregation propensity yet are associated with chronic progressive multiple sclerosis (Additional file 3), which is currently not considered a prion-like disorder. It is possible that, in addition to known prion-like disorders, certain forms of progressive multiple sclerosis (MS) may also involve prion-like aggregation. Intriguingly, the hnRNPA1 PrLD (which overlaps with its M9 nuclear localization signal) is targeted by autoantibodies in MS patients [65], and hnRNPA1 mislocalizes to the cytoplasm and aggregates in patients with MS [66], similar to observations in hnRNPA1-linked prion-like disorders [33].

Many of the high scoring proteins with mutations affecting aggregation propensity have been linked to protein aggregation, yet are not currently considered prion-like. For example, missense mutations in the PrLD of light chain neurofilament protein (encoded by the NEFL gene) are associated with autosomal dominant forms of Charcot-Marie Tooth (CMT) disease [67]. Multiple mutations within the PrLD are predicted to increase aggregation propensity (Fig. 4a and Additional file 3), and a subset of these mutations have been shown to induce aggregation of both mutant and wild-type neurofilament light protein in a dominant manner in mammalian cells [68]. Fibrillin 1 (encoded by the FBN1 gene) is a structural protein of the extracellular matrix that forms fibrillar aggregates as part of its normal function. Mutations in fibrillin 1 are predominantly associated with Marfan Syndrome, and lead to connective tissue abnormalities and cardiovascular complications [69]. While the majority of disease-associated mutations affect key cysteine residues (Additional file 3), a subset of mutations lie within its PrLD and are predicted to increase aggregation propensity (Fig. 4a), which could influence normal aggregation kinetics, thermodynamics, or structure. Multiple mutations within the PrLD of the gelsolin protein (derived from the GSN gene) are associated with Finnish type familial amyloidosis [also referred to as Meretoja syndrome [70–72];] and are predicted to increase aggregation propensity (Fig. 4a). Furthermore, mutant gelsolin protein is aberrantly proteolytically cleaved, releasing protein fragments that overlap with the PrLD and are found in amyloid deposits in affected individuals [for review, see [73]].

For proteins that cross the classical 0.05 aggregation propensity threshold, proteins exhibiting large relative changes in predicted aggregation propensity upon single-amino acid substitution likely reflect changes in intrinsic disorder classification implemented in PAPA via the FoldIndex algorithm. Therefore, these substitutions may reflect the disruption of predicted structural regions, thereby exposing high-scoring PrLD regions normally buried in the native protein. Indeed multiple mutations in the prion-like protein p53 lead to large changes in predicted aggregation propensity (Fig. 4b, Additional file 3), are thought to disrupt p53 structural stability, and result in a PrLD that encompasses multiple predicted aggregation-prone segments [74]. Additionally, two mutations in the Parkin protein (encoded by the PRKN/PARK2 gene), which has been linked to Parkinson's disease, increase its predicted aggregation propensity (Fig. 4b, Additional file 3). Parkin is prone to misfolding and aggregation upon mutation [75, 76] and in response to stress [77, 78]. Indeed, both mutants associated with an increase in predicted aggregation propensity for Parkin were shown to decrease Parkin solubility,

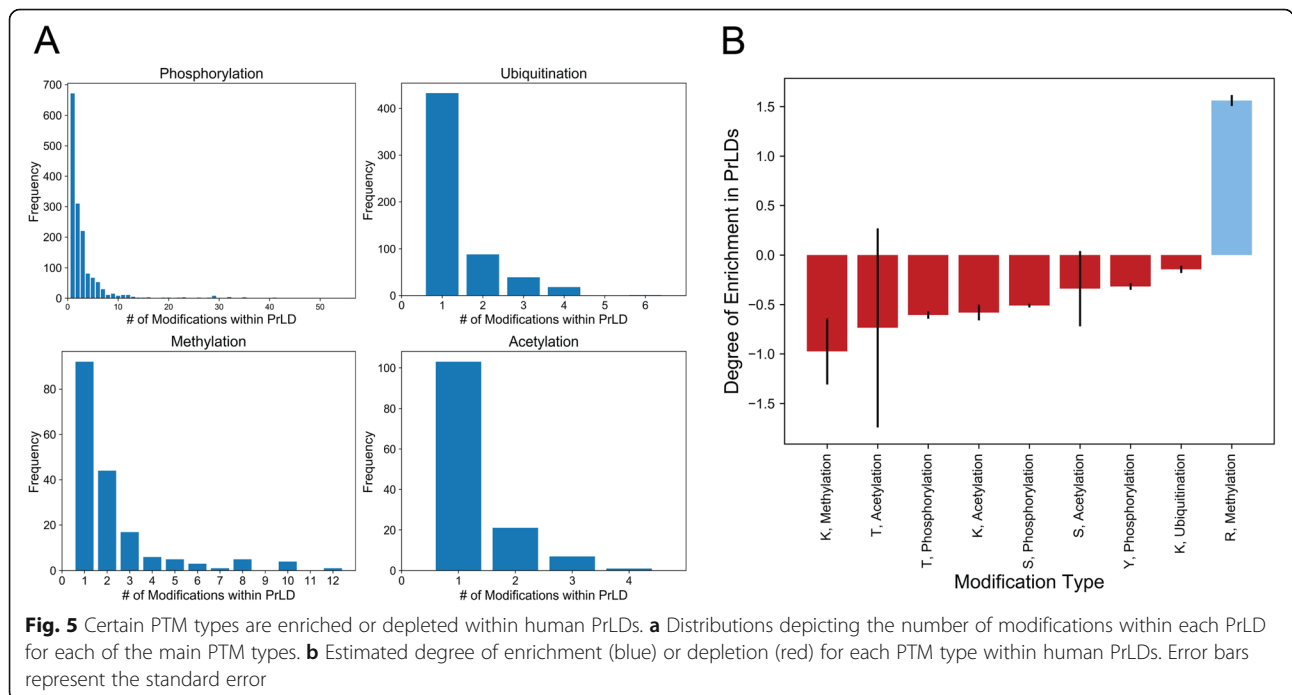
derived from multiple centralized PTM resources, as well as individual studies, have been combined into a single database describing a broad range of PTM sites across the human proteome [44]. PTMs could directly affect protein aggregation by increasing or decreasing inherent aggregation propensity. Indeed, changes in PTMs have been associated with a variety of aggregated proteins in neurodegenerative diseases [79–81], and PTMs can influence liquid-liquid phase separation [82, 83], which has recently been linked to low-complexity domains and PrLDs. Therefore, PTMs likely play an important role in regulating the aggregation propensity of certain PrLDs.

Using centralized PTM databases, we mapped PTMs to human PrLDs. While the contribution of each of the canonical amino acids to aggregation of PrLDs has been fairly well-characterized [7, 84], consistent effects of each type of PTM on aggregation of PrLDs have not been defined. Therefore, we mapped PTMs to PrLDs using a relaxed aggregation propensity threshold (PAPA cutoff = 0.0, rather than the standard 0.05 threshold), which accounts for the possibility that PTMs could increase aggregation propensity or regulate the solubility of proteins whose aggregation propensity is near the standard 0.05 aggregation threshold.

For each PTM type, distributions for the number of modifications per PrLD are shown in Fig. 5a, and PTMs mapped to PrLDs are provided in Additional file 4. Although PTMs are likely important regulators of aggregation for certain PrLDs and should be examined experimentally on a case-by-case basis, we explored whether

any PTMs were globally enriched or depleted within PrLDs. Since PrLDs typically have unusual amino acid compositions (which would affect the gross total for some PTMs within PrLDs), the number of potentially modifiable residues for each type of PTM was first calculated for the whole proteome and for PrLDs and statistically compared (see Methods for detailed description).

Arginine methylation was the only PTM type significantly enriched in human PrLDs (Fig. 5b and Additional file 5). In contrast, serine phosphorylation, threonine phosphorylation, tyrosine phosphorylation, lysine acetylation, lysine methylation, and lysine ubiquitination are significantly depleted within human PrLDs. The global underrepresentation of nearly all PTM types within PrLDs is particularly surprising since PrLDs are typically intrinsically disordered, and many of the PTM types studied here are enriched within intrinsically disordered regions *vis-à-vis* ordered regions [85]. However, it is important to note that the frequency of each PTM within PrLDs may be influenced by the amino acid compositions associated with the flanking regions surrounding PTM sites. For example, regions flanking phosphorylation sites are typically enriched in charged residues and depleted in neutral and aromatic residues [86]. Similarly, the flanking regions of arginine methylation sites are significantly associated with increased net charge and high glycine content (among other properties) and decreased glutamine and glutamic acid content [87]. Regions flanking lysine methylation sites are also enriched in glycine, aromatic residues, and threonine, and depleted in non-aromatic hydrophobic residues, glutamine, and glutamic acid. This highlights an



important point: while these features are consistent with PTM sites occurring preferentially within intrinsically disordered regions, they may be specific for disordered regions of particular amino acid compositions. Therefore, although PrLDs are typically considered intrinsically disordered, the Q/N-richness of most PrLDs may result in fewer PTMs compared to non-Q/N-rich intrinsically disordered regions.

Nevertheless, the global depletion of PTMs within PrLDs does not imply a lack of importance for PTMs that do occur within PrLDs. The mapping of PTMs to PrLDs may catalyze the experimental determination of the effects of each individual PTM on PrLD aggregation.

Sequence variation at the genetic, transcriptional, and posttranslational levels is associated with disease-relevant aggregation of a PrLD-containing protein – a case study of hnRNPA1

We were surprised to find that the hnRNPA1 PrLD is affected by every form of sequence variation examined in the present study, including genetic variation, alternative splicing, multiple disease-associated mutations, and post-translational modification (Fig. 6a). The short isoform, hnRNPA1-A (320 amino acids), scores just below the 0.05 PAPA threshold. Multiple mutations within the hnRNPA1 PrLD increase prion propensity and in vivo aggregation [33]. The long isoform, hnRNPA1-B (372 amino acids), scores substantially higher than the short isoform (PAPA scores are 0.093 and 0.042, respectively), and contains the region affected by the disease-associated mutations. It is possible that mutations within the hnRNPA1 PrLD, in combination with the high scoring isoform, have particularly potent aggregation-promoting effects. Under the current model for prion-like aggregation, the high-scoring protein isoform (which is typically less-abundant than the low-scoring isoform [88, 89]) could “seed” protein aggregates, which may then be capable of recruiting the lower-scoring isoform. Although this is currently speculative, it is supported by a recent study, which showed that mutation in the TDP-43 PrLD and cytoplasmic aggregation of TDP-43 in ALS patients was associated with dysregulation of hnRNPA1 mRNA splicing [89, 90]. This dysregulation led to increased abundance of the high-scoring hnRNPA1-B isoform and subsequent aggregation of the hnRNPA1 protein [89]. Finally, 31 unique posttranslational modifications map to the hnRNPA1 long-isoform PrLD, particularly to sites immediately flanking the highest-scoring PrLD region. It may also be possible that perturbations in posttranslational regulation of hnRNPA1, could influence protein aggregation in vivo. For example, phosphorylation of certain modification sites within the hnRNPA1 PrLD are differentially modified upon osmotic shock, which promotes accumulation of hnRNPA1 in

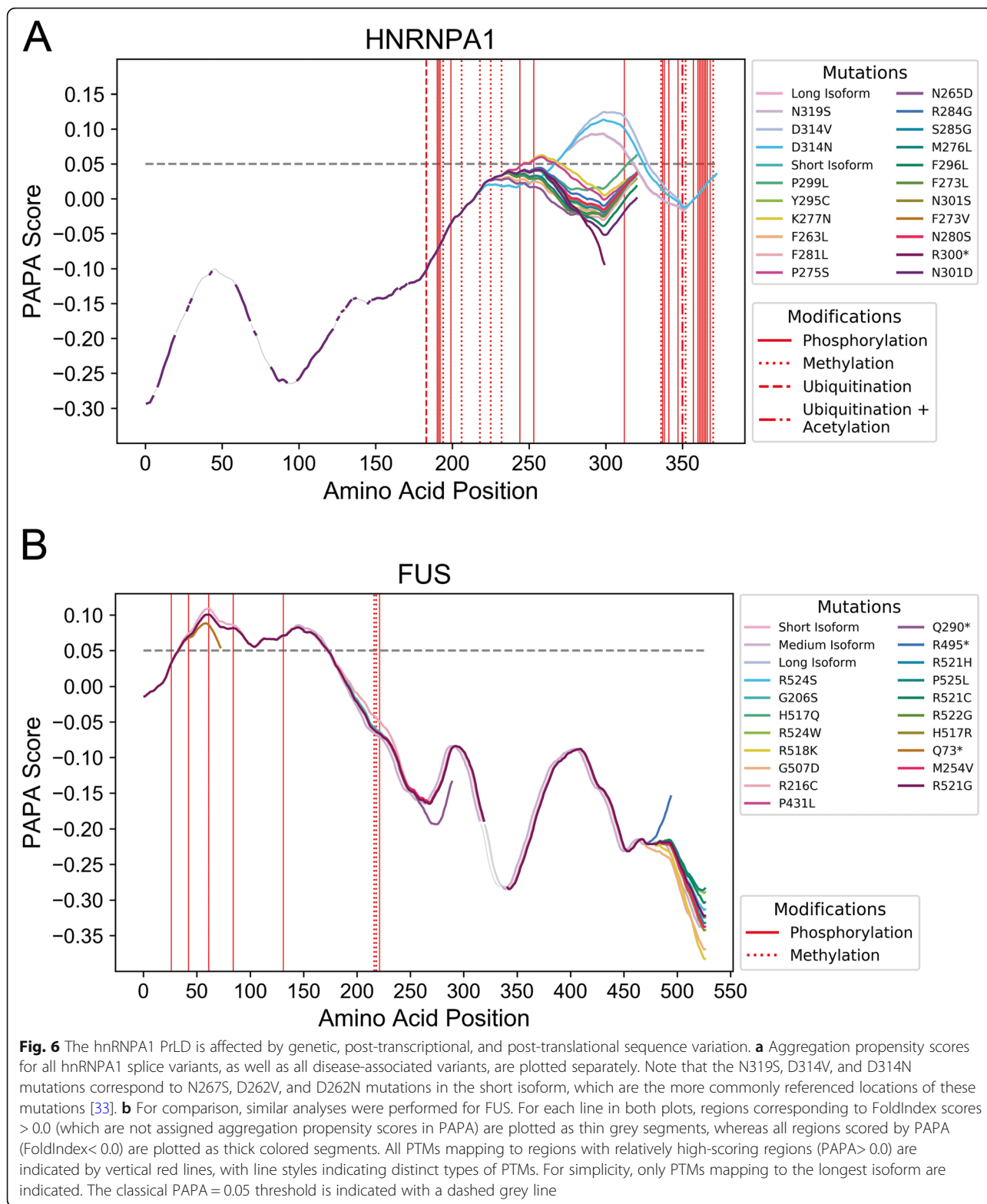
the cytoplasm [91], and a variety of PTMs within the PrLD regulate additional aspects of hnRNPA1 localization and molecular interactions [92]. Together, these observations suggest that multiple types of sequence variation may conspire to simultaneously influence hnRNPA1-related disease phenotypes.

While our study has focused predominantly on how sequence variation directly influences the predicted aggregation propensity of PrLDs, it is important to note that aggregation of PrLD-containing proteins may be contingent upon other domains or conditions. To illustrate, we analyzed FUS in a similar manner. Mutations in FUS have been implicated in ALS, and FUS aggregates are observed in a number of ALS cases [27, 28]. Furthermore, phosphorylation at multiple sites within the FUS PrLD has been shown to decrease FUS phase separation and aggregation in vitro and in vivo [93, 94]. Indeed, PAPA identifies a high-scoring PrLD near the N-terminus of FUS that contains multiple known phosphorylation sites (Fig. 6b). Additionally, one of the mutations in the ClinVar database results in a truncation in the middle of the PrLD, potentially leading to the production of highly aggregation-prone PrLD fragments. However, most disease-associated mutations occur in a nuclear localization sequence at the extreme C-terminus of FUS [95]. These mutations disrupt the nucleocytoplasmic shuttling of FUS and lead to its accumulation in cytoplasmic granules in ALS patients [95]. The FUS PrLD is highly aggregation-prone and is capable of forming aggregates with the parallel in-register β -sheet architecture characteristic of classical prion aggregates [94]. Therefore, aggregation of FUS may be due to a combination of the aggregation-prone PrLD, cytoplasmic mislocalization of FUS, and/or changes in PTM dynamics within the PrLD, as has been proposed recently [96].

Discussion

Numerous studies have explored the pervasiveness of candidate PrLDs across a variety of organisms. Although initial prediction of prion propensity among reference proteomes is an important first step in identifying candidate PrLDs, these predictions do not account for the richness of sequence diversity across individuals of the same species. Here, we complement these studies with an in-depth analysis of human intraspecies sequence variation and its effects on predicted aggregation propensity for PrLDs.

Prion aggregation is strongly (though not exclusively) dependent on the physicochemical characteristics of the aggregating proteins themselves. While analyses of reference proteomes necessarily treat protein sequences as invariable, protein sequence variation can be introduced at the gene, transcript, or protein levels via mutation, alternative splicing, or post-translational modification, respectively. Importantly, these protein changes can



exert biologically-relevant effects on protein structure, function, localization, and physical characteristics, which could influence prion-like behavior.

Broadly, we found that protein sequence variation is common within human PrLDs, and can influence predicted aggregation propensity rather substantially. Using

the frequency of observed single-amino acid variants from a large collection of human exomes (~ 60,700 individuals), we estimated the range of aggregation propensity scores by generating all pairwise combinations of variants for moderately high-scoring proteins. Aggregation propensity score ranges were often remarkably large, indicating that sequence variation could, in theory, have a dramatic effect on the prion-like behavior of certain proteins. However, it is important to note that not all variant combinations may naturally occur. For example, it is possible that certain variants commonly co-occur *in vivo*, or that some variants are mutually exclusive. Indeed, it is likely that aggregation propensity acts as a selective constraint which limits the allowable sequence space that can be viably explored by PrLDs. Conversely, our method conservatively assumed that all single amino acid variants were rare, even though some variants are substantially more common [43]: it is possible that some double, triple, or even quadruple variants may occur in a single individual with some regularity. Therefore, while our method for sampling sequence variants may over- or under-estimate aggregation propensity ranges for some PrLDs, our results nevertheless highlight the sequence diversity within PrLD regions across individuals. In principle, subtle changes in prion-like behavior could have phenotypic consequences, and may explain at least a small portion of human phenotypic diversity, although we emphasize that this is currently speculative.

We also identified a variety of proteins for which alternative splicing influences predicted aggregation propensity, which has a number of important implications. According to the prion model of protein aggregation, it is possible that aggregation of high-scoring isoforms could seed the aggregation of lower-scoring isoforms, assuming at least a portion of the PrLD is present in both isoforms. Importantly, this “cross-seeding” could occur even if the aggregation propensity of the low-scoring isoform is not itself sufficient to promote aggregation. Additionally, tissue-specific expression or splicing of certain proteins could impact prion-like behavior, effectively compartmentalizing or modulating prion-like activity in specific tissues. This also implies that dysregulation of alternative splicing could lead to overproduction of aggregation-prone isoforms. Interestingly, many of the prion-like proteins found in aggregates in individuals with neurological disease are splicing factors, and their sequestration into aggregates may impact the splicing of mRNAs encoding other aggregation-prone proteins [89]. This was recently proposed to produce a “snowball effect”, whereby aggregation of key proteins result in the aggregation of many other proteins via an effect on splicing or expression which could, in-turn, affect the aggregation of additional proteins [97].

Protein sequence variation can be beneficial, functionally inconsequential, or pathogenic. Examination of pathogenic sequence variants specifically (i.e. mutations in PrLDs associated with human disease) yielded a number of new prion-like protein candidates. Many of these new candidates have been associated with protein aggregation in previous studies, yet are not widely classified as prion-like, making them perhaps the most promising candidates for future studies and in-depth experimentation. In addition to candidates with experimental support, a number of candidates have not been previously linked to prion-like activity but may still have yet undiscovered prion-like activity *in vivo*. It is worth noting that, while PAPA and PLAAC predictions often overlap, many of these new candidate PrLDs (when considering disease-associated mutations) were only identified by PAPA, so experimental confirmation of aggregation and prion-like behavior is necessary.

One aspect of sequence variation that our study has not addressed is genomic mosaicism among somatic cells. Although it is convenient to treat individuals as having a fixed genome sequence across all cells, in reality genomic variation is introduced by replication errors during cell division and by DNA damage in dividing and post-mitotic cells [98]. Consequently, in principle, every cell may possess a unique genome, resulting in a “mosaic” of different genotypes, even for closely-related cell types. Genomic mosaicism is particularly important in neurons due to their long lifespan and interconnectivity (for review, see [99, 100]), and somatic cell mutations accumulate in an age-dependent manner in neurons [101]. At present, for some age-dependent prion-like disorders such as ALS, the vast majority of cases are considered “sporadic”, with familial mutations in a limited set of genes accounting for only ~ 5–10% of diagnosed individuals. Genomic mosaicism may have particularly insidious implications in conjunction with the prion-like mechanism proposed for these disorders: if aggregation-promoting somatic cell mutations occur within critical PrLDs, highly-stable aggregates may persist and spread in a prion-like manner even after the original mutation-harboring cell has perished. Therefore, it is possible that apparently sporadic cases may yet have a genomic origin and involve mutation of PrLDs.

Post-translational modification represents the final stage at which cells can modify protein properties and behavior. In a number of cases, PTMs are associated with protein aggregation across a diverse set of neurodegenerative disorders [79–81]. However, the precise effects of PTMs on aggregation propensity and whether they play a causative role in protein aggregation are often unclear. Nevertheless, one could speculate about what the effects of each PTM might be with respect to aggregation of PrLDs based on prion propensities for the

20 canonical amino acids and the physicochemical characteristics of the PTM. For example, charged residues typically inhibit prion aggregation within PrLDs [7, 84], so phosphorylation of serine, threonine, or tyrosine residues may tend to suppress prion-like activity [93]. Conversely, lysine acetylation or N-terminal acetylation neutralizes the charge, increases hydrophobicity, and introduces hydrogen bond acceptors, which may positively contribute to prion activity. Arginine and lysine methylation does not neutralize the charge, but slightly increases the bulkiness and hydrophobicity of the sidechain. Asymmetric dimethylation of arginine is common within proteins with PrLDs [102] and can weaken cation- π interactions with aromatic sidechains within PrLDs [103]. Recent studies implicate arginine methylation (which was the only PTM type significantly enriched within human PrLDs in our study) as an important suppressor of PrLD phase separation and pathological aggregation [for review, see [82, 102]]; together with our data, this suggests that arginine methylation may play a vital role in regulating the aggregation propensity of a multitude of PrLDs. Ubiquitination of lysine residues within PrLDs may sterically hinder PrLD aggregation. There are likely additional considerations that extend beyond the physicochemical properties of PTMs that alter aggregation propensity. For example, the proportion of any particular PrLD-containing protein that is modified at a given time in the cell dictates the effective concentration of each species which may influence the likelihood of forming a stable aggregate, analogous to the apparent resistance to prion disease in humans that are heterozygous at position 129 in the prion protein, PrP [104]. PTMs also regulate subcellular localization, protein-protein interactions, and structural characteristics, which may secondarily influence PrLD aggregation propensity. As with any attempt at generalizing predictions, the effects of PTMs may be highly context-specific, depending on interactions with particular neighboring residues. To facilitate further exploration of PTMs within PrLDs, we mapped PTMs from collated PTM databases to human PrLDs, and provide these maps as resources to encourage case-by-case experimental exploration.

As a final note, we would like to emphasize caution in over-interpreting our observations. As mentioned above, prion-like activity *in vivo* is strongly dependent upon the physicochemical characteristics of PrLDs, which are largely determined by the PrLD sequence. However, prion-like aggregation can be influenced *in vivo* by factors other than inherent sequence characteristics, including expression levels, subcellular localization, protein chaperone activity, and molecular binding partners, among others [105]. Additionally, for certain proteins, non-PrLD regions may be responsible for protein aggregation, or may influence the behavior of PrLDs via

intramolecular interaction. For example, phase separation of FUS relies on interactions between the FUS PrLD and FUS RNA-binding domains [106]. Furthermore, multivalent protein-protein, protein-RNA, and RNA-RNA interactions may contribute to the aggregation or phase separation for some proteins [107]. Many PrLD-containing proteins also contain RNA-binding domains, which may themselves be aggregation-prone [108]. In some cases, PrLDs may even prevent irreversible aggregation by enhancing recruitment of the protein to reversible protein granules induced by stress [109]. The influence of these factors will likely vary on a case-by-case basis; two similarly aggregation-prone PrLDs may be differentially regulated, leading one to aggregate while the other remains functional/soluble. At the same time, our prion prediction algorithm was developed in the context of a eukaryotic model organism [7], thereby incorporating at least some contribution from additional cellular factors and a crowded intracellular environment. Furthermore, prion-like aggregation is one of many possible mechanisms that can affect protein function upon mutation or alternative splicing. We are not advocating for a mutual exclusivity view of prion-like aggregation: protein sequence variation can have multiple concomitant consequences, and prion-like aggregation may simply be one of those consequences. For example, mutations can disrupt native protein sequence, resulting in loss of function of the protein. But those same mutations may also enhance prion-like aggregation, leading to a cytotoxic gain-of-function and a contribution to overall disease pathology. Additionally, while we have focused in this study on mutations that increase predicted aggregation propensity, mutations within PrLDs that decrease predicted aggregation propensity may be equally important. Adaptive, reversible aggregation activity exhibited by some PrLDs may involve a delicate balance in kinetic and thermodynamic parameters, which could be disrupted by mutations that either decrease or increase predicted prion-like behavior. Mutations that decrease predicted aggregation propensity may ultimately lead to PrLD aggregation *in vivo* if the loss in inherent aggregation propensity is ultimately outweighed by an indirect increase in aggregation propensity caused, for example, by disrupted molecular interactions that normally sequester the PrLD. Therefore, sequence variants that affect high-scoring PrLDs yet decrease predicted aggregation propensity may still be of interest and utility, and are retained in all supplementary resources.

Finally, while PrLDs have now been closely linked to liquid-liquid phase separation, the degree of overlap between classically-defined PrLD sequence features and those driving liquid-liquid phase separation of PrLDs has not been explored in great detail. A small subset of features important for phase separation have been

determined experimentally [106, 110, 111]. However, at present, a complete understanding of the effects of each amino acid on liquid-liquid phase separation propensity is currently lacking. Early phase separation prediction algorithms (recently reviewed in [112]), though capable of identify phase separating proteins from whole-proteomes, base their predictions on a limited subset of amino acids and are likely not optimized to resolve the effects of single-amino acid substitutions. It is unclear whether the amino acids that are classically considered prion-promoting or prion-inhibiting will affect PrLD phase separation in a similar manner. Therefore, it will be interesting to delineate the amino acids favoring liquid-liquid phase separation of PrLDs, solid phase aggregation of PrLDs, or both processes.

Conclusions

Our analyses indicate that sequence variation within human PrLDs is pervasive, occurs at each major stage of protein production, and often influences predicted aggregation propensity. Collectively, our results shed new light on the relationship between protein sequence diversity and inherent aggregation propensity, highlight a number of promising new prion-like candidates whose aggregation propensities may be influenced by protein sequence variation, and provide a variety of resources to propel future protein aggregation research.

Methods

Data acquisition and processing

Human protein isoform sequences, along with PTM sites, were acquired from the ActiveDriver database [[44]; <https://www.activedriverdb.org/>; downloaded on 10/5/2018]. Corresponding clinical variants were derived from NCBI's ClinVar database [113, 114] (downloaded in tab-delimited form from ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/ on 10/7/2018). For estimation of the range of theoretical aggregation propensity scores based on observed sequence variants, reference sequences including >6 million annotated single amino acid variants were obtained from the neXtProt database [[42, 115]; <https://www.nextprot.org/>; downloaded on 2/12/2019].

All data processing, including data re-structuring, quantification, calculation, statistical analysis, and plotting was performed using in-house Python scripts. All statistical analyses were performed using the built-in Python stats module with default settings, except that all statistical tests were two-sided. Where applicable, correction for multiple hypothesis testing was implemented via the statsmodels package available for Python. All plotting was performed using the Matplotlib and Seaborn packages. All source code required to reproduce

the analyses in all figures and additional files are available at <https://github.com/RossLabCSU>.

Modifications to the original PAPA method

PAPA source code was downloaded (<http://combi.cs.colostate.edu/supplements/papa/>) and augmented with custom functions scripted in Python. Briefly, the original PAPA algorithm assigns aggregation propensity scores to each position in a protein based on a combined score from 41 consecutive 41-amino acid windows (effectively, an 81-amino acid window for each position) [7, 116]. Our modified PAPA algorithm differs from the original PAPA algorithm in three key ways: 1) PAPA scores are assigned to the last residue of the first sliding window, which improves the scoring of protein termini and is critical for mapping PTM sites to PrLDs; 2) overlapping domains within a single protein that exceed a pre-defined PAPA threshold are merged, which yields precise definitions of predicted PrLD boundaries and accounts for multiple PrLDs within a single protein; and 3) predictions of protein disorder are simplified by calculating the FoldIndex over each full window, rather than the average of 41 consecutive windows. Additionally, for many analyses, a relaxed aggregation propensity threshold of 0.0 was chosen for two main reasons: 1) sequence variation or post-translational modification may increase aggregation propensity in some cases, such that the aggregation propensity may lie beyond our classical 0.05 threshold upon modification or mutation, and 2) this threshold captures ~10% of each proteome, yielding a reasonable set of high-scoring proteins for analysis. The modified version of PAPA (mPAPA) is available at <https://github.com/RossLabCSU/mPAPA>.

Estimation of aggregation propensity ranges via exhaustive pairwise variant combination

All possible pairwise combinations of single amino acid variants (neXtProt database) within the PrLD regions for proteins with a relatively high baseline aggregation propensity (PAPA score > 0.0) were generated computationally and stored as independent sequences. Theoretical sequence variants were then scored using our modified PAPA algorithm, and the minimum, maximum, and reference sequence scores were subsequently compared. By default, PAPA assigns an arbitrary score of -1.0 to proteins lacking a predicted intrinsically disordered region. Therefore, variants with a theoretical minimum PAPA score of -1.0 were excluded from analyses.

Analysis of PTM enrichment/depletion within PrLDs

PrLDs are, by definition, biased in terms of amino acid composition [2, 3]. Without controlling for compositional biases, certain PTMs would be over- or under-represented among PrLDs simply by virtue of the

availability of modifiable residues. Therefore, when comparing protein modifications within PrLDs vs. the remainder of the proteome, non-modified residues were defined as residues capable of being modified by the PTM of interest but with no empirical evidence of modification. For example, serine phosphorylation was analyzed by comparing the number of phosphorylated serine residues within PrLDs to the number of non-phosphorylated serine residues within PrLDs. Calculations were performed similarly for non-PrLD regions (i.e. the remainder of the proteome). The degree of PTM enrichment within PrLDs was then calculated as:

$$E_{PTM} = \ln(OR_{PTM})$$

and

$$OR_{PTM} = \left(\frac{f_{modPrLD}}{1-f_{modPrLD}} \right) / \left(\frac{f_{modnonPrLD}}{1-f_{modnonPrLD}} \right)$$

where $f_{modPrLD}$ and $f_{modnonPrLD}$ represent the fraction of modified residues out of potentially modifiable residues for the given PTM type within PrLD and non-PrLD regions, respectively. PTMs with fewer than 100 known modification sites within the human proteome were excluded from analyses. Statistical enrichment or depletion for each PTM type within PrLDs was evaluated using a two-sided Fisher's exact test, with Benjamini-Hochberg correction for multiple hypothesis testing (with false discovery rate threshold of 0.05).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6425-3>.

Additional file 1. PAPA scores derived from random sampling of sequence variant combinations for proteins with high-scoring PrLDs. For all proteins with a moderately high-scoring PrLD (PAPA > 0.0) and at least one single-amino acid variant, the minimum and maximum aggregation propensity scores obtained from randomly sampling PrLD sequence variants, along with the corresponding reference sequence score, are indicated (see Methods for a complete description of sequence variant calculations).

Additional file 2. Aggregation propensity scores and inter-isoform score comparison for all human protein isoforms. Predicted aggregation propensity for all "high-confidence" human protein isoforms (derived from ActiveDriverDB) was calculated using the modified PAPA algorithm (see Methods section for details). Scores and corresponding full protein sequences are indicated for all isoforms, along with the maximum PAPA score among all isoforms mapping to the same gene, the difference between the PAPA score for the indicated isoform and the maximum PAPA score among related isoforms, and the protein sequence corresponding to the highest-scoring related isoform. Additionally, the PLAAC algorithm was used to analyze the same sequences. A binary variable indicates if the protein contains a PLAAC-predicted PrLD that overlaps with the PAPA-predicted PrLD for high-scoring proteins only and, if so, the position of the PLAAC-predicted PrLD.

Additional file 3. Comparison of wild-type and disease-associated mutant PAPA scores. For all disease associated mutants in the ClinVar database, mutant sequences were generated by incorporating the indicated amino acid substitution at the appropriate position and re-scored using the

modified PAPA algorithm. For each variant, both wild-type and mutant aggregation propensity scores are indicated, as well as the difference between mutant and wild-type scores. For each variant, the associated disease phenotype annotation is also included. PLAAC predictions are also included, as indicated for Additional file 2.

Additional file 4. Comprehensive mapping of PTMs within moderately high-scoring human PrLDs. Human PTMs derived from the ActiveDriverDB were mapped to all human PrLDs with PAPA score > 0.0. For each protein the maximum PAPA score, moderately high-scoring PrLD sequence (corresponding to all overlapping regions with PAPA score > 0.0), amino acid positions bounding the PrLD sequence, and all PTMs mapping to the PrLD region are indicated. PLAAC predictions are also included, as indicated for Additional file 2.

Additional file 5. Statistical analysis of PTM enrichment/depletion within human PrLDs. For each PTM type, statistical enrichment or depletion within PrLDs was evaluated using a two-sided Fisher exact test (see Methods section for detailed description).

Abbreviations

PAPA: Prion Aggregation Prediction Algorithm; PLAAC: Prion-Like Amino Acid Composition algorithm; PrLD: Prion-like domain; PTM: Post-translational modification

Acknowledgements

We would like to thank members of the Ross lab for helpful discussion and insight.

Authors' contributions

Initial conception of the research, data acquisition, computational analyses, statistical analyses, modification of the PAPA algorithm script, data visualization, data interpretation, and preparation of the manuscript were performed by SMC. Essential contributions to analytical methodology, data interpretation, manuscript editing, and funding were provided by EDR. All authors read and approved the final manuscript.

Funding

This work was supported by the National Institute of General Medical Sciences [R35GM130352] to EDR. Funding for open access charge: National Institute of General Medical Sciences. The funding body had no role in the study design, data collection, data analysis, interpretation, and manuscript preparation.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. All source code required to reproduce the analyses within the article are available at <https://github.com/RossLabCSU>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 17 June 2019 Accepted: 23 December 2019

Published online: 08 January 2020

References

- Liebman SW, Chernoff YO. Prions in yeast. *Genetics*. 2012;191:1041–72.
- Cascarina SM, Ross ED. Yeast prions and human prion-like proteins: sequence features and prediction methods. *Cell Mol Life Sci*. 2014;71:2047–63 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24390581>.
- Du Z. The complexity and implications of yeast prion domains. *Prion*. 2011; 5:311–6 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22156731>.
- Wickner RB. Yeast and fungal prions. *Cold Spring Harb Perspect Biol*. 2016;8: a023531.
- Ross ED, Baxa U, Wickner RB. Scrambled prion domains form prions and amyloid. *Mol Cell Biol*. 2004;24:7206–13.

6. Ross ED, Edskes HK, Terry MJ, Wickner RB. Primary sequence independence for prion formation. *Proc Natl Acad Sci U S A*. 2005;102:12825–30.
7. Toombs JA, McCarty BR, Ross ED. Compositional determinants of prion formation in yeast. *Mol Cell Biol*. 2010;30:319–32. <https://doi.org/10.1128/MCB.01140-09> Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2798286&tool=pmcentrez&rendertype=abstract>.
8. Espinosa Angarica V, Ventura S, Sancho J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genomics*. 2013;14:316 Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3654983&tool=pmcentrez&rendertype=abstract>.
9. Michelišch MD, Weissman JS. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci USA*. 2000;97:11910–5 Available from: <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=11050225>.
10. Harrison PM, Gerstein M. A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. *Genome Biol*. 2003;4:R40.
11. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of Prionogenic proteins. *Cell*. 2009;137:146–58. <https://doi.org/10.1016/j.cell.2009.02.044>.
12. Sabate R, Rousseau F, Schymkowitz J, Ventura S. What makes a protein sequence a prion? *PLoS Comput Biol*. 2015;11:e1004013.
13. Afşar Minhas FUA, Ross ED, Ben-Hur A. Amino acid composition predicts prion activity. *PLoS Comput Biol*. 2017;13:1–20.
14. Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED. De novo design of synthetic prion domains. *Proc Natl Acad Sci U S A*. 2012;109:6519–24.
15. Yuan AH, Hochschild A. A bacterial global regulator forms a prion. *Science* (80-). 2017;355:198–201.
16. Fleming E, Yuan AH, Heller DM, Hochschild A. A bacteria-based genetic assay detects prion formation. *Proc Natl Acad Sci*. 2019;116:4605–10.
17. Chakrabortee S, Kayatekin C, Newby GA, Mendillo ML, Lancaster A, Lindquist S. Lumini-dependens (LD) is an Arabidopsis protein with prion behavior. *Proc Natl Acad Sci*. 2016;113:6065–70. <https://doi.org/10.1073/pnas.1604478113>.
18. Tariq M, Wegrzyn R, Anwar S, Bukau B, Paro R. Drosophila GAGA factor polyglutamine domains exhibit prion-like behavior. *BMC Genomics*. 2013;14:374.
19. Tetz G, Tetz V. Prion-like domains in eukaryotic viruses. *Sci Rep*. 2018;8:8931.
20. Nan H, Chen H, Tuite MF, Xu X. A viral expression factor behaves as a prion. *Nat Commun*. 2019;10:359 Cited 2019 Feb 11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30664652>.
21. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun*. 2006;351:602–11 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17084815>.
22. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* (80-). 2006;314:130–3 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17023659>.
23. Mackenzie IR, Nicholson AM, Sarkar M, Messing J, Purice MD, Pottier C, et al. TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. *Neuron*. 2017;95:808–816.e9.
24. Lee Y, Jonson PH, Sarparanta J, Palmio J, Sarkar M, Viñola A, et al. TIA1 variant drives myodegeneration in multisystem proteinopathy with SQSTM1 mutations. *J Clin Invest*. 2018;128:1164–77.
25. Vieira NM, Naslavsky MS, Licinio L, Kok F, Schlesinger D, Vainzof M, et al. A defect in the RNA-processing protein HNRPL causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum Mol Genet*. 2014;23:4103–10.
26. Iglesias V, Paladini L, Juan-Blanco T, Pallarès I, Aloy P, Tosatto SCE, et al. In silico Characterization of Human Prion-Like Proteins: Beyond Neurological Diseases. *Front Physiol Frontiers*. 2019;10:314. Cited 2019 Mar 27. <https://doi.org/10.3389/fphys.2019.00314/full>.
27. Kwiatkowski TJ Jr, Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* (80-). 2009;323:1205–8 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19251627>.
28. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* (80-). 2009;323:1208–11 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19251628>.
29. Couthouis J, Hart MP, Erion R, King OD, Diaz Z, Nakaya T, et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Hum Mol Genet*. 2012;21:2899–911. Available from: <https://doi.org/10.1093/hmg/dd5116>.
30. Neumann M, Bentmann E, Dormann D, Jawaid A, DeJesus-Hernandez M, Ansorge O, et al. FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain*. 2011;134:2595–609 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21856723>.
31. Couthouis J, Hart MP, Shorter J, DeJesus-Hernandez M, Erion R, Oristano R, et al. A yeast functional screen predicts new candidate ALS disease genes. *Proc Natl Acad Sci U S A*. 2011;108:20881–90 Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3248518>.
32. Ticozzi N, Vance C, LeClerc AL, Keagle P, Glass JD, McKenna-Yasek D, et al. Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. *Am J Med Genet Part B Neuropsychiatr Genet*. 2011;156:285–90.
33. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNP A2B1 and hnRNP A1 cause multisystem proteinopathy and ALS. *Nature*. 2013;495:467–73 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23455423>.
34. Klar J, Sobol M, Melberg A, Mabert K, Ameer A, Johansson AC, et al. Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing. *Hum Mutat*. 2013;34:572–7 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23348830>.
35. Navarro S, Marinelli P, Diaz-Caballero M, Ventura S. The prion-like RNA-processing protein HNRPL forms inherently toxic amyloid-like inclusion bodies in bacteria. *Microb Cell Factories*. 2015;14:102.
36. Paul KR, Molliex A, Cascarina S, Boncella AE, Taylor JP, Ross ED. Effects of mutations on the aggregation propensity of the human prion-like protein hnRNP A2B1. *Mol Cell Biol*. 2017;37:e00652–16.
37. Iglesias V, Conchillo-Sole O, Batlle C, Ventura S. AMYCO: evaluation of mutational impact on prion-like proteins aggregation propensity. *BMC Bioinformatics*. 2019;20:24 Cited 2019 Feb 11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30642249>.
38. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics*. 2014;30:2–3 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24825614>.
39. King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res*. 2012;1462: 61–80. Elsevier B.V. <https://doi.org/10.1016/j.brainres.2012.01.016>.
40. Batlle C, De Groot NS, Iglesias V, Navarro S, Ventura S. Characterization of soft amyloid cores in human prion-like proteins. *Sci Rep*. 2017;7:12134.
41. An L, Harrison PM. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biol Direct*. 2016;11:32.
42. Gaudet P, Michel PA, Zahn-Zabal M, Britan A, Cusin I, Domagalski M, et al. The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Res*. 2017;45:D177–82.
43. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536:285–91 Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5018207>.
44. Krassowski M, Paczkowska M, Cullion K, Huang T, Dzieladze I, Ouellette BFF, et al. ActiveDriverDB: human disease mutations and genome variation in post-translational modification sites of proteins. *Nucleic Acids Res*. 2018;46: D901–10.
45. Gueroussov S, Weatheritt RJ, O'Hanlon D, Lin ZY, Narula A, Gingras AC, et al. Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell*. 2017;170:324–339.e23.
46. Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. *Biochem J*. 2017;474:1417–38. Available from: <https://doi.org/10.1042/BCJ20160499>.
47. Lederer CW, Torrisi A, Pantelidou M, Santama N, Cavallaro S. Pathways and genes differentially expressed in the motor cortex of patients with sporadic amyotrophic lateral sclerosis. *BMC Genomics*. 2007;8:26.
48. Morello G, Spampinato AG, Conforti FL, D'Agata V, Cavallaro S. Selection and prioritization of candidate drug targets for amyotrophic lateral sclerosis through a meta-analysis approach. *J Mol Neurosci*. 2017;61:563–80.

49. Kabuta T, Kinugawa A, Tsuchiya Y, Kabuta C, Setsuie R, Tateno M, et al. Familial amyotrophic lateral sclerosis-linked mutant SOD1 aberrantly interacts with tubulin. *Biochem Biophys Res Commun*. 2009;387:121–6.
50. Smith BN, Ticozzi N, Fallini C, Tkazi AS, Topp S, Kenna KP, et al. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron*. 2014;84:324–31.
51. Clark JA, Yeaman EJ, Blizzard CA, Chuckowree JA, Dickson TC. A case for microtubule vulnerability in amyotrophic lateral sclerosis: altered dynamics during disease. *Front Cell Neurosci*. 2016;10:204.
52. Yamanaka T, Miyazaki H, Oyama F, Kurosawa M, Washizu C, Doi H, et al. Mutant Huntingtin reduces HSP70 expression through the sequestration of NF-Y transcription factor. *EMBO J*. 2008;27:827–39.
53. Lee LC, Chen CM, Wang HC, Hsieh HH, Chiu IS, Su MT, et al. Role of the CCAAT-binding protein NFY in SCA17 pathogenesis. *PLoS One*. 2012;7:e35302.
54. Bowser R, Giambrone A, Davies P. FAC1, a novel gene identified with the monoclonal antibody Al250, is developmentally regulated in human brain. *Dev Neurosci*. 1995;17:20–37.
55. Mu X, Springer JE, Bowser R. FAC1 expression and localization in motor neurons of developing, adult, and amyotrophic lateral sclerosis spinal cord. *Exp Neurol*. 1997;146:17–24.
56. Schoonover S, Davies P, Bowser R. Immunolocalization and redistribution of the FAC1 protein in Alzheimer's disease. *J Neuropathol Exp Neurol*. 2008;55:444–55.
57. Fan X, Messaëd C, Dion P, Laganieri J, Brais B, Karpati G, et al. HnRNP A1 and A/B interaction with PABPN1 in oculopharyngeal muscular dystrophy. *Can J Neurol Sci*. 2003;30:244–51. Cited 2019 Apr 25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12945950>.
58. Duchange N, Pidoux J, Camus E, Sauvaget D. Alternative splicing in the human interleukin enhancer binding factor 3 (ILF3) gene. *Gene*. 2000;261:345–53.
59. Saunders LR, Jurecic V, Barber GN. The 90- and 110-kDa human NFAR proteins are translated from two differentially spliced mRNAs encoded on chromosome 19p13. *Genomics*. 2001;71:256–9.
60. Shiina N, Nakayama K. RNA granule assembly and disassembly modulated by nuclear factor associated with double-stranded RNA 2 and nuclear factor 45. *J Biol Chem*. 2014;289:21163–80.
61. Yang-Hartwich Y, Soteras MG, Lin ZP, Holmberg J, Sumi N, Craveiro V, et al. P53 protein aggregation promotes platinum resistance in ovarian Cancer. *Oncogene*. 2015;34:3605–16.
62. Harashima A, Guettouche T, Barber GN. Phosphorylation of the NFAR proteins by the dsRNA-dependent protein kinase PKR constitutes a novel mechanism of translational regulation and cellular defense. *Genes Dev*. 2010;24:2640–53.
63. Rehfeld F, Maticzka D, Gresser S, Knauff P, Eravci M, Vida I, et al. The RNA-binding protein ARPP21 controls dendritic branching by functionally opposing the miRNA it hosts. *Nat Commun*. 2018;9:1235.
64. Silva JL, Gallo CVD, Costa DCF, Rangel LP. Prion-like aggregation of mutant p53 in cancer. *Trends Biochem Sci*. 2014;39:260–7.
65. Lee S, Xu L, Shin Y, Gardner L, Hartzes A, Dohan FC, et al. A potential link between autoimmunity and neurodegeneration in immune-mediated neurological disease. *J Neuroimmunol*. 2011;235:56–69.
66. Salapa HE, Johnson C, Hutchinson C, Popescu BF, Levin MC. Dysfunctional RNA binding proteins and stress granules in multiple sclerosis. *J Neuroimmunol*. 2018;324:149–56.
67. Horga A, Laurà M, Jaunmuktane Z, Jerath NU, Gonzalez MA, Polke JM, et al. Genetic and clinical characteristics of NEFL-related Charcot-Marie-tooth disease. *J Neurol Neurosurg Psychiatry*. 2017;88:575–85.
68. Pérez-Ollé R, López-Toledano MA, Goryunov D, Cabrera-Poch N, Stefanis L, Brown K, et al. Mutations in the neurofilament light gene linked to Charcot-Marie-tooth disease cause defects in transport. *J Neurochem*. 2005;93:861–74.
69. Sakai LY, Keene DR, Renard M, De Backer J. FBN1: The disease-causing gene for Marfan syndrome and other genetic disorders. *Gene*. 2016;591:279–91.
70. Levy E. Mutation in gelsolin gene in Finnish hereditary amyloidosis. *J Exp Med*. 2004;172:1865–7.
71. Maury CPJ, Kere J, Tolvanen R, de la Chapelle A. Finnish hereditary amyloidosis is caused by a single nucleotide substitution in the gelsolin gene. *FEBS Lett*. 1990;276:75–7.
72. de la Chapelle A, Tolvanen R, Boysen G, Santavy J, Bleeker-Wagemakers L, Maury CPJ, et al. Gelsolin-derived familial amyloidosis caused by asparagine or tyrosine substitution for aspartic acid at residue 187. *Nat Genet*. 1992;2:157–60.
73. Solomon JP, Page LJ, Balch WE, Kelly JW. Gelsolin amyloidosis: Genetics, biochemistry, pathology and possible strategies for therapeutic intervention. *Crit Rev Biochem Mol Biol*. 2012;47:282–96.
74. Rangel LP, Costa DCF, Vieira TCRG, Silva JL. The aggregation of mutant p53 produces prion-like properties in cancer. *Prion*. 2014;8(1):75–84.
75. Sriram SR, Li X, Ko HS, Chung KKK, Wong E, Lim KL, et al. Familial-associated mutations differentially disrupt the solubility, localization, binding and ubiquitination properties of parkin. *Hum Mol Genet*. 2005;14:2571–86.
76. Henn IH, Gostner JM, Lackner P, Tatzelt J, Winklhofer KF. Pathogenic mutations inactivate parkin by distinct mechanisms. *J Neurochem*. 2005;92:114–22.
77. Wang C, Ko HS, Thomas B, Tsang F, Chew KCM, Tay SP, et al. Stress-induced alterations in parkin solubility promote parkin aggregation and compromise parkin's protective function. *Hum Mol Genet*. 2005;14:3885–97.
78. Um JW, Park HJ, Song J, Jeon I, Lee G, Lee PH, et al. Formation of parkin aggregates and enhanced PINK1 accumulation during the pathogenesis of Parkinson's disease. *Biochem Biophys Res Commun*. 2010;393:824–8.
79. Sambataro F, Pennuto M. Post-translational modifications and protein quality control in motor neuron and Polyglutamine diseases. *Front Mol Neurosci*. 2017;10:82.
80. Didonna A, Benetti F. Post-translational modifications in neurodegeneration. *AIMS Biophys*. 2015;3:27–49. Cited 2019 Mar 9. <https://doi.org/10.3934/biophys.2016.1.27>.
81. Marcelli S, Corbo M, Iannuzzi F, Negri L, Blandini F, Nistico R, et al. The Involvement of Post-Translational Modifications in Alzheimer's Disease. *Curr Alzheimer Res*. 2017;15:313–35. Cited 2019 Mar 9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28474569>.
82. Hofweber M, Dormann D. Friend or foe — post-translational modifications as regulators of phase separation and RNP granule dynamics. *J Biol Chem*. 2018; jbc.TM118.001189. Available from: <https://doi.org/10.1074/jbc.TM118.001189>.
83. Owen I, Shewmaker F. The role of post-translational modifications in the phase transitions of intrinsically disordered proteins. *Int J Mol Sci*. 2019;20: E5501.
84. MacLea KS, Paul KR, Ben-Musa Z, Waechter A, Shattuck JE, Gruca M, et al. Distinct amino acid compositional requirements for formation and maintenance of the [PSI⁺] prion in yeast. *Mol Cell Biol*. 2015;35:899–911. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4323492&tool=pmcentrez&rendertype=abstract>.
85. Darling AL, Uversky VN. Intrinsic disorder and posttranslational modifications: the darker side of the biological dark matter. *Front Genet*. 2018;9:158.
86. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res*. 2004;32:1037–49.
87. Daily KM, Radivojac P, Dunker AK. Intrinsic Disorder and Prote in Modifications: Building an SVM Predictor for Methylation; 2008. p. 1–7.
88. Buvoli M, Cobianchi F, Bestagno MG, Mangiarotti A, Bassi MT, Biamonti G, et al. Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein. *EMBO J*. 1990;9:1229–35. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC551799>.
89. Deshaies JE, Shkreta L, Moszczynski AJ, Sidibé H, Semmler S, Fouillen A, et al. TDP-43 regulates the alternative splicing of hnRNP A1 to yield an aggregation-prone variant in amyotrophic lateral sclerosis. *Brain*. 2018;141:1320–33.
90. Sivakumar P, De Giorgio F, Ule AM, Neeves J, Nair RR, Bentham M, et al. TDP-43 mutations increase HNRNP A1-7B through gain of splicing function. *Brain*. 2018;141:e83.
91. Allemand E, Guil S, Myers M, Moscat J, Caceres JF, Krainer AR. Regulation of heterogeneous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proc Natl Acad Sci*. 2005;102:3605–10.
92. Jean-Philippe J, Paz S, Caputi M. hnRNP A1: The Swiss Army Knife of gene expression. *Int J Mol Sci*. 2013;14:18999–9024.
93. Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, et al. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J*. 2017;36:2951–67. Available from: <https://doi.org/10.15252/emboj.201696394>.
94. Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, et al. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell*. 2017;171:615–627.e16.
95. Dormann D, Rodde R, Edbauer D, Bentmann E, Fischer I, Hruscha A, et al. ALS-associated fused in sarcoma (FUS) mutations disrupt transport-mediated nuclear import. *EMBO J*. 2010;29:2841–57.

96. Rhoads SN, Monahan ZT, Yee DS, Shewmaker FP. The Role of Post-Translational Modifications on Prion-Like Aggregation and Liquid-Phase Separation of FUS. *Int J Mol Sci.* 2018;19(3):886.
97. Fratta P, Isaacs AM. The snowball effect of RNA binding protein dysfunction in amyotrophic lateral sclerosis. *Brain.* 2018;141:1236–8 Cited 2019 Mar 11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29701791>.
98. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science (80-).* 2015;350:94–8.
99. Rohrbach S, Siddoway B, Liu CS, Chun J. Genomic mosaicism in the developing and adult brain. *Dev Neurobiol.* 2018;78:1026–48.
100. McConnell MJ, Moran JV, Abyzov A, Akbarian S, Bae T, Cortes-Ciriano I, et al. Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science (80-).* 2017;356:eaal1641.
101. Lodato MA, Rodin RE, Bohrsen CL, Coulter ME, Barton AR, Kwon M, et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science (80-).* 2018;359:555–9.
102. Chong PA, Vernon RM, Forman-Kay JD. RGG/RG Motif Regions in RNA Binding and Phase Separation. *J Mol Biol.* 2018;430:4650–65.
103. Qamar S, Wang GZ, Randle SJ, Ruggeri FS, Varela JA, Lin JQ, et al. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell.* 2018;173:720–734.e15.
104. Palmer MS, Dryden AJ, Hughes JT, Collinge J. Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease. *Nature.* 1991;352:340–2.
105. Cascarina SM, Paul KR, Ross ED. Manipulating the aggregation activity of human prion-like proteins. *Prion.* 2017;11:323–31.
106. Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell.* 2018;174:688–699.e16 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867418307311>.
107. Protter DSW, Parker R. Principles and Properties of Stress Granules. *Trends Cell Biol.* 2016;26:668–79.
108. Agrawal S, Kuo PH, Chu LY, Golzarroshan B, Jain M, Yuan HS. RNA recognition motifs of disease-linked RNA-binding proteins contribute to amyloid formation. *Sci Rep.* 2019;9:6171.
109. Franzmann TM, Jahnel M, Pozniakovskiy A, Mahamid J, Holehouse AS, Nüske E, et al. Phase separation of a yeast prion protein promotes cellular fitness. *Science (80-).* 2018;359:eaao5654.
110. Vernon RMC, Chong PA, Tsang B, Kim TH, Bah A, Farber P, et al. Pi-pi contacts are an overlooked protein feature relevant to phase separation. *Elife.* 2018;7:e31486.
111. Bolognesi B, Gotor NL, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, et al. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.* 2016;16:222–31.
112. Vernon RM, Forman-Kay JD. First-generation predictors of biological protein phase separation. *Curr Opin Struct Biol.* 2019;58:88–96.
113. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42:D980–5 Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3965032>.
114. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46:D1062–7.
115. Lane L, Argoud-Puy G, Britan A, Cusin I, Duek PD, Evalet O, et al. NeXtProt: A knowledge platform for human proteins. *Nucleic Acids Res.* 2012;40:D76–83.
116. Ross ED, Toombs JA. The effects of amino acid composition on yeast prion formation and prion domain interactions. *Prion.* 2010;4:60–5.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

