In silico Characterization of Human Prion-Like Proteins: Beyond Neurological Diseases

Valentin Iglesias[1†], Lisanna Paladin[2†], Teresa Juan-Blanco[3], Irantzu Pallarès[1], Patrick Aloy[3,4], Silvio C. E. Tosatto[2,5] and Salvador Ventura[1]*

[1] *Institut de Biotecnologia i de Biomedicina, Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Barcelona, Spain,* [2] *Department of Biomedical Sciences, University of Padua, Padua, Italy,* [3] *Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain,* [4] *Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain,* [5] *CNR Institute of Neuroscience, Padua, Italy*

Prion-like behavior has been in the spotlight since it was first associated with the onset of mammalian neurodegenerative diseases. However, a growing body of evidence suggests that this mechanism could be behind the regulation of processes such as transcription and translation in multiple species. Here, we perform a stringent computational survey to identify prion-like proteins in the human proteome. We detected 242 candidate polypeptides and computationally assessed their function, protein–protein interaction networks, tissular expression, and their link to disease. Human prion-like proteins constitute a subset of modular polypeptides broadly expressed across different cell types and tissues, significantly associated with disease, embedded in highly connected interaction networks, and involved in the flow of genetic information in the cell. Our analysis suggests that these proteins might play a relevant role not only in neurological disorders, but also in different types of cancer and viral infections.

Keywords: prion-like proteins, disease, protein–protein interaction, protein aggregation, amyloid, bioinformatics

INTRODUCTION

Prions were first reported in the context of mammalian neurodegenerative disorders (Prusiner, 1982; van Rheede et al., 2003; Harrison et al., 2010; Sikorska and Liberski, 2012), but it is now clear that different organisms exploit prion conformational conversion for functional purposes (Halfmann and Lindquist, 2010). The most studied organism is *Saccharomyces cerevisiae*, with up to 11 functional prions identified so far (Cascarina and Ross, 2014; Batlle et al., 2017c) Initially, these yeast prions were proposed to be pathological agents (Nakayashiki et al., 2005; McGlinchey et al., 2011), but nowadays they are widely recognized to provide beneficial advantages in changing environments, predominantly by regulating transcription, translation, or RNA processing (Halfmann et al., 2012; Newby and Lindquist, 2013). Yeast prions switch from an initially soluble state through a structural conversion toward an aggregated amyloid conformation. This conversion is encoded in PrDs; long intrinsically disordered regions of low complexity.

A significant number of proteins sharing most, but not all, prion characteristics have been identified in different organisms, and generically named prion-like proteins (Pallares et al., 2015; Si, 2015; Chakrabortee et al., 2016). In higher eukaryotes, prion-like structural conversion plays a central role in diverse functions such as viral response (Hou et al., 2011; Franklin et al., 2014; Xu et al., 2014) or long-term memory acquisition and maintenance (Si et al., 2010;

Abbreviations: GO, gene ontology; LCC, largest connected component; MSC, mean shortest distance; PPIs, protein–protein interactions; PrD, prion domain; PrLD, prion-like domain.

Majumdar et al., 2012; Si and Kandel, 2016). Even though multiple beneficial functions have been assigned to prion-like mechanisms across all kingdoms of life, aggregated proteins in human neurodegenerative diseases such as Alzheimer's and Parkinson's diseases and amyotophic lateral sclerosis also share certain prion-like properties (Aguzzi and Rajendran, 2009; Gitler and Shorter, 2011; Luk et al., 2012; Stohr et al., 2012; Kim et al., 2013; Nomura et al., 2014).

The accumulated knowledge on the determinants of yeast prions conformational conversion has provided strong stimuli for the development of bioinformatics tools to uncover new PrLDs in other organisms (Michelitsch and Weissman, 2000; Harrison and Gerstein, 2003; Toombs et al., 2012; Espinosa Angarica et al., 2014; Lancaster et al., 2014; Afsar Minhas et al., 2017; Batlle et al., 2017c). Previous screenings for PrLDs in the human proteome have targeted the characteristic compositional bias of these protein regions (An and Harrison, 2016). We have recently proposed that, in addition to a distinctive amino acidic composition, PrLDs contain soft amyloidogenic sequence stretches that would contribute to trigger the initial protein self-assembly reaction (Sabate et al., 2015a,b). These cryptic amyloids were not only shown to be present and promote conformational conversion in *bona fide* yeast prions (Sant'Anna et al., 2016), but they also exist in human prion-like proteins (Batlle et al., 2017a) and appear to play key role in the induction, propagation, and inheritance of the prion state in the mammalian cytosol (Duernberger et al., 2018). The amyloid stretches embedded within PrLDs can be identified computationally (Sabate et al., 2015b; Zambrano et al., 2015).

Here we applied to the human proteome the same prediction scheme that allowed us to uncover the first *bona fide* prion-like protein in a bacterial proteome (Pallares et al., 2015; Yuan and Hochschild, 2017). Human proteins were first analyzed for the presence of regions with compositional similitude to yeast PrDs using the prion-like compositional bias (PLAAC) algorithm (Alberti et al., 2009; Lancaster et al., 2014) and afterward these protein domains were individually screened for the presence of soft amyloidogenic sequences using the pWALTZ program (Sabate et al., 2015b). Indeed, we have recently shown that such a combination of compositional and sequential PrLDs prediction provides the best accuracy when forecasting the aggregation propensities of individual human prion-like proteins (Batlle et al., 2017b).

In the present work, we computationally characterized the function, location, expression, PPI networks, and the connection to disease of the human prion-like subproteome. The picture that emerges from this analysis is that prion-like proteins are widespread expressed proteins that function in biological processes tightly associated to disease.

MATERIALS AND METHODS

Data Acquisition

The human reference proteome dataset was obtained from Uniprot (UniProt Consortium, 2015) (Proteome ID UP000005640; release 2016_09) and scanned for PrLDs with PLAAC using as background probability the frequency of human proteome. From the initial 70,940 proteins in the proteome, 431 PrLD containing candidates were identified. Their predicted PrLDs were further evaluated with pWALTZ applying a cutoff of 60.00, as in Batlle et al. (2017a), which resulted in 242 final positive predictions (**Supplementary Table S1**).

Prion-Like Domain Localization Within the Protein Sequence

Each prion-like protein sequence was divided into three segments, the N- and C- terminal, accounted for 25% of the residues each, whereas the resting 50% of the sequence was considered as internal. Each predicted PrLD was located in the sequence and the number of residues mapping in each of the segments counted.

Functional Annotation

The GO annotation of all proteins in the prion-like dataset were collected, excluding the terms Inferred from Electronic Annotation (IEA) and filtering through the Generic GO slim developed by GO Consortium (Gene Ontology Consortium, 2015). All UniProt human proteins were used as background set to infer enrichment. A Fisher's exact test of GO term distributions was performed in the three ontologies separately, to calculate the enrichment/depletion of dataset proteins with respect to the whole UniProt. The Bonferroni correction was applied in performing all the tests. The results are shown in **Figure 3** applying the formula:

$$ E = \log \frac{\text{GO freq. in P}_{PR}}{\text{Tot GO in P}_{PR}} - \log \frac{\text{GO freq. in P}_{Back}}{\text{Tot GO in P}_{Back}} $$

where GO is the GO term, P_{PR} and P_{Back} are the datasets of prion-like proteins and the whole proteome, respectively. The abbreviations freq. and Tot stay for frequency and total.

Pfam Domains

Pfam (Finn et al., 2016) domains annotation in the dataset proteins were collected and compared to the human proteome (from UniProt). Fisher's exact test was used to assess significance.

Tissue and Cellular Localization

Tissue and cellular localization data of human proteins were retrieved from Human Protein Atlas (Uhlen et al., 2015). The prion-like proteins identifiers were converted to Ensembl Gene Ids. Human Protein Atlas reports a textual ranking of protein expression of each coding gene. This ranking ("none," "low," "medium," "high") was converted to numerical expressions, from 0 to 3, and each gene value for each particular tissue was collected. The expression of the complete gene set for the tissue was then averaged.

Association to Diseases

OMIM disease annotation was extracted from the field "diseases" of the UniProt description (Amberger et al., 2015). All information regarding the associated diseases was collected from the OMIM FTP site. DisGeNET data were retrieved from

DisGeNET download section (Pinero et al., 2015). For both databases, the number of proteins associated to at least one disease ID was divided by the total number of proteins, obtaining the fraction of disease-associated proteins. The results were compared to 100 random sampling of sets with the same number of proteins than the one in the database.

Human Network Analysis

The human prion-like protein dataset was curated for duplicities and scanned for PPIs with Interactome3D (2017_06 version) (Mosca et al., 2013). Out of the 121 unique identities, 100 had annotated physical binary interactions. The degree and the number of interactions between prion-like proteins were analyzed and compared to a random distribution by sampling the complete human binary interactome in Interactome3D. Moreover, the sizes of the LCC and the MSD were measured (Menche et al., 2015). The subnetwork of prion-like proteins and their interactors were functionally characterized with DAVID database (Huang da et al., 2009) for GO and KEGG pathways enrichment (n = 1542). The significance of the differences was assessed by Wilcox p-value or empirical p-value.

RESULTS

Human Prion-Like Proteins Prevalence and Modularity

A combination of prion-like compositional bias (PLAAC) and sequential amyloid propensity (pWALTZ) analysis was applied to the complete human proteome. This resulted in the identification of a total of 242 polypeptides (unique UniProt entries) bearing PrLDs (**Supplementary Table S1**). Our list of candidates included all human prion-like proteins shown to behave as such both *in vitro* and *in vivo*: FUS (Ju et al., 2011), TDP-43 (Wang et al., 2012), EWS (Couthouis et al., 2012), hnRNP A1 and hnRNP A2 (Kim et al., 2013), TIA1 (Li et al., 2014), and TAF15 (Couthouis et al., 2011) proteins, reinforcing the suitability of our dataset for the further evaluation of the global properties of human prion-like sequences.

According to our predictions, prion-like proteins account for a 0.34% of the human proteome. This is in line with two previous independent surveys for human prion-like proteins that exploited compositional bias alone for their detection; both studies predicting that the prevalence of these proteins is <1% (An and Harrison, 2016). Despite the percentage of proteins with PrLDs in the proteomes of different organisms seems to differ significantly (Michelitsch and Weissman, 2000; Espinosa Angarica et al., 2013; Malinovska et al., 2015; Chakrabortee et al., 2016; Pallares et al., 2018), their presence in all evolutionary lineages analyzed so far suggests that these regions might play conserved functional roles (Michelitsch and Weissman, 2000; Malinovska et al., 2015; Batlle et al., 2017c).

Yeast prion proteins tend to be modular (Li and Lindquist, 2000; Alberti et al., 2009). PrDs being generally located near the N- or C-terminal ends of the sequence (Baxa et al., 2007; Zambrano et al., 2015). In our dataset, 195 proteins; an 80.6% of the putative human prion-like proteins, presented their PrLDs located in any of the protein's ends (**Figures 1**, **2** and **Supplementary Table S1**). PrLDs were 1.67 times more frequent at the protein C-terminus. This was the case for 122 proteins, while in 73 of them the PrLDs were located at the N-terminus. This statistically significant imbalance between the presence of PrLDs at C- and N- in human proteins (p-value < 0.005, Z-test), contrasts with that found in *bona fide* yeast PrDs. In SUP35, URE2, NEW1, MOT3, and SWI1 proteins, the PrD is placed at the N-terminus, whereas only in RNQ1, it is located near the carboxyl end (Baxa et al., 2007; Zambrano et al., 2015). The modular architecture of prion-like proteins would allow the self-assembly of the PrLDs without disturbing the structure and productive associations of the adjacent globular moieties. This is likely facilitated by the predicted disordered nature of these protein segments (**Supplementary Table S1**).

Human Prion-Like Proteins Play a Major Role in Nucleic Acid Binding

As a first step to gain insights into the biological role of the candidate human prion-like proteins, we used a GO term analysis. GO terms were collected for biological process, molecular function, and cellular component categories and their enrichment with respect to the human proteome calculated (**Figure 3**). When we analyzed the "biological process" category for the set of candidate proteins, we found a statistically significantly enriched cluster of GO terms related to RNA and DNA associated processes, including positive regulation of transcription from RNA polymerase II promoter (p-value < 1.20E-16, 30 proteins), positive regulation of transcription DNA-templated (p-value < 6.92E-14, 22 proteins), mRNA splicing (p-value < 2.27E-9, 13 proteins), transcription DNA-templated (p-value < 5.26E-8, 36 proteins), RNA processing (p-value < 7.5E-8, 10 proteins), and negative regulation of transcription from RNA polymerase II promoter (p-value < 6.28E-4, 11 proteins) (**Figure 3A**). This result is consistent with the observation that the prion-like subproteomes identified in organisms belonging to different taxonomic divisions are usually enriched in proteins associated to the regulation of the flux of genetic information in the cell (Iglesias et al., 2015; Pallares et al., 2018).

With respect to the "molecular function," the most enriched GO terms are all involved in essential activities related with nucleic acid binding and transcription processes, such as transcription coactivator activity (p-value < 5.63E-17, 20 proteins), nucleotide binding (p-value < 4.96 E-17, 37 proteins), poly(A)RNA-binding (p-value < 3.94E-15, 30 proteins), RNA-binding (p-value < 2.99E-14, 31 proteins), chromatin binding (p-value < 3.34E-14, 14 proteins), transcription factor activity-sequence-specific DNA binding (p-value < 9.79E-6, 29 proteins), and ATP binding (p-value < 1.14E-4, 13 proteins) (**Figure 3B**). The conformational plasticity of PrLDs has been shown to be behind certain transcription factors ability to bind to many different targets and to play a role in the formation of chromatin regulatory complexes (Boulay et al., 2017; Kataoka and Mochizuki, 2017; Cho et al., 2018). Moreover, it is becoming increasingly clear that PrLDs are crucial for the
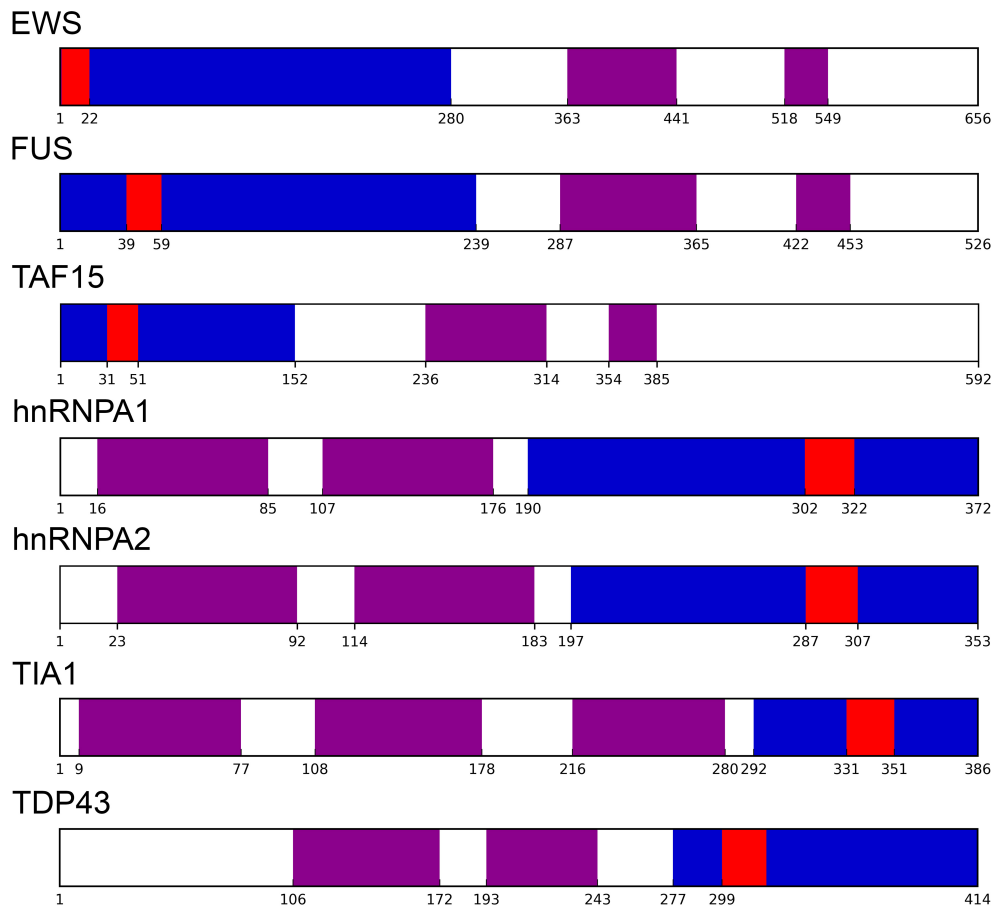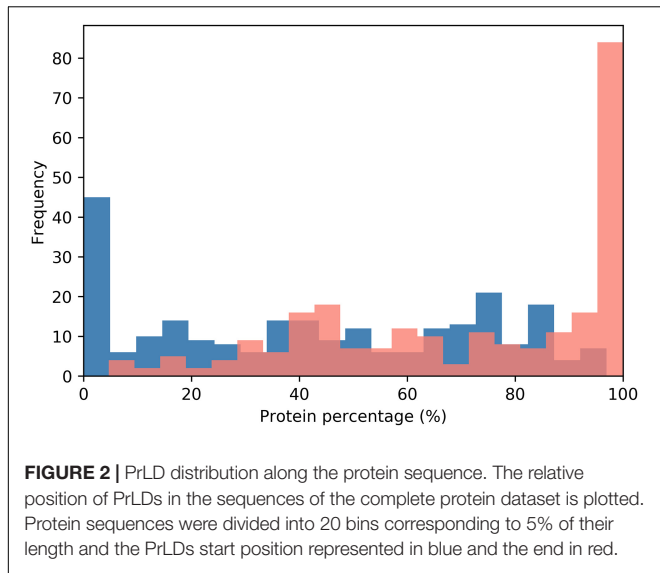
FIGURE 1 | Prion-like proteins modularity. Well-characterized prion-like human proteins have their PrLD (as identified by PLAAC in blue) and soft amyloid core (as identified by pWALTZ in red) at the protein edges, separated from their respective globular domains (retrieved from Pfam database in violet).

formation of membraneless organelles, since they enable RNA-binding proteins (RBPs) to undergo liquid–liquid transition, confining their RNA cargos (Villarroya-Beltri et al., 2013; Wang et al., 2018).

When we analyzed the cellular components populated by our protein subset, the most enriched GO terms were the nucleoplasm, nucleus, and the intracellular ribonucleoprotein complex (**Figures 3C,D**). As expected, all these compartments correspond to locations were the binding between nucleic acids and proteins occur frequently. Of particular interest is the so-called ribonucleoprotein complex which includes cellular structures like the stress granules, or P-bodies, which are sites for mRNA decay as well as for mRNA storage and therefore act as important cell regulatory centers in determining levels of gene expression (Anderson et al., 2015). The RBPs associated to those membrane-less organelles are key determinants in the control of the organelle function and have been implicated not only in adaptation to stress but also in tumor biology and the pathogenesis of neurodegenerative, immunological, and infectious diseases (Loomis et al., 1990; Villarroya-Beltri et al., 2013; Anderson et al., 2015; Harrison and Shorter, 2017).

We extended our analysis to look for the role of the constituent functional domains in the collection of PrLDs containing proteins. In agreement with the above presented results, Pfam domain clustering rendered DNA/RNA binding as the most enriched functional group (**Figure 3E**). Among them, the canonical RNA recognition motif (RRM) is by far the most statistically enriched, with 14% of the detected proteins harboring an RRM. This observation is line with previous studies (King et al., 2012) and consistent with the fact that the RRM is the most abundant domain in RBPs, conserved from bacteria to higher eukaryotes (Reddy et al., 2015). This set of RRM-bearing prion-like proteins includes FUS, TDP-43, TIA1, or hnRNP A1, all involved in the formation of dynamic membraneless intracellular compartments and associated to disease (Cascarina and Ross, 2014; March et al., 2016; Wang et al., 2018).

The second most enriched domain in our data set is the FoxP coiled-coil (p-value $< 2.95E$-19, 10 proteins). It corresponds to a coiled–coil domain involved in the modulation of the dimeric associations of the forkhead box family of transcription factors FoxP. There are multiple lines of evidence suggesting the biological relevance of domain swapping in FoxP functionality being important not only for their function regulation but

FIGURE 2 | PrLD distribution along the protein sequence. The relative position of PrLDs in the sequences of the complete protein dataset is plotted. Protein sequences were divided into 20 bins corresponding to 5% of their length and the PrLDs start position represented in blue and the end in red.
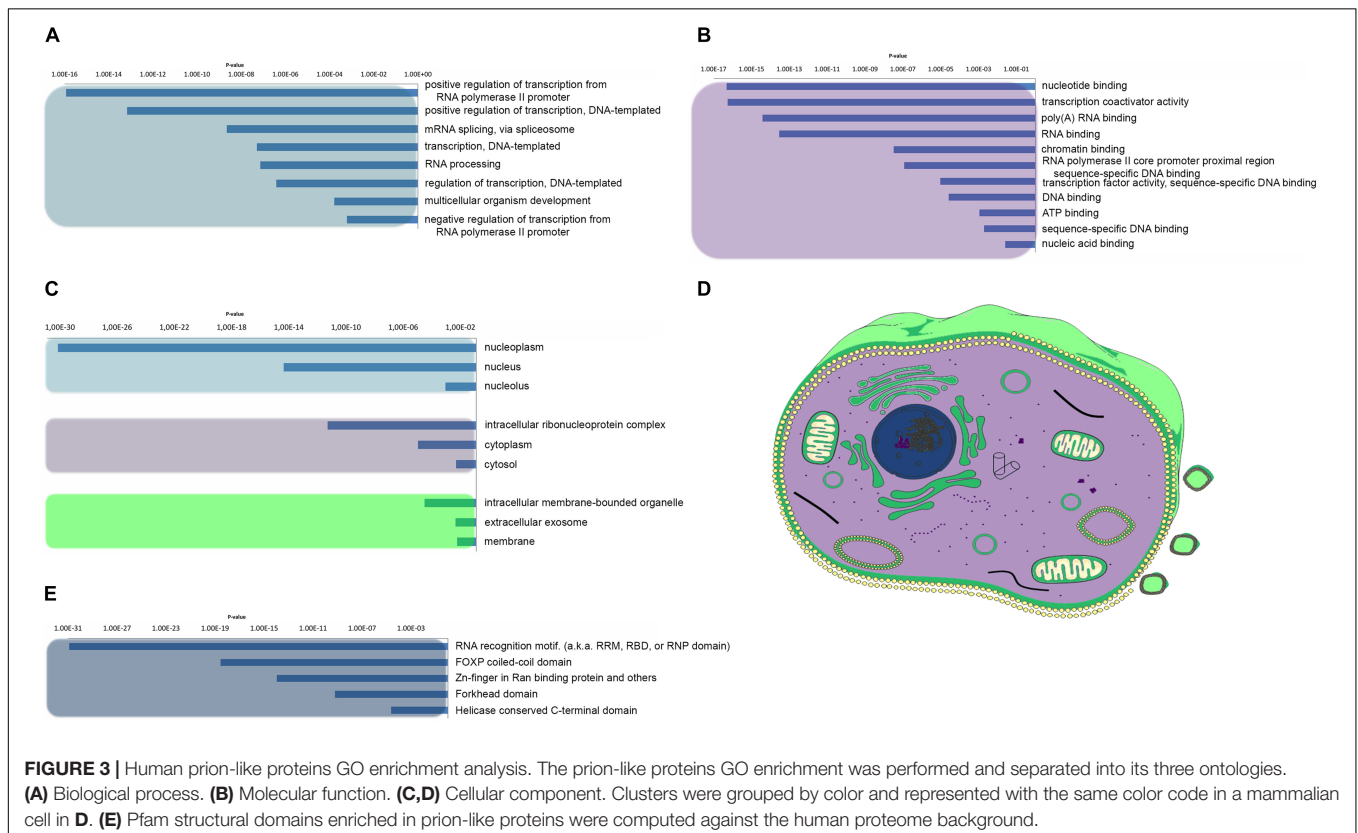
also linked to disease onset (Hafner-Bratkovic et al., 2011; Medina et al., 2016).
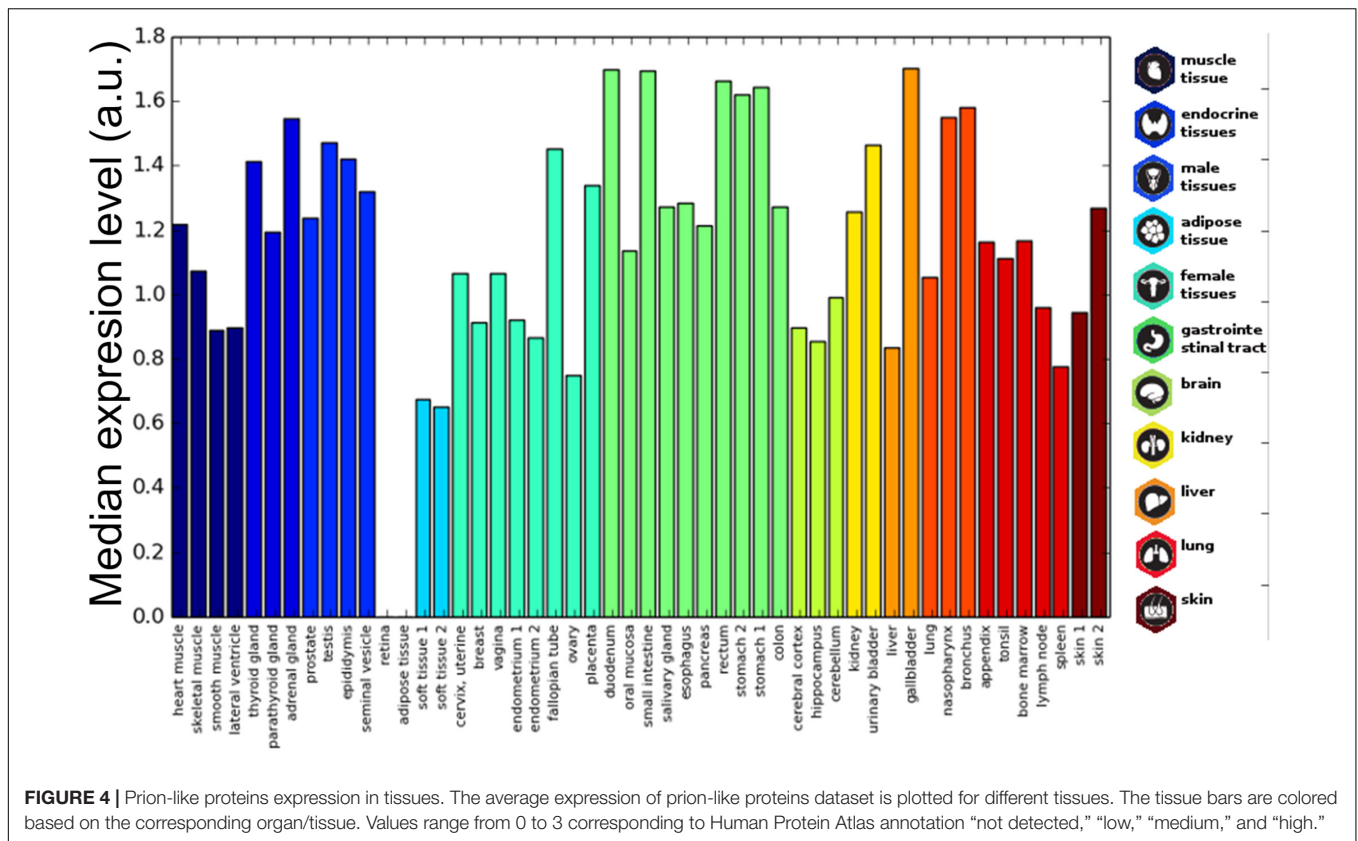
The other two enriched Pfam families include Zinc-fingers in Ran binding proteins (Zn_RanBP) (*p*-value < 1.14E-14, 9 proteins) and the Helicase conserved C-terminal domain (*p*-value < 2.47E-05, 7 proteins). Zinc Finger domains are a very versatile group of small protein domains which are evolutionary

conserved. Interestingly, RBPs with PrLDs such as FUS or EWS accommodate in their structure a Zn_RanBP domain in close proximity to an RRM domain. The Helicase conserved C-terminal domain is found at the C-terminus of DEAD-box helicases. Helicases function in the separation of double-stranded RNA, DNA, and RNA/DNA structures in an energy-dependent manner and therefore it is clear their role in RNA metabolism. Interestingly, the first prion-like protein identified in bacteria corresponds to the transcription terminator Rho, a helicase that can undergo a prion-state that results in genome-wide changes at the transcriptome level, contributing to rapid bacterial adaptation to fluctuating environments (Pallares et al., 2015; Yuan and Hochschild, 2017). The multitasking transcriptional regulators DDX5 and DDX17 included in our dataset contain an helicase domain in their structure reported to be associated with cancer development and cell proliferation (Mazurek et al., 2012; Fuller-Pace, 2013).

Prion-Like Proteins Are Widespread Among Tissues

The histological localization of human prion-like proteins was assayed by retrieving data from the Human Protein Atlas. To compare the expression levels, proteins were mapped to Ensemble gene annotations (121 genes). The expression data were collected for each cell type and averaged by tissue and organ. The result illustrates that prion-like proteins are widely distributed in human tissues (**Figure 4**). Importantly, the data indicate that, globally, the expression of these proteins in the

FIGURE 3 | Human prion-like proteins GO enrichment analysis. The prion-like proteins GO enrichment was performed and separated into its three ontologies. **(A)** Biological process. **(B)** Molecular function. **(C,D)** Cellular component. Clusters were grouped by color and represented with the same color code in a mammalian cell in **D**. **(E)** Pfam structural domains enriched in prion-like proteins were computed against the human proteome background.

FIGURE 4 | Prion-like proteins expression in tissues. The average expression of prion-like proteins dataset is plotted for different tissues. The tissue bars are colored based on the corresponding organ/tissue. Values range from 0 to 3 corresponding to Human Protein Atlas annotation "not detected," "low," "medium," and "high."

brain is not higher than in most organs or tissues, being more represented in endocrine tissues, in the gastrointestinal tract, the kidney, or the lung.
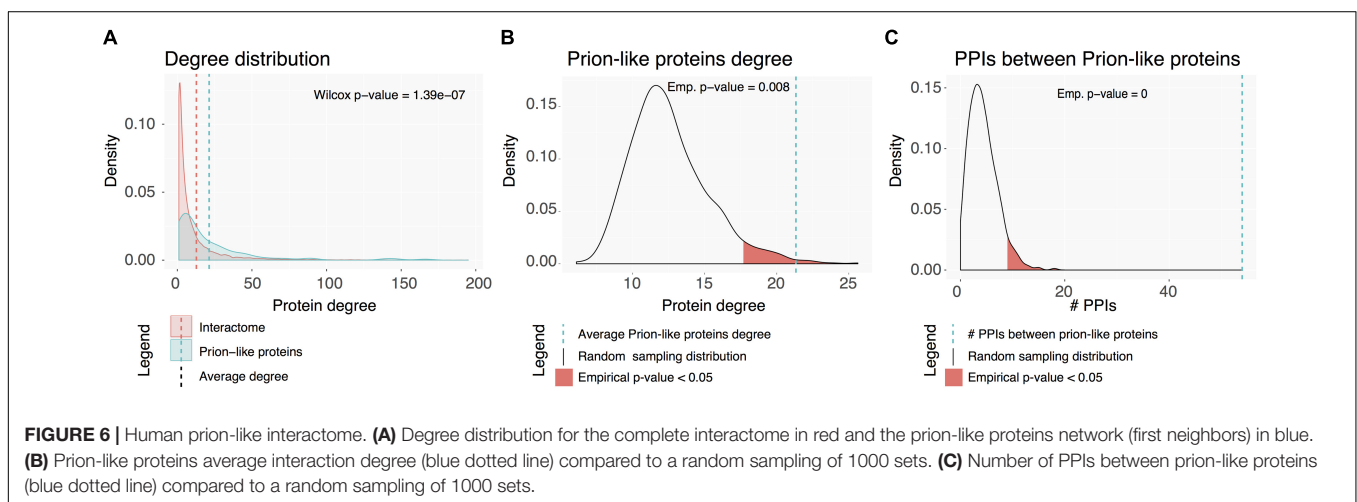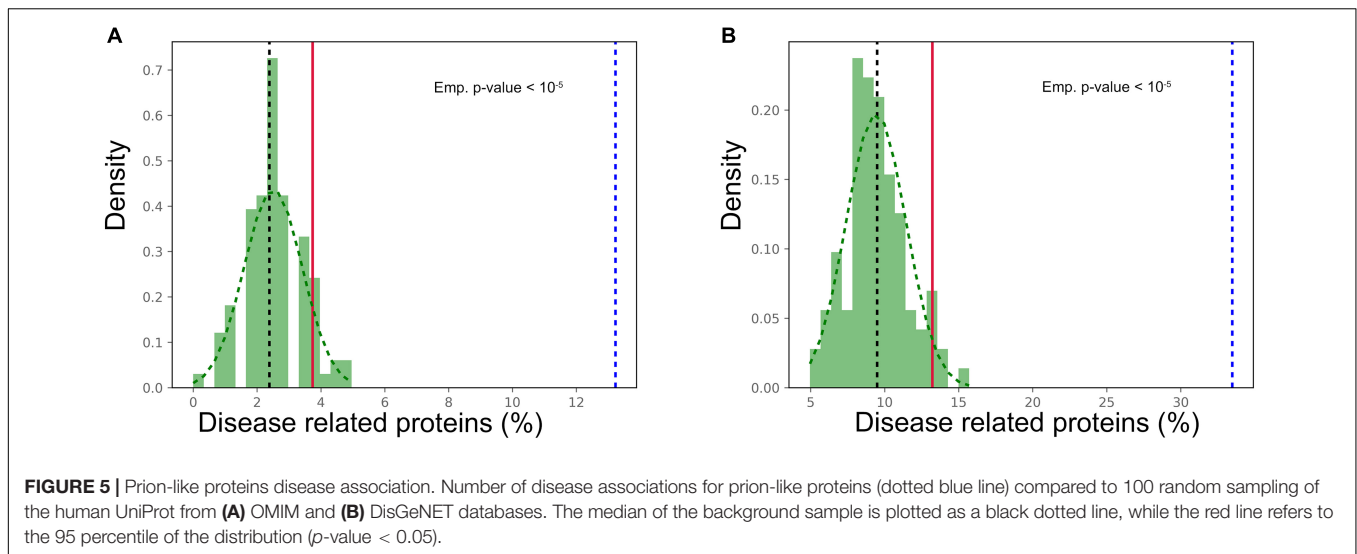
In order to identify interesting cases, we clustered the dataset by representing each gene as a vector of the difference of its expression with respect to the proteome-level tissue average ($V_g = [(E - \bar{E})_1 \ldots (E - \bar{E})_n]$ where V_g: the vector of gene expressions; E: gene expression in tissue n, and $\bar{E}$: average expression of all human proteome in tissue n). The clustering was performed through k-means algorithm implementation of scikit-learn Python module, which uses Euclidean distances by default. We tested cluster numbers from 3 to 10 and chose 6 as the most discriminative one (Average silhouette, **Supplementary Figure S1**). Thus, the highest expression level cluster represents a group of prion-like proteins that are generally over-expressed and remarkably includes most of the human prion-like proteins for which it has been already demonstrated their direct involvement in disease: FUS, TDP-43, hnRNP A1, hnRNP A2/B1, hnRNP A3, hnRNP U, hnRNP H1, and EWS. Many of these proteins have already been described to be spread throughout most tissues and identified at different developmental stages (Bastian et al., 2008; Uhlen et al., 2015).

Prion-Like Proteins Are Disease Related

Given the widespread tissue distribution of the prion-like proteins and the link to disease of proteins in the most expressed cluster, we explored whether, globally, genes encoding for these polypeptides were connected to pathological processes.

Their association to diseases was retrieved separately from the Online Mendelian Inheritance in Man (OMIM) (Amberger et al., 2015) and the database of gene-disease association (DisGeNET) (Pinero et al., 2017). The percentage of genes with disease annotations was calculated and compared with that in the complete human UniProt dataset, which was used as background. According to the OMIM database, 13.22% of the prion-like proteins encoding genes are disease-related against a 2.39% for the UniProt dataset, whereas values of 33.47 and 9.49% were obtained in the case of DisGeNET (p-value < 1.0E-5 for both databases, Z-test). Thus, the association with disease of prion-like proteins was threefold and fivefold higher than the one in the complete human proteome, according to DisGeNET and OMIM, respectively. To assess the significance of this enrichment, 100 random samples with the same size that the prion-like proteins dataset were selected from the background, the percentage of proteins associated to a disease in each sample was counted and the distribution of the percentages calculated (**Figure 5**). For both OMIM and DisGeNET, the prion-like dataset proportion is clearly above the 95 percentile of the distribution, which implies a significant over-representation of disease-associated proteins among human prion-like proteins. At this point, it is important to underline that the prion-like protein identification pipeline is sequence-based and totally blind with respect to the protein annotation.

Prion-like proteins have been associated to the onset of neurological disorders (Harrison and Shorter, 2017). The 9% of genes encoding for prion-like proteins, 11 out of 121, are linked

FIGURE 5 | Prion-like proteins disease association. Number of disease associations for prion-like proteins (dotted blue line) compared to 100 random sampling of the human UniProt from **(A)** OMIM and **(B)** DisGeNET databases. The median of the background sample is plotted as a black dotted line, while the red line refers to the 95 percentile of the distribution (*p*-value < 0.05).

FIGURE 6 | Human prion-like interactome. **(A)** Degree distribution for the complete interactome in red and the prion-like proteins network (first neighbors) in blue. **(B)** Prion-like proteins average interaction degree (blue dotted line) compared to a random sampling of 1000 sets. **(C)** Number of PPIs between prion-like proteins (blue dotted line) compared to a random sampling of 1000 sets.

to neurological diseases, according to OMIM (**Supplementary Table S3**). This constitutes a significant enrichment, relative to the complete proteome (*p*-value < 1.5 E-8). However, it is important to note that, despite proteins connected with neurological disorders are over enriched by 1.4-fold within the disease associated prion-like protein subgroup, this enrichment is not statistically significant (*p*-value > 0.11). It is clear from the results presented above that many of the detected proteins are ubiquitous regulators involved in a wide range of signaling pathways; which suggests that perturbations affecting their function may have a great impact in multiple disorders and not exclusively in neurological diseases, as it is usually assumed.

Prion-Like Proteins' Role in Highly Interconnected Subnetworks

Proteins rarely perform their functions independently; but mostly rely on complexes to carry them out. The connectivity of human prion-like proteins and the properties of their interactors were analyzed. As above, prion-like proteins were first mapped

to genes to obtain unique entities. Out of the 121 resulting genes, 100 had annotated physical binary interactions (physical interactions between two individual proteins). Overall, prion-like dataset and the proteins they interact with establish a subnetwork of 1544 proteins with 2079 PPIs between them. Both the prion-like dataset and the complete subnetwork have higher average interaction degrees than the human interactome (**Figure 6A**). To uncover whether prion-like proteins interact more than expected by chance, the average degree of interactions of the prion-like protein set was compared with 1000 random sets of proteins of the same size (**Figure 6B**). This analysis confirms that prion-like proteins exhibit a significant higher number of interactions than the average human interactome. Next, we assessed whether prion-like proteins interact more between them than expected by chance, by comparing the number of intra-set interactions with that in 1000 random sets, as before. The results showed that prion-like proteins establish more interactions – one order of magnitude higher – between them than expected randomly (**Figure 6C**). To further describe the human prion-like subnetwork, it was tested to what extent prion-like proteins

	Prion-like protein set	Random expectation	Z-score	P-value
LCC size	32	3.16	16.8	$<1.10^{-5}$
MSD	1.64	2.2	−6.78	$<1.10^{-5}$

cluster into specialized interactome neighborhoods. The size of the LCC and the MSD was measured and compared to 1000 random sets (**Table 1**). The results clearly show that prion-like proteins share a higher interactomic vicinity than expected randomly, providing support to the concept that they exist well-defined interaction networks for human prion-like proteins.

To functionally characterize this subnetwork of prion-like proteins and their interactors, the 1544 proteins were analyzed for GO and KEGG pathways enrichment. GO enrichment analysis are consistent with the results obtained for the prion-like proteins dataset alone, as it highlights regulation of gene expression through DNA and RNA binding as the main biological role played by this protein subset (**Supplementary Table S4**).

When we examined the statistically enriched pathways obtained from KEGG analysis, we observed that they can be grouped into two main clusters. Remarkably, the largest cluster collects pathways involved in different types of cancer, such as transcriptional misregulation in cancer (p-value $< 9.86E-15$, 53 proteins), pancreatic cancer (p-value $< 2.01E-12$, 29 proteins), prostate cancer (p-value $< 1.31E-11$, 33 proteins), or colorectal cancer (p-value $< 1.88E-7$, 22 proteins) among others; 12 prion-like proteins (10% of the total unique entries) and 122 (8.4%) of their interactors were found in these cancer related pathways. These interactors include cornerstones in mitogenesis, growth factor signaling, apoptotic attenuation, cell cycle progression, angiogenesis, cell invasion, immune regulation, and microenvironment alterations.

The second group encompasses pathways associated with viral infection, such as viral carcinogenesis (p-value $< 1.42E-15$, 61 proteins), Epstein–Barr virus infection (p-value $< 3.91E-14$, 56 proteins), herpes simplex infection (p-value $< 6.6E-12$, 51 proteins), or hepatitis C (p-value $< 1.07E-6$, 38 proteins). This is consistent with the involvement of RBPs, helicases, and splicing-related proteins in the control of viral assembly and trafficking of the viral genomic RNA from the nucleus. Prion-like candidates such as DDX17 (Moy et al., 2014), DDX5 (Cheng et al., 2018), and hnRNP A2B2 (Levesque et al., 2006) have been already described to play key roles in these processes.

DISCUSSION

In the present work, we used a stringent computational approach that considers that PrLDs should not be only disordered and compositionally biased, but also encode for short sequences with moderate, but significant, amyloid propensity (Sabate et al., 2015b). We concluded that 242 polypeptides in the human proteome fulfill the requirements to potentially behave

as prion-like proteins. This accounts for less than 1% of the human proteins, which implies that, compared with organisms like *Plasmodium* or *Dictyostelium* where 10–25% of their proteins are predicted be prionogenic (Singh et al., 2004), the prionic load of the human proteome is low. The dataset included several widely studied proteins with prion-like behavior, such as FUS, TIA1, TDP-43, EWS, and several hnRNPs, but also previously undescribed proteins with very important cellular functions: members of the mediator complex, nucleoporins, chromatin remodeling proteins, and transcription factors.

As their counterparts in yeast (Santoso et al., 2000; Alberti et al., 2009), human prion-like proteins, locate their PrLDs mostly at their ends; with a slight preference for the amino terminus. This might imply that the position of the PrLD within the protein sequence might be relevant for its function. Indeed, previous analyses on proteins containing low complexity regions already suggested that these terminal positions would allow them to act as act as promiscuous interfaces for protein binding, without steric interferences by the adjacent globular domains (Coletta et al., 2010). In a similar manner, prion-like modularity and the preference for terminal regions are likely maintained in order to delimit a flexible region which can switch its conformation and assemble, modulating in this way the activity of folded domains without impacting their native 3D structure.

According to the GO terms analysis, a highly significant fraction of prion-like proteins are involved in functions related to nucleic acid binding and transcription and translation activities. This includes proteins of the Mediator complex, implicated in the regulated transcription of nearly all RNA polymerase II-dependent genes (Zhu et al., 2015; Cho et al., 2018), proteins recruited in chromatin-remodeling complexes (Boulay et al., 2017; Kataoka and Mochizuki, 2017), and a significant number of transcription factors. The dataset also includes the large majority of RBPs already described to behave as prion-like in humans, such as FUS which is implicated in transcription, DNA repair, and RNA biogenesis (Patel et al., 2015), TIA1 which functions in mRNA turnover and regulation of translation (Li et al., 2014), TDP-43 which is involved in transcriptional regulation and RNA processing (Buratti and Baralle, 2008; King et al., 2012), EWS which is implicated in RNA binding and processing, or diverse hnRNPs involved in the packaging of pre-mRNA into RNP particles (He and Smith, 2009). Not surprisingly, we found that a high proportion of these proteins map into the nucleus and intracellular ribonucleoprotein complex. This last observation is consistent the extensive literature identifying prion-like sequences as drivers of liquid–liquid phase separation in membrane-less cellular compartments (Patel et al., 2015; Banani et al., 2017).

Our data reveal that human prion-like proteins are multifunctional proteins involved in important regulatory processes. Indeed, 50% of the proteins in our dataset carry at least two different Pfam domains. As expected from the molecular functions in which these proteins are involved, the most statistically enriched domains correspond to RNA and DNA binding domains such as the canonical RRM, the Zn finger domain, the forkhead domain, or the helicase domain. All of them present in well-characterized transcription factors and

RNPs. These are evolutionary conserved domains in which, because of their functional relevance, genetic mutations are often linked to disease (King et al., 2012; Cascarina and Ross, 2014).

We assessed the expression of genes coding for prion-like proteins for each human tissue, to try to rationalize why, so far, these proteins have been mostly related to neurological diseases. Human prion-like protein expression was not restricted to nervous tissue but ubiquitously spread among tissues; also, they are not especially abundant in the brain, relative to other organs of the human body. This suggests that they play a physiological role in different cellular types, although it raises the question of why most prion-like proteins related diseases are tissue-specific. This situation is not unique for prion-like proteins but common to other proteins involved in neurodegenerative disorders, i.e., a-synuclein the protein responsible for Parkinson's disease, is abundantly expressed in both the cerebral cortex and the bone marrow, but only aggregates in the brain (Spillantini et al., 1997; Barbour et al., 2008). The protein quality control machinery has an active role in managing protein misfolding and aggregation. Cellular aging impacts cell homeostasis and leads to proteostatic-compromised cells in which misfolding and aggregation events cannot be compensated (Aguzzi and Altmeyer, 2016). It has been proposed that the low efficacy of replacing dying neurons, relative to other cells types, could be one of the underlying reasons why the malfunction of prion-like proteins is more often associated to neurological conditions. One important finding here is that many of the human prion-like proteins that have been convincingly associated to disease are among the most expressed polypeptides in the dataset. This fits very well with the so-called "life at the edge" hypothesis, which states that, because protein aggregation is extremely dependent on concentration, abundant proteins are, on the average, at highest risk of misfolding and aggregation (Tartaglia et al., 2007).

Independently of their tissue distribution, what becomes clear from the analysis of the OMIM and DisGeNet databases is that human prion-like proteins are strongly connected to disease. Two complementary properties might explain, at least in part, this strong association. First, the propensity of PrLDs to establish intermolecular interactions together with the presence of regions with significant amyloid propensity, exposed to solvent within large disordered regions, impose an inherent risk to aggregate to these polypeptides. In fact, genetic mutations that increase the aggregation propensity of PrLDs have been shown to be directly associated with disease (Harrison and Shorter, 2017). Second, according to the "centrality-lethality rule" (Jeong et al., 2001), the highest the number of interactions for a protein is, the largest is the impact of its disruption on cell function. Thus, the high connectivity of prion-like proteins networks might well

account for their strong link to human diseases. Importantly, KEGG pathway enrichment analysis of the prion-like proteins interactome allowed us to uncover a highly significant association with two previously undescribed set of devastating pathological processes: cancer and viral infections.

Overall, despite the present study constitutes only a first theoretical approach to the function of human prion-like proteins, our results indicate that this subproteome exert important regulatory functions in different biological pathways, thanks to both their protein-protein and protein-nucleic acids binding capabilities, two properties that seem to be favored by their modular architecture. The analysis suggests that in the forthcoming years, we can expect the discovery of a connection between prion-like proteins malfunction and other pathologies apart from neurological disorders.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and/or the **Supplementary Files**.

AUTHOR CONTRIBUTIONS

IP, PA, ST, and SV conceived the experiments and analyzed the results. VI, LP, and TJ-B conducted the experiments and prepared the figures. VI, IP, and SV wrote the main manuscript text. All authors reviewed the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2019.00314/full#supplementary-material

REFERENCES

Afsar Minhas, F. U., Ross, E. D., and Ben-Hur, A. (2017). Amino acid composition predicts prion activity. *PLoS Comput. Biol.* 13:e1005465. doi: 10.1371/journal.pcbi.1005465

Aguzzi, A., and Altmeyer, M. (2016). Phase separation: linking cellular compartmentalization to disease. *Trends Cell Biol.* 26, 547–558. doi: 10.1016/j.tcb.2016.03.004

Aguzzi, A., and Rajendran, L. (2009). The transcellular spread of cytosolic amyloids, prions, and prionoids. *Neuron* 64, 783–790. doi: 10.1016/j.neuron.2009.12.016

Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137, 146–158. doi: 10.1016/j.cell.2009.02.044

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2015). OMIM.org: online mendelian inheritance in man (OMIM(R)),

an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 43, D789–D798. doi: 10.1093/nar/gku1205

An, L., and Harrison, P. M. (2016). The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biol. Direct* 11:32. doi: 10.1186/s13062-016-0134-5

Anderson, P., Kedersha, N., and Ivanov, P. (2015). Stress granules, P-bodies and cancer. *Biochim. Biophys. Acta* 1849, 861–870. doi: 10.1016/j.bbagrm.2014. 11.009

Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298. doi: 10.1038/nrm.2017.7

Barbour, R., Kling, K., Anderson, J. P., Banducci, K., Cole, T., Diep, L., et al. (2008). Red blood cells are the major source of alpha-synuclein in blood. *Neurodegener. Dis.* 5, 55–59. doi: 10.1159/000112832

Bastian, F., Parmentier, G., Roux, J., Moretti, S., Laudet, V., and Robinson-Rechavi, M. (2008). "Bgee: Integrating and Comparing Heterogeneous Transcriptome Data Among Species," in *Data Integration in the Life Sciences: 5th International Workshop, DILS 2008, Evry, France, June 25-27, 2008. Proceedings*, eds A. Bairoch, S. Cohen-Boulakia, and C. Froidevaux (Berlin: Springer), 124–131. doi: 10.1007/978-3-540-69828-9_12

Batlle, C., de Groot, N. S., Iglesias, V., Navarro, S., and Ventura, S. (2017a). Characterization of soft amyloid cores in human prion-like proteins. *Sci. Rep.* 7:12134. doi: 10.1038/s41598-017-09714-z

Batlle, C., Fernandez, M. R., Iglesias, V., and Ventura, S. (2017b). Perfecting prediction of mutational impact on the aggregation propensity of the ALS-associated hnRNPA2 prion-like protein. *FEBS Lett.* 591, 1966–1971. doi: 10. 1002/1873-3468.12698

Batlle, C., Iglesias, V., Navarro, S., and Ventura, S. (2017c). Prion-like proteins and their computational identification in proteomes. *Expert Rev. Proteo.* 14, 335–350. doi: 10.1080/14789450.2017.1304214

Baxa, U., Wickner, R. B., Steven, A. C., Anderson, D. E., Marekov, L. N., Yau, W. M., et al. (2007). Characterization of beta-sheet structure in Ure2p1-89 yeast prion fibrils by solid-state nuclear magnetic resonance. *Biochemistry* 46, 13149–13162. doi: 10.1021/bi700826b

Boulay, G., Sandoval, G. J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., et al. (2017). Cancer-specific retargeting of baf complexes by a prion-like domain. *Cell* 171, 163.e119–178.e119. doi: 10.1016/j.cell.2017.07.036

Buratti, E., and Baralle, F. E. (2008). Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. *Front. Biosci.* 13:867–878. doi: 10.2741/ 2727

Cascarina, S. M., and Ross, E. D. (2014). Yeast prions and human prion-like proteins: sequence features and prediction methods. *Cell Mol. Life Sci.* 71, 2047–2063. doi: 10.1007/s00018-013-1543-6

Chakrabortee, S., Kayatekin, C., Newby, G. A., Mendillo, M. L., Lancaster, A., and Lindquist, S. (2016). Luminidependens (LD) is an *Arabidopsis* protein with prion behavior. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6065–6070. doi: 10.1073/pnas. 1604478113

Cheng, W., Chen, G., Jia, H., He, X., and Jing, Z. (2018). DDX5 RNA helicases: emerging roles in viral infection. *Int. J. Mol. Sci.* 19:E1122. doi: 10.3390/ ijms19041122

Cho, W. K., Spille, J. H., Hecht, M., Lee, C., Li, C., Grube, V., et al. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415. doi: 10.1126/science.aar4199

Coletta, A., Pinney, J. W., Solis, D. Y., Marsh, J., Pettifer, S. R., and Attwood, T. K. (2010). Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* 4:43. doi: 10.1186/1752-0509-4-43

Couthouis, J., Hart, M. P., Erion, R., King, O. D., Diaz, Z., Nakaya, T., et al. (2012). Evaluating the role of the FUS/TLS-related gene EWSR1 in amyo-trophic lateral sclerosis. *Hum. Mol. Genet.* 21, 2899–2911. doi: 10.1093/hmg/ dds116

Couthouis, J., Hart, M. P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., et al. (2011). A yeast functional screen predicts new candidate ALS disease genes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20881–20890. doi: 10.1073/pnas. 1109434108

Duernberger, Y., Liu, S., Riemschoss, K., Paulsen, L., Bester, R., Kuhn, P. H., et al. (2018). Prion replication in the mammalian cytosol: functional regions within a prion domain driving induction, propagation and inheritance. *Mol. Cell Biol.* 38, e111–e118. doi: 10.1128/MCB.00111-18

Espinosa Angarica, V., Angulo, A., Giner, A., Losilla, G., Ventura, S., and Sancho, J. (2014). PrionScan: an online database of predicted prion domains in complete proteomes. *BMC Genomics* 15:102. doi: 10.1186/1471-2164-15-102

Espinosa Angarica, V., Ventura, S., and Sancho, J. (2013). Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genomics* 14:316. doi: 10.1186/1471-2164-14-316

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Franklin, B. S., Bossaller, L., De Nardo, D., Ratter, J. M., Stutz, A., Engels, G., et al. (2014). The adaptor ASC has extracellular and 'prionoid' activities that propagate inflammation. *Nat. Immunol.* 15, 727–737. doi: 10.1038/ni.2913

Fuller-Pace, F. V. (2013). The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochim. Biophys. Acta* 1829, 756–763. doi: 10.1016/j.bbagrm.2013.03.004

Gene Ontology Consortium (2015). Gene ontology consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056. doi: 10.1093/nar/gku1179

Gitler, A. D., and Shorter, J. (2011). RNA-binding proteins with prion-like domains in ALS and FTLD-U. *Prion* 5, 179–187. doi: 10.4161/pri.5.3.17230

Hafner-Bratkovic, I., Bester, R., Pristovsek, P., Gaedtke, L., Veranic, P., Gaspersic, J., et al. (2011). Globular domain of the prion protein needs to be unlocked by domain swapping to support prion protein conversion. *J. Biol. Chem.* 286, 12149–12156. doi: 10.1074/jbc.M110.213926

Halfmann, R., Jarosz, D. F., Jones, S. K., Chang, A., Lancaster, A. K., and Lindquist, S. (2012). Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature* 482, 363–368. doi: 10.1038/nature10875

Halfmann, R., and Lindquist, S. (2010). Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. *Science* 330, 629–632. doi: 10.1126/science.1191081

Harrison, A. F., and Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438. doi: 10.1042/ BCJ20160499

Harrison, P. M., and Gerstein, M. (2003). A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. *Genome Biol.* 4:R40. doi: 10.1186/gb-2003-4-6-r40

Harrison, P. M., Khachane, A., and Kumar, M. (2010). Genomic assessment of the evolution of the prion protein gene family in vertebrates. *Genomics* 95, 268–277. doi: 10.1016/j.ygeno.2010.02.008

He, Y., and Smith, R. (2009). Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol. Life Sci.* 66, 1239–1256. doi: 10.1007/s00018-008-8532-1

Hou, F., Sun, L., Zheng, H., Skaug, B., Jiang, Q. X., and Chen, Z. J. (2011). MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* 146, 448–461. doi: 10.1016/j.cell.2011.06.041

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Iglesias, V., de Groot, N. S., and Ventura, S. (2015). Computational analysis of candidate prion-like proteins in bacteria and their role. *Front. Microbiol.* 6:1123. doi: 10.3389/fmicb.2015.01123

Jeong, H., Mason, S. P., Barabasi, A. L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* 411, 41–42. doi: 10.1038/35075138

Ju, S., Tardiff, D. F., Han, H., Divya, K., Zhong, Q., Maquat, L. E., et al. (2011). A yeast model of FUS/TLS-dependent cytotoxicity. *PLoS Biol.* 9:e1001052. doi: 10.1371/journal.pbio.1001052

Kataoka, K., and Mochizuki, K. (2017). Heterochromatin aggregation during DNA elimination in tetrahymena is facilitated by a prion-like protein. *J. Cell Sci.* 130, 480–489. doi: 10.1242/jcs.195503

Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 495, 467–473. doi: 10.1038/ nature11922

King, O. D., Gitler, A. D., and Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* 1462, 61–80. doi: 10.1016/j.brainres.2012.01.016

Lancaster, A. K., Nutter-Upham, A., Lindquist, S., and King, O. D. (2014). PLAAC: a web and command-line application to identify proteins with prion-like

amino acid composition. *Bioinformatics* 30, 2–3. doi: 10.1093/bioinformatics/btu310

Levesque, K., Halvorsen, M., Abrahamyan, L., Chatel-Chaix, L., Poupon, V., Gordon, H., et al. (2006). Trafficking of HIV-1 RNA is mediated by heterogeneous nuclear ribonucleoprotein A2 expression and impacts on viral assembly. *Traffic* 7, 1177–1193. doi: 10.1111/j.1600-0854.2006.00461.x

Li, L., and Lindquist, S. (2000). Creating a protein-based element of inheritance. *Science* 287, 661–664. doi: 10.1126/science.287.5453.661

Li, X., Rayman, J. B., Kandel, E. R., and Derkatch, I. L. (2014). Functional role of Tia1/Pub1 and Sup35 prion domains: directing protein synthesis machinery to the tubulin cytoskeleton. *Mol. Cell* 55, 305–318. doi: 10.1016/j.molcel.2014.05.027

Loomis, P. A., Howard, T. H., Castleberry, R. P., and Binder, L. I. (1990). Identification of nuclear tau isoforms in human neuroblastoma cells. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8422–8426. doi: 10.1073/pnas.87.21.8422

Luk, K. C., Kehm, V., Carroll, J., Zhang, B., O'Brien, P., Trojanowski, J. Q., et al. (2012). Pathological alpha-synuclein transmission initiates parkinson-like neurodegeneration in nontransgenic mice. *Science* 338, 949–953. doi: 10.1126/science.1227157

Majumdar, A., Cesario, W. C., White-Grindley, E., Jiang, H., Ren, F., Khan, M. R., et al. (2012). Critical role of amyloid-like oligomers of drosophila orb2 in the persistence of memory. *Cell* 148, 515–529. doi: 10.1016/j.cell.2012.01.004

Malinovska, L., Palm, S., Gibson, K., Verbavatz, J. M., and Alberti, S. (2015). Dictyostelium discoideum has a highly Q/N-rich proteome and shows an unusual resilience to protein aggregation. *Proc. Natl. Acad. Sci. U.S.A.* 112, E2620–E2629. doi: 10.1073/pnas.1504459112

March, Z. M., King, O. D., and Shorter, J. (2016). Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* 1647, 9–18. doi: 10.1016/j.brainres.2016.02.037

Mazurek, A., Luo, W., Krasnitz, A., Hicks, J., Powers, R. S., and Stillman, B. (2012). DDX5 regulates DNA replication and is required for cell proliferation in a subset of breast cancer cells. *Cancer Discov.* 2, 812–825. doi: 10.1158/2159-8290.CD-12-0116

McGlinchey, R. P., Kryndushkin, D., and Wickner, R. B. (2011). Suicidal [PSI+] is a lethal yeast prion. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5337–5341. doi: 10.1073/pnas.1102762108

Medina, E., Cordova, C., Villalobos, P., Reyes, J., Komives, E. A., Ramirez-Sarmiento, C. A., et al. (2016). Three-Dimensional domain swapping changes the folding mechanism of the forkhead domain of foxP1. *Biophys. J.* 110, 2349–2360. doi: 10.1016/j.bpj.2016.04.043

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., et al. (2015). Disease networks. uncovering disease-disease relationships through the incomplete interactome. *Science* 347:1257601. doi: 10.1126/science.1257601

Michelitsch, M. D., and Weissman, J. S. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11910–11915. doi: 10.1073/pnas.97.22.11910

Mosca, R., Ceol, A., and Aloy, P. (2013). Interactome3D: adding structural details to protein networks. *Nat. Methods* 10, 47–53. doi: 10.1038/nmeth.2289

Moy, R. H., Cole, B. S., Yasunaga, A., Gold, B., Shankarling, G., Varble, A., et al. (2014). Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell* 158, 764–777. doi: 10.1016/j.cell.2014.06.023

Nakayashiki, T., Kurtzman, C. P., Edskes, H. K., and Wickner, R. B. (2005). Yeast prions [URE3] and [PSI+] are diseases. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10575–10580. doi: 10.1073/pnas.0504882102

Newby, G. A., and Lindquist, S. (2013). Blessings in disguise: biological benefits of prion-like mechanisms. *Trends Cell Biol.* 23, 251–259. doi: 10.1016/j.tcb.2013.01.007

Nomura, T., Watanabe, S., Kaneko, K., Yamanaka, K., Nukina, N., and Furukawa, Y. (2014). Intranuclear aggregation of mutant FUS/TLS as a molecular pathomechanism of amyotrophic lateral sclerosis. *J. Biol. Chem.* 289, 1192–1202. doi: 10.1074/jbc.M113.516492

Pallares, I., de Groot, N. S., Iglesias, V., Sant'Anna, R., Biosca, A., Fernandez-Busquets, X., et al. (2018). Discovering putative prion-like proteins in plasmodium falciparum: a computational and experimental analysis. *Front. Microbiol.* 9:1737. doi: 10.3389/fmicb.2018.01737

Pallares, I., Iglesias, V., and Ventura, S. (2015). The Rho Termination factor of clostridium botulinum contains a prion-like domain with a highly amyloidogenic core. *Front. Microbiol.* 6:1516. doi: 10.3389/fmicb.2015.01516

Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., et al. (2015). A liquid-to-solid phase transition of the ALS Protein FUS accelerated by disease mutation. *Cell* 162, 1066–1077. doi: 10.1016/j.cell.2015.07.047

Pinero, J., Bravo, A., Queralt-Rosinach, N., Gutierrez-Sacristan, A., Deu-Pons, J., Centeno, E., et al. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 45, D833–D839. doi: 10.1093/nar/gkw943

Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., et al. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* 2015:bav028. doi: 10.1093/database/bav028

Prusiner, S. B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science* 216, 136–144. doi: 10.1126/science.6801762

Reddy, B. P., Shrestha, S., Hart, K. J., Liang, X., Kemirembe, K., Cui, L., et al. (2015). A bioinformatic survey of RNA-binding proteins in PLasmodium. *BMC Genomics* 16:890. doi: 10.1186/s12864-015-2092-1

Sabate, R., Rousseau, F., Schymkowitz, J., Batlle, C., and Ventura, S. (2015a). Amyloids or prions? That is the question. *Prion* 9, 200–206. doi: 10.1080/19336896.2015.1053685

Sabate, R., Rousseau, F., Schymkowitz, J., and Ventura, S. (2015b). What makes a protein sequence a prion? *PLoS Comput. Biol.* 11:e1004013. doi: 10.1371/journal.pcbi.1004013

Sant'Anna, R., Fernández, M. R., Batlle, C., Navarro, S., de Groot, N. S., Serpell, L., et al. (2016). Characterization of amyloid cores in prion domains. *Sci. Rep.* 6:34274. doi: 10.1038/srep34274

Santoso, A., Chien, P., Osherovich, L. Z., and Weissman, J. S. (2000). Molecular basis of a yeast prion species barrier. *Cell* 100, 277–288. doi: 10.1016/S0092-8674(00)81565-2

Si, K. (2015). Prions: what are they good for? *Annu. Rev. Cell Dev. Biol.* 31, 149–169. doi: 10.1146/annurev-cellbio-100913-013409

Si, K., Choi, Y. B., White-Grindley, E., Majumdar, A., and Kandel, E. R. (2010). Aplysia CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation. *Cell* 140, 421–435. doi: 10.1016/j.cell.2010.01.008

Si, K., and Kandel, E. R. (2016). The role of functional prion-like proteins in the persistence of memory. *Cold Spring Harb. Perspect. Biol.* 8:a021774. doi: 10.1101/cshperspect.a021774

Sikorska, B., and Liberski, P. P. (2012). Human prion diseases: from kuru to variant creutzfeldt-jakob disease. *Subcell. Biochem.* 65, 457–496. doi: 10.1007/978-94-007-5416-4_17

Singh, G. P., Chandra, B. R., Bhattacharya, A., Akhouri, R. R., Singh, S. K., and Sharma, A. (2004). Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in *Plasmodium falciparum. Mol. Biochem. Parasitol.* 137, 307–319. doi: 10.1016/j.molbiopara.2004.05.016

Spillantini, M. G., Schmidt, M. L., Lee, V. M., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997). Alpha-synuclein in lewy bodies. *Nature* 388, 839–840. doi: 10.1038/42166

Stohr, J., Watts, J. C., Mensinger, Z. L., Oehler, A., Grillo, S. K., DeArmond, S. J., et al. (2012). Purified and synthetic Alzheimer's amyloid beta (Abeta) prions. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11025–11030. doi: 10.1073/pnas.1206555109

Tartaglia, G. G., Pechmann, S., Dobson, C. M., and Vendruscolo, M. (2007). Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* 32, 204–206. doi: 10.1016/j.tibs.2007.03.005

Toombs, J. A., Petri, M., Paul, K. R., Kan, G. Y., Ben-Hur, A., and Ross, E. D. (2012). De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6519–6524. doi: 10.1073/pnas.1119366109

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419. doi: 10.1126/science.1260419

UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989

van Rheede, T., Smolenaars, M. M., Madsen, O., and de Jong, W. W. (2003). Molecular evolution of the mammalian prion protein. *Mol. Biol. Evol.* 20, 111–121. doi: 10.1093/molbev/msg014

Villarroya-Beltri, C., Gutierrez-Vazquez, C., Sanchez-Cabo, F., Perez-Hernandez, D., Vazquez, J., Martin-Cofreces, N., et al. (2013). Sumoylated

hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat. Commun.* 4:2980. doi: 10.1038/ncomms3980

Wang, I. F., Chang, H. Y., Hou, S. C., Liou, G. G., Way, T. D., and James Shen, C. K. (2012). The self-interaction of native TDP-43 C terminus inhibits its degradation and contributes to early proteinopathies. *Nat. Commun.* 3:766. doi: 10.1038/ncomms1766

Wang, J., Choi, J. M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., et al. (2018). A molecular grammar governing the driving forces for phase separation of prion-like rna binding proteins. *Cell* 174, 688.e616–699.e616. doi: 10.1016/j.cell.2018.06.006

Xu, H., He, X., Zheng, H., Huang, L. J., Hou, F., Yu, Z., et al. (2014). Structural basis for the prion-like MAVS filaments in antiviral innate immunity. *eLife* 3:e01489. doi: 10.7554/eLife.01489

Yuan, A. H., and Hochschild, A. (2017). A bacterial global regulator forms a prion. *Science* 355, 198–201. doi: 10.1126/science.aai7776

Zambrano, R., Conchillo-Sole, O., Iglesias, V., Illa, R., Rousseau, F., Schymkowitz, J., et al. (2015). PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res.* 43, W331–W337. doi: 10.1093/nar/gkv490

Zhu, X., Chen, L., Carlsten, J. O., Liu, Q., Yang, J., Liu, B., et al. (2015). Mediator tail subunits can form amyloid-like aggregates in vivo and affect stress response in yeast. *Nucleic Acids Res.* 43, 7306–7314. doi: 10.1093/nar/gkv629