

# Deep conservation of prion-like composition in the eukaryotic prion-former Pub1/Tia1 family and its relatives

Wan-Chun Su and Paul M. Harrison

Department of Biology, McGill University, Montreal, QC, Canada

## ABSTRACT

Pub1 protein is an important RNA-binding protein functional in stress granule assembly in budding yeast *Saccharomyces cerevisiae* and, as its co-ortholog Tia1, in humans. It is unique among proteins in evidencing prion-like aggregation in both its yeast and human forms. Previously, we noted that Pub1/Tia1 was the only protein linked to human disease that has prion-like character and has demonstrated such aggregation in both species. Thus, we were motivated to probe further into the evolution of the Pub1/Tia1 family (and its close relative Nam8 and its orthologs) to gain a picture of how such a protein has evolved over deep evolutionary time since the last common ancestor of eukaryotes. Here, we discover that the prion-like composition of this protein family is deeply conserved across eukaryotes, as is the prion-like composition of its close relative Nam8/Ngr1. A sizeable minority of protein orthologs have multiple prion-like domains within their sequences (6–20% depending on criteria). The number of RNA-binding RRM domains is conserved at three copies over >86% of the Pub1 family (>71% of the Nam8 family), but proteins with just one or two RRM domains occur frequently in some clades, indicating that these are not due to annotation errors. Overall, our results indicate that a basic scaffold comprising three RNA-binding domains and at least one prion-like region has been largely conserved since the last common ancestor of eukaryotes, providing further evidence that prion-like aggregation may be a very ancient and conserved phenomenon for certain specific proteins.

Submitted 5 November 2019

Accepted 30 March 2020

Published 17 April 2020

Corresponding author

Paul M. Harrison,  
paul.harrison@mcgill.ca

Academic editor

Joseph Gillespie

Additional Information and  
Declarations can be found on  
page 10

DOI [10.7717/peerj.9023](https://doi.org/10.7717/peerj.9023)

© Copyright

2020 Su and Harrison

Distributed under

Creative Commons CC-BY 4.0

**OPEN ACCESS**

**Subjects** Bioinformatics, Evolutionary Studies

**Keywords** Prion, Pub1, Tia1, Nam8, RNA-binding, Compositional bias, Intrinsic disorder, Yeast, Human, Evolution

## INTRODUCTION

Prions in eukaryotes have been linked to diseases, evolutionary capacitance, large-scale genetic control and long-term memory formation. Prion formation and propagation have been studied extensively in the budding yeast *Saccharomyces cerevisiae*. The yeast *S. cerevisiae* has >200 prion-like proteins that have N/Q-rich domains of the sort observed in >10 known prion-formers (An, Fitzpatrick & Harrison, 2016). A domain of the yeast protein Pub1 exhibits prion-like protein aggregation and propagation in *S. cerevisiae* (Alberti et al., 2009). Pub1 in *S. cerevisiae* is a poly-(A)<sup>+</sup>-binding protein that functions in

stress granule assembly, and is required for mRNA stability for many transcripts (Anderson, Paddy & Swanson, 1993; Matunis, Matunis & Dreyfuss, 1993; Rayman & Kandel, 2017). Pub1 protein abundance is linked to increased DNA replication stress (Tkach et al., 2012). There is evidence that the N/Q-rich prion-like domain of Pub1 is functional in recruiting protein synthesis machinery to the cytoskeleton, which is critical for maintaining cytoskeletal integrity (Li et al., 2014). In doing so, Pub1 interacts with the (PSI<sup>+</sup>)-prion-forming Sup35 protein to form a two-protein self-propagating state (Li et al., 2014). Pub1 can act as an accessory translation factor that appears to fine-tune translation termination efficiency (Urakov et al., 2017). Similarly to Sup35 (Franzmann et al., 2018), Pub1 forms condensates upon cellular stress (Kroschwald et al., 2018). Condensates induced by different stresses have different biophysical characteristics, i.e., some are gel-like, while others are more solid-like (Kroschwald et al., 2018). Another dimension to the relationship between Sup35 and Pub1 is that the latter appears to be involved in a cellular mechanism to alleviate toxicity of the (PSI<sup>+</sup>) prion, which is formed from Sup35 proteins (Urakov et al., 2018). Pub1 is one of a small number of prion-formers that serve as protein interaction hubs for other prion-like proteins (Harbi & Harrison, 2014b). The Pub1 prion-forming domain is one of the most rapidly evolving prion-forming domains sequence-wise (Su & Harrison, 2019). The Pub1 paralog Nam8 (which also contains a prion-like domain) is a part of the U1 snRNP protein complex; mutants of Nam8 are defective in meiotic recombination (Puig et al., 1999; Qiu, Schwer & Shuman, 2011). Nam8 was one of a small panel of genes that were discovered in a genome-wide genetic screen of yeast to rescue the cytotoxicity of FUS/TLS, a prion-like protein that causes some familial amyotrophic lateral sclerosis (Ju et al., 2011). Nam8 is non-essential in the U1 snRNP protein complex, but demonstrates synthetic-lethal genetic interactions with other non-essential interactors (Agarwal, Schwer & Shuman, 2019).

The Pub1 human homologue, Tia1, is also functional through prion-like aggregation in stress granule assembly in mammalian cells (Gilks et al., 2004; Rayman & Kandel, 2017). The prion-like domain of Tia1 can be substituted with that of Sup35 and still maintain its stress granule assembly role, and forms protease-resistant aggregates in mammalian cells (Gilks et al., 2004). Full-length Tia1 can form self-perpetuating prion-like aggregates heterologously in *S. cerevisiae* (Li et al., 2014). Tia1 and its paralog Tiar (also called TiaL1) together regulate stress response and cell-cycle progression mediated by the protein kinase EIF2AK2 (Meyer et al., 2018). Reduction of Tia1 levels protects against tau-mediated neurodegeneration in the brain, rescues behavioural deficits and improves survival (Apicco et al., 2018). Single-nucleotide polymorphisms (SNPs) in Tia1 are linked at very low frequency to amyotrophic lateral sclerosis and frontotemporal dementia in Chinese cohorts but not in European ones (Baradaran-Heravi et al., 2018; Gu et al., 2018; Zhang et al., 2018). SNPs in Tia1 are linked to a multisystem proteinopathy that involves disruption of stress granule formation (Lee et al., 2018), and to Welander distal myopathy (Hackman et al., 2013; Klar et al., 2013).

All of these proteins (Pub1, Nam8, etc.) are examples of proteins that contain RNA-binding RRM domains. The RRM domain provides RNA-binding functionality for a diverse set of proteins across the eukaryotic clade, including dozens in *S. cerevisiae*.

It comprises a very conserved alpha–beta protein fold that can bind to a large number of diverse RNA sequences (*Daubner, Clery & Allain, 2013*).

Previously, we noted that Pub1/Tia1 was the only protein linked to human disease that has prion-like character and aggregation in both humans and *S. cerevisiae* (*An & Harrison, 2016*). Here, motivated by these observations, we have probed further into the evolution of the Pub1/Tia1 family, including the close relative Nam8 and its orthologs. We discover that prion-like composition is deeply conserved across a wide diversity of eukaryotes, with some clades of proteins conserving multiple prion-like domains within their individual sequences.

## METHODS

### Data

Sets of orthologs for the following protein families were downloaded from OrthoDB database version 10.1 at the Eukaryota level (*Kriventseva et al., 2019*): Pub1/Tia1 (OrthoDB group 1066369at2579), Nam8/Ngr1 (group 775799at2579) and Pab1 (group 1027234at2759). These data also contain paralogs. In addition, lists of reciprocal-best-hits (RBH) orthologs were made for the same proteins across sets of reference eukaryotic proteomes downloaded from UniProt version 2019\_1 (*UniProt Consortium, 2019*). These were cross-referenced with the OrthoDB sets to determine the subset of OrthoDB annotations that are such RBH orthologs. Also, only OrthoDB proteins that are from the UniProt reference proteomes were considered for further analysis. FASTA-format files of the Pub1/Tia1/Tiar and Nam8/Ngr1 families of proteins are provided as Supplemental Data ([Datas S1](#) and [S2](#)).

### Multiple sequence alignment and phylogenetic trees

Multiple sequence alignments were constructed using KMAD (*Lange, Wyrwicz & Vriend, 2016*), using default parameters. KMAD is designed to handle sequences with intrinsic disorder, a trait expected characteristically of prion-like proteins (*Harbi & Harrison, 2014b*). Phylogenetic trees were made using PHYML with Bayesian information criterion and aBayes branch support (*Guindon et al., 2005*), with the exception of a tree including extra paralogs described below. The aBayes branch support is a Bayesian-like transformation of the standard approximate likelihood ratio test implemented in PhyML, that is very fast and has high statistical power (*Anisimova et al., 2011*).

To construct a phylogenetic tree with extra paralogs, we used a stepwise procedure of tree construction. Examination of BLASTP (*Altschul et al., 1997*) output indicates that Pab1 and its orthologs were the next most similar RRM-domain-containing proteins in yeast that are widely conserved across eukaryotes ([Table S3](#)). Pab1 is a polyadenylate-binding protein that is an important interactor at various stages of messenger mRNA stability, biogenesis and translation. We used similarities to Pab1 to help us to filter for further paralogs to the Pub1/Tia1 and Nam8/Ngr1 proteins as follows. BLASTP was used to compare each ortholog from the Pub1/Tia1, Nam8/Ngr1 and Pab1 protein sets against their own proteomes, and any significant matches (with  $e$ -value  $\leq 1E-04$ ) against the majority (>50% length) of a proteome sequence that matched a Pub1/Tia1 or Nam8/Ngr1 sequence in preference to a

Pab1 sequence were labelled as potential paralogs of Pub1/Tia1 or Nam8/Ngr1, formed more recently in specific eukaryotic clades. Then, a phylogenetic tree was constructed using the RaxML programme (using the PROTGAMMAJTTF mutation parameters and the ‘-f a’ flag for rapid bootstrap analysis and 100 bootstrap replicates) (*Stamatakis, 2015*) with the Pub1/Tia1, Nam8/Ngr1 and Pab1 ortholog sets plus the additional potential paralogs proteins as input. The tree was pruned at an appropriate node to remove all of the Pab1 sequences and associated paralogs to leave just those paralogs that were formed more recently than the last common ancestor of the Pub1/Tia1 and Nam8 proteins.

### **Annotation of RNA-binding (RRM) domains**

Positions of RRM RNA-binding domains were taken from annotations in InterPro version 66.0 from the database Pfam (*El-Gebali et al., 2019; Mitchell et al., 2019*). We checked these annotations against those also for the PROSITE and SMART databases (*Letunic & Bork, 2018; Sigrist et al., 2010*), and they differ by <1% of extra or fewer domain annotations, which gave us confidence to proceed with the Pfam set. Since there were some OrthoDB sequences that do not have Pfam annotations, and other annotations may be incomplete, we generated additional RRM domain annotations. Firstly, the sequences for the RRM domain annotations were extracted and then compared against the ASTRAL SCOPE database version 2.07 (*Fox, Brenner & Chandonia, 2014*) using BLASTP (*Altschul et al., 1997*) with *e*-value threshold  $1E-04$ , to pull out the identities of protein domain structures that are for RRM domains. Then the ASTRAL SCOPE protein domains plus the sequence fragments for the RRM-domain annotations were then formed into a BLASTP database to compare against the OrthoDB sequences to augment the existing list of RRM-domain annotations (again using *e*-value threshold  $1E-04$ ). These were then reduced for overlap by sorting them in decreasing order of domain length and progressively flagging overlappers further down the list for deletion.

### **Annotation of prion-like domains and intrinsic disorder**

Prion-like domains were annotated using the PLAAC programme with default parameters (*Lancaster et al., 2014*). Prion-like domains were counted using two thresholds for the PLAAC log likelihood ratio (LLR) score:  $>0.0$  and  $\geq 15.0$ . Any score  $>0.0$  indicates some predicted propensity to form a prion-like domain; the second threshold of 15.0 was picked since the lowest LLR value for a known fungal prion-forming domain that was picked out for experimental analysis is  $\sim 16.0$  in the fission yeast prion-former Ctr4 (*Sideri et al., 2017*). The programme fLPS was used to detect compositional bias for glutamine and asparagine residues, which is a bias characteristic of the substantial majority of the known prion-forming domains in budding yeast and other organisms (*Harrison, 2006; Harrison, 2017*). Expected fractional frequencies = 0.05 for both glutamine and asparagine were used in running fLPS. fLPS Binomial *P*-values indicate the degree of compositional bias, and here we picked the most biased subsequence for either Q or N or Q + N for each sequence. The log<sub>10</sub> of these *P*-values is investigated here for all the data analysis.

To examine for multiple prion-like domains, the sequences were split up into fragments with the ends delimited by RRM domains and other protein domains (plus a

10 amino-acid-residue buffer added onto both ends), which were then input into the PLAAC programme. Prion-forming N/Q-rich domains are highly intrinsically disordered (Harbi & Harrison, 2014b). Intrinsic disorder was annotated using the IUPRED2a and DISOPRED3 programmes with default parameters (Dosztanyi, 2018; Ward et al., 2004). The total percentage of intrinsic disorder was calculated for each whole protein sequence, since there is not an intrinsic disorder score for a specific tract that can be extracted from the output of the disorder annotation programmes. Intrinsic disorder assignments by IUPRED2a and DISOPRED3 are strongly correlated ( $R = 0.43$  for Pub1,  $R = 0.47$  for Nam8,  $P$ -values  $< 1E-20$ ).

## RESULTS

### Evolution within *Saccharomyces*

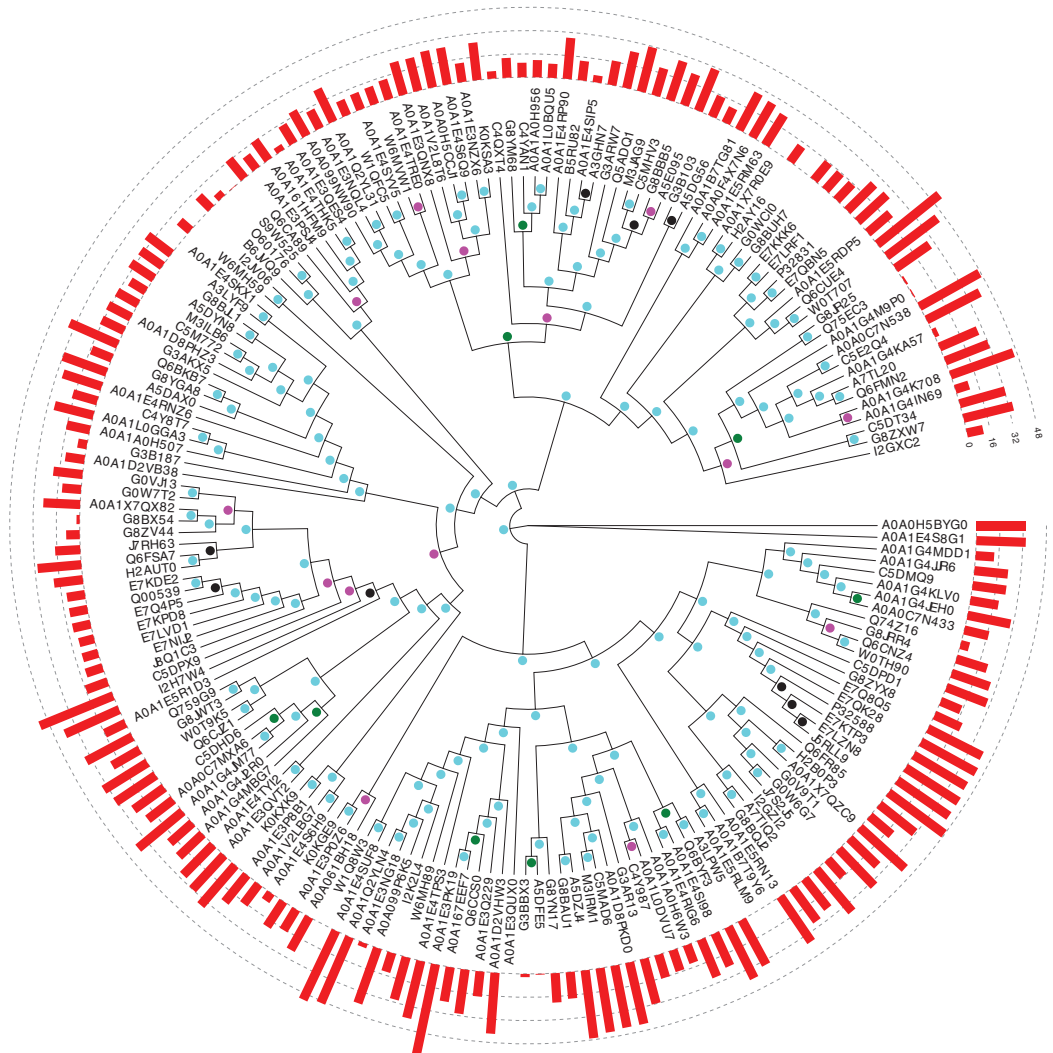
In *Saccharomyces*, the three main proteins from the protein families under study are Pub1, Nam8 and Ngr1. Preliminary analysis of phylogenetic trees indicated that Ngr1 formed by duplication from Nam8 in the last common ancestor of *Saccharomyces*. Firstly, we assessed the conservation of prion-like composition across the *Saccharomyces* clade for Pub1, and for its paralogs Nam8 and Ngr1. Prion-like composition is largely maintained for all three paralogs across this clade (Fig. 1). The substantial majority of Pub1 orthologs have PLAAC prion-like region scores  $\geq 15.0$  (53/65, 82%), and almost all  $>0.0$  (61/65, 94%); less so for Nam8 and Ngr1 (73/114, 64%  $\geq 15.0$ ; 113/114, 99%  $>0.0$ ).

### Evolution across eukaryotes

Furthermore, this prion-like composition is deeply conserved across large diverse clades within the whole eukaryotic domain for the Pub1/Tia1 and Nam8/Ngr1 families (Fig. 2). Overall, 90% of sequences have PLAAC LLR scores  $>0.0$ , with 58%  $\geq 15.0$ . Specifically for Pub1, 84% have scores  $>0.0$  and 60% have scores  $\geq 15.0$ . Prion-forming N/Q-rich domains are highly intrinsically disordered (Harbi & Harrison, 2014b). There is some degree of correlation of LLR score with overall percentage annotated intrinsic disorder in the sequences, particularly for Pub1 ( $P < 2E-16$ ; Fig. 3B), but some clades on the phylogenetic tree are observed that have low prion-like character and relatively high intrinsic disorder (Fig. 2).

Prion-like character as judged by the PLAAC programme is maintained in a similar fashion for both the Pub1/Tia1 and Nam8/Ngr1 families (Fig. 3A). PLAAC LLR scores are strongly correlated with the fLPS compositional biases for Q and N residues ( $R = 0.70$  for Nam8, and  $R = 0.52$  for Pub1;  $P < 1E-20$ ), but substantial clades can be seen on the phylogenetic tree that have low fLPS Q/N residue bias, but high PLAAC LLR scores, such as some of the fungal clades clustering near *Saccharomyces*.

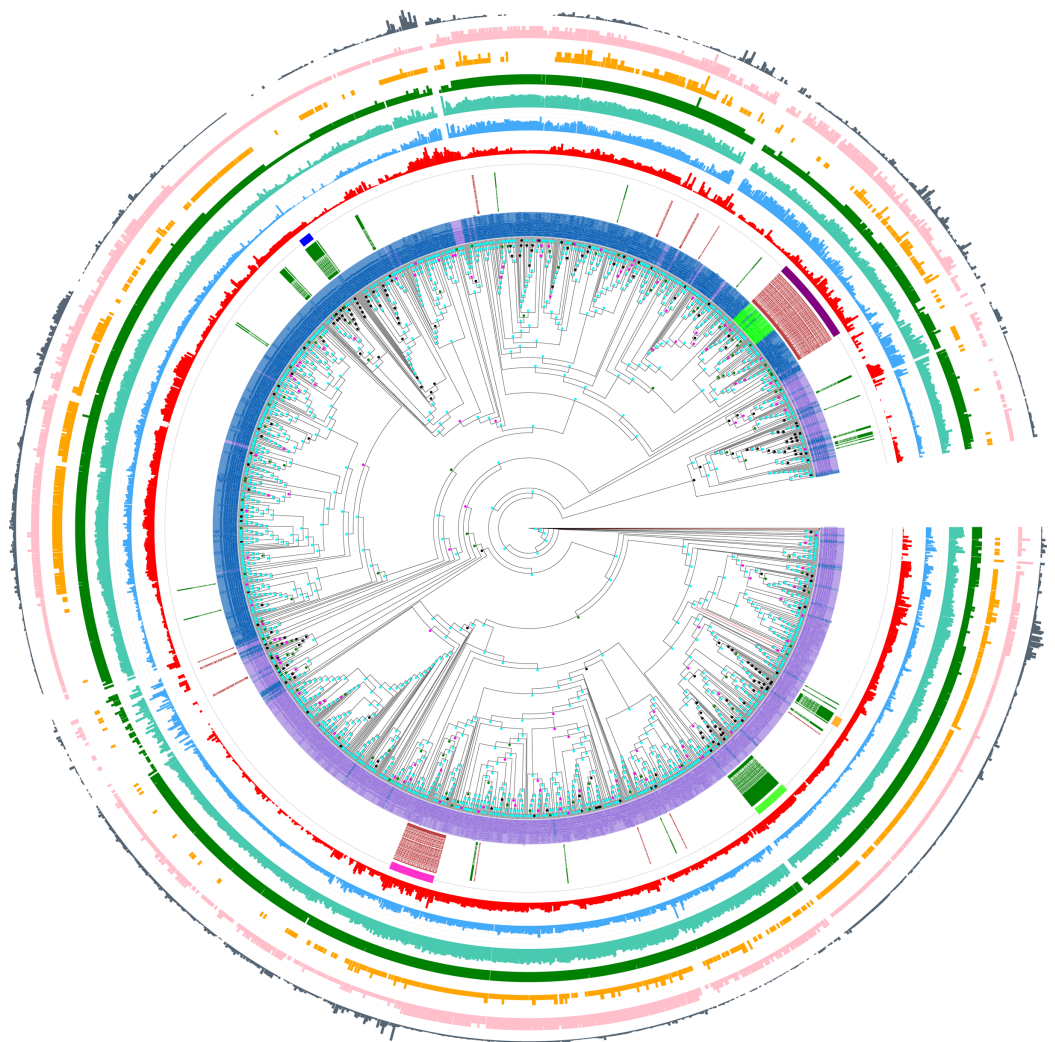
The larger tree with additional paralogs whose construction is described in 'Methods', indicates that no further major paralogs have formed during eukaryotic speciation, since these would appear as intermittent larger blocks in the tree figure (Fig. S1). Such paralogs have however formed sporadically in evolutionarily more recent clades; some of these appear to have lost the prion-like domains, or intrinsic disorder generally (Fig. S1).



**Figure 1** Phylogenetic tree of Pub1 and Nam8 in *Saccharomyces*. This is made using PhyML (Guindon *et al.*, 2005), as described in ‘Methods’, and drawn using Evolview (Subramanian *et al.*, 2019). The UniProt (UniProt Consortium, 2019) names for proteins are used. The aBayes support values are colour-coded black for <math>< 0.5</math>, green for 0.5–0.65, magenta for 0.65–0.8 and cyan for 0.8–1.0. The PLAAC (Lancaster *et al.*, 2014) log likelihood ratio (LLR) prion-like composition scores are depicted as a histogram around the circumference of the tree. The UniProt identifier for Pub1 is P32588, for Nam8 it is Q00539 and for Ngr1 it is P32831. [Full-size](#) DOI: 10.7717/peerj.9023/fig-1

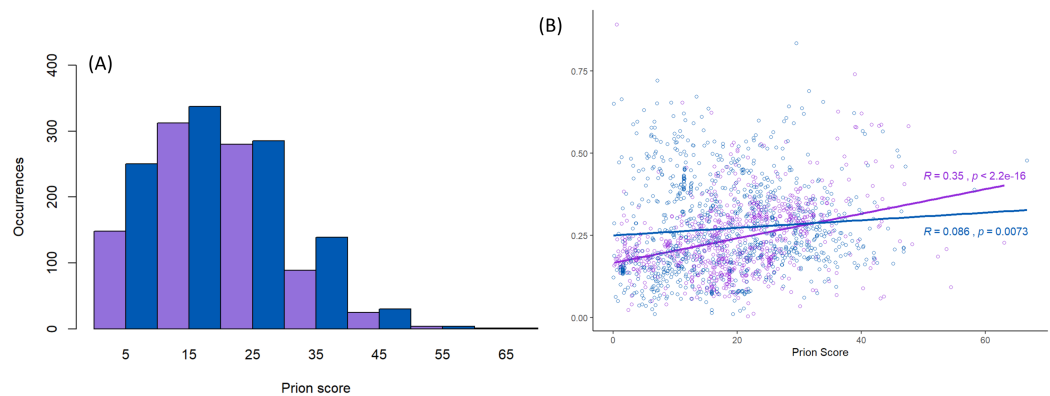
## Multiple prion-like domains occur in several diverse clades

Prion-forming proteins are usually characterised as having one distinct prion-forming area (Harbi & Harrison, 2014a; Harbi *et al.*, 2012). Likewise, prion-like proteins are usually annotated by picking out a single region with the highest adjudged propensity for forming prions; the algorithms that have been derived for annotating prion-like proteins are all based on this principle (Harrison, 2017; Lancaster *et al.*, 2014; Ross *et al.*, 2013; Zambrano *et al.*, 2015). However, it is possible that multiple prion-forming regions can occur within the same protein chain. We decided to check for this in the protein families under study here, since from spot-checking they often contain more than one annotated



**Figure 2** Phylogenetic tree of Pub1/Tia1 and Nam8/Ngr1 for the eukaryotic domain. This is made using PhyML (Guindon *et al.*, 2005), as described in ‘Methods’, and drawn using Evolvie (Subramanian *et al.*, 2019). The aBayes support values are colour-coded black for <0.5, green for 0.5–0.65, magenta for 0.65–0.8 and cyan for 0.8–1.0. The UniProt (UniProt Consortium, 2019) names for proteins are used if available, but otherwise the protein identifier from the OrthoDB database is used (Kriventseva *et al.*, 2019). Pub1/Tia1/Tiar orthologs are coloured purple, Nam8 orthologs blue and Ngr1 green. The circles of annotations are as follows in order from inside to outside: (i) labelling of Mammalian and *Saccharomyces* proteins (green and brown respectively); (ii) colour-coded bars for the Pub1 *Saccharomyces* clade (magenta), mammalian Tia1 (bright green), mammalian Tiar (gold), *Saccharomyces* Nam8/Ngr1 (dark purple), mammalian Nam8 (dark blue); (iii) PLAAC (Lancaster *et al.*, 2014) LLR prion-like composition score (red); (iv) percentage of residues annotated as disordered by IUPred2a (Dosztanyi, 2018) (blue); (v) percentage of residues annotated as disordered by DISOPRED3 (Ward *et al.*, 2004) (teal); (vi) number of RRM domains in each protein (dark green); (vii) number of prion-like domains in each protein, using the threshold of >0.0 (pink); (viii) number of prion-like domains in each protein, using the threshold of  $\geq 15.0$  (gold); (ix)  $\log_{10}$  of the fLPS binomial *P*-value for compositional bias for Q and N residues (Harrison, 2017) (dark grey). The data for this figure are in Tables S2 and S3.

Full-size  DOI: 10.7717/peerj.9023/fig-2



**Figure 3 Scores for prion-like composition.** (A) Distribution of prion-like PLAAC log-likelihood ratio scores (Lancaster *et al.*, 2014) corresponding to the data in Fig. 2, with the same sequence colour-coding except Nam8 and Ngr1 are merged into one set labelled blue (Pub1/Tia1/Tiar and Nam8 + Ngr1). The intervals of prion-like scores are labelled with the value of their midpoints, that is 5 for 0–10, etc. (B) Scatter plot of intrinsic disorder versus prion-like scores with the same sequence colour-coding, with linear regression lines fitted. The *P*-values are two-tailed. [Full-size](#) DOI: 10.7717/peerj.9023/fig-3

intrinsically disordered region delimited by the structured protein domains (specifically here most often the RRM RNA-binding domain). We parsed each protein sequence into fragments delimited by the positions of protein domains and annotated these fragments for prion-like composition, as described in ‘Methods’. We find that it is not uncommon to have multiple prion-like regions in one protein chain, particularly in the Nam8/Ngr1 family (11%  $\geq 2$  prion-like regions for the Nam8/Ngr1 family and 1% for the Pub1/Tia1 family, Figs. 4B and 4C). Indeed, there are specific large clades that maintain two prion-like regions (Fig. 2). Only with a lower PLAAC LLR score threshold of  $>0.0$  does the Pub1/Tia1 family display a sizeable minority of protein sequences with multiple prion-like regions (10% with  $\geq 2$  prion-like domains; Figs. 2, 4B and 4C). A large fraction of these extra domains with prion-like composition are internal in the sequence (46% for Pub1/Tia1 and 29% for Nam8/Ngr1 for PLAAC LLR threshold  $>0.0$ ).

### The number of RNA-binding domains is largely fixed at three

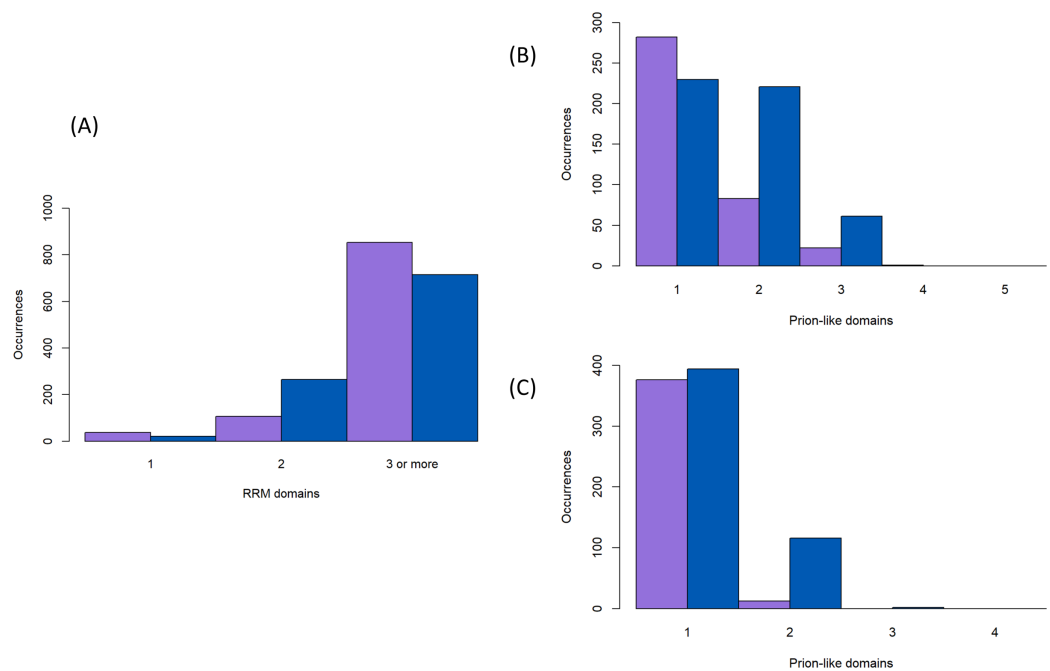
The number of RNA-binding RRM domains is largely fixed at three for both the Nam8/Ngr1 and Pub1/Tia1 families of proteins (70% having three RRM domains, with just 17 sequences having  $>3$ ), but proteins with just one or two RRM domains occur frequently in some specific clades (Figs. 2 and 4A).

## DISCUSSION

### Evolution across eukaryotes

Analysis of the evolution across eukaryotes of prion-like composition for the Pub1/Tia1 and Nam8 families, as judged by the PLAAC programme, indicated that such composition is deeply conserved. We can therefore surmise that this composition is important for the protein function. However there are some clades that: (i) have low prion-like character and maintain relatively high annotated intrinsic disorder; (ii) have high prion-like character as judged by PLAAC, but low bias for N and Q residues, as judged by the fLPS





**Figure 4** Numbers of RNA-binding and prion-like domains. (A) The number of RNA-binding domains for the proteins in Fig. 2 (sequence colour-coding same as Fig. 3). (B and C) The number of prion-like domains for the proteins in Fig. 2 (sequence colour-coding same as Fig. 3), for a PLAAC LLR score of  $>0.0$  and  $\geq 15.0$ .  
[Full-size !\[\]\(fcc3264021d438d9732560e78099f674\_img.jpg\) DOI: 10.7717/peerj.9023/fig-4](https://doi.org/10.7717/peerj.9023/fig-4)

algorithm. Sequences from such clades may be good candidates for checking for alternative compositions for prion-like domains in Pub1 that fulfil either its prion-forming ability or its role in stress granule formation. Sequences of type (ii) above would particularly be interesting candidates for further experimental study for prion-like aggregation, since the prion-like composition is made up more from non-Q + N residues that contribute to prion-forming biases in the known *S. cerevisiae* cases.

### Multiple prion-like domains occur in several diverse clades

The occurrence of multiple prion-like domains in proteins raises the possibility of competing prion-like aggregation by the different prion-like domains within the same protein sequence. However, a large fraction of these extra domains with prion-like composition are internal in the sequence, indicating that they would have greater difficulty forming amyloid fibrils. Also, multiple prion-like domains might affect the character of intracellular condensates, that is as either solid-like or gel-like, perhaps making them more solid-like because of the increased prion-like composition relative to other types of intrinsic disorder propensity in the equivalent regions in other orthologs (Kroschwald *et al.*, 2018).

### The number of RNA-binding domains is largely fixed at three

We found very strong conservation of a three-domain RRM RNA-binding domain pattern in both the Pub1/Tia1 and Nam8 families. Proteins with just one or two RRM domains occur frequently in a conserved pattern in some smaller clades, which indicates that these

are not due to annotation errors and are real evolutionary trends (Figs. 2 and 4A). There are a handful of sequences with four or more RRM domains, but it is unclear whether these may be annotation errors, due to their sporadic occurrence (Fig. 2). One notable cluster that has either as one or two RRM domains (upper left corner of Fig. 2, near clade labelled with blue band), consists mainly of vertebrate Nam8 orthologs that are assigned function in selenocysteine biosynthesis, including TRNAU1AP (transfer RNA selenocysteine 1—associated protein 1) the human RBH ortholog of *S. cerevisiae* Nam8, a protein that has essential functions in both early and later steps of selenocysteine biosynthesis (Small-Howard et al., 2006). Although no human disorders have been directly linked to mutations in the encoding gene of TRNAU1AP, faults in selenocysteine biosynthesis have been linked to complex disorders of metabolism and the development of muscles, skin, the nervous system and immune system (Schweizer & Fradejas-Villar, 2016). These proteins also tend to have lower annotated intrinsic disorder content and lower prion-like composition (Fig. 2).

## CONCLUSIONS

Previously, we noted that Pub1/Tia1 was unique in being linked to human disease while having prion-like character aggregation in both humans and the budding yeast *S. cerevisiae*. Thus, we were motivated to probe further into the evolution of the Pub1/Tia1 family and its close relatives, to gain a picture of how such a protein has evolved over deep evolutionary time since the last common ancestor of eukaryotes. We discovered that the prion-like composition of the Pub1/Tia1 prion-former family, and its closely related Nam8/Ngr1 family is widely conserved across eukaryotes, with 90% of sequences having PLAAC LLR scores >0.0, with 58%  $\geq 15.0$ . Also, the number of RNA-binding domains in these sequences is widely conserved, with 70% of these sequences having three RRM domains. A sizeable minority of the proteins in these ortholog families contain multiple prion-like domains within them, particularly for certain clades of the Nam8/Ngr1 family (~11% of this family having  $\geq 2$  prion-like domains with PLAAC LLR scores  $\geq 15.0$ ).

Overall, these results suggest that a basic scaffold comprising three RNA-binding domains and at least one prion-like region has been widely conserved since the last common ancestor of eukaryotes, and provide some support for the notion that prion-like aggregation may be an anciently conserved phenomenon for certain specific proteins.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This research was funded completely by the Natural Sciences Engineering Research Council of Canada, except that some of the calculations were performed on a desktop computer funded by both the Canada Foundation for Innovation and the Natural Sciences Engineering Research Council of Canada. Computing resources supplied by Compute Canada were also used. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Grant Disclosures

The following grant information was disclosed by the authors:  
Natural Sciences Engineering Research Council of Canada.  
Canada Foundation for Innovation.  
Compute Canada.

## Competing Interests

The authors declare that they have no competing interests.

## Author Contributions

- Wan-Chun Su analysed the data, prepared figures and/or tables, and approved the final draft.
- Paul M. Harrison analysed the data, authored or reviewed drafts of the paper, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

The sequences used to perform the phylogenetic analysis and data generated are available in [Supplemental Files](#).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.9023#supplemental-information>.

## REFERENCES

- Agarwal R, Schwer B, Shuman S. 2019. Domain requirements and genetic interactions of the mud1 subunit of the *Saccharomyces cerevisiae* U1 snRNP. *G3 (Bethesda)* **9(1)**:145–151 DOI 10.1534/g3.118.200781.
- Alberti S, Halfmann R, King O, Kapila A, Lindquist S. 2009. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137(1)**:146–158 DOI 10.1016/j.cell.2009.02.044.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25(17)**:3389–3402 DOI 10.1093/nar/25.17.3389.
- An L, Fitzpatrick D, Harrison PM. 2016. Emergence and evolution of yeast prion and prion-like proteins. *BMC Evolutionary Biology* **16(1)**:24 DOI 10.1186/s12862-016-0594-3.
- An L, Harrison PM. 2016. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biology Direct* **11(1)**:32 DOI 10.1186/s13062-016-0134-5.
- Anderson JT, Paddy MR, Swanson MS. 1993. PUB1 is a major nuclear and cytoplasmic polyadenylated RNA-binding protein in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **13(10)**:6102–6113 DOI 10.1128/MCB.13.10.6102.
- Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology* **60(5)**:685–699 DOI 10.1093/sysbio/syr041.

- Apicco DJ, Ash PEA, Maziuk B, LeBlang C, Medalla M, Al Abdullatif A, Ferragud A, Botelho E, Ballance HI, Dhawan U, Boudeau S, Cruz AL, Kashy D, Wong A, Goldberg LR, Yazdani N, Zhang C, Ung CY, Tripodis Y, Kanaan NM, Ikezu T, Cottone P, Leszyk J, Li H, Luebke J, Bryant CD, Wolozin B. 2018. Reducing the RNA binding protein TIA1 protects against tau-mediated neurodegeneration in vivo. *Nature Neuroscience* 21(1):72–80  
DOI 10.1038/s41593-017-0022-z.
- Baradaran-Heravi Y, Dillen L, Nguyen HP, Van Mossevelde S, Baets J, De Jonghe P, Engelborghs S, De Deyn PP, Vandenbulcke M, Vandenberghe R, Van Damme P, Cras P, Salmon E, Synofzik M, Heutink P, Wilke C, Simon-Sanchez J, Rojas-Garcia R, Turon-Sans J, Lleó A, Illán-Gala I, Clarimón J, Borroni B, Padovani A, Pastor P, Diez-Fairen M, Aguilar M, Gelpi E, Sanchez-Valle R, Borrego-Ecija S, Matej R, Parobkova E, Nacmias B, Sorbi S, Bagnoli S, De Mendonça A, Ferreira C, Fraidakis MJ, Diehl-Schmid J, Alexopoulos P, Almeida MR, Santana I, Van Broeckhoven C, Van der Zee J, BELNEU Consortium, EU EOD Consortium. 2018. No supportive evidence for TIA1 gene mutations in a European cohort of ALS-FTD spectrum patients. *Neurobiology of Aging* 69:293.e9–293.e11.
- Daubner GM, Clery A, Allain FH-T. 2013. RRM–RNA recognition: NMR or crystallography... and new findings. *Current Opinion in Structural Biology* 23(1):100–108  
DOI 10.1016/j.sbi.2012.11.006.
- Dosztanyi Z. 2018. Prediction of protein disorder based on IUPred. *Protein Science* 27(1):331–340  
DOI 10.1002/pro.3334.
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* 47(D1):D427–D432 DOI 10.1093/nar/gky995.
- Fox NK, Brenner SE, Chandonia J-M. 2014. SCOPe: structural classification of proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research* 42(D1):D304–D309 DOI 10.1093/nar/gkt1240.
- Franzmann TM, Jahnel M, Pozniakovskiy A, Mahamid J, Holehouse AS, Nuske E, Richter D, Baumeister W, Grill SW, Pappu RV, Hyman AA, Alberti S. 2018. Phase separation of a yeast prion protein promotes cellular fitness. *Science* 359(6371):eaa05654  
DOI 10.1126/science.aao5654.
- Gilks N, Kedersha N, Ayodele M, Shen L, Stoecklin G, Dember LM, Anderson P. 2004. Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Molecular Biology of the Cell* 15(12):5383–5398 DOI 10.1091/mbc.e04-08-0715.
- Gu X, Chen Y, Wei Q, Cao B, Ou R, Yuan X, Hou Y, Zhang L, Liu H, Chen X, Shang H-F. 2018. Mutation screening of the TIA1 gene in Chinese patients with amyotrophic lateral sclerosis/frontotemporal dementia. *Neurobiology of Aging* 68:161.e1–161.e3  
DOI 10.1016/j.neurobiolaging.2018.04.010.
- Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research* 33(Web Server):W557–W559 DOI 10.1093/nar/gki352.
- Hackman P, Sarparanta J, Lehtinen S, Vihola A, Evilä A, Jonson PH, Luque H, Kere J, Screen M, Chinnery PF, Åhlberg G, Edström L, Udd B. 2013. Welander distal myopathy is caused by a mutation in the RNA-binding protein TIA1. *Annals of Neurology* 73(4):500–509  
DOI 10.1002/ana.23831.
- Harbi D, Harrison PM. 2014a. Classifying prion and prion-like phenomena. *Prion* 8(2):161–165  
DOI 10.4161/pri.27960.

- Harbi D, Harrison PM. 2014b.** Interaction networks of prion, prionogenic and prion-like proteins in budding yeast, and their role in gene regulation. *PLOS ONE* **9(6)**:e100615 DOI [10.1371/journal.pone.0100615](https://doi.org/10.1371/journal.pone.0100615).
- Harbi D, Parthiban M, Gendoo DM, Ehsani S, Kumar M, Schmitt-Ulms G, Sowdhamini R, Harrison PM. 2012.** PrionHome: a database of prions and other sequences relevant to prion phenomena. *PLOS ONE* **7**:e31785.
- Harrison PM. 2006.** Exhaustive assignment of compositional bias reveals universally prevalent biased regions: analysis of functional associations in human and *Drosophila*. *BMC Bioinformatics* **7(1)**:441 DOI [10.1186/1471-2105-7-441](https://doi.org/10.1186/1471-2105-7-441).
- Harrison PM. 2017.** fLPS: fast discovery of compositional biases for the protein universe. *BMC Bioinformatics* **18**:476 DOI [10.1186/s12859-017-1906-3](https://doi.org/10.1186/s12859-017-1906-3).
- Ju S, Tardiff DF, Han H, Divya K, Zhong Q, Maquat LE, Bosco DA, Hayward LJ, Brown RH Jr, Lindquist S, Ringe D, Petsko GA. 2011.** A yeast model of FUS/TLS-dependent cytotoxicity. *PLOS Biology* **9(4)**:e1001052 DOI [10.1371/journal.pbio.1001052](https://doi.org/10.1371/journal.pbio.1001052).
- Klar J, Sobol M, Melberg A, Mabert K, Ameer A, Johansson AC, Feuk L, Entesarian M, Orlen H, Casar-Borota O, Dahl N. 2013.** Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing. *Human Mutation* **34**:572–577 DOI [10.1002/humu.22282](https://doi.org/10.1002/humu.22282).
- Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simao FA, Zdobnov EM. 2019.** OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* **47**:D807–D811 DOI [10.1093/nar/gky1053](https://doi.org/10.1093/nar/gky1053).
- Kroschwald S, Munder MC, Maharana S, Franzmann TM, Richter D, Ruer M, Hyman AA, Alberti S. 2018.** Different material states of Pub1 condensates define distinct modes of stress adaptation and recovery. *Cell Reports* **23(11)**:3327–3339 DOI [10.1016/j.celrep.2018.05.041](https://doi.org/10.1016/j.celrep.2018.05.041).
- Lancaster AK, Nutter-Upham A, Lindquist S, King OD. 2014.** PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30(17)**:2501–2502 DOI [10.1093/bioinformatics/btu310](https://doi.org/10.1093/bioinformatics/btu310).
- Lange J, Wyrwicz LS, Vriend G. 2016.** KMAD: knowledge-based multiple sequence alignment for intrinsically disordered proteins. *Bioinformatics* **32(6)**:932–936 DOI [10.1093/bioinformatics/btv663](https://doi.org/10.1093/bioinformatics/btv663).
- Lee Y, Jonson PH, Sarparanta J, Palmio J, Sarkar M, Vihola A, Evila A, Suominen T, Penttila S, Savarese M, Johari M, Minot M-C, Hilton-Jones D, Maddison P, Chinnery P, Reimann J, Kornblum C, Kraya T, Zierz S, Sue C, Goebel H, Azfer A, Ralston SH, Hackman P, Bucelli RC, Taylor JP, Weihl CC, Udd B. 2018.** TIA1 variant drives myodegeneration in multisystem proteinopathy with SQSTM1 mutations. *Journal of Clinical Investigation* **128(3)**:1164–1177 DOI [10.1172/JCI97103](https://doi.org/10.1172/JCI97103).
- Letunic I, Bork P. 2018.** 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* **46(D1)**:D493–D496 DOI [10.1093/nar/gkx922](https://doi.org/10.1093/nar/gkx922).
- Letunic I, Bork P. 2019.** Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47(W1)**:W256–W259 DOI [10.1093/nar/gkz239](https://doi.org/10.1093/nar/gkz239).
- Li X, Rayman JB, Kandel ER, Derkatch IL. 2014.** Functional role of Tia1/Pub1 and Sup35 prion domains: directing protein synthesis machinery to the tubulin cytoskeleton. *Molecular Cell* **55(2)**:305–318 DOI [10.1016/j.molcel.2014.05.027](https://doi.org/10.1016/j.molcel.2014.05.027).
- Matunis MJ, Matunis EL, Dreyfuss G. 1993.** PUB1: a major yeast poly(A)<sup>+</sup> RNA-binding protein. *Molecular and Cellular Biology* **13(10)**:6114–6123 DOI [10.1128/MCB.13.10.6114](https://doi.org/10.1128/MCB.13.10.6114).

- Meyer C, Garzia A, Mazzola M, Gerstberger S, Molina H, Tuschl T. 2018. The TIA1 RNA-binding protein family regulates EIF2AK2-mediated stress response and cell cycle progression. *Molecular Cell* 69(4):622–635.e6 DOI 10.1016/j.molcel.2018.01.011.
- Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong S-Y, Finn RD. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research* 47(D1):D351–D360 DOI 10.1093/nar/gky1100.
- Puig O, Gottschalk A, Fabrizio P, Seraphin B. 1999. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes & Development* 13(5):569–580 DOI 10.1101/gad.13.5.569.
- Qiu ZR, Schwer B, Shuman S. 2011. Defining the Mer1 and Nam8 meiotic splicing regulons by cDNA rescue. *RNA* 17(9):1648–1654 DOI 10.1261/rna.2792011.
- Rayman JB, Kandel ER. 2017. TIA-1 is a functional prion-like protein. *Cold Spring Harbor Perspectives in Biology* 9(5):a030718 DOI 10.1101/cshperspect.a030718.
- Ross ED, Maclea KS, Anderson C, Ben-Hur A. 2013. A bioinformatics method for identifying Q/N-rich prion-like domains in proteins. *Methods in Molecular Biology* 1017:219–228 DOI 10.1007/978-1-62703-438-8\_16.
- Schweizer U, Fradejas-Villar N. 2016. Why 21? The significance of selenoproteins for human health revealed by inborn errors of metabolism. *FASEB Journal* 30(11):3669–3681 DOI 10.1096/fj.201600424.
- Sideri T, Yashiroda Y, Ellis D, Rodriguez-Lopez M, Yoshida M, Tuite M, Bahler J. 2017. The copper transport-associated protein Ctr4 can form prion-like epigenetic determinants in *Schizosaccharomyces pombe*. *Microbial Cell* 4(1):16–28 DOI 10.15698/mic2017.01.552.
- Sigrist CJA, Cerutti L, De Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research* 38(Suppl. 1):D161–D166 DOI 10.1093/nar/gkp885.
- Small-Howard A, Morozova N, Stoytcheva Z, Forry EP, Mansell JB, Harney JW, Carlson BA, Xu XM, Hatfield DL, Berry MJ. 2006. Supramolecular complexes mediate selenocysteine incorporation in vivo. *Molecular and Cellular Biology* 26(6):2337–2346 DOI 10.1128/MCB.26.6.2337-2346.2006.
- Stamatakis A. 2015. Using RA×ML to infer phylogenies. *Current Protocols in Bioinformatics* 51(1):6.14.1–6.14.14 DOI 10.1002/0471250953.bi0614s51.
- Su T-Y, Harrison PM. 2019. Conservation of prion-like composition and sequence in prion-formers and prion-like proteins of *Saccharomyces cerevisiae*. *Frontiers in Molecular Biosciences* 6:54 DOI 10.3389/fmolb.2019.00054.
- Subramanian B, Gao S, Lercher MJ, Hu S, Chen W-H. 2019. Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Research* 47(W1):W270–W275 DOI 10.1093/nar/gkz357.
- Tkach JM, Yimit A, Lee AY, Riffle M, Costanzo M, Jaschob D, Hendry JA, Ou J, Moffat J, Boone C, Davis TN, Nislow C, Brown GW. 2012. Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature Cell Biology* 14(9):966–976 DOI 10.1038/ncb2549.

- UniProt Consortium. 2019.** UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47(D1):D506–D515** DOI [10.1093/nar/gky1049](https://doi.org/10.1093/nar/gky1049).
- Urakov VN, Mitkevich OV, Dergalev AA, Ter-Avanesyan MD. 2018.** The Pub1 and Upf1 proteins act in concert to protect yeast from toxicity of the [PSI<sup>+</sup>] prion. *International Journal of Molecular Sciences* **19(11):3663** DOI [10.3390/ijms19113663](https://doi.org/10.3390/ijms19113663).
- Urakov VN, Mitkevich OV, Safenkova IV, Ter-Avanesyan MD. 2017.** Ribosome-bound Pub1 modulates stop codon decoding during translation termination in yeast. *FEBS Journal* **284(12):1914–1930** DOI [10.1111/febs.14099](https://doi.org/10.1111/febs.14099).
- Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004.** Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* **337(3):635–645** DOI [10.1016/j.jmb.2004.02.002](https://doi.org/10.1016/j.jmb.2004.02.002).
- Zambrano R, Conchillo-Sole O, Iglesias V, Illa R, Rousseau F, Schymkowitz J, Sabate R, Daura X, Ventura S. 2015.** PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Research* **43(W1):W331–W337** DOI [10.1093/nar/gkv490](https://doi.org/10.1093/nar/gkv490).
- Zhang K, Liu Q, Shen D, Tai H, Fu H, Liu S, Wang Z, Shi J, Ding Q, Li X, Liu M, Cui L, Zhang X. 2018.** Genetic analysis of TIA1 gene in Chinese patients with amyotrophic lateral sclerosis. *Neurobiology of Aging* **67:201.e9–201.e10** DOI [10.1016/j.neurobiolaging.2018.03.010](https://doi.org/10.1016/j.neurobiolaging.2018.03.010).