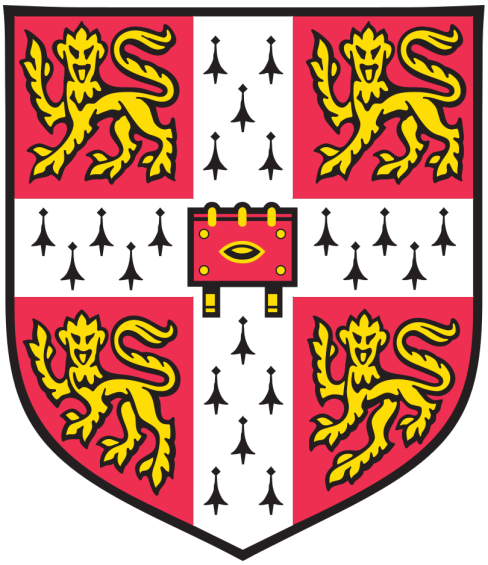
# The Life of Prion: an investigation into the physiological role of a prion-like protein in the nematode *Caenorhabditis elegans*



**Joshua Newman**

MRC Laboratory of Molecular Biology

University of Cambridge

This thesis is submitted for the degree of
*Doctor of Philosophy*

Downing College

April 2020

## Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text (see section 6.0 Experimental Contributions).

I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed 60,000-word limit for the Faculty of Biology Degree Committee, exclusive of tables, footnotes, bibliography, and appendices.

In the Introduction, Figure 1.1B is adapted from Ruttkay-Nedecky et al. (2015), whilst Figures 1.4A-E contain panels adapted from Fitzpatrick et al. (2018), Falcon et al. (2019a,b), Wasmer et al. (2008), and Wille et al. (2009), respectively. Figures 1.6 and 1.7 are adapted from WormAtlas. Additionally, Figure 1.8A was adapted from Rieckher et al. (2013), whilst Figure 1.10 and 1.11 are adapted from BioRender templates. In Chapter 3, Figure 3.2 is adapted from Schwartz and Jorgenson (2016), whilst Figures 3.7C and 3.7D contain graphs adapted from Molliex et al. (2015).

In Chapter 2, the lifespan analysis in Figure 2.4 was conducted by Dr. Rebecca Taylor. In Chapter 3, the bacterial purification from BL21 *E. coli* and *Drosophila* S2 cells, the preparation of imaging chambers, and the subsequent *in vitro* phase separation assays were conducted in collaboration with Joseph Watson. The mass spectrometry in Chapter 4 was performed by the Mass Spectrometry facility at the MRC LMB

# The Life of Prion: an investigation into the physiological role of a prion-like protein in the nematode *Caenorhabditis elegans*

*Joshua Newman, MRC Laboratory of Molecular Biology*

Summary

For centuries, the threat of prion disease has plagued populations – whether it be in the form of scrapie ravaging through the sheep populations of Spain in the eighteenth century, fatal familial insomnia afflicting families in Italy, or an outbreak of Creutzfeldt-Jakob disease in the UK triggered by the consumption of contaminated beef. The ability of prions to spread with such pathogenic intent whilst remaining incurable fuels the fire of public concern - yet delving beyond that classical view of prions has led to an enlightened appreciation of the role of such proteins throughout the biota of life.

Prion-like proteins subvert the dogma of the one-to-one structure-function relationship traditionally held onto when considering the physiology of proteins. Instead, these proteins display structural polymorphism, capable of adopting different conformations under different conditions; this expands the capabilities of protein function and represents a mechanism for epigenetic coding, with protein conformation possessing the ability convey and propagate specific states for biological advantage without the need for DNA modification. The properties observed in many of these proteins, including those of multivalency and low complexity, predispose them towards phase separation, whereby localised concentration of biomolecules compartmentalise cellular activity. This ability confers upon an organism the ability to respond to stimuli in dynamic and transient fashions without needing to use cumbersome membrane-bound organelles.

Using the Prion-Like Amino Acid Composition program, we have identified the stress-responsive protein ABU-13 as a potential prion in the nematode species, *Caenorhabditis elegans*. Phenotypic analysis of ABU-13 knockout animals has identified a non-redundant role for this protein in ER stress and innate immune responses – in keeping with previous evidence that this family of proteins is involved with a non-canonical unfolded protein response pathway of the endoplasmic reticulum. These knockout animals do not display defects in the activation of the stress-response gene *hsp-4* upon tunicamycin treatment, suggesting that these effects occur independently of the canonical UPR pathway.

*In vivo* and *in vitro* characterisation of ABU-13 has demonstrated a propensity for coalescence, reminiscent of a phase-separated organelle. This is further supported by the ability of this protein to bind RNA – a common feature of prion-like proteins involved with such transitions. Co-immunoprecipitation of ABU-13 confirmed an enrichment of ER and RNA-related protein interactions, supporting a role for this protein in ER stress responses. In addition to this, a number of glycosylation related terms were identified, pointing towards a role for ABU-13 in the folding quality control of such proteins.

These punctated structures do, however, appear less mobile than traditional liquid-like condensates, as demonstrated by a slow FRAP recovery, indicative of a more hydrogel-like structure. To further investigate this, we extracted hydrogel-forming proteins from N2 animals using a biotinylated isoxazole precipitation, identifying a number of novel proteins that may be involved with in the formation of such solid-like structures. ABU-13, however, was not amongst this dataset, suggesting that if it were indeed involved in a physiological hydrogel, this transition is driven by another protein component.

Overall, we propose that ABU-13 represents a phase-separating prion-like protein capable of modulating ER stress and immune responses, potentially via the regulation of RNA processing.

## Acknowledgements

There are, of course, many people who have contributed to the completion of this thesis, both in a professional and a personal capacity. First and foremost, I would like to extend my deepest gratitude to my supervisor, Dr. Rebecca Taylor, who has guided and advised me throughout the project, as well as encouraging and supporting my independent scientific pursuits as they developed.

I would also like to thank my fellow lab members, Dr. Soudabeh Imanikia, Dr. Ming Sheng, Dr. Neşem Özbey, Dr. Nadia Cummins and Dr. Max Thompson, for their supervision and guidance throughout my time at the MRC LMB. They have also proved to be excellent friends and a much-needed support system, especially during the last few months as my thesis drew to a close amongst the chaos of an emerging pandemic. On a much more practical level, I would like to thank Soudabeh for her help and teaching, particularly when I first began my PhD. In addition to this, I would like to thank all of those at the MRC LMB, particularly those within the scientific support facilities and to those I consider good friends, without whom I could not have finished this thesis.

Outside of the lab, I have also been extremely lucky to be supported by my family - in particular by my Mum and two sisters, as well as my partner, Will. Without their unconditional patience, love and encouragement, I would have struggled to find that extra motivation and perseverance to keep pushing during those particularly stressful times.

The pursuit of a PhD extends well beyond an academic challenge, but has proved itself to be a formative time within my life on a personal level. During the course of the last three and a half years in Cambridge, I have had incredible experiences and made friends that will last a lifetime, all of whom I am very grateful for. In particular, I'd like to thank my housemates whose friendship has provided welcome, and often comedic, relief. In the summer of 2019, however, a beautiful and beloved friend passed away suddenly and unexpectedly. His intellectualism, frustrating persistence, infectious excitement, and irresistible charm are deeply missed. One of our last conversations was in the Downing College MCR in Easter Term last year, where we discussed and debated recent developments in the field of neurodegenerative disease. This acknowledgement could not pass without mention of him. As such, I wanted to dedicate this thesis to my dear friend, Dominic Hamlyn (1994 – 2019), whose life and death act as an ever-present reminder to us all of the brevity of life. We should all try to 'be more Dom'.

This thesis is also dedicated to my father, Colour Sergeant Phillip Newman (1971 – 2007), who always used to tell me when I was younger that I should work harder in the present to avoid having to work as hard later in life. I think he was wrong when he said that, however, rather than having to work hard to survive, I have instead found myself working hard within a field I'm passionate about, and I am very grateful for the opportunity I continue to be given to pursue my scientific interests.

# Table of Contents

# List of Figures

# List of Tables

## Abbreviations

| | |
|---|---|
| ABU | Activated in blocked UPR |
| AD | Alzheimer's disease |
| ALS | Amyotrophic lateral sclerosis |
| CB | Cajal bodies |
| CBD | Cortical basal degeneration |
| CLIP | Crosslinking and immunoprecipitation |
| CNS | Central nervous system |
| CPEB | Cytoplasmic polyadenylation element binding protein |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| CTFC | Corrected cell total fluorescence |
| DAPI | 4′,6-diamidino-2-phenylindole |
| DMSO | Dimethyl sulphoxide |
| dsODN | Double-stranded donor oligonucleotides |
| *E. coli* | *Escherichia coli* |
| ER | Endoplasmic reticulum |
| ERAD | Endoplasmic reticulum-associated degradation |
| FRAP | Fluorescence recovery after photobleaching |
| GFP | Green fluorescence protein |
| IDR | Intrinsically disordered region |
| IPTG | Isopropyl β- d-1-thiogalactopyranoside |
| IRE-1 | Inositol-requiring enzyme 1 |
| LC-MS/MS | Liquid Chromatography with tandem mass spectrometry |
| LLPS | Liquid-liquid phase separation |
| NTR | Nuclear transport receptor |
| OD | Oligomerisation domain |
| PA14 | *Pseudomonas aeruginosa* |
| PAGE | Polyacrylamine gel electrophoresis |
| PAM | Protospacer adjacent motif |
| PCR | Polymerase chain reaction |
| PD | Parkinson's disease |
| PEG | Polyethylene glycol |
| PLAAC | Prion-like amino acid composition |
| PML NB | Promyelocytic leukaemia |
| PQN | Prion-like glutamine[Q]/asparagine[N]-rich domain-bearing protein |
| PrLD | Prion-like domain |
| $PrP^{C}$ | Soluble form of the prion protein |
| $PrP^{Sc}$ | Insoluble, pathogenic form of the prion protien (Sc = scrapie) |
| PSP | Progressive supranuclear palsy |
| RRM | RNA recognition motif |
| RT-QuIC | Real-Time Quaking-Induced Conversion |
| sDMA | Symmetrical dimethylarginines |
| SDS | Sodium dodecyl sulphate |
| SEC | Self-excising cassette |

| | |
|---|---|
| ssODN | Single-stranded donor oligonucleotides |
| UPR, $UPR^{ER}$ | Unfolded protein response of the endoplasmic reticulum |
| XBP-1 | X-box binding protein-1 |

# Chapter 1

# Introduction

## 1.0 Introduction

For centuries, the threat of prion disease has plagued populations – whether it be in the form of scrapie ravaging the sheep populations of Spain in the eighteenth century, fatal familial insomnia afflicting families in Italy, or an outbreak of Creutzfeldt-Jakob disease in the UK triggered by the consumption of contaminated beef. The ability of prions to spread with such pathogenic intent whilst remaining incurable fuels the fire of public concern - yet delving beyond that classical view of prions has led to an enlightened appreciation of the role of such proteins throughout the biota of life.

Prion-like proteins subvert the dogma of the one-to-one structure-function relationship traditionally held onto when considering the physiology of proteins. Instead, these proteins display structural polymorphism, capable of adopting different conformations under different conditions; this expands the capabilities of protein function and represents a mechanism for epigenetic coding, with protein conformation possessing the ability to convey and propagate specific states for biological advantage without the need for DNA modification. The properties observed in many of these proteins, including those of multivalency and low complexity, predispose them towards phase separation, whereby localised concentration of biomolecules compartmentalises cellular activity. This ability confers upon an organism the ability to respond to stimuli in dynamic and transient fashions without needing to use cumbersome membrane-bound organelles.

In this thesis, the role of a novel prion-like protein will be explored in the nematode species *Caenorhabditis elegans*. During this investigation, the function of this protein in cellular stress and the potential molecular mechanisms underlying these functions will be examined.

## 1.1 Prions

Whilst the physiological function of the mammalian prion protein remains elusive, its role in the development of neurodegenerative pathology is better understood. This protein, encoded by the *Prnp* gene, is endogenously expressed within the nervous system where it is believed to play roles in synaptic maintenance, potentially by modulating the voltage-gated potassium channel Kv4.2 and reducing neuronal excitability[1], or through the inhibition of glutamatergic NMDA receptor subunits[2]. Under healthy conditions, this protein remains in its soluble conformation ($PrP^{C}$). However, in disease, this soluble protein is capable of a conversion into an insoluble, aggregation-prone structure ($PrP^{Sc}$)[3–7]. In this insoluble state, it is then able to seed the conversion of other soluble forms of the protein to the same conformation – a transition known as conformational templating (Figure 1.1A-B).

In this insoluble conformation, these proteins expose hydrophobic residues normally retained within the core of the protein – as a result, as these misfolded species build up, they begin to fibrilise, forming large intracellular deposits that bury these hydrophobic domains within the fibril itself, protected from

the solvent environment of the cytoplasm. These oligomeric and fibrillar structures are able to splinter, resulting in the production of new seeds capable of triggering the formation of further fibrils within that environment.

These seeds are also capable of moving intercellularly, spreading pathology between cells and tissues within an organism. A number of pathways have been implicated in this process including tunnelling nanotubes between cells[8], exosomal release from donor cells[9], and macropinocytosis uptake into acceptor cells[10]. Whilst the exact mechanism of transcellular spreading is not fully understood, it is clear that these prionogenic seeds can trigger fibril formation in the acceptor cell. It is this affinity for intracellular seeding and intercellular spreading that partly gives rise to the canonical properties that define the prion protein.

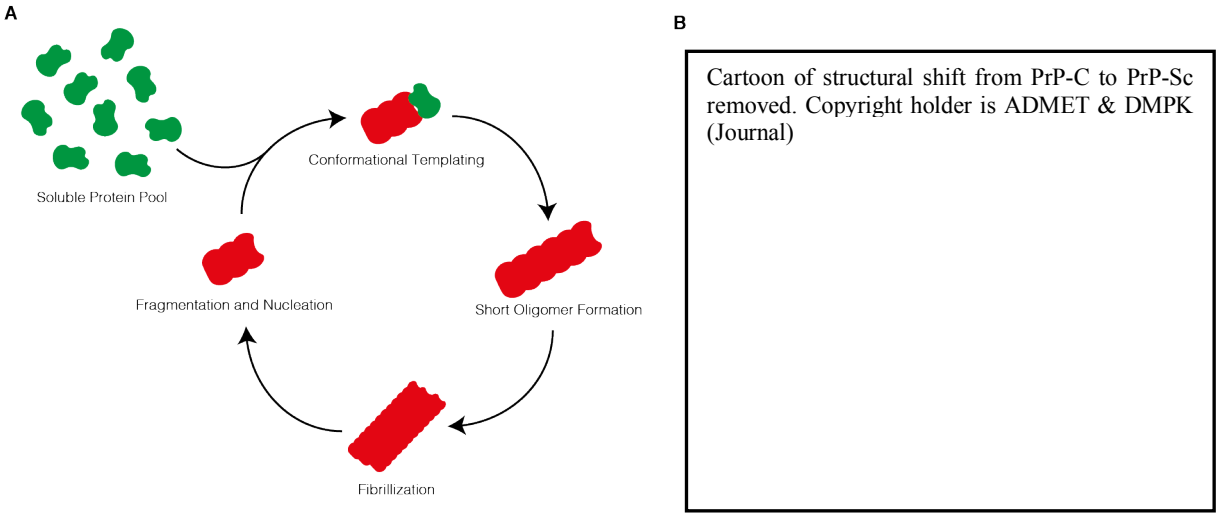



**Figure 1.1 Prion-like propagation** (A) schematic depicting the prion propagation life cycle from soluble protein to insoluble fibrils; (B) structural differences between the soluble $PrP^{C}$ and insoluble $PrP^{Sc}$ conformations (adapted from *Ruttkay-Nedecky et al., 2015*)

Whilst in this aggregated, insoluble state, $PrP^{Sc}$ becomes resistant to degradation, as illustrated by its resistance to heat denaturation, Proteinase K digestion[7] and UV irradiation[4]. It is also difficult for the cell to clear these deposits, and they are therefore able to accumulate with time. This build up disrupts neuronal cell function and ultimately leads to cell death - however the mechanisms underlying this damage are unclear. There are some suggestions that these species are toxic in their oligomeric form and thus their fibrillisation is protective[12]. Whilst this is yet to be fully understood, genetic evidence from knockout animals illustrates the necessity for $PrP^{Sc}$ aggregation in the development of neuropathology including cortical vacuolation and astrogliosis[13] – both hallmarks of prion disease.

Akin to traditional infectious agents, prion diseases come in different 'strains' and, when propagated onto the next host, this strain is then maintained within that organism. In traditional agents, such as viruses and bacteria, this strain variation derives from evolutionary selection pressures favouring particular nucleotide differences that manifest as advantageous phenotypes in specific environments. For prions, however, strains are an entirely distinct concept to this – instead, a variety of misfolded conformational states of the prion protein can be found within a single host. Some of these

conformations will make up the majority of the misfolded protein load, whilst others will constitute a minor proportion of these. The proportional contributions of each conformation to the overall mix of prion structures is what defines a prion 'strain' – and these conformational proportions are carried through when the infection spreads to another organism[14]. This strain diversity extends beyond the biochemistry of the prion protein too, as different conformations exhibit differential abilities to seed, spread and cause disease; this effect is best illustrated by both the varying lag times between infection and clinical disease between prion strains, and the maintenance of similar seeding abilities within a strain, as measured using *in vitro* methods such as protein misfolding cyclic amplification[15] (PMCA) or real-time quaking-induced conversion (RT-QuIC) assays[16]. As such, although the molecular determinants of 'strain' variation differ between prions and other infectious agents, they share the ability to trigger pathology in different ways.

### *1.1.1 Prion-like proteins in pathology*

Beyond the actual prion protein, however, there are a range of other proteins that display prion-like characteristics. These often encompass aggregation-prone proteins involved with the development of neurodegenerative diseases, including tau in Alzheimer's disease[17], and TDP-43 and FUS in amyotrophic lateral sclerosis (ALS)[18]. The term 'prion-like' is used to indicate that these proteins display some of the distinguishing properties that the bona fide prion protein possesses. These characteristics include: the ability to change conformations from a soluble to an insoluble state; the capacity to self-template this conformation onto naïve proteins; and the capability for cell-to-cell transmission. Prion-like proteins exist on a spectrum, with some exhibiting minor prionogenic capabilities, and others robustly displaying these features.

#### *1.1.1.1 Alpha synuclein*

There is some evidence that alpha synuclein, the pathological protein involved in synucleinopathies such as Parkinson's disease (PD), possesses some prion-like properties, including the ability to spread intercellularly. This was demonstrated in human PD patients following neural transplantation trials. Foetal ventral mesencephalic tissue grafted into the striatum could integrate into the lost circuit and recapitulate lost dopaminergic inputs. However, upon death, these grafted cells showed immunoreactivity for Lewy body and Lewy neurite pathology[19,20]. While there are a number of potential causes of this pathology, it is possible that this represents a prionic mechanism at work, with pathological alpha synuclein templating its conformation onto the acceptor cells within the graft. Alongside PD, alpha synuclein is also involved in the development of multiple systems atrophy (MSA); recent evidence has identified that, while brain homogenates from PD patients were unable to stimulate the spread of pathology when inoculated into the brains of mice transgenic for human A53T alpha synuclein, the brain homogenates from MSA patients were capable of this[21,22]. This therefore suggests that different conformational states of the same aggregation prone protein can determine how prionogenic a protein is, even when the secondary structure is identical. This provides support for the

cloud hypothesis of prion strains that, as mentioned previously, suggests different conformational states alter the prion-like abilities of the protein.

#### *1.1.1.2 Tau*

Tauopathies, including Alzheimer's disease and progressive supranuclear palsy (PSP), arise from the aggregation of misfolded tau, which consequently results in neurodegeneration. This protein, encoded by the *MAPT* gene, normally functions to stabilise microtubules; however, upon disease this function is lost. There are 6 isoforms of tau within the CNS produced as a result of alternative splicing of exons 2, 3 and 10, with exons 2 and 3 found in the N-terminal region, whilst exon 10 represents a 30 amino acid repeat sequence within the microtubule binding region. Whilst exons 2 and 3 do not appear to influence pathology, alternative splicing of exon 10 - producing either 3 repeat (3R) or 4 repeat (4R) tau - does, with evidence pointing towards the latter predisposing individuals to both aggregation and neurodegeneration. This is evident in diseases whose tau-positive inclusions have been shown to contain the 4R tau isoform only, such as corticobasal degeneration (CBD)[17,23,24].

There has been emerging evidence that the tau protein acts as a prion-like protein when involved in pathology. This is supported, for example, by the trans-synaptic spread of tau in mouse models of Alzheimer's disease, where induction of pathology in the entorhinal cortex resulted in the spread of pathology along anatomically connected circuits, into regions including the subiculum, the CA1 hippocampal pyramidal neurons, and granule cells of the dentate gyrus[25]. There is also evidence that tau is capable of conformational templating; *in vitro* data suggests that tau aggregates applied externally can be internalised and seed polymerisation in the acceptor cells[26]. This is further backed up by *in vivo* data showing that when the brain homogenates of animals with different tauopathies are injected into the brains of mice expressing non-aggregation-prone forms of human tau, pathology spreads in the injected animal, with the resulting inclusions resembling those of the original brain homogenate[27]. This suggests that these proteins are capable of seeding aggregation and templating their specific misfolded conformation onto correctly folded counterparts in a prion-like fashion.

Whilst tau does not possess what would be canonically recognised as a prion-like domain, recent work on the structure of Alzheimer's disease filaments has identified a specific beta solenoid fold at the core of the fibril. This is of interest as it bears a striking resemblance to the core of a fungal prion, HET-s; this beta-solenoid structure has been postulated to act as a prionogenic scaffold, facilitating fibrillisation[28]. If this were indeed the case, it is possible that such a structural conformation could act as a non-canonical prion-like domain for the tau protein, driving its prionogenic activity. It is worth noting, however, that whilst this structure is seen in Alzheimer's-related filaments, comprised of both 3R and 4R tau, this is not the case for other 3R and 4R tauopathies, such as PSP and CBD respectively[29–31].

### *1.1.2 Physiological Prion-like Proteins*

Disease-related proteins, however, are not the only proteins with prion-like properties – there are many proteins throughout biology that utilise these prion-like characteristics to evolutionarily advantageous effect. Many of these were first identified in yeast, where switching to a prionogenic, insoluble state can promote increased survival in response to challenging environments. It is interesting to note, however, that some of the proteins that possess these functional prion-like properties also overlap with those involved with disease, such as TDP-43 in ALS; this suggests that a switch between physiological and pathological state may be the driver of pathogenesis.

#### *1.1.2.1 [URE3]*

Under basal conditions, the Ure2p protein in *S. cerevisiae* acts as a negative regulator of nitrogen metabolism[32]. It achieves this through the regulation of uptake of allantoate[33], an intermediate in the nitrogen metabolism pathway, by inhibiting the transcription factor Gln3p[34,35] and sequestering it in the cytoplasm[36]. In an environment rich with preferable nitrogen sources, such as glutamic acid or ammonium sulphate, Ure2p is disinhibited as a downstream consequence of Mks1p inhibition[37], thus allowing it to bind to this transcription factor and prevent its translocation into the nucleus. However, when faced with a non-preferable nitrogen source, such as proline, Mks1p is disinhibited, consequently leading to the inhibition of Ure2p[38]. Due to this inhibition, Gln3p is free to translocate into the nucleus where it drives the transcription of Dalp5[39]. This is a transmembrane protein involved with the co-transport of allantoate and ureidosuccinate into the cell; when its expression is increased, more of these intermediates can be imported from the environment into the organism, thus allowing it to continue with its nitrogen metabolism cycle, even in the presence of poor nitrogen sources.

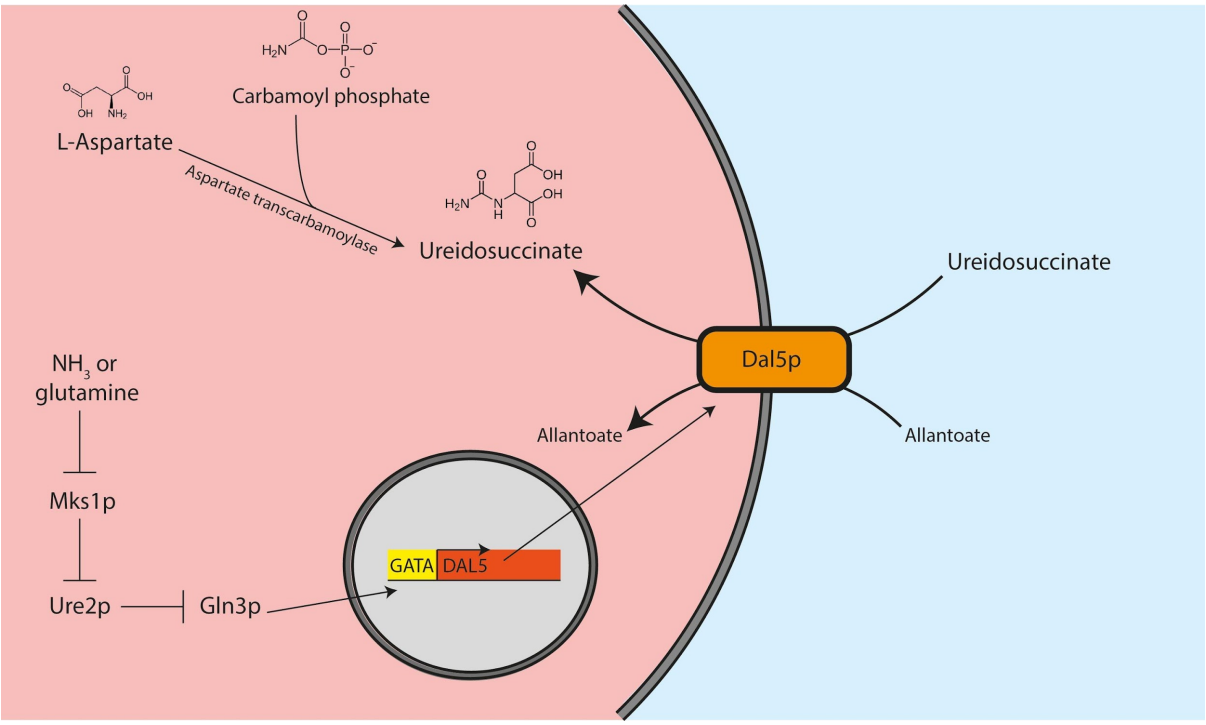



**Figure 1.2** Summary of the Ure2p related pathways in *S. cerevesiae*

Following a UV mutagenesis screen, several mutants were identified that exhibited urediosuccinate-related phenotypes, a useful readout for mutations affecting both the pyramidine biosynthesis and the nitrogen metabolism pathways, due to the co-transport of intermediates in both of these systems[40]. The majority of those identified displayed recessive, Mendellian genetic inheritance patterns when crossed into wild type strains. One of these, *ure-1*, showed a lack of growth on urediosuccinate-supplemented media under nitrogen-rich conditions[41]; these mutants, shown to derive from a mutation in the *ura*-2 gene, coding for the aspartate transcarbamoylase protein, cannot produce their own USA and thus rely on its uptake from the environment. However, when ammonium ions or glutamine are present in the media, Dal5p production is blocked, inhibiting this uptake from occurring and ultimately preventing growth (Figure 1.2).

Further work to delineate this regulatory pathway led to the identification of a further set of mutants following mutagenesis of this *ura-2*-deficient strain. Of these, one mutant, known as *ure-3*, displayed dominant, non-Mendellian genetic traits – this mutant was capable of utilising USA within the supplemented media, even in the presence of nitrogen-rich sources. This phenotype, known as [URE3], has been shown to be the prionogenic form of Ure2p, conferring evolutionarily beneficial effects by triggering the aggregation of the protein and thus preventing it from inhibiting its downstream target, Gln3p. As a result, Dalp5 expression is stimulated, and both allantoate and ureidosuccinate can be imported from the surroundings, even in a nitrogen-rich environment.

The resulting prion species exhibits many of the same traits as mammalian prions. The propagation of [URE3], for example, is dependent on the expression of wild type Ure2p, thus satisfying the need for conformational templating – this can be illustrated via cytoduction, whereby the cytoplasmic contents of two organisms are mixed without nuclear fusion; in such conditions, the [URE3] phenotype persists only when Ure2p is expressed and not in the *ure-2* mutants[42] thus demonstrating this is a non-genetic mode of transmission. Furthermore, overexpression of Ure2p leads to a nearly 100-fold increase in the frequency of spontaneous [URE3] development – even when this overexpression is removed, the [URE3] phenotype is able to endure thus demonstrating that the phenotype does not rely on a high copy number for maintenance. Finally, this conversion to a prion-state has been shown to be reversibly curable, as illustrated by treatment of [URE3] strains with 5mM guanidine chloride - however, when the conditions are conducive, this phenotype can re-emerge later[43].

The molecular driver for this conversion has also been investigated biochemically through targeted mutagenesis of the Ure2p protein. By mutating specific residues – namely a single nucleotide deletion in codon 44 alongside an addition in 80 – it is possible to frame shift an entire domain of the protein, whilst leaving the RNA species relatively unchanged. This manipulation was targeted at the so-called prion domain and results in the ability of Ure2p to convert to [URE3] being lost. This is further supported by truncation mutations; mutants expressing only the N-terminus, where the prion-domain is located,

were sufficient for [URE3] propagation following cytoduction from other [URE3] strains. Those with just C-terminus expression, however, were unable to facilitate this propagation[44].

[URE3] is one of the first examples of a prion-like protein shown to confer physiological benefits onto an organism – in this case, by enabling the uptake and metabolism of ureidosuccinate as part of the pyramidine biosynthesis pathway, and allantoate as part of the nitrogen metabolism pathway. This represents a species of prion that facilitates its effects through loss of function of the soluble protein. There are, however, further examples of prion-like proteins that confer advantageous effects through novel functionality when in their prion state.

#### *1.1.2.2 [HET-s]*

Amongst the filamentous fungi family, heterokaryon formation is a process that allows individual fungal cells to form a syncytium of connected cells with a shared continuous cytoplasm known as a mycelium[45]. This functional network of fungal hyphae can spread throughout an environment, absorbing and sharing nutrients from the surroundings. This process, however, is tightly regulated with a mechanism in place to detect self from non-self – this is beneficial as many homogenous heterokaryon combinations are not viable, even amongst different wild type isolates of the same species. In the ascomycete fungi *Podospora anserina*, this heterokaryon incompatibility is detected using the prion, HET-s[46].

*P. anserina* can exist in one of two forms – either as [Het-s*] or as [Het-s][47], representing the non-prion and prion-states respectively. In addition to these states, HET-s also has a polymorphic variant known as HET-S that differs by just 13 amino acids[48]. Early evidence demonstrated that, whilst fusion of [Het-s*] and [Het-S] was indeed heterokaryon viable, the fusion of [Het-s] and [Het-S] was not and led to cell death. Site-specific mutagenesis of individual residues determined that this incompatibility was a consequence of a single amino acid change at position 33 of the HET-s protein.

Similar to other prion-like proteins, the [Het-s] state can only be propagated when there is expression of a functional *het-s* gene, and overexpression of this locus results in an increased frequency of the [Het-s*] to [Het-s] transition. Furthermore, the protein content of the [Het-s] state displays increased proteinase K resistance than that of the [Het-s*] state, reminiscent of insoluble prion states described previously. The prionogenic nature of the HET-s protein has been illustrated using HET-s-GFP fusions; when expressed at wild-type levels, the protein appears diffuse and soluble in the [Het-s] state however, when overexpressed, aggregate formation is visible. Conversely, when the polymorphic variant HET-S-GFP is visualised, no aggregation formation can be seen in either the wild type or overexpression conditions. This is a similar case for HET-s proteins with a double amino acid substitution – namely D23A and P33H; these substitutions are known to abolish prion infectivity and, when visualised, show a suppression of aggregation *in vivo*[49].

Using these GFP fusions, the mechanism of action of these prion states was further elucidated. By co-expression of HET-s and HET-S, transgenic strains effectively became 'self incompatible', characterised phenotypically by slow growth and aberrant morphology. This phenotype was variable between cells however, with those exhibiting normal growth also showing diffuse cytoplasmic GFP, whilst those with incompatibility phenotypes showing a combination of cytoplasmic and vacuolar GFP, with no HET-s-GFP aggregates[49]. It has been suggested that this results from HET-s-GFP aggregation being partially suppressed in the presence of HET-S. The mode of toxicity of this incompatibility response however, derives from the structural differences between HET-s and HET-S; whilst both have the same domain structure, with an N-terminal α-helical globular domain - known as the HeLo domain – and a C-terminal prion-forming domain[50], there are different functional consequences to their fibrillisation. The conversion of HET-s to its prion-state has no specific adverse effects on the cell however, when HET-S is converted, it triggers a structural change in the prion-forming domain which goes on to induce the refolding and activation of the HeLo domain[51]. This activation triggers the relocalisation of HET-S to the periphery of the cell – a process that is toxic to the cell, with this relocation correlating with cell death[52].

Akin to other prion-like proteins, these HET-s aggregates form functional amyloid states – as demonstrated by biolistic incorporation of these aggregates into [Het-s*] fungi[47]. Unlike their amorphous counterparts, only the amyloidogenic forms of these inclusions are capable of triggering a transformation of [Het-s*] to [Het-s]. This also lends credence to the notion of conformational templating within this system, with HET-s amyloids triggering a conformational switch within the soluble pool of HET-s monomers, as opposed to the presence of aggregate species causing a general proteotoxic collapse within the cell and thus leading to spontaneous conversion of the soluble monomers to an insoluble state due to dysregulation of protein homeostasis.

Hence, HET-s represents a species of prion-like proteins that confer a specific advantage to *Podospora anserina* through gain of functionality – specifically, by discerning self from non-self and thus promoting survival during heterokaryon formation.

#### *1.1.2.3 Cytoplasmic Polyadenylation Element Binding Protein (CPEB)*

Prionic mechanisms have also been implicated in the persistence of long-term memory in a range of species, as they are able to facilitate a self-perpetuating state at the level of individual synapses, allowing for spatially and temporally regulated potentiation. Early evidence from *Aplysia* suggested a role for a neuronal isoform of the cytoplasmic polyadenylation element binding protein (ApCPEB) in this facilitation through the regulation of mRNA translation[53]. This role is conserved across phyla, with similar mechanisms identified in the *Drosophila*[54] and mammalian orthologs[55], Orb2a and CPEB3 respectively.

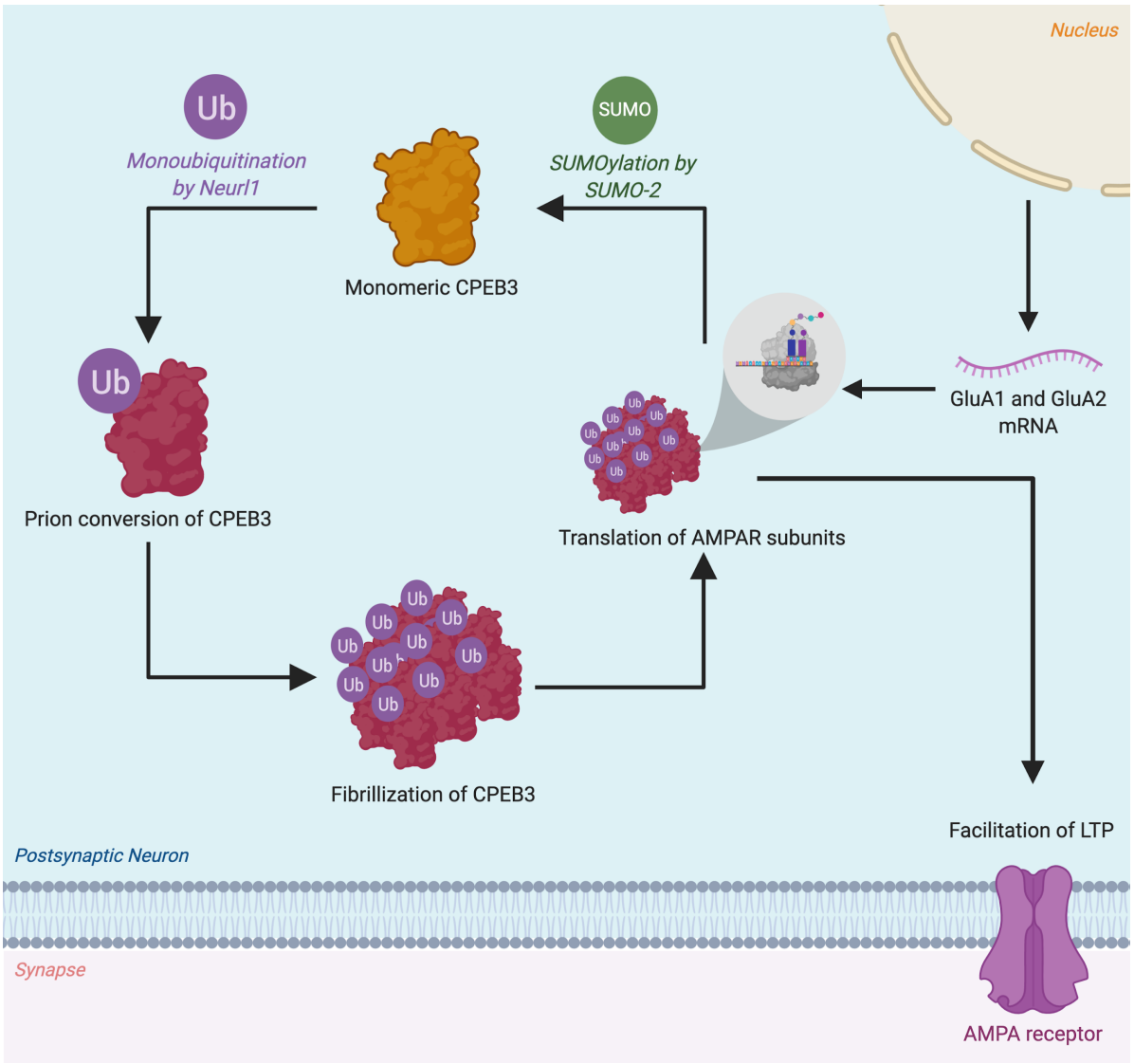



**Figure 1.3** Summary of the CPEB3 prion-like mechanism of action, resulting in the translation of AMPA receptor subunits and the facilitation of long-term potentiation (LTP)

Under basal conditions, this protein is expressed at a low level and is found in its monomeric form, however, upon stimulation, its expression is increased and is able to form a multimeric, self-propagating active state, as illustrated both *in vitro* with the formation of GFP-positive puncta in fluorescent reconstitution assays[59] and *in vitro* in murine hippocampal extracts following contextual fear conditioning and spatial memory tasks[60]. The formation of such prionogenic puncta is regulated by the E3 ubiquitin ligase, Neuralized1 (Neurl1), which monoubiquitinates the N-terminal PrLD of monomeric CPEB3[61], stimulating its prionogenic conversion. Inhibition of this ligase results in a failure to maintain long term memory, mimicking the phenotype observed in both ApCPEB inhibition studies[57] and in CPEB3 conditional knockout studies[60] suggesting that, without this ubiquitination, CPEB cannot be activated. Conversely, this protein is also negatively regulated via SUMOylation, as illustrated in CPEB3, in which activation is correlated with a decrease in SUMOylation. Additionally, chimeric fusion proteins of SUMO-2 and CPEB3 are unable to stimulate mRNA translation or form aggregates, neither

can they stimulate the same morphological changes observed synaptically downstream of CPEB3 activation in wild type conditions[62]. These regulatory mechanisms are particularly interesting as, unlike other prion-like proteins that adopt prion states sporadically, the prion-like conversion of CPEB3 has a very specific mode of activation.

The activation and subsequent prion-like oligomerisation of CPEB drives RNA binding and polyadenylation activity of specific dormant mRNAs[53], driving their translation. This mechanism ensures that mRNA translation of target transcripts only occurs locally, as a consequence of the spatially restricted prion-like activity of CPEB. Of particular interest amongst these transcripts are the AMPA receptor subunits GluA1 and GluA2, the increased translation of which help to facilitate the long-term persistence of synaptic potentiation (Figure 1.3)[60,61,63].

This isoform of CPEB is an example of a 'functional prion' as it gains functionality when in its prion state – it is particularly interesting due to its tight regulation, with prionogenic conversions being triggered in response to specific stimuli, resulting in complex behavioural states.

### *1.1.3 Prion-like domains*

The discovery of a diverse range of prion-like proteins has facilitated a greater understanding of what drives prionogenicity. This activity seemingly derives from specific domains within the protein, known as prion-like domains, as demonstrated by their fusion onto other, non-prionogenic proteins. For example, fusion of the NM domains of Sup35, which have been demonstrated to be vital for formation and maintenance of prion activity in [PSI+] isolates, onto the rat glucocorticoid receptor (GR) confers onto the chimera the ability to self-propagate heritable changes within a [PSI+] expression system[64]. This illustrates that the molecular determinant of prion state is both discrete and transferable between proteins.

The modular nature of this domain has allowed sequence-specific factors facilitating prion ability to be identified. These domains tend to have a low complexity, an intrinsic disorder, and are often rich in low hydrophobicity, polar residues, particularly glutamine and asparagine. In fact, if they are scrambled in a way as to maintain the same residue frequencies within the region but simply rearranging their order, these prion domains are still functional, as illustrated with the scrambled Q/N-rich N-terminus of Ure2p[65]. It is important to note, however, that whilst these proteins maintain their ability to seed and propagate their prionogenic states, the frequencies of *de novo* seeding and efficiencies of propagation are altered. This suggests that whilst residue composition is a major factor in driving prion activity, residue position within the domain can modulate the efficiency of this system.

Interestingly, there is a significant underrepresentation of hydrophobic residues in prion-like domains thought to be due to the ability of such residues to favour amyloid formation. This is reflected in their

overrepresentation in the hexapeptide core of many amyloids[66]. The *limited* presence of hydrophobic amino acids, however, has been shown to enhance prionic activity without promoting amyloid formation – for example, insertion of additional residues into the prion domain of Sup35 increases prion propensity, whilst removal of the infrequently occurring tyrosine residues decreases this propagation ability[67]. There are, however, differences between the influence that aromatic and non-aromatic hydrophobic amino acids exert on these prion domains. Specifically, whilst aromatic hydrophobic residues promote prionogenicity by increasing the efficiency of fragmentation and propagation, non-aromatic hydrophobic residues work instead to promote prion formation. This difference has been shown in the Sup35 prion domain, which can be split into subdomains; a nucleation domain (ND) and an oligopeptide repeat domain (ORD). Whilst non-aromatic hydrophobic residues are overrepresented in the ND, they are underrepresented in the ORD. The differential effects of these residues have been demonstrated in amino acid substitution studies, illustrating that non-aromatic hydrophobic residues promote prion formation and aggregation, to the detriment of propagation, whilst aromatic residues promote propagation[68]. For the aromatic hydrophobic residues, this is thought to be as a consequence of fragmentation ability; fibril fragmentation is key in the life cycle of prion propagation as it allows for the formation of new seeds. The presence of these aromatic amino acids, such as tyrosine, facilitates the interaction of the chaperone Hsp104 with prion-like polymers, which has been shown to be vital for the propagation of Sup35 prions in [PSI+] states[69,70]. This has been illustrated in tetracycline-inducible repression models in yeast, whereby inhibition of *HSP104* following doxycycline treatment subsequently results in an increase in Sup35 polymer size, and thus a decrease in propagation efficiency[71]. Similar effects have also been shown for other prions as well - for example, there is evidence that Hsp104 can reverse the active form of ApCPEB[58]. This demonstrates the powerful influence these underrepresented residues can have on prion-like activity.

The predictable composition of these regions has allowed methods for the bioinformatic identification of novel prions to be developed, with mixed success. Initial algorithms, such as DIANA, focused on the Q/N-rich nature of these domains but neglected the influence of other residues, thus making it difficult to distinguish between prion and non-prion[72]. Such distinctions have been improved on since the development of DIANA. Alberti et al., for example, undertook a bioinformatic proteome-wide screen for prionogenic proteins in yeast, identifying and characterising around 100 novel candidates. By replacing the Sup35 prion domain with that of the candidate prions, they were able to demonstrate that 19 of these candidates indeed functioned as prions, capable of transmitting their conformation onto naïve counterparts in a [PSI+] state[73]. Whilst this improved on previous predictions, the false positive rate observed in this screen remained high, suggesting that there are other factors influencing prion activity – possibility as a result of the systematic exclusion of bioinformatic algorithms to consider the influences of hydrophobic residues. Using mutagenesis data, Toombs et al. were able to reduce the false positive rate in their PAPA algorithm by quantifying the influence of the residues within these domains[74]. The

success of this algorithm was demonstrated *in silico,* with the design of synthetic prion forming domains (sPFDs), a number of which were shown to reconstitute prion activity to Sup35[75].

As a consequence of their stereotyped amino acid composition, it is possible to predict novel prion-like domains across a range of species. Thus, such approaches could also be utilised in *Caenorhabditis elegans* to identify and understand the physiology of novel prion-like proteins in a variety of biological contexts.

### *1.1.4 Structure of amyloids and prion-like proteins*

Amongst the broad spectrum of amyloidogenic proteins, there are some commonalities between their structures upon fibrillisation, with many adopting ordered β-sheet conformations. Early evidence from X-ray diffraction studies demonstrated that amyloid fibrils extracted from the spleens of patients with myeloma-associated amyloidosis had a cross-β conformation[76]. Similar cross-β conformations have also been identified in fibrils extracted from patients with a range of other amyloidosis diseases, as well as from synthetic transthyretin fibrils[77]. Such structural motifs can be recognised by a characteristic 4.7 – 4.8Å meridional diffraction pattern, which denotes hydrogen bonding between β-sheets, as well as a 10Å equatorial diffraction resulting from the stacking of these sheets. This common structure has led to the hypothesis that such conformations are characteristic of amyloidogenic fibrils. Following on from their work demonstrating that poly-L-glutamine fibres, akin to those involved in polyglutamine expansion disorders such as Huntington's disease, also had this cross-β conformation, forming a nanotube of β-sheets surrounding a water filled cavity, Perutz et al. proposed a universal structure for amyloids, based on the thermodynamic properties of these disordered proteins[78]; they hypothesised that the formation of such β-sheet conformations and their subsequent fibrillisation was a result of entropic factors, with the high internal degrees of freedom favouring a random coil structure rather than a folded state. Within such proteins, stability is gained as hydrogen bonds form between main chain amides as the structure turns and creates these nanotube conformations – with each successive turn resulting in the formation of another hydrogen bond, stabilising the structure further and reducing its degrees of freedom. Eventually, the freedom lost becomes negligible to that of the positive entropy gained by freed water in the solvent environment. Such biophysical properties, they speculate, favour a cylindrical β-sheet conformation for amyloidogenic fibrils.

Recent evidence has also determined the structure of tau filaments using cryoelectron microscopy (cryo-EM) of fibrils extracted from the post-mortem brain homogenate of a patient with Alzheimer's disease (AD)[29]. The structures of both straight (SF) and paired-helical (PHF) filaments were shown to contain two protofilaments arranged into a cross-β/β-helical conformation (Figure 1.4A). These protofilament cores, isolated from the fibril using pronase treatment to remove the non-core 'fuzzy coat' without structurally damaging the core, were shown to be highly similar between the two types of filament, with both containing a series of eight β-sheets arranged in a C-shaped morphology. Within these β-sheets,

three formed a triangular β-helix, whilst another two were arranged anti-parallel, forming a cross-β structure. The difference between these fibrils derives from their interface interactions, with the protofilaments of PHFs forming an anti-parallel stack with helical symmetry, whilst SFs pack asymmetrically. In a similar fashion to other prion-like proteins, the core of these amyloids remains consistent. It is important to note, however, that further work using both cryo-EM and immune-gold negative strain electron microscopy has determined that, whilst these folds are common amongst both sporadic and inherited forms of Alzheimer's disease[79], they are not universal for all tauopathies. Pick's disease, for example, adopts a very distinct fold that lacks these β-helix and cross-β motifs[30] whilst the fibrils that form in chronic traumatic encephalopathy (CTE) do form a β-helix structure albeit conformationally different to those seen in Alzheimer's disease (Figure 1.4B-C)[31]. The CTE filaments form a much more open β-helix structure, with a hydrophobic cavity not seen in the AD fibrils – this cavity is filled by an unidentified cofactor that, once identified, may hint at the differential pathogenesis of these two conditions.

The amyloids formed in prion-disease also adopt similar structures – for example, the core of Sup35, which forms the prion [PSI] in yeast, is made up of tightly packed β-sheets when in its fibrillar state[80,81]. There is also evidence that this core structure is maintained between prion strains[82]; residue specific information derived from quenched hydrogen/deuterium exchange NMR demonstrates that strain divergence between two [PSI] prion strains – namely Sc4 and Sc37 – results from the length of the 'core' region. Whilst the two had an overlapping core of around 40 residues, vital for amyloid formation and conformational templating, Sc37 had a much larger expansion of this region. This explains some of the molecular differences observed between these strains, with the expanded core of Sc37 potentially increasing its stability by occluding chaperone recognition sites. In turn, this limits the formation of new seeds and reduces propagation.

Cross β-sheet structures are not the only β-sheet structures observed in these proteins. Solid-state NMR of HET-s amyloid fibrils, for example, demonstrated that the prion-forming domain forms a left-handed β-solenoid structure made up of a twofold-repeating motif of 4 short β-strands, enclosing a triangular hydrophobic core (Figure 1.4D)[28]. Such a structure has also been observed in other prion species too – of particular interest in this context is the FgHET-s prion of the fungal species of *Fusarium graminearium*; despite having only a 38% sequence homology with the *P. anserina* HET-s in the prion-forming domain, these proteins are able to cross-seed and reciprocally trigger conformational templating and fibrillisation in the other *in vitro*[83]. In addition to this, substitution of the FgHET-s prion domain into the HET-s protein *in vivo* is sufficient to functionally recapitulate its prionogenic activity in *P. anserina*, demonstrating the potential universality of this structural motif. It was shown in a yeast co-expression system however that, whilst able to cross-seed fibrillisation, the fibrils forming after this initial seeding mostly segregated out into individual prion species, with preferential formation of

homotypic fibrils. This suggests that, even though this amyloid core plays an important role in seeding and templating of prion states, other factors are also involved.

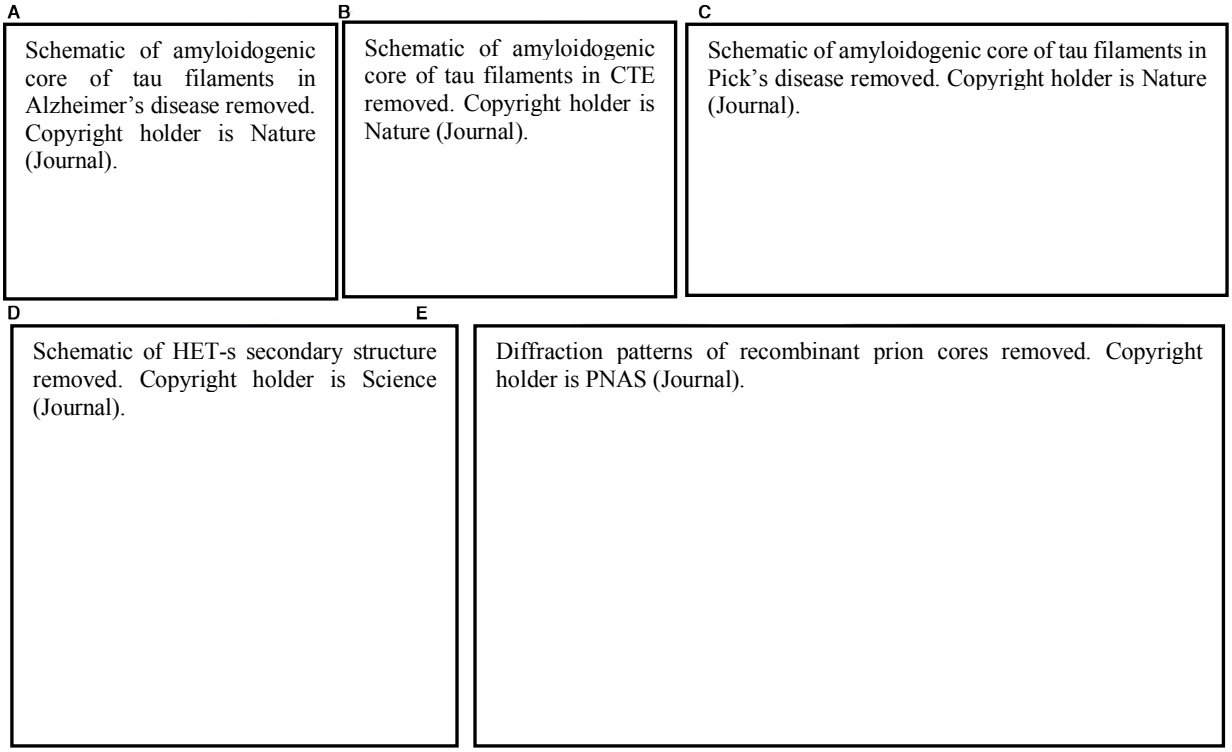



**Figure 1.4 Structure of amyloid-forming domains** (A-C) Schematic of secondary structure of amyloidogenic core of tau filaments derived from (A) Alzheimer's disease (adapted from *Fitzpatrick et al., 2018*); (B) chronic traumatic encephalopathy (adapted from *Falcon et al., 2019*); and (C) Pick's disease (adapted from Falcon et al., 2019); (D) Secondary structure of the fungal prion HET-s (adapted from *Wasmer et al., 2008*); (E) X-ray diffraction patterns showing cross β structures from (ii) prion cores derived from Syrian hamster samples, showing a 4.8Å meridional diffraction (black arrow); (ii) recombinant prion cores, showing a 4.8Å meridional diffraction (black arrow) and a 10Å equatorial diffraction (white arrow) (adapted from *Wille et al., 2009*)

This β-solenoid structure has also been implicated in mammalian $PrP^{Sc}$ fibrils, with recent cryo-EM of truncated PrP peptides extracted from mouse brain showing a 4.8Å meridional diffraction, but lacking any equatorial diffraction, supporting previous X-ray diffraction work (Figure 1.4Ei)[84,85]. This led to the hypothesis that such fibrils form four-rung β-solenoid structures. Interestingly, recombinant fibrils did show a 10.5Å deflection, suggesting that the structure of these fibrils differs from those found *in vivo* (Figure 1.4Eii).

The recurrence of these β-helical and β-solenoid structures in amyloids and prion-like aggregates suggests that such motifs are important for conferring prionogenic abilities onto a protein. Whilst prion-like proteins such as HET-s in *P. anserina* and tau in humans do not have a canonical prion-like domain, they do possess this structural motif – a motif that seems to be acting as a prionogenic scaffold. As such, it is possible that these represent a species of prion-like proteins with non-canonical prion-like domains, thus demonstrating that low complexity and intrinsic disorder are not necessarily the only molecular determinants of prion ability.

## 1.2 Phase Separation

Many proteins containing prion-like domains have been identified within dynamic, liquid-like assemblies within the cell, coalescing in a heterotypic fashion with many other protein and RNA components. This process, known as phase separation, compartmentalises cellular function into non-membrane-bound structures. Instead, weak intermolecular forces between proteins and RNA hold these condensates together, separating them out from the surrounding aqueous environment. These structures tend to be dynamic and are capable of responding to specific cellular stimuli – for example, the formation of stress granules is triggered in response to cellular stress, after which they are dissolved[86].

This process of liquid-liquid phase separation (LLPS) is driven by certain proteins that act as scaffolds and can nucleate assembly whilst other proteins act as clients – these clients are unable to induce LLPS alone, unlike the scaffold proteins, and instead interact with phase separated organelles after their biogenesis. This difference is further demonstrated within these condensates, as client proteins show more rapid diffusion than scaffold proteins, as evidenced by the increased recovery time of scaffolds compared to clients after FRAP, suggesting that scaffolds interact more transiently with clients than they do with other scaffolds[87].

The existence of these multiphase states is acutely dependent upon component concentration, with an increased diffuse concentration of a scaffold protein driving the formation of these condensates. This has been illustrated in numerous overexpression models, whereby the increased expression of a scaffold protein can lead to LLPS or even solid-like aggregates. The overexpression of TIA-1, or the artificial induction of endogenous hnRNPA1 expression can result in the formation of stress granules, for example[88,89]. Similarly, overexpression of the Processing (P) body component, Lsm4, can drive the formation of amyloid-like inclusions[90]. This concentration dependency has been elegantly investigated in the oocytes of *C. elegans*; these oocytes form as a consequence of the cellularisation of the syncytial gonad cytoplasm and show an inverse scaling of nucleolar size (i.e. the larger the oocyte, the smaller the nucleolus, and vice versa), as illustrated by manipulating oocyte size via RNAi treatment. This inverse correlation derives from the fixed number, rather than concentration, of nucleolar components loaded into each maturing oocyte. As a result, larger oocytes have a lower concentration of these components and thus smaller nucleoli, whilst smaller oocytes have a higher concentration and larger nucleoli.

As alluded to above, these biomolecular condensates exist on a heterogeneous spectrum, with some exhibiting liquid-like properties and others forming more solid-like structures. Such differences have important physiological implications, but also hint at the pathological potential of such transitions.

### *1.2.1 Liquid-like Assemblies*

Morphology is a useful readout for the material state of a phase-transitioned system. Liquid-like assemblies tend to be spherical, have rapid internal dynamics, and show an ability to fuse with one

another and subsequently relax into spherical structures once again[91]. P-bodies and stress granules represent two stimulus-responsive membraneless organelles that are involved in the regulation of mRNA following cellular stress events. In mammalian systems, both of these structures fulfil the aforementioned criteria, with live imaging of fluorescently-labelled marker proteins highlighting fusion events whilst the fast intra-organelle recovery of fluorescence exhibited following an internal 'half-bleach' demonstrates their rapid kinetics[90]. They can also be dissolved after treatment with 1,6-hexanediol; this aliphatic alcohol has been shown to disrupt the weak interactions most commonly observed in LLPS, and thus has been co-opted as an experimental approach to differentiate between liquid and solid transition states. The dissolution of these phase-separated organelles upon treatment indicates that they indeed do represent liquid-like assemblies.

Many of these organelles also demonstrate fast exchange of components between the body and the surrounding environment, in addition to a high mobility within the structure itself. The nuclear condensates, Cajal bodies (CBs) and promyelocytic leukaemia nuclear bodies (PML NBs), exemplify these characteristics, with fluorescence recovery after photobleaching (FRAP) illustrating the rapid recovery of GFP-fused CB and PML-NB components *in vivo*[92–94]. Intriguingly such approaches have also been able to determine kinetic differences between components within these condensates; DAXX and BLM, for example, are both proteins found within PML nuclear bodies – unlike the PML protein itself, which has slow recovery dynamics, these components show rapid exchange dynamics with the surrounding nucleoplasm, indicative of the high degree of movement of such components between these spaces[94]. The slower recovery curve of PML, an isoform of which is thought to be a scaffold for the formation and maintenance of these structures, suggests it is more constrained within the body, possibly as a result of the greater multivalency exhibited by phase separating scaffold proteins compared to their clients.

### *1.2.2 Hydrogel Assemblies*

Hydrogels represent the next stage in the phase transition spectrum, with such structures showing more solid-like properties and often being less transient than their liquid-like counterparts. These characteristics can be demonstrated by a slow and minimal recovery after photobleaching, demonstrating their low mobility and reduced turnover of protein. Many phase separating proteins are capable of such transitions: for some, this has been proposed to represent a step in the process towards pathological fibril formation[95]; for others, this hydrogel formation is an end unto itself, with biological function being derived from this state.

The nuclear pore complex, for example, is an intelligent example of how hydrogels have been utilised in biology to physiological benefit. In this case, nucleoporins that make up this complex, such as Nsp1 are able to form hydrogels around the nucleus and act as sieve-like meshes, allowing for the selective movement of molecules between the nucleus and the cytoplasm. As defined by the 'selective phase

model', this permeability barrier allows small inert molecules to pass, but prevents the translocation of large macromolecules unless they are in complex with an appropriate nuclear transport receptor (NTR)[96]. The gelation behaviour observed is driven by intrinsically disordered[97] phenylalanine-glycine (FG)-rich repeat regions in the N-terminal region of Nsp1, as a consequence of a number of different interactions. Hydrophobic interactions between Phe residues of FG repeats are critical for both binding to NTRs and for gel formation, and mutation of these residues to serine prevents any gel formation[98]. In addition to this, solid state NMR has also identified a role for NQST-rich inter-repeat spacer regions in gelation – these regions can drive the formation of amyloid-like β-sheet structures. As mentioned previously, such structures are akin to the those observed within prion-like protein assemblies; in fact, co-incubation of Nsp1 with the prion domain of the yeast prion Sup35, which exhibits a similar enrichment of NQ residues, shows incorporation of the Sup35 PrLD into FG-repeat region derived hydrogel structures[99]. Thus, on the whole, this demonstrates how the biophysical properties of low complexity, prion-like proteins can also influence phase separation biology to drive hydrogel transitions.

### *1.2.3 Functional Aggregates*

Whilst aggregation plays a pivotal role in the pathogenesis of many diseases of proteostasis, it can also have important physiological functions. Yeast stress granules possess very different material characteristics to their mammalian counterparts – they cannot, for example, be dissolved by 1,6-hexanediol treatment. They are also capable of coalescing with misfolding-prone proteins following heat shock[90]. Such data suggests that these structures have more solid characteristics, reminiscent of aggregate-like structures - an idea that is further evidenced by the ability of misfolded proteins to nucleate the formation of these stress granules. They do not, however, show any thioflavin-T-positive staining or SDS-resistance, suggesting that these aggregates are amorphous rather than amyloids. Despite their potential similarities to pathological aggregates, their formation can be readily reversed and their components show a greater mobility than their colocalised misfolding-prone protein counterparts, thus fulfilling their functionality as dynamic stress-responsive assemblies.

### *1.2.4 Biophysical properties of phase separated structures*

#### *1.2.4.1 Prion-like Domains and Intrinsic Disorder*

There are suggestions that prion-like domains are able to act as scaffolds for the formation of phase-separated bodies[100]. This is supported by the huge enrichment of PrLD-containing proteins within such structures, pointing towards a vital role for such domains in the formation and maintenance of these transitioned states.

The nucleolus contains many proteins with disordered regions, including nucleophosmin (NPM1), which has an N-terminal oligomerisation domain and a central domain both exhibiting disorder[101]. This protein has structural polymorphism and is capable of adopting either a disordered state when

monomeric or a folded state when pentameric. This transition from disorder to order increases the valency of the oligomer, stabilising the structure and promoting phase transitions. Evidence from circular dichroism studies suggests that these N-terminal regions form a β-sheet-rich secondary structure[102], thus bearing some resemblance to the structures formed by prion-like proteins, as previously described.

As low complexity and intrinsic disorder are hallmarks of prion-like domains, many of the proteins involved with phase separation share similarities with these proteins. This similarity extends to their propensity for disease; hnRNPA1, for example, is a stress granule component that has been shown to transition into fibrillar disease states as a result of mutations in its low complexity, prion-like domain[103]. Akin to NPM1, hnRNPA1 also exhibits a predisposition to amyloid formation with a cross-β structure, reminiscent of the prion-like protein structures discussed earlier[104]. This aggregation-prone nature points towards a physiological role for IDRs and low complexity domains in state transitions – and indeed this appears to be the case. The purified IDR of hnRNPA1, for instance, is sufficient to drive LLPS alone in an RNA-independent fashion[89]. This coalescence of hnRNPA1 results in a high local protein concentration, thus biasing these phase-separated structures towards fibrillisation if insufficiently regulated. In this instance, this comes as a consequence of a steric zipper motif in the centre of the IDR that can drive this fibrillisation, aligning to form the spine of the fibril; this ability is abrogated if the steric zipper motif is removed[103].

These prion-like properties are also observed in a range of components involved in paraspeckle biogenesis. FUS and RBM14, for example, both contain prion-like domains that are essential for their formation; siRNA knockdown of either of these proteins in HeLa cells results in a dramatic reduction in paraspeckle number – a reduction that can only be rescued by the re-expression of FUS or RBM14 in their wild type forms. Expression of PrLD mutant variants of these proteins is unable to rescue this phenotype[105]. The importance of these domains can be further illustrated *in vivo*, with mutations to the prion-like domain of FUS resulting in diseases as a consequence of the mislocalisation of these protein components[106], for example.

On the whole, this points towards an important role for prion-like domains in both orchestrating protein interactions and helping to drive phase-transitioned states. These domains are often paired with RNA interacting domains, many of which fail to facilitate LLPS alone, unlike their PrLD counterparts – this suggests that, whilst prion-like domains may drive phase separation, RNA interactions might modulate it.

#### *1.2.4.2 Multivalency*

Within a polymerising system, as the degree of bonding increases, the propensity for a sharp state transition is also increased, as formulated by theories of condensation polymerisation. In a biophysical

system, this translates to the valency of a biological molecule – with a greater valency predisposing the system towards phase transition. This can be demonstrated artificially by exploiting the modular nature of protein interaction domains; by engineering a host of proteins to contain different numbers of repeats of either the domain SRC homology 3 (SH3) or its proline-rich motif (PRM) ligand, the influence of multivalent protein interactions can be examined. In this case, as the number of repeats is increased, the critical concentration for LLPS is reduced[107]. This has been further validated in similar studies using constructs of repeating SUMO3 domains and its interaction partner, SIM, which similarly showed that an increased repetition of these domains led to an increase in the formation of phase separated cellular bodies[87]. Thus, overall, these experiments demonstrate that the physics of polymer condensation proposed for material states can indeed be applied to biophysical systems as well, confirming there is a role for multivalency in driving biological phase transitions.

This phenomenon is observed *in vivo* as well, aided by the heterogeneous mix of biological molecules within cellular space. In particular, the importance of both protein and RNA interactions within a system can be intricately illustrated by the nucleolus. Nucleophosmin (NPM1), a key driver of nucleolar formation - specifically of the granular component – is highly multivalent. In its N terminus, this protein contains an oligomerisation domain (OD), followed by a central disordered region, and finally an RNA recognition motif (RRM). Removal of either the N or C-termini renders the protein incapable of phase transitioning, demonstrating the interdependence of these regions on one another[101,108]. Under wild type conditions, acidic tracts within the N-terminus can bind to short, linear, arginine (R)-rich motifs allowing NPM1 to form cross-linked homopentameric structures[102]. When in this pentameric form, the multivalency of NPM1 is increased, allowing for recruitment of further R-rich nucleolar proteins, as well as increasing the affinity of the RRM for rRNA[101]. This multivalent pentamer can facilitate phase transitions, whilst the interaction of the C-terminus with rRNA is vital for the correct localisation of these structures within the granular component of nucleolus. This reliance of NPM1 on both RNA and proteins interactions is an elegant example of how mutlivalency can powerfully modulate the formation of liquid condensates, this is particularly the case when factoring in the influence that posttranslational modifications can have on the formation of these phase transitioning pentamers (see section 1.2.3).

#### *1.2.4.3 RNA*

Beyond localising organelle components to the correct phase, RNA also plays an important modulatory role in the formation and dissolution of droplets, particularly in the nucleus. This is especially the case for prion-like RNA binding proteins that are often predisposed to transitioning into solid states and causing disease; RNA helps to mitigate these risks. The influence of nuclear conditions in this modulation can be demonstrated in a yeast model of ALS in which FUS is fused to a nuclear localisation signal, causing it to relocate from the cytoplasm into the nucleus. This results in a dramatic reduction in aggregate formation, suggesting that the nuclear environment does not promote phase transitions in the same way that the cytoplasm does[109]. *In vitro* phase separation assays have demonstrated that FUS,

amongst other prion-like RBPs, can undergo LLPS at nuclear concentrations – this is despite their presence in mostly diffuse states within the nucleus. Hence there must be another factor preventing the formation of these droplets *in vivo*. Indeed, this effect comes as a consequence of the high RNA concentration within the nucleus. Addition of RNA to the *in vitro* system, particularly of small RNA species, results in the dissolution of these droplets, an effect that can be easily reversed with the reformation of droplets following RNase A treatment. This solubilising effect can also be observed *in vivo* with injection of RNase A into the nuclei of HeLa cells containing GFP-fused FUS, showing a sudden formation of FUS-positive condensates[106].

This ability of RNA to buffer phase separation also reduces the ability of any phase separating structures that do form in this environment to transition into fibrillar states. Removal of RNA binding motifs from FUS results in a reduction in its mobility, thus, RNA functions to maintain the dynamic behaviour of prion-like RBPs, preventing their aberrant transition into a solid aggregate.

Even within this highly solubilising environment, however, phase transitions can still occur, driven by specific RNAs capable of triggering LLPS. In the case of FUS, the long noncoding RNA *Neat1* fulfils this role[110], which can be recapitulated *in vitro*[106], clearly illustrating the sufficiency of *Neat1* RNA alone in nucleating this phase transition in the backdrop of high RNA solubilising conditions. This is enabled by the ability of *Neat1* to form large RNA assemblies capable of acting as a scaffold for nucleation, with protein components associating with these assemblies and the process of phase separation thus being initiated. This ability of RNA to seed such transitions is not only seen in the nucleus, there is also evidence of such a mechanism at play in the formation of cytoplasmic stress granules, with *trans*-RNA-RNA interactions between long transcripts acting as this scaffold. This ability has been demonstrated within protein-free lysates containing all cellular RNA; when incubated in conditions mimicking cellular stress, RNA can form liquid-like droplets and even more solid-like RNA tangles. Akin to the *in vivo* setting, these droplets are enriched for longer RNA, suggesting that length plays a role in facilitating the formation of these scaffold structures[111]. This is further evidenced in repeat expansion disorders – whilst traditionally thought of as diseases of protein disorder, these repetitive sequences also have an effect on RNA biology. As the number of repeats within an RNA transcript increases, their multivalent base-pairing ability also increases, thus with it, the propensity for a phase transition[112]. It is likely that longer RNA transcripts are more likely to form phase-separated scaffolds due to the increased valency this length affords them.

The importance of such RNA-RNA interactions in the formation of ribonucleoprotein complexes and their roles in normal physiology can be illustrated in *Drosophila* – for example, *Drosophila* embryogenesis, anterior patterning is reliant upon the asymmetric localisation of Bicoid mRNA. Its association with an RNP complex containing the protein Staufen facilitates this localisation. The nucleation of this RNP complex is dependent upon the oligomerisation of loop regions within the 3’

UTR of Bicoid mRNA, without which, the association with Staufen is disrupted[113]. Thus, even beyond phase separating systems, it is commonplace for intermolecular RNA interactions to provide a platform for further protein interaction within RNP complexes.

#### *1.2.4.4 Multiphase Coexistence*

The microenvironment facilitated by liquid liquid phase separation can best be illustrated by the existence of multiple phases within a single system. These phases are capable of coexistence whilst also exhibiting immiscibility. Such behaviour has functional implications – for example, in the nucleolus, a sequential progression of liquid phases facilitates the multistep coordination of rRNA processing in preparation for ribosome biogenesis. The nucleolus itself has a tripartite structure, with three distinct sub-compartments localising different protein and RNA components within them. Namely, these are; (1) the fibrillar centre [FC]; (2) the dense fibrillar component [DFC]; and (3) the granular component [GC]. The immiscibility of these phases, and the intrinsic biophysical mechanisms driving this, can be demonstrated *in vitro* as nucleolar proteins purified from different phases will not coalesce[101]. Surface tension differences between these compartments prevent this coalescence into one large liquid organelle, whilst actin networks help to maintain structural integrity and minimise homotypic fusion events. This can be clearly demonstrated by disrupting these networks, leading to the fusion of all nucleolar sub-compartments into a homogenous phase[101,114].

This core-shell architecture, whereby there is a denser, less liquid-like centre surrounded by a more fluid outer layer, is also seen in other phase separated organelles as well. Superresolution microscopy has identified stable substructures within both mammalian stress granules that fit this description. These core structures contain a greater density of protein and RNA interactions, resulting in a less mobile phase than the shell surrounding it. Isolation and proteomic analysis of these cores has allowed for components of the core to be identified, including many ATPase-dependent complexes that play regulatory roles in the assembly and disassembly of these granules[115]. The shell, however, is much more dynamic, allowing for a greater exchange of components from the cytoplasm. This further supports the existence of a spectrum of phases, with the immiscibility of each of these phases pointing at highly regulated mechanisms differentiating between these layers.

### *1.2.5 Regulation of phase separated structures*

The functional significance of these phase-separated compartments is further validated by the tight physiological regulation of their assembly and dissolution. This process is facilitated via a multitude of mechanisms, and can lead to very precise spatiotemporal control of these organelles, allowing for targeted and effective responses to specific stimuli.

#### *1.2.5.1 Spatial Regulation*

The ability to spatially regulate the formation of these condensates is vital for their effective functionality. This control often relies on the regulation of organelle components in their diffuse state – for example, protein localisation, or RNA synthesis. The best example of this exquisite spatial regulation comes from P granules in the developing *C. elegans* embryo – these begin uniformly distributed, however, after symmetry breaking, they begin to segregate along the anterior-posterior axis, with a bias towards formation posteriorly[116]. This process is tightly modulated by a PAR-1 kinase-dependent MEX-5 concentration gradient that runs in the opposite direction to P granule asymmetry[117]. This regulation originates with the asymmetric cortical localisation of PAR-1, leading to preferential phosphorylation of MEX-5 in the posterior via serine 404. MEX-5 is an RNA-binding protein that is capable of associating with poly-U tract-containing RNA throughout the embryo[118,119]. Such associations with large RNA-containing complexes result in the slow diffusivity of this protein[120] – however, upon phosphorylation of this C terminal serine, it is able to dissociate from such complexes and diffuse more freely, leading to greater mobility of posterior MEX-5 compared to anterior, leading to an enrichment in the anterior[121]. The intrinsically disordered protein, MEG-3, acts as a scaffold for RNA-induced phase separation and the formation of P granules. Co-incubation of purified MEG-3 with both MEX-5 and RNA demonstrates that ability of MEX-5 to inhibit the phase separation of MEG-3 in an RNA dependent manner, with MEX-5 outcompeting MEG-3 for RNA interaction[122]. This shows that the localisation of unphosphorylated MEX-5 triggers the preferential dissolution of MEG-3-mediated P granules in the anterior, whilst its absence in the posterior results in their preferential formation. The asymmetry observed in P granules illustrates the fine spatial control that an organism can exert on organelle dynamics – and is a clear demonstration of the functional consequences of this control; in this case, such spatial regulation positions P granules for their selective preferential inclusion in the posterior germline blastomere, $P_1$, and exclusion from anterior blastomere, AB, following asymmetric cell division[121]. There are examples of similar spatial patterning in other phase separated systems as well, such as the formation of Cajal bodies around RNA during transcription[123] – the requirement for RNA to phase transition means that their spatial regulation relies on the localisation of RNA synthesis, thus linking functionality and spatial control.

#### *1.2.5.2 ATP dependency*

The role of ATP in phase transitions is two-fold: as would be expected, it is capable of driving energy-dependent processes that may be involved these transitions; however, beyond this, it has also been hypothesised that ATP can act as a biological hydrotrope, solubilising hydrophobic molecules within the aqueous environment of the cell. Akin to other hydrotropes, ATP is amphiphillic, containing both hydrophilic phosphates and a hydrophobic adenosine ring, lending support to this hypothesis. *In vitro*, this has indeed been shown to be the case, with ATP increasing the solubility of the hydrophobic compound fluorescein diacetate in a concentration dependent manner. Similar effects are observed with biologically aggregating species, with ATP preventing aggregate formation of FUS, amyloid-beta and the prion-like domain of Mot-3. It was even capable of dissolving preformed FUS aggregates[124]. Thus

this suggests that ATP can indeed act as a biological hydrotrope, potentially providing an explanation as to the magnitudinous surplus of ATP within the cell, which far exceeds its energy demands.

In addition to this, the presence of ATP-dependent protein and RNA related complexes within many of these phase-separated organelles has provided a mechanism for their regulated assembly and dissolution. Proteomic analysis of isolated core regions of stress granules, for instance, identified a number of components involved in these processes; for example, the chaperonin-containing T (CCT) complex is involved in the inhibition of stress granule assembly, and manipulation of its ATPase domain results in a greater number of granules forming. Conversely, the mini-chromosome maintenance (MCM) complex and the RvuB-like helicase complex are both involved in inhibiting their disassembly; siRNA knockdown of these complexes results in a faster loss of granules during stress recovery. The reliance of these processes on ATP can be demonstrated in cell culture through the inhibition of glycolysis and oxidative phosphorylation, leading to a depletion of ATP. When applied concurrently with sodium arsenite, a stress granule inducer, formation of these stress granules is inhibited. Similarly, if applied after formation, the resulting stress granules lose their dynamic nature and become static, low mobility structures[115]. Thus, this illustrates the essential role that ATP plays in the biogenesis and dissolution of these granules, as well as in the maintenance of their liquid-like microenvironments.

Due to the properties of many phase separation components, with their intrinsic disorder, prion-like nature, and their propensity for aggregate-like behaviour, it is perhaps unsurprising that chaperones can regulate the activity of these compartments, often in an ATP-dependent fashion. The AAA+ ATPase, Hsp104, already discussed in relation to yeast prion-like proteins, is of particular interest in this context. Working in conjunction with Hsp70 and Hsp40, this chaperone is able to dissolve aggregate-like formations and restore the constituent components to their functional, folded state[125]. Similar mechanisms are at play with the regulation of phase-transitioned organelles; stress granules, for example, show an increased assembly and decreased dissolution when Hsp104 is deleted. P-body formation, on the other hand, is not affected by Hsp104 deletion under basal conditions however, upon the application of an acute stress, P-body components such as Edc3 and Lsm4 mislocalise with stress granule components, forming irregularly shaped aggregate-like structures. The level of mislocalisation and aggregation of these assemblies correlates with the level of Hsp104 expression, demonstrating that this chaperone is required to maintain the liquid-like properties of P bodies[90]. On a mechanistic level, the action of Hsp104 relies on ATP hydrolysis to elicit a dramatic conformational change in its hexameric structure, resulting in a cyclic peristaltic pump motion[126]. This pumping motion drives the translocation of prion-like aggregates into the central channel of this hexamer – a process that is key for its disaggregase activity; this can be demonstrated by N-terminal domain deletion mutations that perturb this mechanism allowing substrates to escape this central channel. Without this mechanism, prion-like protein assemblies can be fragmented but not fully dissolved, as illustrated in the yeast prion Sup35, with mutant Hsp104 improving prionogenesis by driving fragmentation without dissolution[126,127].

Although many chaperones have been identified as having regulatory influences over the life cycle of a phase separated entity, Hsp104 has some of the most diverse and wide-ranging effects, clearly demonstrating how ATP can be used to modulate these activities.

On the whole, ATP can modulate the formation, maintenance and dissolution of phase-separated compartments in a multitude of ways. This diversity of action allows for such structures to be tightly controlled in response to specific physiological stimuli.

#### *1.2.5.3 Post-translational modifications*

Post-translational modifications (PTM) are a powerful mechanism through which cellular events can be temporally regulated. There are numerous types of PTM, often working in concert to fully orchestrate the coordinated subcellular environment. Such modifications have been implicated in the regulation of phase transitions.

SUMOylation regulates the formation and dissolution of PML nuclear bodies[128,129]. The scaffold protein, PML, contains three lysine residues that allow for its modification by the ubiquitin-like modifier, SUMO-1. Mutagenesis of these SUMO sites results in NB disruption, demonstrating their importance for the formation and maintenance of these condensates[130,131]. This effect can be shown in a physiological context as well, with infection by the herpes simplex virus triggering the desumoylation of PML and, consequently, the dissolution of these phase separated structures[132]. The role of this modification in PML NB formation revolves around the ability of SUMO-modified proteins to interact with and recruit other proteins. This is best demonstrated by the presence of a SUMO binding motif within the PML protein – this domain can non-covalently interact with SUMOylated proteins and has been hypothesised to be the nucleation trigger for PML NB formation, with non-modified PML interacting with modified PML via its SUMO binding motif[133]. After this point, other SUMO interacting proteins, including Daxx, can interact and be recruited to the condensate[134]. Daxx is an interesting example of a client protein whose function is directly related to the SUMOylation state of other proteins. When diffuse, it can interact with SUMOylated transcription factors, preventing their transcriptional activity – however, upon recruitment to these inclusions via SUMOylated PML, this transcriptional activity can be restored and gene expression increased[135]. Thus, SUMO modifications are a useful mechanism to influence both formation and dissolution, as well as to modulate the activity of client proteins. Beyond these phase-separated structures, SUMOylation has also been implicated in the regulation of prion-like activity in proteins such as CPEB-3[62], suggesting some commonalities between their biology.

The methylation of arginine residues has also been shown to contribute towards the regulation of phase-separated organelles – specifically, in the regulated formation of Cajal bodies and nuclear gems. Coilin, a core component of CBs, contains symmetrical dimethylarginines (sDMAs) allowing it to interact with

complexes containing the protein SMN, retaining them in Cajal bodies as a consequence of their preferential affinity for sDMA-containing proteins[136,137]. However, when methylation is artificially inhibited, or there is a physiological hypomethylation, coilin can no longer efficiently interact with SMN complexes[138,139], resulting in the formation of parallel condensates, known as nuclear gems. This has functional consequences, with reductions in pre-mRNA splicing and spliceosomal complex formation being observed in hypomethylated nuclear extracts as a result of this interaction loss[137]. On a clinical level, this also has profound effects; patients with mutations in SMN that prevent its interaction with the RG-rich, sDMA-containing sites of its partners develop spinal muscular atrophy, underscoring the vital role that this regulated interaction can have[136,140]. Thus, on the whole this demonstrates the importance of post-translational modifications, in this case methylation, on the regulation of a phase-separated organelle, namely Cajal bodies.

Beyond these modifications, phosphorylation can also play a role in this regulatory process. Stress granule formation, for example, is triggered in response to an environmental stressor; the elicitation of this response involves a phosphorylation cascade ultimately resulting in the phosphorylation of the translation initiation factor, eIF2α[141]. This modification disrupts the normal activity of the protein by preventing the formation of pre-initiation complexes. Under basal conditions, this protein is unphosphorylated and interacts with GTP and $tRNA^{Met}$ to form a component of the complex involved in the loading of the initiator tRNA onto the 40s ribosomal subunit, thus driving translation. Upon phosphorylation, the affinity of eIF2α for eIF-2B, a guanidine nucleotide exchange factor, is increased – this interaction prevents the exchange of GDP for GTP and thus inhibits the formation of eIF2α-GTP-$tRNA^{Met}$ and, subsequently, the formation of these pre-initiation complexes[142]. These stalled complexes are then involved in the nucleation of stress granule formation, via their interactions with the scaffold proteins TIA-1 and TIAR[88]. The direct influence that this phosphorylation event can have on stress granule formation can be demonstrated in phosphomimetic mutants of eIF2α – these mutants result in the constitutive formation of stress granules, even in basal conditions[141]. Conversely, mutants that cannot be phosphorylated show no stress granule formation before or after stress in most cases – though it is important to note that phosphorylation-independent reductions in eIF2α-GTP-$tRNA^{Met}$ can also elicit formation, albeit at a much lower frequency[143]. On the whole, however, this demonstrates the influence that phosphorylation can have on driving phase transitions in response to very specific stimuli. The reversibility of such modifications demonstrates a clear mechanism for reverse phase transitions in these contexts.

### *1.2.6 Interactions driving phase transitions*

#### *1.2.6.1 Dipole-dipole*

The influence that polar residue-rich intrinsically disordered regions (i.e. Q/N-rich domains) have on phase transitions can be modelled mathematically if the polymer-solvent environment is considered a

lattice, with each geospatial position ($z$) within the lattice occupied by either a solvent or polymer molecule (Figure 1.5A)[86,144]. The interactions of these residues along the polypeptide backbone of a protein result in a 'chain of dipoles'. As the aqueous solvent environment is also dipolar, a number of potential dipolar interactions are possible; (1) solvent-polymer, $u_{ps}$; (2) polymer-polymer, $u_{pp}$; and (3) solvent-solvent, $u_{ss}$. These potential interactions are in constant competition, with the interaction energies between such dipoles accounting for the enthalpic contribution to the free energy within the system. The effect that these dipoles have on the free energy state of the system can be approximated by a single mean-field value (Flory parameter, χ), averaging out the individual energies of each potential dipole. This can be calculated by the following equation.

$$\chi = \frac{z}{k_B T}\left[u_{ps} - \frac{1}{2}\left(u_{pp} + u_{ss}\right)\right]$$

Using these variables, the Flory-Huggin's model of homopolymer physics can be applied[145,146], resulting in a simple approximation of the energy state of mixing within the system and can describe simple phase separation dynamics, highlighting the importance of factors including polymer length ($N$), thermal energy ($k_B T$), and ratio of polymer to solvent within the environment ($1$- $ø$).

$$\frac{F}{k_B T} = \frac{ø}{N}\ln ø + (1 - ø)\ln(1 - ø) + \chi ø(1 - ø)$$

This equation illustrates the energetic cost involved in these polymer-solvent interactions. The greater the χ, the larger the energetic cost of mixing. Phase separation is favoured when $\chi > 0$, indicating the polymer is in a poor solvent; past a critical threshold of χ, this energetic cost overcomes that of the mixing entropy, resulting in a phase transition.

#### *1.2.6.2 Charge-charge*

Other modes of interaction cannot be modelled with such a simple lattice, however. Charge-charge interactions, for example, result in a type of phase separation known as complex coacervation, driven by the various electrostatic properties of charged amino acids (Figure 1.5B). Such interactions are observed with polyampholytic intrinsically disordered proteins. In these cases, the effect of the short-range interactions of neutral groups is significantly outweighed by the long-distance interactions of charged groups within the protein, thus rendering variables such as χ entirely negligible.

Unlike before, where the fraction of polymer-to-solvent within a given environment was important, this time, charge-charge interactions consider the proportion of polycations ($ø^+$) to polyanions ($ø^-$) within a system. These electrostatic interactions can be modelled using the Overbeek-Voorn mean field theory[147], quantified as follows;

$$\frac{F}{k_B T} = \frac{ø}{N}\ln\frac{ø}{2} + (1 - ø)\ln(1 - ø) - \alpha(\sigma ø)^{3/2}$$

This equation summarises the factors that influence the phase separation dynamics of electrostatic-induced transitions. In particular, the α value denotes a variable that is influenced by; (a) charge at the lattice site; (b) thermal energy ($k_B T$); and (C) partial molar volume of the solvent. The $\sigma$ variable, on the other hand, represents the linear charge density on each polyion. This theory hypothesises that when $\sigma^3 N > 0.5$ then phase separation will occur, with a condensation of these polyions.

Electrostatic interactions within the nucleolar protein NPM1 play important roles in modulating the equilibrium between monomeric and oligomeric states[148]. As previously described, this protein undergoes LLPS when in a homopentameric conformation[108]; this conformation is stabilised by its configuration of electrostatic charges, with negatively charged residues arranged on one face of the oligomer, and mixed electrostatic residues on the other[102]. The influence of these electrostatic interactions can be demonstrated *in vitro*, with analytical ultracentrifugation studies demonstrating that NPM1 exists as a monomer in low-salt conditions, whilst pentamer formation is observed in high-salt conditions. In addition to this, the presence of $Mg^{2+}$ and $Ca^{2+}$ has also been shown to drive oligomerisation – possibly due to a proposed role of cations in shielding negative charges from repulsive electrostatic forces, thus stabilising the pentameric conformation[148].

The charge asymmetry of NPM1 can be exploited to allow for the dynamic modulation of monomeric to pentameric state via posttranslational modification. Phosphorylation of solvent exposed residues adds a negative charge to the pentamer, causing a thermodynamic destabilisation of NPM1 structure. This phosphorylation continues sequentially, with the destabilised structure exposing further sites for phosphorylation until eventually there is a structural switch and the pentamer is driven into its monomeric state[102]. Thus, charge-charge interactions can be utilised as a tool for both maintaining and regulating phase transitions due to the charged nature of certain residues and posttranslational modifications.

#### *1.2.6.3 Cation-π*

Other electrostatic interactions can also modulate phase separation behaviour. Cation-π interactions, for instance, occur between aromatic rings and charged residues; these aromatic rings contain six *p*-orbitals corresponding with each of the six carbon atoms of the ring. These *p*-orbitals conjugate, resulting in one delocalised system, with *p*-orbital electrons forming a cloud above and below the aromatic ring plane, rather than being confined to their specific orbitals. In the presence of a positively charged residue, the electron cloud becomes polarised and the side facing the residue is attracted thus resulting in a cation-π interaction (Figure 1.5C).

Such interactions can be investigated in phase separating bodies by identifying components that are enriched in one of the three amino acids to contain such aromatic ring side chains – namely phenylalanine, tryptophan or tyrosine. FUS, for example, shows a large overrepresentation of tyrosine in its low complexity domain. Substitution experiments, in which tyrosine is replaced with alanine or phenylalanine, demonstrate the involvement of cation-π interactions in driving the phase separation properties of this protein. Alanine substitutions abolish the ability of FUS to phase separate either *in vitro* or *in vivo* – however, as phenylalanine also contains an aromatic ring, the phase separation behaviour of FUS is maintained in this replacement. Similarly, if the positively charged residue arginine is replaced with alanine or lysine, the phase separation ability is only observed with lysine substitutions, as this residue also has a cationic side chain, allowing for the maintenance of charge-π interactions. This has fascinating physiological repercussions, as these interactions can be disrupted by the methylation of arginine residues, thus reducing the propensity of FUS for phase separation – if this methylation is prevented, FUS more readily coalesces, forming more gel-like condensates[149].

This phenomenon has also been illustrated in Ddx4, an intrinsically disordered protein constituent of the phase-separated organelle, nuage. Similarly to FUS, this ability of this protein to form droplets is dependent upon the presence of repeating blocks of charged residues and enrichment of F or R containing dipeptides. Removal of either of these eliminates their phase transition abilities. Arginine methylation also plays a role in the regulation of these bodies as well, with this posttranslational modification triggering the dissolution of these droplets[150]. Thus, this demonstrates the importance of cation-π interactions in not only facilitating the formation of these condensates, but also for allowing a mechanism of regulation due to their sensitivity to reversible protein modification.

#### *1.2.6.4 π-π interactions*

In addition to this, π-bonds have also been implicated in the process of phase separation. Such interactions occur between the *p*-orbitals of neighbouring atoms, which run perpendicular to σ bonds formed between the same molecules. The formation of these bonds can occur following $sp^2$ hybridisation, whereby the *s*-orbital and *p*-orbitals are combined; this results in the creation of three new $sp^2$-orbitals, involved in the formation of the sigma bond, and a spare *p*-orbital that goes on to form the π-bond (Figure 1.5D). Proteomic analysis has identified an enrichment of residues capable of such interactions in low complexity, intrinsically disordered regions and in RNA binding domains, both of which are commonly associated with phase transitions, suggesting that they might indeed play a role in driving the formation of these assemblies[151].

Arginine is one of these overrepresented residues, with $sp^2$-hybridisation observed within the guanidium group of its side chain. The importance of this has been demonstrated using the N-terminal IDR of the phase separating protein Ddx4, which is enriched with arginine-glycine repeats. Due to the positive-charge of arginine, it has been previously proposed that such transitions are driven by charge-charge

interactions, however, upon substitution experiments in which arginine was replaced with lysine, this was shown not be the case. Lysine, which holds the same positive-charge as arginine, is unable to form π-bonds due to its $sp^3$-hybridised atoms – unlike the $sp^2$-hybridised atoms of arginine, they do not have a spare *p*-orbital capable of forming such bonds. This R to K replacement prevents the phase separation of Ddx4, thus suggesting this transition is driven by π-interactions, rather than charge-interactions. Similar findings have been made for pathological fibril formation in Alzheimer's disease, with π-stacking of arginine side-chains thought to be involved tau fibril interactions[152]. This once again points towards a similar mechanism of assembly between pathological and physiological phase transitions.

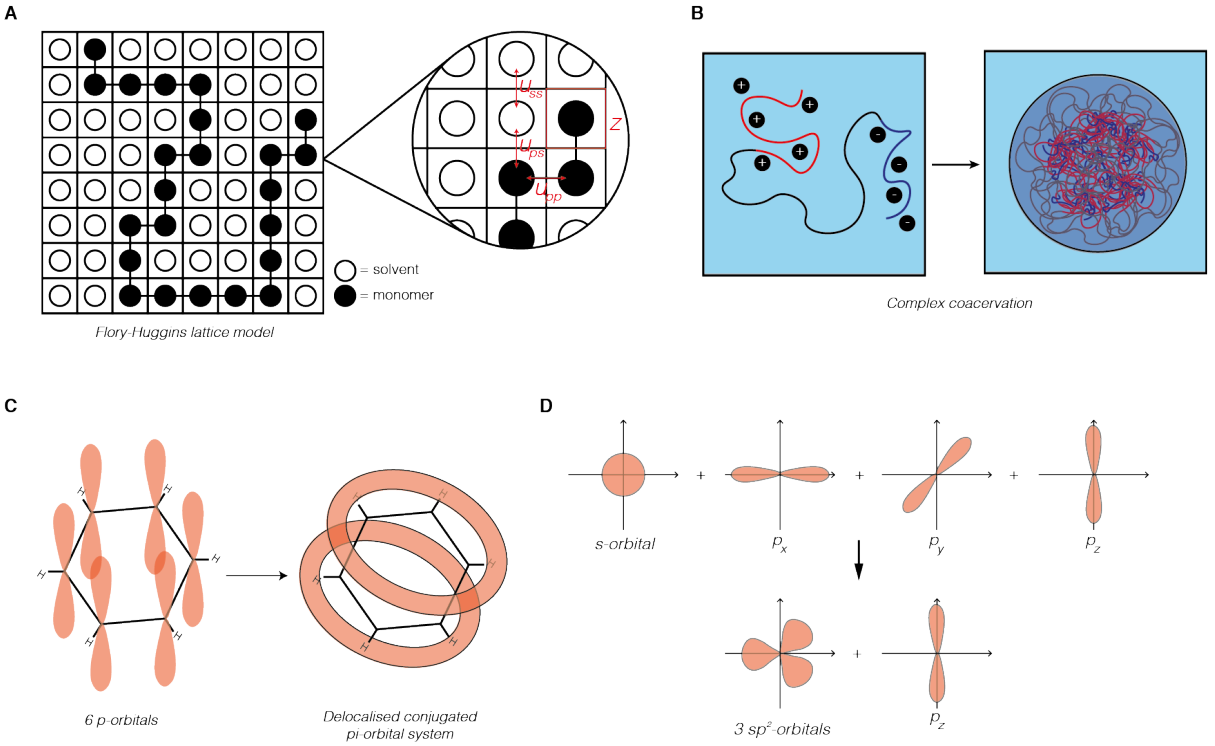



**Figure 1.5 Interactions driving phase transitions** (A) schematic of the Flory-Huggins model of dipole-dipole interactions; (B) schematic of the charge-charge interactions driving complex coacervation; (C) depiction of the fusion of adjacent orbital groups forming a conjugated pi-orbital system, with delocalised pi electrons forming clouds that become polarised during charge-pi interactions; (D) schematic depicting the process of $sp^2$ hybridization that ultimate drives pi-pi interactions

This was further investigated in studies assessing the influence of mixed charge domains (MCDs) within intrinsically disordered regions[153]; using similar R to K substitution experiments, synthetic dipeptide combinations were used to demonstrate the increased propensity for condensation resulting from an increase in net-positive charge driven by arginine residues. This effect was not observed for domains with a lysine-derived net-positive charge demonstrating that the guanidium side chain of arginine is the determinant of this phase behaviour. The multivalent nature of this guanidium ion, compared to the monovalent nature of the side chain amide of lysine, may be the driving force for this ability, as the arginine side chain is capable of π-π, cation-π and charge-charge interactions.

It is important to note that, whilst many of these biophysical determinants of phase separation can individually have large influences on the transitional behaviour of individual components, it is likely

that a combination of all the above can influence these abilities. Within the example just mentioned, for instance, even within R-enriched MCDs, which show an increased propensity for phase separation, a net-negative charge can reduce condensation and incorporation into phase-separated bodies[153]. This is a clear demonstration of the complementary nature of these interactions in finely modulating the behaviour of these constituent components.

## 1.3 Stress responses

The ability to adapt to environmental challenges is vital for survival and, as such, a number of cellular stress response pathways have evolved to deal with these potential threats. As has been demonstrated, many phase-separated organelles are involved in dynamic responses to such environmental stimuli. Much of this relates to stress, such as the formation of stress granules and P bodies in response to protein misfolding or heat stress. There is increasing evidence that a whole host of novel phase separation events are able to facilitate responses to stressful stimuli. The existence of such a variety of structures points towards a much broader role for phase separation in physiology than perhaps first thought.

The proteasome, for example, is a key mechanism in the maintenance of cellular proteostasis, with ubiquitinated substrates being targeted for protein degradation by proteasomal machinery. Recent evidence has identified a role for ubiquitin-chain-dependent phase separation in this process, condensing proteasomal components and their target proteins in transient, liquid-like foci, driven by the multivalent interactions made by the substrate-shuttling factor, RAD23B, with ubiquitin chains[154].

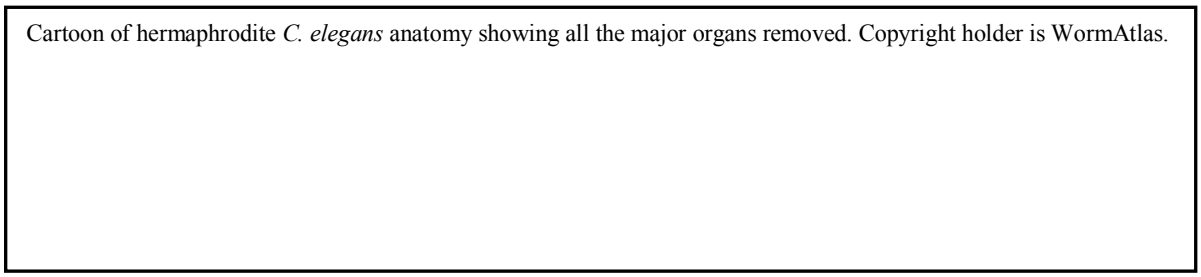
A role for phase separation has also been identified in modulating the response to tissue damage in *C. elegans*; the prion-like protein TIAR-2 forms granules capable of facilitating the inhibition of regeneration following axonal injury. This inhibition relies on the formation of these liquid-like assemblies, as demonstrated by the abolition of puncta formation via removal of the prion-like domain, which results in loss of this inhibition. As with all dynamic stress responses, the formation of these TIAR-2 condensates can be regulated – in this case, such regulation comes in the form of serine phosphorylation, with phospho-deficient mutants losing both the ability to form granules and the ability to inhibit regeneration[155].

An emerging role for phase transitions in the dynamic response to environmental stimuli is becoming increasingly clear – even amongst well-characterised stress responses. The following section explores stress response pathways within the nematode species *Caenorhabditis elegans*, although in most cases they represent highly evolutionarily conserved processes. Due to the often transient nature of these responses, it would be interesting to see if a further role for phase separation may be identified within these pathways.

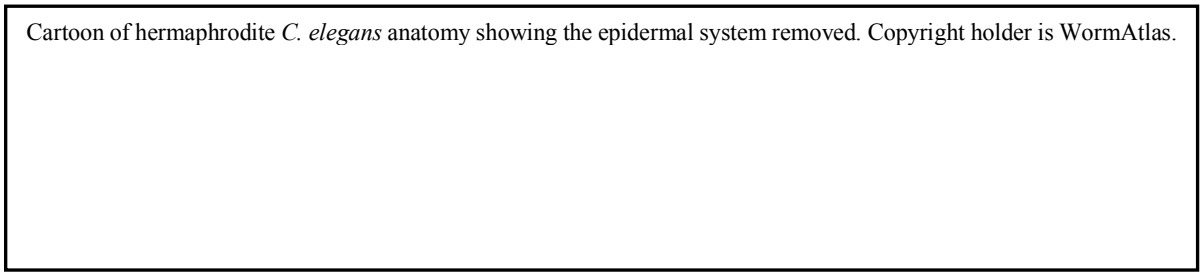
### *1.3.1 C. elegans as a model system for studying stress responses*

*Caenorhabditis elegans* is a soil-dwelling nematode species made up of a mere 959 somatic cells in the hermaphrodite. The simplicity of these animals is useful for investigating cellular processes and pathways within the physiologically relevant context of a multicellular, living organism (Figure 1.6A-C). This is important as many *in vivo* influences cannot be recapitulated within an *in vitro* setting such as cell culture[156–159].

A




B




C

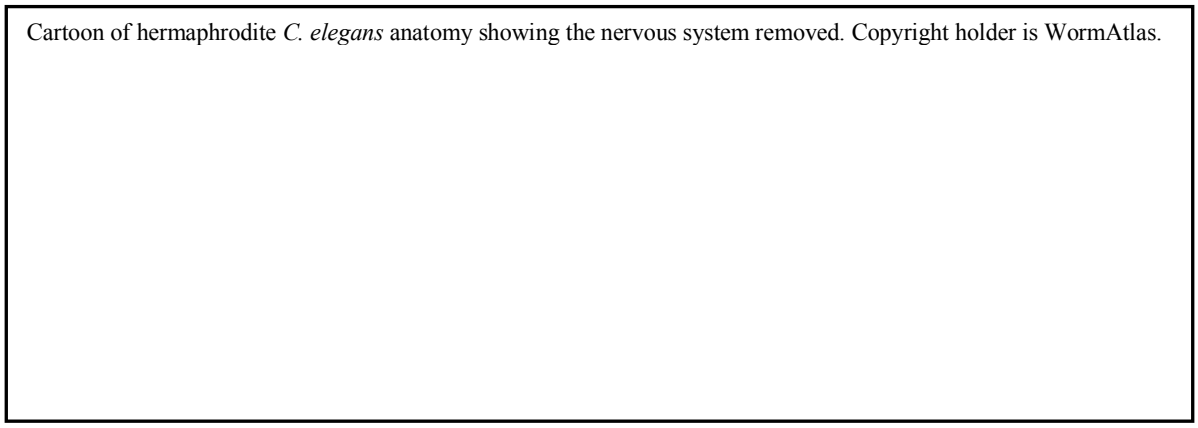



**Figure 1.6** An introduction to the anatomy of the adult *C. elegans* hermaphrodite showing (A) major organs including the pharynx, intestine, uterus and gonad; (B) the epidermal (hypodermal) system, including the seam cells; (C) the nervous system (adapted from WormAtlas)

These animals are capable of a diverse range of behaviours in response to the external environment. On an organismal level, this requires sensory processing and global motor coordination following reception of a cacophony of stimuli. This is facilitated by a small yet sophisticated nervous system made up of just 302 neurons in the hermaphrodite[158,159], and 381 in the male[160]. Within the animal, neuronal cell bodies generally cluster in ganglia, often located in the head and tail regions, with neuronal processes extending out from there[161]. These neurons can be categorised by function, with circuits of sensory, motor and interneurons communicating and enabling appropriate stimuli-specific responses[162]. These neurons communicate both synaptically and extrasynaptically, utilising a similar diversity of neurotransmitters to that seen in mammalian species, with some exceptions – the biogenic amine,

octopamine, for example, is a tyramine-derivative that is much less abundant in vertebrate systems than in the nematode[163]. The nervous system is capable of the systemic coordination of stress responses – for example, splicing of the nematode ER stress-responsive transcription factor XBP-1 in neurons results in a global activation of cellular stress pathways via cell non-autonomous signaling[164].

Whilst the nervous system plays an important role in coordinating responses, it is not the only tissue that influences global states. The intestine, for example, is a highly metabolically active tissue, capable of altering global metabolic conditions in response to environmental stimuli and stress conditions[165,166]. In addition to this, intestinal tissue plays roles that are traditionally associated with the gut, such as secretion of digestive enzymes into the lumen for bacterial digestion, leading to uptake of appropriate nutrients[167]. In terms of stress responses, the intestine is also thought to be involved in the initiation of the innate immune response, due to it being one of the first sites of pathogenic invasion[168].

Cartoon of *C. elegans* lifecycle removed. Copyright holder is WormAtlas.

**Figure 1.7** A schematic depicting the life cycle of *C. elegans*, beginning with *in utero* development of fertilised eggs in gravid adult animals, to gastrulation, embryonic development, larval development, and ultimately adulthood (adapted from WormAtlas)

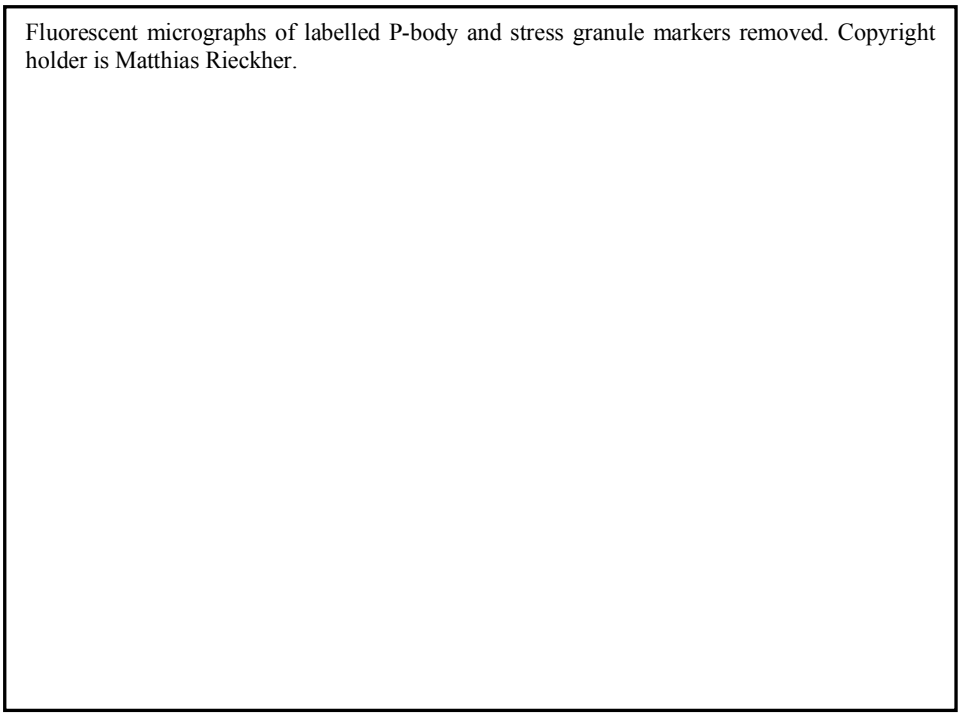
The anterior portion of the intestine is anchored to the pharynx. This epithelial organ is responsible for the initial bacterial uptake step via the buccal cavity. Within the pharynx, this bacteria is then ground and transported into the intestine for digestion. The pharyngeal grinder that facilitates this is made up of

3 pairs of muscle cells – contraction of these muscles causes a rotation motion that acts to disrupt the bacteria[169]. These muscle cells are interspersed by marginal cells that act as a repository for nutrients in a sphingomyelin-dependent fashion[170]. In addition to this, the pharynx has a dedicated nervous system, with 20 neurons regulating both contractility of the pharyngeal muscles, but also communicating nutritional status and relaying signals from the external environment to the rest of the organism[171,172].

Surrounding these organs is a tough cuticle rich in structural extracellular matrix components, such as collagen[173,174], that are secreted by epithelial cells including the hypodermis, the seam cells and the interfacial cells. This structure acts as an exoskeletal, aiding motility of animals, as well as providing a vital barrier function, protecting the animal from the external environment – as a result, it acts as a primary line of defence against environmental challenges such as pathogenic invasion. This ability to protect animals from pathogens is aided by the presence of a negatively charged epicuticular coating rich in lipids and glycoproteins[175], with animals defective for genes encoding these epicuticular components displaying increased susceptibility to bacterial infection[176,177]. Beyond these functions, the cuticle is also important for the maintenance of body shape[178], with a new cuticle grown after each larval stage upon shedding of the old[179].

The epithelial system that secretes these cuticular components is comprised of two different classes of epithelial tissue – either the hypodermis or specialised epithelial cells. The hypodermis encapsulates the entire organism, with the main hypodermal syncytium, hyp-7, spanning the main body of the animal, while smaller hypodermal cells lie in the anterior and posterior regions[180]. Throughout the life of an organism, these tissues take on different roles; during development, for example, they are involved in establishing the basic body plan and regulating cell fate determination, as well as guiding axonal migration[181]. In adulthood, however, the hypodermis adopts roles in nutrient storage and maintenance of cuticle integrity.

There are a number of cell types that fall under the branch of specialised epithelial cell, including the seam cells, which form rows of 16 cells running longitudinally along both the left and right sides of the animal. Akin to the hypodermis, following larval development, these cells fuse and form a multinucleate syncytium[159] that also plays a role in the secretion of cuticular components[182]. In addition to the main cuticle components, seam cells also produce cuticular ridges along the length of the animal known as alae[158]; formation of these structures is often used as a readout of seam cell terminal differentiation. Prior to this point, however, these cells are important in the specification of some neuronal and glial populations, with PVW, PVN and PHC neurons, for example, deriving from seam cell divisions[159].




| Compartment | Localization | Role |
|---|---|---|
| Processing bodies | Cytoplasm | *Regulation of mRNA translation and decay* |
| P granules | Perinuclear | *Germlines compartments regulating posttranscriptional activity of RNA* |
| Stress granules | Cytoplasm | *Form under stress, sequestering specific RNAs, stalling their translation* |
| Nucleoli | Nuclear | *Formation of ribosomes, densely packed with RNA* |
| Paraspeckles | Nuclear | *Unclear, but thought to localize specific proteins within the nucleus* |
| Cajal bodies | Nuclear | *Sites of snRNP formation, involved with mRNA processing* |

**Figure 1.8** Phase-separated organelles within *C. elegans* (A-E) animals expressing IFE-2::GFP and DCAP-1::RFP as markers for stress granules and processing bodies respectively, before and after heat shock, demonstrating a degree of colocalization after this stress stimuli (adapted from *Rieckher et al., 2013*)

**Table 1.1** A summary of a number of conserved phase-separated structures observed in these nematode species, their localization within the cell, and their function

This microscopic nematode, with a lifespan of only around 20 days, undergoes a number of recognisable developmental processes post-fertilisation, including periods of embryonic and larval development, before reaching fully fertile adulthood (Figure 1.7). As this process is relatively rapid, the generation time is short – this allows for large populations sizes to monitored in a relatively short space of time.

The tractability of genetic manipulation in *C. elegans* makes this model organism attractive to work with, particularly in light of recent technological advancements that have allowed for the CRISPR-mediated editing of endogenous loci. Unlike higher organisms, the short lifecycle of these animals allows age-related phenotypes to be investigated in real time – this is particularly useful when

investigating stress-related phenotypes, as these are often exacerbated in an age-dependent manner. In addition to this, these animals are transparent – a powerful characteristic for monitoring transient and dynamic effects visually, via fluorescence tagging for example, as these processes can be monitored in real time within an *in vivo* setting. When considering proteins structures that aggregate, coalesce, fuse and propagate, the ability to observe these processes in a living organism is invaluable.

There are a number of phase-separated organelles that have already been identified and investigated in *C. elegans* (Figure 1.8A-B). Many of these are have homologs across biology and thus, information gained from this model organism can be translated into higher organisms. A number of these organelles have known roles in stress responses – understanding how these fit in with known stress response pathways will aid our understanding of the role that these structures play within those pathways, and how they are regulated as a part of coordinated responses to stress.

### *1.3.2 Innate Immune Response*

In order to protect against the invasion of pathogens, *C. elegans* possess evolutionarily conserved innate immune pathways including the p38 MAPK and the DAF-2/DAF-16 pathways (Figure 1.9). A genetic screen for mutants that showed increased susceptibility to the pathogenic bacteria *Pseudomonas aeruginosa* (PA14) first led to the identification of genes involved in the phosphorylation cascade pathway of p38 MAPK. This screen identified *sek-1*, a MAP kinase kinase, and *nsy-1*, a MAP kinase kinase kinase, as being involved in this response[183]. Knockouts of either of these genes prevent the subsequent phosphorylation of PMK-1, the MAP kinase itself, thus demonstrating that this acts downstream from these proteins[184]. This pathway elicits an immune response by altering gene expression patterns to combat the pathogenic threat – upstream components of NSY-1 are able to detect such threats, driving this phosphorylation cascade and thus stimulating a remodelling of gene expression. For example, the murine homolog of NSY-1, ASK1, is activated downstream of lipopolysaccharide detection, thus demonstrating its direct stimulation by pathogen associated molecular patterns.

The DAF-2/DAF-16 pathway has also been implicated in innate immune responses, with *daf-2* mutants showing resistance to PA14 killing[184]. This is a DAF-16 dependent effect as *daf-2;daf-16* double mutants do not have this reduced susceptibility[185]; this suggests that the immune response is facilitated via the canonical DAF-2/DAF-16 pathway, whereby DAF-2 activation results in DAF-16 phosphorylation, inhibiting its movement into the nucleus where it would normally induce expression of target genes. Thus, antagonism of the DAF-2 receptor in innate immune responses can elicit DAF-16-mediated changes in gene expression.

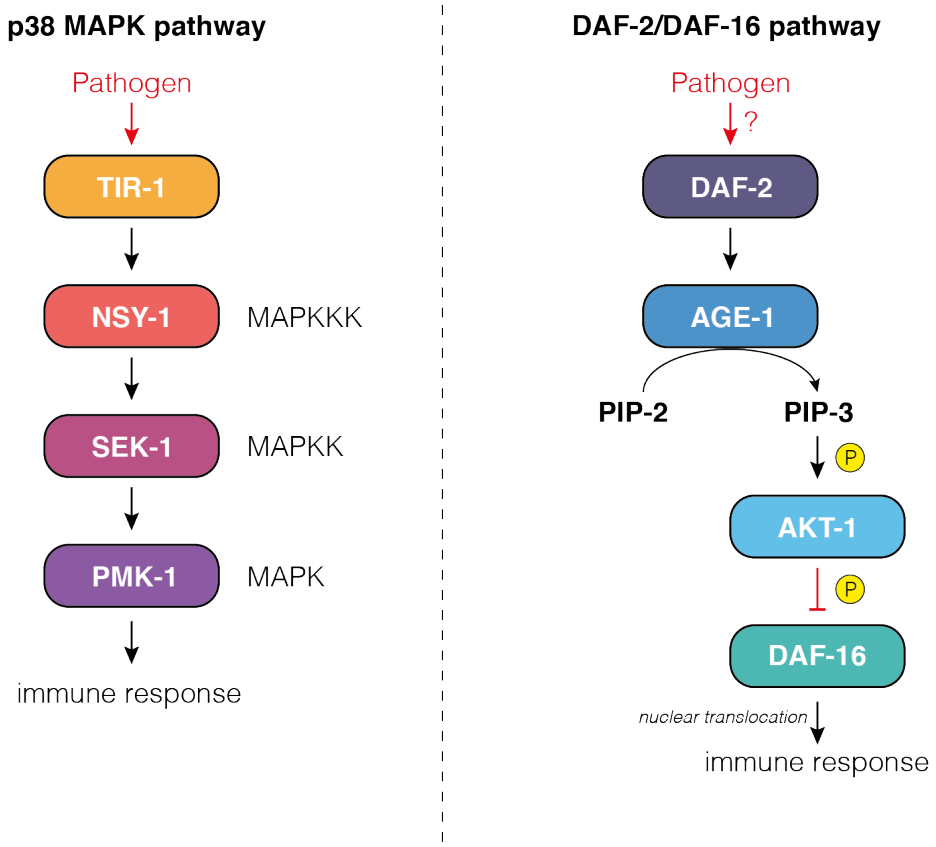



**Figure 1.9** Innate immune response pathways in *C. elegans* (A) the p38 MAPK pathway; (B) the DAF-2/DAF-16 pathway

### *1.3.3 Unfolded Protein Response*

Under normal functioning, the endoplasmic reticulum facilitates the proper folding of newly translated proteins via an array of molecular chaperones found within the ER lumen. An aberrant accumulation of misfolded or unfolded proteins within this compartment, however, can lead to the activation of a tripartite response system known as the unfolded protein response of the endoplasmic reticulum ($UPR^{ER}$), involving one of three transmembrane ER stress sensors; IRE1, PERK and ATF6, corresponding to IRE-1, PEK-1 and ATF-6 in *C. elegans* respectively (Figure 1.10)[186]. The molecular chaperone BiP is involved in the regulation of both IRE1 and PERK, functioning as an indirect readout of protein folding state within the ER. During unstressed, basal conditions, BiP associates with the luminal domains of both IRE1 and PERK, maintaining them in their monomeric forms and thus repressing their activation. Upon the accumulation of unfolded proteins, however, BiP can reversibly dissociate from these transmembrane proteins and bind to the unfolded protein instead. This dissociation derepresses both IRE1 and PERK, leading to their oligomerisation and activation[187].

In the case of IRE1, this activation is stimulated by the autophosphorylation of the dimer. In this state, the endonuclease activity of IRE1 is activated, resulting in the splicing of X box-binding protein 1 (*xbp-1*) mRNA[188]. In its spliced form, *xbp-1s* can be translated into a functional transcription factor that translocates into the nucleus, driving the expression of $UPR^{ER}$ target genes with the ultimate aim of resolving that ER stress[186].

PERK, meanwhile, triggers a translational attenuation, thus reducing the folding demands of the ER. Following its activation, PERK is able to phosphorylate the translation initiation complex eIF-2α, which, as previously described, prevents the formation of pre-initiation complexes involved with the loading of the initiator tRNA onto the 40s ribosomal subunit[142]. Consequently, protein synthesis is halted, allowing ER resident proteins to clear the accumulated unfolded proteins without further translation adding to the protein-folding burden. This effect can be demonstrated clearly in $PERK^{-/-}$ mouse embryonic stem cell lines – eIF-2α in these cells cannot be phosphorylated after treatment with pharmacological inducers of ER stress, such as the N-linked glycosylation inhibitor tunicamycin, and as a result display increased sensitivity to this stress[189].

ATF6, on the other hand, works in a similar fashion to XBP1 by regulating the expression of $UPR^{ER}$ response genes. The N-terminal cytoplasmic domain of this protein contains a basic leucine zipper motif (bZIP) that protrudes into the cytoplasm, whilst the remainder of the protein is embedded in the ER membrane. Upon stress, this N terminal fragment undergoes a proteolytic cleavage, releasing it into the cytoplasm, from which point it is then able to translocate into the nucleus and induce the expression of genes including BiP[190].

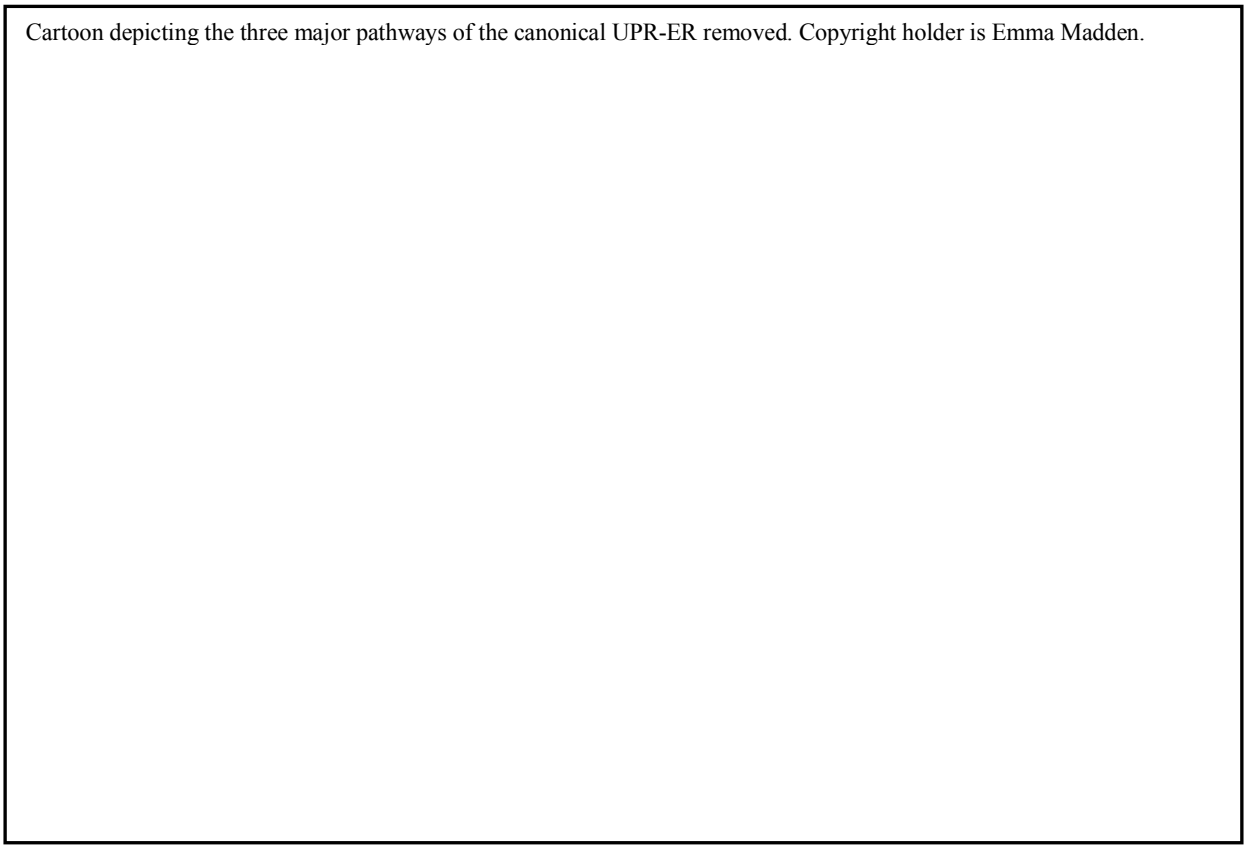



**Figure 1.10** Schematic outlining the three major pathways of the canonical unfolded protein response (UPR) of the endoplasmic reticulum, including the ATF-6. PEK-1 and IRE-1 branches (adapted from BioRender, courtesy of E. Madden)

Although these branches are very much distinct in their own right, they do display a degree of cross talk, compensating for one another in case of deficiency. This can be demonstrated in mutants of *ire-1/xbp-*

*1* paired with *pek-1* or *atf-6* mutations; whilst single mutant animals are viable, double mutants display larval lethality[191,192]. The underlying molecular underpinnings of this overlap have been investigated in animals with XBP-1-deficiencies. These *xbp-1* mutant animals show an increased level of IRE-1 activity and PEK-1-mediated eIF-2α phosphorylation in unstressed conditions compared to wild type animals[193]. This suggests that *xbp-1* mutation can trigger a constitutive activation of ER stress responses in basal conditions, thus allowing for alternative branches of the UPR to coordinately compensate for the loss of the IRE-1/XBP-1 axis.

### *1.3.4 The "Non-Canonical" Unfolded Protein Response*

Upon inhibition of the *xbp-1* branch of the $UPR^{ER}$, a non-canonical pathway can be stimulated to compensate for the loss of this response (Figure 1.11). The protein family involved in this pathway, known as the ABU family of proteins (**A**ctivated in **B**locked **U**PR), is capable of eliciting a stress response in *C. elegans* under these conditions, mitigating against the loss of this key UPR component. Nine of these non-canonical components, namely *abu-1* to *abu-9*, were first identified following a screen of genes upregulated in ER-stressed versus unstressed animals in an *xbp-1* mutant background[194]. These nine genes, which all bear a close resemblance to one another with an N-terminal signal sequence and a transmembrane domain, can be knocked down simultaneously using RNAi against a highly related 3' stretch of sequence. In keeping with this initial evidence, RNAi knockdown of these genes in an *xbp-1* or *ire-1* mutant background reduces survivorship in cadmium-treated ER-stressed animals compared to those on control RNAi – this suggests that the *abu* gene family does indeed play a role in responding to ER stress when traditional pathways are compromised. However, the mechanisms underlying this protection are as yet unknown, with the function of individual ABU proteins still to be elucidated.

### *1.3.5 UPR and Immunity*

A role for XBP-1 has been identified in protecting against pathogenic insult, thus highlighting a link between the UPR and the innate immune system. Under non-pathogenic conditions, *xbp-1* mutants develop normally whilst those exposed to *Pseudomonas aeruginosa* (PA14) develop abnormal ER morphology and display larval lethality; these phenotypes are suppressed when PMK-1-mediated immunity is lost, and are exacerbated when this pathway is hyperactivated, demonstrating that this effect is mediated via the p38-MAPK pathway. The link between these two stress response systems is further validated by the expression of *hsp-4p::gfp*, a marker for ER stress, which is upregulated following PA14 exposure, an effect that is lost in *pmk-1* mutants[195]. This effect does not, however, protect against the immune threat itself but rather against the immune response; this has been demonstrated in *xbp-1;pmk-1* double mutants, which exhibit increased survival and development on PA14 compared to single *xbp-1* mutants. This suggests that the PMK-1-mediated immune response is having a detrimental impact upon the animal – an effect that XBP-1 mitigates against. It has been proposed that this is as a consequence of the increased demands placed upon the endoplasmic reticulum upon activation of such responses, with around half of the roughly 300 genes induced by the innate immune system shuttling

through the endomembrane system[184]. This large increase in nascent protein production can have negative effects on an organism due to the excessive strain it puts on the folding capacity of the system; the XBP-1 branch of the UPR can protect against this innate immune response-induced increase in ER stress[195].

Evidence has also emerged of a role for the ABU proteins in the innate immune response, with a set of *abu/pqn* (prion-like glutamine[Q]/asparagine[N]-rich domain-bearing protein]) genes being downregulated in mutants of the phagocytic receptor CED-1, pointing towards a cross over between these stress response pathways[196,197]. In keeping with its role in immunity, *ced-1* mutants are deficient in bacterial clearance and are rapidly killed upon exposure to live bacteria. Genome expression analysis identified a group of genes highly upregulated by CED-1 – this group contained an enrichment of *abu/pqn* genes, suggesting that CED-1 can activate genes of the non-canonical UPR response. This has functional consequences, with abrogation of these genes reducing survival when exposed to live, but not dead, pathogenic bacteria. This can be visualised *in vivo* by the increased invasion of GFP-labelled *S. enterica* observed in *ced-1* mutants when treated with RNAi against either *abu-1* or *abu-11*. Hence, this demonstrates the role of CED-1 in activating the non-canonical UPR pathway to elicit an immune response.

### *1.3.6 Regulation by OCTR-1*

Neuronal control of the innate immune response is vital for ensuring that such responses are not under- or over-activated; under-activation would exacerbate infection whilst over-activation could result in tissue damage. The p38 MAPK pathway, for example, is finely tuned to minimise such damage – this can be demonstrated in gain-of-function mutations in which positive regulators of this pathway, such as NSY-1, trigger an aberrant and heightened response, ultimately resulting in toxicity[198]. In *C. elegans*, a neuronal circuit involving the G-protein coupled receptor, NPR-1, has been shown to suppress innate immune responses, with loss of this receptor reducing survival upon exposure to *Pseudomonas aeruginosa* (PA14)[199]. This is a clear demonstration of the level of regulation that can be exerted on the immune system by the nervous system. Such regulatory effects have also been identified for OCTR-1, another neuronal GPCR. Unlike NPR-1, however, *octr-1* mutants show a decreased susceptibility to PA14 killing compared to wild type animals, suggesting that this GPCR acts to inhibit the innate immune response. This improved survival phenotype of *octr-1* mutants can be suppressed when treated with RNAi against *pmk-1*, thus demonstrating that OCTR-1 negatively regulates the innate immune response by inhibiting the p38 MAPK pathway[200]. Rescue of this phenotype can be elicited by expression of *octr-1* under control of the *sra-6* promoter, which restricts expression to the neurons ASH, ASI and PVQ, although it is important to note that PVQ neurons do not endogenously express *octr-1,* thus suggesting that the expression of OCTR-1 in solely ASH and ASI drives this inhibition of the innate immune response downstream. This is further corroborated by the degeneration of the ASH neuron using neuron-

specific polyQ expression, which recapitulated the enhanced resistance to PA14 observed in *octr-1* mutants.

Analysis of the genes upregulated in *octr-1* mutants compared to wild type animals highlights a subset of genes upregulated by CED-1, including some of the *abu/pqn* genes previously identified as being regulated by this phagocytic receptor[196]. Indeed, *ced-1* mutants can inhibit the enhanced pathogen resistance of *octr-1* mutants, implying that the inhibitory effect exerted by OCTR-1 is facilitated via CED-1. Similar suppression phenotypes are observed following RNAi knockdown of *abu-1, -7, -8, -12* and *-13*, although no additive effect was seen when this RNAi was paired with *octr-1;ced-1* double mutants; this illustrates that, in the context of enhancing the pathogen resistance of *octr-1* mutants, *ced-1* and the *abu* genes are in the same genetic pathway[200].

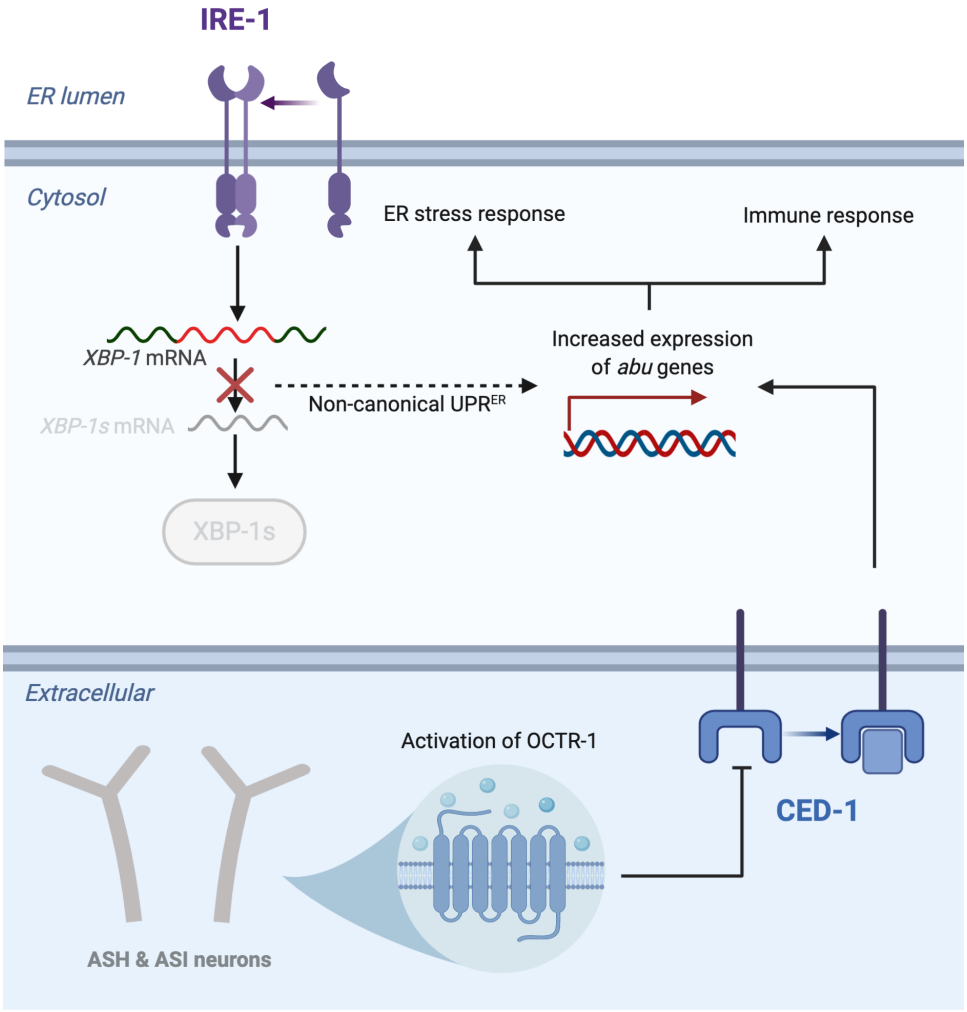



**Figure 1.11** Schematic outlining the function of the ABU family in both a non-canonical ER stress response pathway and within the innate immune system, in response to CED-1 activation (created with BioRender)

In addition to neuronal regulation of the non-canonical pathway via OCTR-1, there is also evidence of its role in the regulation of the canonical XBP-1 pathway in adulthood. During postembryonic development, *octr-1* mutants show no difference in the XBP-1 splicing activity of IRE-1 compared to wild type animals, suggesting that OCTR-1 activity does not influence this pathway prior to adulthood.

In post-developmental stages, however, this does not seem to be the case, with expression of a number of XBP-1-target genes upregulated in *octr-1* mutants following PA14 exposure. Similarly, these mutants also show a greater *hsp-4p::gfp* expression after exposure than wild type animals, suggesting an increased activity of the XBP-1 branch when OCTR-1 activity is lost[201].

Thus, on the whole, this demonstrates that both the canonical and non-canonical UPR pathways can be negatively regulated by the neuronal GPCR, OCTR-1, with the ABU family of proteins regulated via the phagocytic receptor CED-1.

## 1.4 Aims of thesis

Thus far, the role of prion-like proteins in the facilitation of stimuli-induced phase separation has been explored, but a further understanding of the diverse roles that these proteins play physiologically remains to be investigated. It is clear that the characteristics of prion-like domains - such as multivalency, low complexity and intrinsic disorder - incline these proteins towards such transitions. Identifying and understanding the function that these proteins play in phase separation has been enlightening in the pursuit of a mechanistic understanding of protein aggregation disorders. As has been illustrated, many of these aggregation disorders result from aberrant transitions of liquid- or hydrogel-like condensates as a consequence of misregulation of these structures. Developing our understanding of how these membraneless organelles are regulated, and in response to which stimuli, may help us to understand the mechanics of this prion-like aggregation both physiologically and pathologically.

Using *Caenorhabditis elegans* as a model system, we will investigate a novel prion-like protein and characterise its functionality. This model system provides a great deal of physiological and molecular flexibility in this process, whilst maintaining the complexity of a multicellular, multi-tissue organism.

In this thesis, I will aim to:

1. Identify and functionally characterise a novel prion-like protein in *C. elegans*
2. Investigate the characteristics of this prion-like protein *in vitro* and *in vivo*
3. Understand both the protein and RNA interactions made by this protein, and the physiological implications of such interactions

## 1.5 References


1. Mercer, R. C. C. *et al.* The Prion Protein Modulates A-type K+ Currents Mediated by Kv4.2 Complexes through Dipeptidyl Aminopeptidase-like Protein 6. *J. Biol. Chem.* **288**, 37241–37255 (2013).
2. Khosravani, H. *et al.* Prion protein attenuates excitotoxicity by inhibiting NMDA receptors. *J. Cell Biol.* **181**, 551 – 565 (2008).
3. Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science.* **216**, 136 – 144 (1982).
4. Alper, T., Cramp, W. A., Haig, D. A. & Clarke, M. C. Does the Agent of Scrapie Replicate without Nucleic Acid ? *Nature* **214**, 764–766 (1967).
5. Griffith, J. S. Nature of the Scrapie Agent: Self-replication and Scrapie. *Nature* **215**, 1043–1044 (1967).
6. Pan, K. M. *et al.* Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci.* **90**, 10962 – 10966 (1993).
7. Caughey, B. W. *et al.* Secondary structure analysis of the scrapie-associated protein PrP 27-30 in water by infrared spectroscopy. *Biochemistry* **30**, 7672–7680 (1991).
8. Gousset, K. *et al.* Prions hijack tunnelling nanotubes for intercellular spread. *Nat. Cell Biol.* **11**, 328–336 (2009).
9. Fevrier, B. *et al.* Cells release prions in association with exosomes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9683 – 9688 (2004).
10. Wadia, J. S., Schaller, M., Williamson, R. A. & Dowdy, S. F. Pathologic Prion Protein Infects Cells by Lipid-Raft Dependent Macropinocytosis. *PLoS One* **3**, e3314 (2008).
11. Ruttkay-Nedecky, B. *et al.* Prion protein and its interactions with metal ions (Cu2+, Zn2+, and Cd2+) and metallothionein 3. *ADMET DMPK* **3**, (2015).
12. Simoneau, S. *et al.* In Vitro and In Vivo Neurotoxicity of Prion Protein Oligomers. *PLOS Pathog.* **3**, e125 (2007).
13. Büeler, H. *et al.* Mice devoid of PrP are resistant to scrapie. *Cell* **73**, 1339–1347 (1993).
14. Collinge, J. & Clarke, A. R. A General Model of Prion Strains and Their Pathogenicity. *Science.* **318**, 930 – 936 (2007).
15. Saborio, G. P., Permanne, B. & Soto, C. Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* **411**, 810–813 (2001).
16. Atarashi, R., Sano, K., Satoh, K. & Nishida, N. Real-time quaking-induced conversion: a highly sensitive assay for prion detection. *Prion* **5**, 150–153 (2011).
17. Goedert, M. & Spillantini, M. G. A Century of Alzheimer's Disease. *Science.* **314**, 777 – 781 (2006).
18. Lagier-Tourenne, C. & Cleveland, D. W. Rethinking ALS: The FUS about TDP-43. *Cell* **136**, 1001–1004 (2009).

19. Kordower, J. H., Chu, Y., Hauser, R. A., Freeman, T. B. & Olanow, C. W. Lewy body–like pathology in long-term embryonic nigral transplants in Parkinson's disease. *Nat. Med.* **14**, 504–506 (2008).
20. Kordower, J. H. *et al.* Neuropathological Evidence of Graft Survival and Striatal Reinnervation after the Transplantation of Fetal Mesencephalic Tissue in a Patient with Parkinson's Disease. *N. Engl. J. Med.* **332**, 1118–1124 (1995).
21. Watts, J. C. *et al.* Transmission of multiple system atrophy prions to transgenic mice. *Proc. Natl. Acad. Sci.* **110**, 19555 – 19560 (2013).
22. Prusiner, S. B. *et al.* Evidence for α-synuclein prions causing multiple system atrophy in humans with parkinsonism. *Proc. Natl. Acad. Sci.* **112**, E5308–E5317 (2015).
23. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D. & Crowther, R. A. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron* **3**, 519–526 (2017).
24. Spillantini, M. G. & Goedert, M. Tau pathology and neurodegeneration. *Lancet Neurol.* **12**, 609–622 (2013).
25. Liu, L. *et al.* Trans-Synaptic Spread of Tau Pathology In Vivo. *PLoS One* **7**, e31302 (2012).
26. Kfoury, N., Holmes, B. B., Jiang, H., Holtzman, D. M. & Diamond, M. I. Trans-cellular Propagation of Tau Aggregation by Fibrillar Species. *J. Biol. Chem.* **287**, 19440–19451 (2012).
27. Sanders, D. W. *et al.* Distinct Tau Prion Strains Propagate in Cells and Mice and Define Different Tauopathies. *Neuron* **82**, 1271–1288 (2014).
28. Wasmer, C. *et al.* Amyloid Fibrils of the HET-s(218–289) Prion Form a β Solenoid with a Triangular Hydrophobic Core. *Science.* **319**, 1523 – 1526 (2008).
29. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
30. Falcon, B. *et al.* Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).
31. Falcon, B. *et al.* Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
32. Coschigano, P. W. & Magasanik, B. The URE2 gene product of *Saccharomyces cerevisiae* plays an important role in the cellular response to the nitrogen source and has homology to glutathione s-transferases. *Mol. Cell. Biol.* **11**, 822 – 832 (1991).
33. Turoscy, V. & Cooper, T. G. Ureidosuccinate is transported by the allantoate transport system in *Saccharomyces cerevisiae*. *J. Bacteriol.* **169**, 2598 – 2600 (1987).
34. Courchesne, W. E. & Magasanik, B. Regulation of nitrogen assimilation in *Saccharomyces cerevisiae*: roles of the URE2 and GLN3 genes. *J. Bacteriol.* **170**, 708 – 713 (1988).
35. Mitchell, A. P. & Magasanik, B. Regulation of glutamine-repressible gene products by the GLN3 function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 2758 – 2766 (1984).

36. Blinder, D., Coschigano, P. W. & Magasanik, B. Interaction of the GATA factor Gln3p with the nitrogen regulator Ure2p in *Saccharomyces cerevisiae*. *J. Bacteriol.* **178**, 4734 – 4736 (1996).
37. Edskes, H. K., Hanover, J. A. & Wickner, R. B. Mks1p Is a Regulator of Nitrogen Catabolism Upstream of Ure2p in *Saccharomyces cerevisiae*. *Genetics* **153**, 585 – 594 (1999).
38. Watson, T. G. Inhibition of Proline Utilization by Glutamate during Steady-state Growth of *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **103**, 123–126 (1977).
39. Rai, R., Genbauffe, F., Lea, H. Z. & Cooper, T. G. Transcriptional regulation of the DAL5 gene in *Saccharomyces cerevisiae*. *J. Bacteriol.* **169**, 3521 – 3524 (1987).
40. Lacroute, F. Non-Mendelian Mutation Allowing Ureidosuccinic Acid Uptake in Yeast. *J. Bacteriol.* **106**, 519 – 522 (1971).
41. Drillien, R. & Lacroute, F. Ureidosuccinic Acid Uptake in Yeast and Some Aspects of Its Regulation. *J. Bacteriol.* **109**, 203 – 208 (1972).
42. Aigle, M. & Lacroute, F. Genetical aspects of [URE3], a non-mitochondrial, cytoplasmically inherited mutation in yeast. *Mol. Gen. Genet.* **136**, 327–335 (1975).
43. Wickner, R. B. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science.* **264**, 566 – 569 (1994).
44. Masison, D. C., Maddelein, M.-L. & Wickner, R. B. The prion model for [URE3] of yeast: Spontaneous generation and requirements for propagation. *Proc. Natl. Acad. Sci.* **94**, 12503 – 12508 (1997).
45. Saupe, S. J. Molecular Genetics of Heterokaryon Incompatibility in Filamentous Ascomycetes. *Microbiol. Mol. Biol. Rev.* **64**, 489 – 502 (2000).
46. Coustou, V., Deleu, C., Saupe, S. & Begueret, J. The protein product of the het-s heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proc. Natl. Acad. Sci.* **94**, 9773 – 9778 (1997).
47. Maddelein, M.-L., Dos Reis, S., Duvezin-Caubet, S., Coulary-Salin, B. & Saupe, S. J. Amyloid aggregates of the HET-s prion protein are infectious. *Proc. Natl. Acad. Sci.* **99**, 7402 – 7407 (2002).
48. Deleu, C., Clavé, C. & Bégueret, J. A single amino acid difference is sufficient to elicit vegetative incompatibility in the fungus *Podospora anserina*. *Genetics* **135**, 45 – 52 (1993).
49. Coustou-Linares, V., Maddelein, M.-L., Bégueret, J. & Saupe, S. J. In vivo aggregation of the HET-s prion protein of the fungus *Podospora anserina*. *Mol. Microbiol.* **42**, 1325–1335 (2001).
50. Saupe, S. J. & Daskalov, A. The [Het-s] Prion, an Amyloid Fold as a Cell Death Activation Trigger. *PLOS Pathog.* **8**, e1002687 (2012).
51. Greenwald, J. *et al.* The Mechanism of Prion Inhibition by HET-S. *Mol. Cell* **38**, 889–899 (2010).
52. Mathur, V., Seuring, C., Riek, R., Saupe, S. J. & Liebman, S. W. Localization of HET-S to the

Cell Periphery, Not to [Het-s] Aggregates, Is Associated with [Het-s]–HET-S Toxicity. *Mol. Cell. Biol.* **32**, 139 – 153 (2012).

53. Hake, L. E. & Richter, J. D. CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. *Cell* **79**, 617–627 (1994).
54. Mastushita-Sakai, T., White-Grindley, E., Samuelson, J., Seidel, C. & Si, K. Drosophila Orb2 targets genes involved in neuronal growth, synapse formation, and protein turnover. *Proc. Natl. Acad. Sci.* **107**, 11987 – 11992 (2010).
55. Theis, M., Si, K. & Kandel, E. R. Two previously undescribed members of the mouse CPEB family of genes and their inducible expression in the principal cell layers of the hippocampus. *Proc. Natl. Acad. Sci.* **100**, 9602 – 9607 (2003).
56. Martin, K. C. *et al.* Synapse-Specific, Long-Term Facilitation of Aplysia Sensory to Motor Synapses: A Function for Local Protein Synthesis in Memory Storage. *Cell* **91**, 927–938 (1997).
57. Si, K. *et al.* A Neuronal Isoform of CPEB Regulates Local Protein Synthesis and Stabilizes Synapse-Specific Long-Term Facilitation in Aplysia. *Cell* **115**, 893–904 (2003).
58. Si, K., Lindquist, S. & Kandel, E. R. A Neuronal Isoform of the Aplysia CPEB Has Prion-Like Properties. *Cell* **115**, 879–891 (2003).
59. Si, K., Choi, Y.-B., White-Grindley, E., Majumdar, A. & Kandel, E. R. Aplysia CPEB Can Form Prion-like Multimers in Sensory Neurons that Contribute to Long-Term Facilitation. *Cell* **140**, 421–435 (2010).
60. Fioriti, L. *et al.* The Persistence of Hippocampal-Based Memory Requires Protein Synthesis Mediated by the Prion-like Protein CPEB3. *Neuron* **86**, 1433–1448 (2015).
61. Pavlopoulos, E. *et al.* Neuralized1 Activates CPEB3: A Function for Nonproteolytic Ubiquitin in Synaptic Plasticity and Memory Storage. *Cell* **147**, 1369–1383 (2011).
62. Drisaldi, B. *et al.* SUMOylation Is an Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. *Cell Rep.* **11**, 1694–1702 (2015).
63. Huang, Y.-S., Kan, M.-C., Lin, C.-L. & Richter, J. D. CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *EMBO J.* **25**, 4865–4876 (2006).
64. Li, L. & Lindquist, S. Creating a Protein-Based Element of Inheritance. *Science.* **287**, 661 – 664 (2000).
65. Ross, E. D., Baxa, U. & Wickner, R. B. Scrambled Prion Domains Form Prions and Amyloid. *Mol. Cell. Biol.* **24**, 7206 – 7213 (2004).
66. Maurer-Stroh, S. *et al.* Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* **7**, 237–242 (2010).
67. Gonzalez Nelson, A. C. *et al.* Increasing Prion Propensity by Hydrophobic Insertion. *PLoS One* **9**, e89286 (2014).
68. MacLea, K. S. *et al.* Distinct Amino Acid Compositional Requirements for Formation and

Maintenance of the [PSI+] Prion in Yeast. *Mol. Cell. Biol.* **35**, 899 – 911 (2015).

69. Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. *Science.* **268**, 880 – 884 (1995).
70. Alexandrov, I. M., Vishnevskaya, A. B., Ter-Avanesyan, M. D. & Kushnirov, V. V. Appearance and propagation of polyglutamine-based amyloids in yeast: tyrosine residues enable polymer fragmentation. *J. Biol. Chem.* **283**, 15185–15192 (2008).
71. Kryndushkin, D. S., Alexandrov, I. M., Ter-Avanesyan, M. D. & Kushnirov, V. V. Yeast [PSI+] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104. *J. Biol. Chem.* **278**, 49636–49643 (2003).
72. Michelitsch, M. D. & Weissman, J. S. 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**, (2000).
73. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* **137**, 146–158 (2009).
74. Toombs, J. A., McCarty, B. R. & Ross, E. D. Compositional Determinants of Prion Formation in Yeast. *Mol. Cell. Biol.* **30**, 319 – 332 (2010).
75. Toombs, J. A. *et al.* De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci.* **109**, 6519–6524 (2012).
76. Bonar, L., Cohen, A. S. & Skinner, M. M. Characterization of the Amyloid Fibril as a Cross-β Protein. *Proc. Soc. Exp. Biol. Med.* **131**, 1373–1375 (1969).
77. Sunde, M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction11Edited by F. E. Cohen. *J. Mol. Biol.* **273**, 729–739 (1997).
78. Perutz, M. F., Finch, J. T., Berriman, J. & Lesk, A. Amyloid fibers are water-filled nanotubes. *Proc. Natl. Acad. Sci.* **99**, 5591 – 5595 (2002).
79. Falcon, B. *et al.* Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol.* **136**, 699–708 (2018).
80. Serio, T. R. *et al.* Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science.* **289**, (2000).
81. Diaz-Avalos, R. *et al.* Cross-beta Order and Diversity in Nanocrystals of an Amyloid-forming Peptide. *J. Mol. Biol.* **330**, 1165–1175 (2003).
82. Toyama, B. H., Kelly, M. J. S., Gross, J. D. & Weissman, J. S. The structural basis of yeast prion strain variants. *Nature* **449**, 233–237 (2007).
83. Benkemoun, L. *et al.* Two structurally similar fungal prions efficiently cross-seed in vivo but form distinct polymers when coexpressed. *Mol. Microbiol.* **82**, 1392–1405 (2011).
84. Vázquez-Fernández, E. *et al.* The Structural Architecture of an Infectious Mammalian Prion Using Electron Cryomicroscopy. *PLOS Pathog.* **12**, e1005835 (2016).
85. Wille, H. *et al.* Natural and synthetic prion structure from X-ray fiber diffraction. *Proc. Natl.*

*Acad. Sci.* **106**, 16990 – 16995 (2009).

86. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science.* **357**, eaaf4382 (2017).
87. Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016).
88. Gilks, N. *et al.* Stress Granule Assembly Is Mediated by Prion-like Aggregation of TIA-1. *Mol. Biol. Cell* **15**, 5383–5398 (2004).
89. Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).
90. Kroschwald, S. *et al.* Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *Elife* **4**, e06807 (2015).
91. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
92. Platani, M., Goldberg, I., Swedlow, J. R. & Lamond, A. I. In Vivo Analysis of Cajal Body Movement, Separation, and Joining in Live Human Cells. *J. Cell Biol.* **151**, 1561–1574 (2000).
93. Dundr, M. *et al.* In vivo kinetics of Cajal body components . *J. Cell Biol.* **164**, 831–842 (2004).
94. Weidtkamp-Peters, S. *et al.* Dynamics of component exchange at PML nuclear bodies. *J. Cell Sci.* **121**, 2731 – 2743 (2008).
95. Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88**, 678–690 (2015).
96. Ribbeck, K. & Görlich, D. Kinetic analysis of translocation through nuclear pore complexes. *EMBO J.* **20**, 1320–1330 (2001).
97. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci.* **100**, 2450 – 2455 (2003).
98. Frey, S., Richter, R. P. & Görlich, D. FG-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties. *Science.* **314**, 815 – 817 (2006).
99. Ader, C. *et al.* Amyloid-like interactions within nucleoporin FG hydrogels. *Proc. Natl. Acad. Sci.* **107**, 6281 – 6285 (2010).
100. March, Z. M., King, O. D. & Shorter, J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* **1647**, 9–18 (2016).
101. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).
102. Mitrea, D. M. *et al.* Structural polymorphism in the N-terminal oligomerization domain of NPM1. *Proc. Natl. Acad. Sci.* **111**, 4466 – 4471 (2014).
103. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause

multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).

104. Kato, M. *et al.* Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767 (2012).
105. Hennig, S. *et al.* Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **210**, 529–539 (2015).
106. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science.* **360**, 918 – 921 (2018).
107. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
108. Mitrea, D. M. *et al.* Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *Elife* **5**, e13571 (2016).
109. Sun, Z. *et al.* Molecular Determinants and Genetic Modifiers of Aggregation and Toxicity for the ALS Disease Protein FUS/TLS. *PLOS Biol.* **9**, e1000614 (2011).
110. Fox, A. H., Nakagawa, S., Hirose, T. & Bond, C. S. Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends Biochem. Sci.* **43**, 124–135 (2018).
111. Van Treeck, B. *et al.* RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc. Natl. Acad. Sci.* **115**, 2734 – 2739 (2018).
112. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
113. Ferrandon, D., Koch, I., Westhof, E. & Nüsslein–Volhard, C. RNA–RNA interaction is required for the formation of specific bicoid mRNA 3′ UTR–STAUFEN ribonucleoprotein particles. *EMBO J.* **16**, 1751–1758 (1997).
114. Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proc. Natl. Acad. Sci.* **108**, 4334 – 4339 (2011).
115. Jain, S. *et al.* ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487–498 (2016).
116. Brangwynne, C. P. *et al.* Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science.* **324**, 1729 – 1732 (2009).
117. Gallo, C. M., Wang, J. T., Motegi, F. & Seydoux, G. Cytoplasmic Partitioning of P Granule Components Is Not Required to Specify the Germline in *C. elegans*. *Science.* **330**, 1685 – 1689 (2010).
118. Tenlen, J. R., Molk, J. N., London, N., Page, B. D. & Priess, J. R. MEX-5 asymmetry in one-cell *C. elegans* embryos requires PAR-4- and PAR-1-dependent phosphorylation. *Development* **135**, 3665 – 3675 (2008).
119. JM, P., Farley BM FAU - McCoig, L. M., McCoig LM FAU - Ryder, S. P. & SP, R. Molecular basis of RNA recognition by the embryonic polarity determinant MEX-5. *Journal of Biological Chemistry* **282,** 8883-8894 (2007)

120. Daniels, B. R., Dobrowsky, T. M., Perkins, E. M., Sun, S. X. & Wirtz, D. MEX-5 enrichment in the *C. elegans* early embryo mediated by differential diffusion. *Development* **137**, 2579 – 2585 (2010).
121. Griffin, E. E., Odde, D. J. & Seydoux, G. Regulation of the MEX-5 Gradient by a Spatially Segregated Kinase/Phosphatase Cycle. *Cell* **146**, 955–968 (2011).
122. Smith, J. *et al.* Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3. *Elife* **5**, e21337 (2016).
123. Platani, M., Goldberg, I., Lamond, A. I. & Swedlow, J. R. Cajal Body dynamics and association with chromatin are ATP-dependent. *Nat. Cell Biol.* **4**, 502–508 (2002).
124. Patel, A. *et al.* ATP as a biological hydrotrope. *Science.* **356**, 753 – 756 (2017).
125. Glover, J. R. & Lindquist, S. Hsp104, Hsp70, and Hsp40: A Novel Chaperone System that Rescues Previously Aggregated Proteins. *Cell* **94**, 73–82 (1998).
126. Sweeny, E. A. *et al.* The Hsp104 N-Terminal Domain Enables Disaggregase Plasticity and Potentiation. *Mol. Cell* **57**, 836–849 (2015).
127. Hung, G.-C. & Masison, D. C. N-Terminal Domain of Yeast Hsp104 Chaperone Is Dispensable for Thermotolerance and Prion Propagation but Necessary for Curing Prions by Hsp104 Overexpression. *Genetics* **173**, 611 – 620 (2006).
128. Müller, S., Matunis, M. J. & Dejean, A. Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *EMBO J.* **17**, 61–70 (1998).
129. Zhong, S. *et al.* Role of SUMO-1–modified PML in nuclear body formation. *Blood* **95**, 2748–2752 (2000).
130. Kamitani, T. *et al.* Identification of three major sentrinization sites in PML. PG - 26675-82.
131. Kamitani, T., Nguyen HP FAU - Kito, K., Kito K FAU - Fukuda-Kamitani, T., Fukuda-Kamitani T FAU - Yeh, E. T. & ET, Y. Covalent modification of PML by the sentrin family of ubiquitin-like proteins. PG - 3117-20.
132. Everett, R. D. *et al.* The Disruption of ND10 during Herpes Simplex Virus Infection Correlates with the Vmw110- and Proteasome-Dependent Loss of Several PML Isoforms. *J. Virol.* **72**, 6581 – 6591 (1998).
133. Shen, T. H., Lin, H.-K., Scaglioni, P. P., Yung, T. M. & Pandolfi, P. P. The Mechanisms of PML-Nuclear Body Formation. *Mol. Cell* **24**, 331–339 (2006).
134. Ishov, A. M. *et al.* Pml Is Critical for Nd10 Formation and Recruits the Pml-Interacting Protein Daxx to This Nuclear Structure When Modified by Sumo-1. *J. Cell Biol.* **147**, 221–234 (1999).
135. Lin, D.-Y. *et al.* Role of SUMO-Interacting Motif in Daxx SUMO Modification, Subnuclear Localization, and Repression of Sumoylated Transcription Factors. *Mol. Cell* **24**, 341–354 (2006).
136. Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A. & Dreyfuss, G. SMN, the Product of the Spinal Muscular Atrophy Gene, Binds Preferentially to Dimethylarginine-Containing Protein Targets. *Mol. Cell* **7**, 1111–1117 (2001).

137. Boisvert, F.-M. *et al.* Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing . *J. Cell Biol.* **159**, 957–969 (2002).
138. MD, H., Szymczyk PW FAU - Shpargel, K. B., Shpargel KB FAU - Matera, A. G. & AG, M. Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. PG - 2720-9.
139. Hebert, M. D., Shpargel, K. B., Ospina, J. K., Tucker, K. E. & Matera, A. G. Coilin Methylation Regulates Nuclear Body Formation. *Dev. Cell* **3**, 329–337 (2002).
140. Pellizzoni, L., Charroux, B. & Dreyfuss, G. SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. *Proc. Natl. Acad. Sci.* **96**, 11167 – 11172 (1999).
141. Kedersha, N. L., Gupta, M., Li, W., Miller, I. & Anderson, P. RNA-Binding Proteins Tia-1 and Tiar Link the Phosphorylation of Eif-2α to the Assembly of Mammalian Stress Granules. *J. Cell Biol.* **147**, 1431–1442 (1999).
142. Murtha-Riel, P. *et al.* Expression of a phosphorylation-resistant eukaryotic initiation factor 2 alpha-subunit mitigates heat shock inhibition of protein synthesis. PG - 12946-51.
143. Kedersha, N. *et al.* Evidence That Ternary Complex (eIF2-GTP-tRNAi Met)–Deficient Preinitiation Complexes Are Core Constituents of Mammalian Stress Granules. *Mol. Biol. Cell* **13**, 195–210 (2001).
144. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
145. Flory, P. J. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.* **10**, 51–61 (1942).
146. Huggins, M. L. Some Properties of Solutions of Long-chain Compounds. *J. Phys. Chem.* **46**, 151–158 (1942).
147. Overbeek, J. T. G. & Voorn, M. J. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *J. Cell. Comp. Physiol.* **49**, 7–26 (1957).
148. Herrera, J. E., Correia, J. J., Jones, A. E. & Olson, M. O. J. Sedimentation Analyses of the Salt- and Divalent Metal Ion-Induced Oligomerization of Nucleolar Protein B23. *Biochemistry* **35**, 2668–2673 (1996).
149. Qamar, S. *et al.* FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **173**, 720-734.e15 (2018).
150. Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **57**, 936–947 (2015).
151. Vernon, R. M. *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **7**, e31486 (2018).
152. Ferrari, L. *et al.* Arginine π-stacking drives binding to fibrils of the Alzheimer protein Tau. *Nat. Commun.* **11**, 571 (2020).
153. Greig, J. A. *et al.* Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol. Cell* (2020). doi:10.1016/j.molcel.2020.01.025
154. Yasuda, S. *et al.* Stress- and ubiquitylation-dependent phase separation of the proteasome.

*Nature* **578**, 296–300 (2020).

155. Andrusiak, M. G. *et al.* Inhibition of Axon Regeneration by Liquid-like TIAR-2 Granules. *Neuron* **104**, 290-304.e8 (2019).
156. Brenner, S. The Genetics of Behaviour. *Br. Med. Bull.* **29**, 269–271 (1973).
157. Brenner, S. The Genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71 – 94 (1974).
158. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
159. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
160. Emmons, S. W. & Lipton, J. Genetic basis of male sexual behavior. *J. Neurobiol.* **54**, 93–110 (2003).
161. Ward, S., Thomson, N., White, J. G. & Brenner, S. Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* **160**, 313–337 (1975).
162. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. London. B, Biol. Sci.* **314**, 1–340 (1986).
163. Alkema, M. J., Hunter-Ensor, M., Ringstad, N. & Horvitz, H. R. Tyramine Functions Independently of Octopamine in the *Caenorhabditis elegans* Nervous System. *Neuron* **46**, 247–260 (2005).
164. Taylor, R. C. & Dillin, A. XBP-1 Is a Cell-Nonautonomous Regulator of Stress Resistance and Longevity. *Cell* **153**, 1435–1447 (2013).
165. Imanikia, S., Sheng, M., Castro, C., Griffin, J. L. & Taylor, R. C. XBP-1 Remodels Lipid Metabolism to Extend Longevity. *Cell Rep.* **28**, 581-589.e4 (2019).
166. Imanikia, S., Özbey, N. P., Krueger, C., Casanueva, M. O. & Taylor, R. C. Neuronal XBP-1 Activates Intestinal Lysosomes to Improve Proteostasis in *C. elegans*. *Curr. Biol.* (2019). doi:10.1016/J.CUB.2019.06.031
167. McGhee, J. D. *et al.* The ELT-2 GATA-factor and the global regulation of transcription in the *C. elegans* intestine. *Dev. Biol.* **302**, 627–645 (2007).
168. Schulenburg, H., Léopold Kurz, C. & Ewbank, J. J. Evolution of the innate immune system: the worm perspective. *Immunol. Rev.* **198**, 36–58 (2004).
169. Albertson, D. G., Thompson, J. N. & Brenner, S. The pharynx of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. London. B, Biol. Sci.* **275**, 299–325 (1976).
170. Kamal, M. *et al.* The marginal cells of the *Caenorhabditis elegans* pharynx scavenge cholesterol and other hydrophobic small molecules. *Nat. Commun.* **10**, 3938 (2019).
171. Avery, L. & Horvitzt, H. R. Pharyngeal pumping continues after laser killing of the pharyngeal nervous system of *C. elegans*. *Neuron* **3**, 473–485 (1989).
172. Franks, C. J., Holden-Dye, L., Bull, K., Luedtke, S. & Walker, R. J. Anatomy, physiology and

pharmacology of *Caenorhabditis elegans* pharynx: a model to define gene function in a simple neural system. *Invertebr. Neurosci.* **6**, 105–122 (2006).

173. Kramer, J. M., Johnson, J. J., Edgar, R. S., Basch, C. & Roberts, S. The *sqt-1* gene of *C. elegans* encodes a collagen critical for organismal morphogenesis. *Cell* **55**, 555–565 (1988).
174. Cox, G. N., Kusch, M. & Edgar, R. S. Cuticle of *Caenorhabditis elegans*: its isolation and partial characterization. *J. Cell Biol.* **90**, 7–17 (1981).
175. Blaxter, M. L., Page, A. P., Rudin, W. & Maizels, R. M. Nematode surface coats: Actively evading immunity. *Parasitol. Today* **8**, 243–247 (1992).
176. Höflich, J. *et al.* Loss of srf-3-encoded Nucleotide Sugar Transporter Activity in *Caenorhabditis elegans* Alters Surface Antigenicity and Prevents Bacterial Adherence. *J. Biol. Chem.* **279**, 30440–30448 (2004).
177. Gravato-Nobre, M. J. *et al.* Multiple Genes Affect Sensitivity of *Caenorhabditis elegans* to the Bacterial Pathogen *Microbacterium nematophilum*. *Genetics* **171**, 1033 – 1045 (2005).
178. von Mende, N., Bird, D. M., Albert, P. S. & Riddle, D. L. *dpy-13*: A nematode collagen gene that affects body shape. *Cell* **55**, 567–576 (1988).
179. Cox, G. N., Staprans, S. & Edgar, R. S. The cuticle of *Caenorhabditis elegans*: II. Stage-specific changes in ultrastructure and protein composition during postembryonic development. *Dev. Biol.* **86**, 456–470 (1981).
180. Witze, E. & Rothman, J. H. Cell Fusion: An EFFicient Sculptor. *Curr. Biol.* **12**, R467–R469 (2002).
181. Michaux, G., Legouis, R. & Labouesse, M. Epithelial biology: lessons from *Caenorhabditis elegans*. *Gene* **277**, 83–100 (2001).
182. Thein, M. C. *et al. Caenorhabditis elegans* exoskeleton collagen COL-19: An adult-specific marker for collagen modification and assembly, and the analysis of organismal morphology. *Dev. Dyn.* **226**, 523–539 (2003).
183. Kim, D. H. *et al.* A Conserved p38 MAP Kinase Pathway in *Caenorhabditis elegans* Innate Immunity. *Science.* **297**, 623 – 626 (2002).
184. Troemel, E. R. *et al.* p38 MAPK Regulates Expression of Immune Response Genes and Contributes to Longevity in *C. elegans*. *PLOS Genet.* **2**, e183 (2006).
185. Garsin, D. A. *et al.* Long-Lived *C. elegans daf-2* Mutants Are Resistant to Bacterial Pathogens. *Science.* **300**, 1921 – 1921 (2003).
186. Mori, K. Tripartite Management of Unfolded Proteins in the Endoplasmic Reticulum. *Cell* **101**, 451–454 (2000).
187. Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P. & Ron, D. Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat. Cell Biol.* **2**, 326–332 (2000).
188. Tirasophon, W., Welihinda, A. A. & Kaufman, R. J. A stress response pathway from the endoplasmic reticulum to the nucleus requires a novel bifunctional protein

kinase/endoribonuclease (Ire1p) in mammalian cells. *Genes Dev.* **12**, 1812–1824 (1998).

189. Harding, H. P., Zhang, Y., Bertolotti, A., Zeng, H. & Ron, D. Perk is Essential for Translational Regulation and Cell Survival during the Unfolded Protein Response. *Mol. Cell* **5**, 897–904 (2000).
190. Haze, K., Yoshida, H., Yanagi, H., Yura, T. & Mori, K. Mammalian Transcription Factor ATF6 Is Synthesized as a Transmembrane Protein and Activated by Proteolysis in Response to Endoplasmic Reticulum Stress. *Mol. Biol. Cell* **10**, 3787–3799 (1999).
191. Shen, X. *et al.* Complementary Signaling Pathways Regulate the Unfolded Protein Response and Are Required for *C. elegans* Development. *Cell* **107**, 893–903 (2001).
192. Shen, X., Ellis, R. E., Sakaki, K. & Kaufman, R. J. Genetic Interactions Due to Constitutive and Inducible Gene Regulation Mediated by the Unfolded Protein Response in *C. elegans*. *PLOS Genet.* **1**, e37 (2005).
193. Richardson, C. E., Kinkel, S. & Kim, D. H. Physiological IRE-1-XBP-1 and PEK-1 Signaling in *Caenorhabditis elegans* Larval Development and Immunity. *PLOS Genet.* **7**, e1002391 (2011).
194. Urano, F. *et al.* A survival pathway for *Caenorhabditis elegans* with a blocked unfolded protein response. *J. Cell Biol.* **158**, 639 – 646 (2002).
195. Richardson, C. E., Kooistra, T. & Kim, D. H. An essential role for XBP-1 in host protection against immune activation in *C. elegans*. *Nature* **463**, 1092–1095 (2010).
196. Haskins, K. A., Russell, J. F., Gaddis, N., Dressman, H. K. & Aballay, A. Unfolded Protein Response Genes Regulated by CED-1 Are Required for *Caenorhabditis elegans* Innate Immunity. *Dev. Cell* **15**, 87–97 (2008).
197. Lamitina, T. & Cherry, S. Dangerous Liaisons: The Apoptotic Engulfment Receptor CED-1 Links Innate Immunity to the Unfolded Protein Response. *Dev. Cell* **15**, 3–4 (2008).
198. Cheesman, H. K. *et al.* Aberrant Activation of p38 MAP Kinase-Dependent Innate Immune Responses Is Toxic to *Caenorhabditis elegans*. *G3 Genes|Genomes|Genetics* **6**, 541 – 549 (2016).
199. Styer, K. L. *et al.* Innate Immunity in *Caenorhabditis elegans* Is Regulated by Neurons Expressing NPR-1/GPCR. *Science.* **322**, 460 – 464 (2008).
200. Sun, J., Singh, V., Kajino-Sakamoto, R. & Aballay, A. Neuronal GPCR Controls Innate Immunity by Regulating Noncanonical Unfolded Protein Response Genes. *Science.* **332**, 729 – 732 (2011).
201. Sun, J., Liu, Y. & Aballay, A. Organismal regulation of XBP-1-mediated unfolded protein response during development and immune activation. *EMBO Rep.* **13**, 855–860 (2012).

# Chapter 2

## Identifying a novel prion-like protein in *Caenorhabditis elegans*

## 2.0 Abstract

Whilst the role that many prion-like proteins play in pathology is well characterised, their involvement in cellular physiology is understood to a much lesser extent. Due to the stereotyped architecture of prion-like domains (PrLDs), which often contain low complexity and intrinsically disordered sequences, it is possible to identify candidate prion-like proteins computationally. Using the Prion-Like Amino Acid Composition program, we have identified the stress-responsive protein ABU-13 as a potential prion in the nematode species, *Caenorhabditis elegans*. This protein scores more highly than many human and yeast prions across multiple comparable algorithms, lending credence to the possibility that this proposed PrLD does indeed confer prionogenic properties upon ABU-13. Analysis of predicted prion-like sequences across species highlights an intraspecies clustering of prion-like protein prediction scores, suggesting that there may be differences in the frequency and sequence content of prion-like proteins between species.

In general, the ABU family of proteins exhibit higher prediction scores than known non-prionogenic proteins, implying that a degree of low complexity and disorder may be an intrinsic feature of many of these proteins. Despite this, ABU-13 scores much higher than the rest of its family and thus remains the strongest candidate for further investigation. Phenotypic analysis of ABU-13 knockout animals has identified a non-redundant role for this protein in ER stress and innate immune responses. Whilst this suggests an interplay with the canonical unfolded protein response pathways, these knockout animals do not display defects in the activation of the ER stress response gene *hsp-4* upon tunicamycin treatment, suggesting that these effects occur independently of this pathway.

## 2.1 Introduction

### *2.1.1 Prion-like domains*

The properties of intrinsic disorder and low complexity that define prion-like domains allow for the conformational flexibility required for their activity. This ability to structurally switch between states is key for conformational templating, and thus for the propagation of prion-like states. Without this flexibility, prion-like states could not exist in their heritable form. Early studies demonstrated that such domains are discrete entities, capable of conferring these prion-like properties if the domain is transferred onto a non-prion[1].

The intrinsic ability of these domains to confer such properties is largely driven by amino acid composition - these abilities are not lost, for example, when the prion domain is scrambled, retaining the relative proportions of each amino acid constituent but in a different order, clearly demonstrating that residue position within the domain is less important than the basic composition[2]. Commonly, an overrepresentation of polar residues such as glutamine and asparagine (Q/N) is observed within such domains[3,4]. Some amino acids do, however, exhibit a very strong positional effect resulting from specific-residue properties; hydrophobic residues, for example, show a high degree of position dependency. These residues are underrepresented within these domains due to their strong propensity to trigger aggregation – however, their limited presence is beneficial in promoting prionogenicity[5,6].

### *2.1.2 Domain Predictions*

The predictable composition of these regions has allowed algorithms for the bioinformatic identification of novel prions to be developed, with mixed success. Initial algorithms, such as DIANA, focused on the Q/N-rich nature of these domains but neglected the influence of other residues, making it difficult to distinguish between prion and non-prion[7]. Such distinctions have been improved on since the development of DIANA. Alberti et al., for example, undertook a bioinformatic proteome-wide screen for prionogenic proteins in yeast, identifying and characterising around 100 novel candidates. By replacing the Sup35 prion domain with that of the candidate prions, they were able to demonstrate that 19 of these candidates indeed functioned as prions, capable of transmitting their conformation onto naïve counterparts in a [PSI+] state[8].

Whilst this improved on previous predictions, the false positive rate observed in this screen remained high, suggesting that there are other factors influencing prion activity – possibly as a result of the systematic exclusion of underrepresented residues, such as those with hydrophobic properties, from these bioinformatic algorithms. Using mutagenesis data, Toombs et al. were able to reduce the false positive rate in their PAPA algorithm by quantifying the influence of the residues within these domains[9].

The success of this algorithm was demonstrated *in silico,* with the design of synthetic prion forming domains (sPFDs), a number of which were shown to reconstitute prion activity to Sup35[10].

In this chapter, we utilise the Prion-Like Amino Acid Composition (PLAAC) algorithm to predict the presence of prion-like domains within the *C. elegans* proteome[11]. This algorithm relies primarily on the hidden Markov Model developed by Alberti et al., based on the composition of known yeast prions including Sup35p, Ure2p, Rnq1p and New1p[8]; it is important to note, however, that prions with dissimilar amino acid composition were not included in this model, as the mechanisms used by these domains to confer prionogenic activity are thought to differ. This HMM parses sequences into one of two hidden states - either a prion-like or a non-prion state. Within the proteome, each amino acid is ascribed a transition probability, indicating the likelihood that the presence of that amino acid within the context of a pre-defined region would drive the protein in one of these two states. As standard, PLAAC defines prion-like domains as containing a minimum of 50 amino acids, using a region of this length as its pre-defined contextual query window for prion-like state in comparison to a background of 1000 amino acids. These 50 amino acid query windows then undergo either a Viterbi or MAP parse – both of which are computational approaches to uncovering hidden states – to identify the cumulative probability of these regions representing prion-like or non-prion states.

The PLAAC algorithm graphically depicts these hidden Markov Model results, plotting which protein regions are predicted to be prion-like and which are not. Beyond this, it also depicts the output from other algorithms, allowing for these readouts to be directly compared – strengthening the overall predictive power of these algorithms. A scaled version of the PAPA scores derived from the Toombs et al. algorithm[12], for example, are plotted as well, following an adjustment of each score by a factor of -4 – this allows lower values to become predictive of prionogenicity within the same numerical scale as the other readouts.

In this chapter, we will apply and validate the use of the PLAAC algorithm on the *C. elegans* proteome, showing that, whilst there do appear to be species-specific differences in the prediction scores, the algorithm still strongly predicts the presence of prion-like domains in certain proteins. Specifically, ABU-13, a protein implicated in the unfolded protein response of the endoplasmic reticulum, scores particularly highly, leading to its further investigation and phenotypic characterisation.

## 2.2 Results

### *2.2.1 Species-specific validation of the PLAAC algorithm*

The PLAAC algorithm combines a number of different predictions of prion-like characteristics to identify potential PrLDs within the proteome of an organism. To normalise against the differential frequencies of amino acid usage between species, the algorithm uses a background frequency parameter for all proteomic data derived from non-yeast species. As the *C. elegans* proteome was used, the background frequency was set to 50 (out of 100), as recommended, compared to the 100 used for the yeast proteome, representing a total reliance on the background amino acid frequency of *S. cerevisiae*.

| Yeast Prions | | | | | Yeast Non-Prions | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Protein** | **CORE Score** | **PAPA Prop** | **Fold Index** | **HMM Viterbi** | **Protein** | **CORE Score** | **PAPA Prop** | **Fold Index** | **HMM Viterbi** |
| SUP35 | 51.215 | 0.1 | -0.423 | 30.358 | PBS2 | N/A | -0.051 | -0.001 | -0.716 |
| RNQ1 | 47.459 | 0.141 | -0.324 | 79.598 | HAC1 | N/A | -0.167 | -0.076 | -0.286 |
| MOT3 | 40.751 | 0.102 | -0.116 | 152.584 | GCN2 | N/A | 0.015 | -0.04 | -1.708 |
| URE2 | 29.225 | 0.103 | -0.13 | 107.777 | BCK1 | N/A | -0.069 | -0.119 | -1.527 |
| MOD5 | N/A | -0.045 | -0.082 | 28.91 | CDC28 | N/A | N/A | N/A | -0.346 |
| CYC8 | 57.104 | -0.004 | -0.49 | -0.476 | HCM1 | N/A | -0.134 | -0.003 | -0.687 |
| STD1 | N/A | -0.025 | -0.13 | 116.93 | SSA2 | N/A | -0.134 | -0.003 | -0.687 |
| PMA1 | N/A | -0.141 | -0.007 | 0.691 | IRE1 | N/A | -0.016 | -0.007 | -1.163 |
| SWI1 | 46.292 | 0.055 | -0.02 | -0.749 | HSF1 | N/A | -0.012 | -0.09 | 9.299 |
| NEW1 | 42.458 | 0.137 | -0.546 | 133.073 | | | | | |

| *C. elegans* Prion-like Proteins | | | | | *C. elegans* Non-Prion-like Proteins | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Protein** | **CORE Score** | **PAPA Prop** | **Fold Index** | **HMM Viterbi** | **Protein** | **CORE Score** | **PAPA Prop** | **Fold Index** | **HMM Viterbi** |
| LSM-4 | N/A | -0.099 | 0 | -0.171 | SEK-1 | N/A | -0.066 | -0.002 | -0.384 |
| TIAR-2 | 33.996 | 0.081 | -0.269 | 26.867 | CED-1 | N/A | -0.108 | -0.003 | -1.159 |
| HRP-1 | 28.649 | 0.014 | -0.219 | 28.649 | XBP-1 | N/A | 0.074 | -0.038 | -0.383 |
| MEG-3 | N/A | 0.063 | -0.059 | 1.307 | PEK-1 | N/A | -0.086 | -0.003 | -1.125 |
| LAF-1 | 24.835 | -0.011 | -0.244 | 28.705 | ATF-6 | N/A | -0.059 | -0.151 | -0.637 |
| PAB-1 | 28.58 | 0.058 | -0.22 | 42.691 | NSY-1 | N/A | -0.068 | -0.004 | -1.547 |
| FUST-1 | 23.329 | -0.099 | -0.219 | 19.639 | DAF-2 | N/A | -0.052 | -0.004 | -1.895 |
| CAR-1 | 22.331 | 0.02 | -0.49 | 24.189 | DAF-16 | N/A | -0.01 | -0.002 | -0.589 |
| | | | | | HSP-4 | N/A | -0.164 | -0.004 | -0.705 |
| | | | | | IRE-1 | N/A | -0.019 | -0.002 | -1.015 |
| | | | | | OCTR-1 | N/A | -0.012 | -0.017 | -0.456 |
| | | | | | HSF-1 | 9.87 | -0.008 | -0.003 | 8.89 |

| Human Prion-like Proteins | | | | | Human Non-Prion-like Proteins | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Protein** | **CORE Score** | **PAPA Prop** | **Fold Index** | **HMM Viterbi** | **Protein** | **CORE Score** | **PAPA Prop** | **Fold Index** | **HMM Viterbi** |
| FUS | 37.794 | 0.101 | -0.21 | 91.609 | MAP2K6 | N/A | -0.21 | -0.008 | -0.382 |
| TIA1 | 23.224 | 0.131 | -0.089 | 20.805 | MEGF10 | N/A | 0.041 | 0 | -1.188 |
| RBM14 | 16.066 | -0.134 | -0.235 | 28.253 | XBP1 | N/A | -0.048 | -0.001 | -0.309 |
| TDP-43 | 26.454 | 0.043 | -0.001 | 30.133 | EIF2AK3 | N/A | -0.054 | -0.018 | -1.164 |
| HNRNPA1 | 28.36 | 0.093 | -0.091 | 19.715 | ATF6 | N/A | 0.041 | -0.009 | -0.718 |
| CPEB3 | N/A | -0.071 | -0.007 | 2.046 | MAP3K15 | N/A | -0.002 | -0.002 | -1.361 |
| PrP | N/A | 0.02 | -0.139 | -0.301 | IGF1R | N/A | -0.024 | -0.01 | -1.415 |
| HNRNPA2 | 30.064 | 0.043 | -0.208 | 38.562 | FOXO1 | 13.046 | -0.013 | -0.068 | 2.244 |
| TIAR | 13.565 | 0.11 | -0.096 | 10.265 | FOXO3 | N/A | -0.058 | -0.001 | -0.721 |
| DDX4 | N/A | -0.089 | -0.161 | -0.986 | FOXO4 | N/A | -0.109 | -0.003 | -0.553 |
| | | | | | HSPA5 (BiP) | N/A | -0.165 | -0.003 | -0.702 |
| | | | | | IRE1 | N/A | -0.06 | -0.011 | -1.025 |
| | | | | | ADRA2A | N/A | -0.106 | -0.001 | -0.513 |
| | | | | | HSF1 | N/A | -0.066 | -0.002 | -0.577 |

**Table 2.1** Summary of PLAAC results from known *S. cerevisiae*, human and *C. elegans* prions, with readouts from the Alberti algorithm (CORE score), Toombs algorithm (PAPA Prop), disorder prediction (FoldIndex) and Viterbi-parsed Hidden Markov Model scores

To assess whether this resulted in an over- or underestimate of prion-like domains within these animals, we compared the PLAAC results from a number of known prion-like proteins from the yeast, human and *C. elegans* proteome, working on the assumption that such proteins would share similar results across these computational readouts (Table 2.1). Each of these proteins had previously been shown to have an intrinsically disordered region and/or a prion-like domain, or displayed prion-like capabilities such as the ability to act as a protein scaffold, nucleating phase separation.

Comparison of scores across these three species demonstrates some of the differences arising from these algorithms. The prion propensity (PAPA prop) values, for example, show no significant differences between prion groups, regardless of species – however, only the yeast prions show any significant difference when compared to non-prion-like proteins of any other species (yeast, $p$=0.0167; human, $p$=0.0115; nematode, $p$=0.0202), whilst neither the human nor nematode prions showed any significant difference from non-prion proteins (Figure 2.1A). The CORE scores of yeast prions, on the other hand, are significantly greater than both human ($p$=0.0005) and nematode ($p$=0.0019) prions (Figure 2.1B). As would be expected, due to the lack of a predicted prion-like domain in many non-prion-like proteins, no CORE score could be determined and thus the difference between these results could not be assessed. Similarly, the Viterbi-parsed Hidden Markov Model shows a difference between yeast prions and the human and nematode prions, as well as between the yeast prions and all three species of non-prion-like protein (Figure 2.1C).

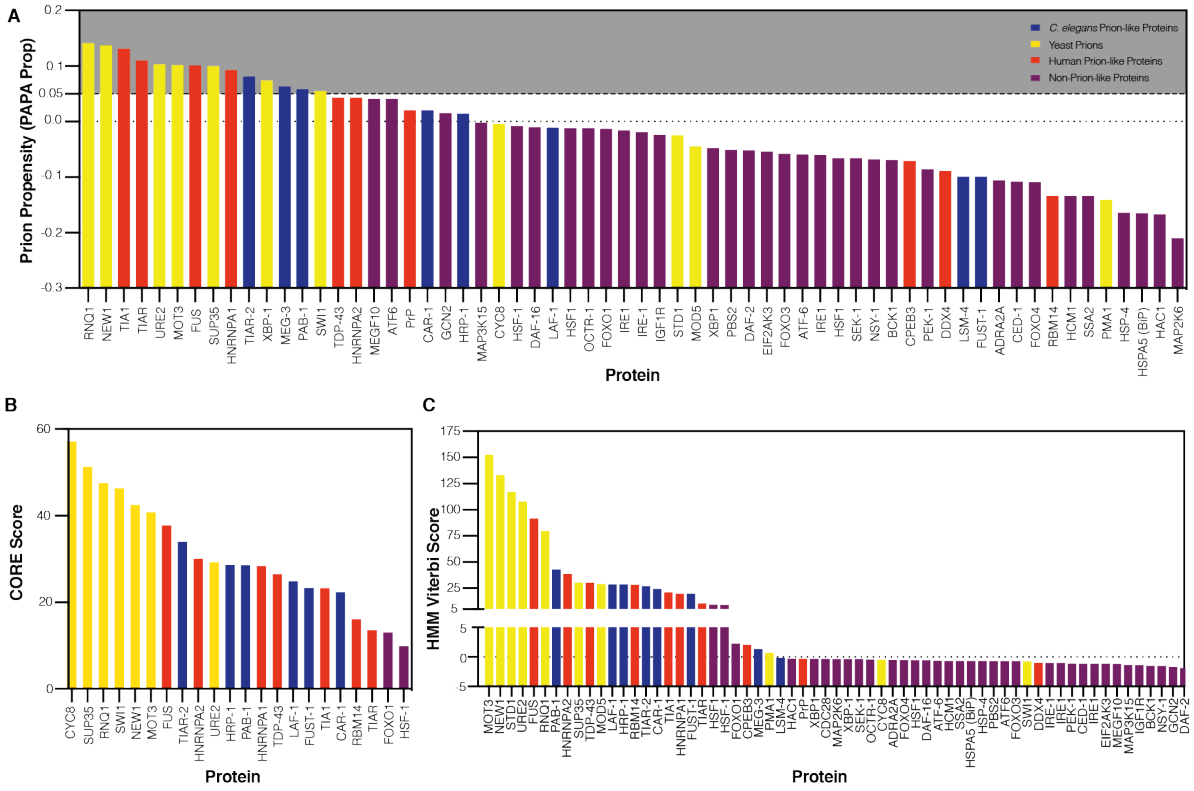



**Figure 2.1** Comparisons of PLAAC output scores between known yeast, human and nematode prions, as well as non-prion like proteins from all three species (A) Prion propensity (PAPA prop) values [Two way ANOVA, Yeast Prions v Yeast Non-Prions, $p$=;0.0167; Yeast Prions v Human Non-Prion, $p$=0.0115*;* Yeast Prion v *C. elegans* Non-Prion, $p$=0.0202, *C. elegans* Prion v *C. elegans* Non-Prion, $p$=0.4657, ns; Human Prion v Human Non-Prion, $p$=0.0841; grey = significant PAPA value] (B) CORE scores [Two way ANOVA, Yeast Prions v Human Prions, $p$=0.0005; Yeast Prions v *C. elegans* Prions, $p$=0.0019; Yeast Prions v Yeast Non-Prion, $p$=0.000] (C) Viterbi-parsed HMM scores One way ANOVA, Yeast Prions v *C. elegans* prions, $p$=00140; Yeast Prions v Yeast Non-Prions, $p$=<0.0001; Yeast Prions v *C. elegans* Prions, $p$=<0.0001; Yeast Prions v Human Non-Prion, $p$<0.0001]

### *2.2.2 ABU-13 is a strong candidate prion-like protein*

Application of the PLAAC algorithm to the full proteome of N2 animals – representing the laboratory standard wild type *C. elegans* strain - highlighted a number of potential prion-like candidates scoring highly across multiple readouts. The top hits amongst these candidates include UNC-10, the ortholog of the human RIMS1 protein, involved in synaptic transmission within the cholinergic system, and CEST-12, a protein thought to function in the hydrolysis of carboxylic esters. In keeping with their overrepresentation amongst prion-like proteins, many of the top hits were PQN proteins previously identified by the DIANA algorithm, such as PQN-31, PQN-71 and PQN-84. It is interesting to observe that many of the PQN proteins identified in this computational screen do not have defined functions beyond domain-based predictions – this is likely a reflection of the technical difficulty of working with such proteins.

Of these hits, the top-scoring candidate overall was ABU-13 (also known as PQN-46), a member of a protein family that has been previously been shown to be differentially regulated in response a number of environmental stress conditions, including activation of the innate immune system and the ER stress response. This protein has a CORE score of 73.571, far exceeding the next ranked candidate, PQN-32, with a rank of 59.467 (Table 1.2, Figure 2.2B), suggesting that ABU-13 has a significant glutamine-

asparagine (Q/N)-rich domain that may confer prionogenic properties onto the protein. In support of this, ABU-13 also ranks highly and significantly on other readouts of prion propensity - in the Toombs algorithm (PAPA propensity), for example, it scores 0.206. This algorithm works in concert with the Alberti algorithm to fine-tune the identification of potentially prion-like domain proteins by assessing the relative contributions of residue interactions within these proposed domains, reducing the false positive rate. Within this system, any value above 0.05 is classed as significant, thus predicting that the proposed prion-like domain of ABU-13 does indeed confer prionogenicity to the protein.

| Rank | Protein | CORE Score | PAPA Propensity | FoldIndex | HMM Viterbi |
|---|---|---|---|---|---|
| 1 | ABU-13 | 73.571 | 0.206 | -0.581 | 103.79 |
| 2 | PQN-32 | 59.467 | 0.057 | -0.326 | 143.97 |
| 3 | PQN-37 | 59.069 | 0.053 | -0.28 | 99.584 |
| 4 | PQN-31 | 58.069 | 0.137 | 0 | 175.317 |
| 5 | CEST-12 | 56.978 | 0.167 | -0.005 | 167.064 |
| 6 | PQN-71 | 56.781 | 0.143 | -0.484 | 60.489 |
| 7 | PQN-84 | 55.072 | 0.074 | -0.227 | 82.28 |
| 8 | PQN-88 | 54.645 | 0.14 | -0.297 | 83.581 |
| 9 | UNC-10 | 53.467 | 0.031 | -0.524 | 149.479 |
| 10 | PQN-8 | 52.97 | 0.143 | -0.239 | 120.172 |
| 11 | PQN-42 | 51.475 | 0.018 | -0.48 | 47.009 |
| 12 | PQN-83 | 51.228 | 0.077 | -0.464 | 73.12 |
| 13 | PQN-27 | 51.034 | 0.101 | -0.477 | 80.718 |
| 14 | PQN-15 | 50.923 | 0.072 | -0.33 | 304.481 |
| 15 | ZC21.3 | 49.849 | 0.132 | -0.293 | 149.056 |
| 16 | HGRS-1 | 49.653 | 0.104 | -0.417 | 91.1 |
| 17 | TLK-1 | 49.234 | 0.107 | -0.471 | 70.538 |
| 18 | IMPH-1 | 48.989 | 0.027 | -0.149 | 112.117 |
| 19 | F13H8.5 | 48.682 | 0.124 | -0.014 | 64.091 |
| 20 | F55A12.6 | 47.906 | 0.272 | -0.276 | 47.808 |

**Table 2.2** Summary of top PLAAC results from the *C. elegans* N2 proteome, with readouts from the Alberti algorithm (CORE score), Toombs algorithm (PAPA Prop), disorder prediction (FoldIndex) and Viterbi-parsed Hidden Markov Model

This is further supported by the prediction of an amyloidogenic core in the middle of this proposed prion-like domain, spanning residues 199 to 219 (Figure 2.2A). This motif could represent the core set of amino acids necessary for fibrillisation of ABU-13, if such an aggregation occurs.

### *2.2.3 Structural predictions of ABU-13*

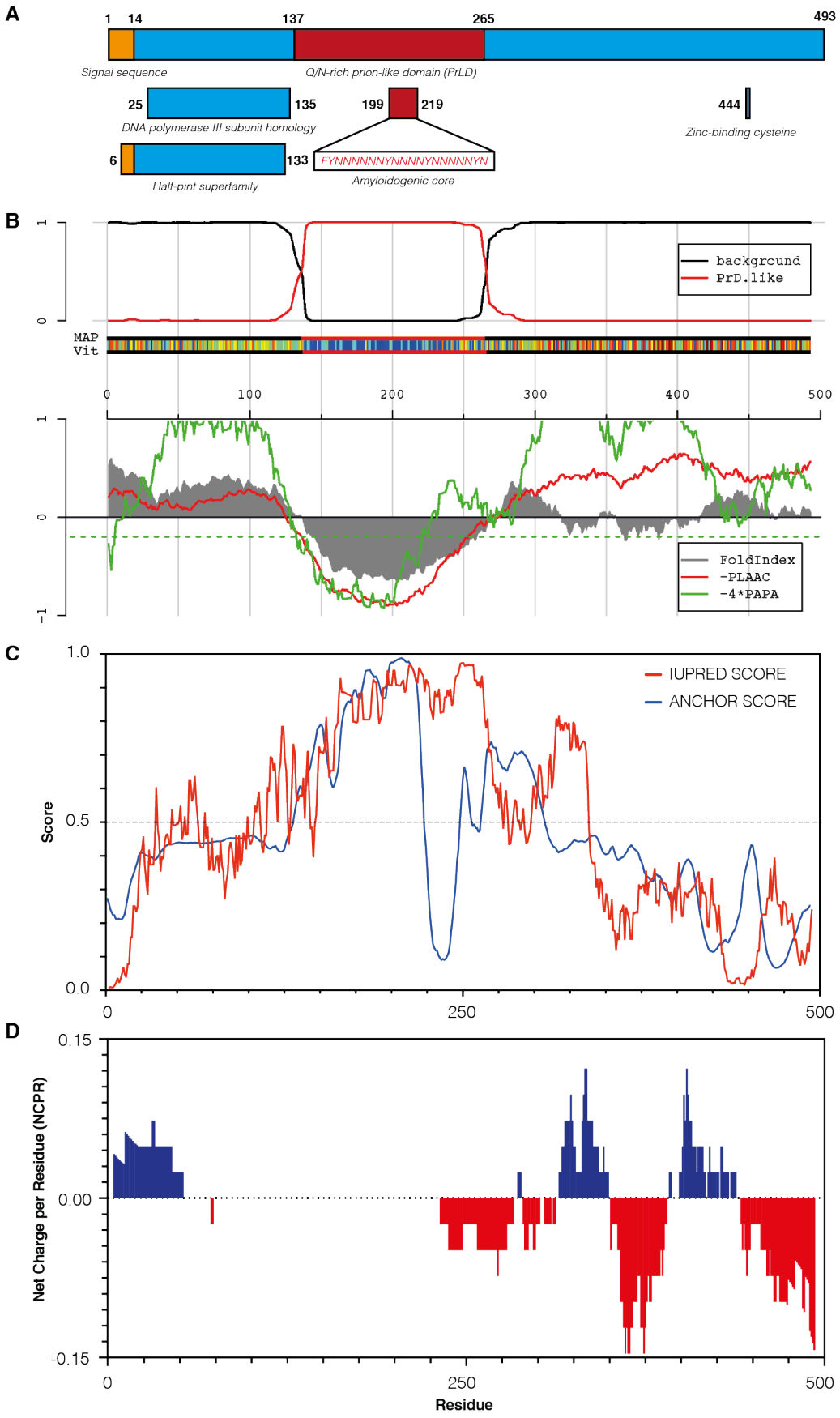



**Figure 2.2** The structural features of ABU-13 (A) a schematic highlighting a number of conserved domains features including a signal peptide, a Q/N-rich prion-like domain and a zinc-binding residue; (B) graphical representation of the PLAAC results, showing the two-state results of the hidden Markov Model (background vs. prion), with a residue by residue pictorial representation of the HMM results, as well as a map of the FoldIndex and PLAAC results across the whole protein; (C) IUPRED2A and ANCHOR2 results of disorder within ABU-13; (D) residue-by-residue charge distribution

As is true for most of the other PQN proteins, there is little known about the specific function of ABU-13 itself. Analysis of conserved domains within this protein highlights an N-terminal region immediately adjacent to the proposed prion-like domain that shares a degree of similarity to a number of proteins involved with RNA and DNA binding (Figure 2.2A). This includes the gamma and tau subunits of DNA polymerase III (accession references PRK12323, $p$=4.79e$^{-4}$; PRK14971, $p$=8.36e$^{-4}$; PRK07764, $p$=1.44e$^{-3}$) and the half-pint superfamily (accession reference TIGR01645, $p$=3.39e$^{-3}$), which contains a number of RNA recognition motifs involved in the binding and splicing of poly-U sequences. This suggests that the N-terminal region of ABU-13 may be involved in RNA or DNA interactions. The presence of a zinc-binding cysteine in the C-terminus of the protein also points towards further interactions as well, potentially with a structurally stabilising effect suggesting that this terminus is more ordered and structured than its N-terminal counterpart.

In addition to this, there is also a predicted signal peptide at the N-terminus of the protein (likeness probability = 0.9858), targeting ABU-13 to the Sec translocon before cleavage by Signal Peptidase I. This cleavage site is predicted to be between residues 14 and 15 ($p$=0.9477). Whilst this suggests that ABU-13 is targeted to the secretory pathway, it is unclear where the processed mature protein is ultimately localised – whether retained in the endoplasmic reticulum or Golgi apparatus, secreted into the extracellular environment, or inserted into a target membrane. The latter is unlikely, as, unlike the other members of the ABU protein family, ABU-13 does not contain a transmembrane domain.

As would be hypothesised for proteins of this nature, applying the IUPRED and ANCHOR algorithms for prediction of unstructured domains identifies a large intrinsically disordered region within the centre of ABU-13, in alignment with the FoldIndex per residue scores and coinciding with the proposed prion-like domain (Figure 2.2B, 2.2C). As previously discussed, these IDRs allow for conformational flexibility, facilitating the switch between different states required for prion formation and propagation. The ANCHOR algorithm suggests that within this unstructured region, there are a number of binding domains, including between residues 133 and 220, and 264 and 305. The nature of these binding regions, however, is unclear at this stage.

The C-terminal region of ABU-13 contains blocks of alternating charge, extending all the way from the centre of the protein. These charge differences could be representative of charge-charge electrostatic interactions.

### *2.2.4 Comparing ABU-13 with the wider ABU family of proteins*

The ABU family of proteins were first identified in *C. elegans* as a group of genes that were upregulated in response to treatment with tunicamycin, an inhibitor of N-linked glycosylation that causes ER stress, in an *xbp-1* mutant background. The genes identified were named the *abu* genes due to their **a**ctivation

in **b**locked **U**PR. In order to determine whether or not this predicted prion-like domain is unique to ABU-13 or whether it is an intrinsic feature of this protein family as a whole, we compared the ABU-13 PLAAC results to those from the wider family and with the set of known *C. elegans* prion-like proteins. Interesting, whilst the known prion-like protein scores were wide ranging, many members of the ABU family score highly across a number of readouts (Table 2.3, Figures 2.3A-C). ABU-13, -12, -2, -8, and -15, for example, all score above the 0.05 threshold for prion propensity, suggesting that they contain some inherent prionogenic ability (Figure 2.2B). Whilst there is no significant difference between the scores of known prion-like proteins and either the ABU family or the non-prion like proteins, the ABU family score significantly higher than the non-prion-like proteins ($p$=0.0005). This suggests that, in general, this family possesses prionogenic capacities with properties akin to those seen in yeast prions.

| Protein | CORE Score | PAPA Prop | FoldIndex | HMM Viterbi |
|---|---|---|---|---|
| ABU-13 | 73.571 | 0.206 | -0.581 | 103.79 |
| ABU-12 | 45.179 | 0.108 | -0.318 | 87.355 |
| ABU-8 | 30.014 | 0.07 | -0.231 | 39.051 |
| ABU-11 | 27.361 | 0.021 | -0.002 | 35.752 |
| ABU-7 | 25.847 | 0.042 | -0.185 | 32.369 |
| ABU-6 | 25.645 | 0.059 | -0.187 | 26.917 |
| ABU-15 | 23.771 | 0.059 | -0.193 | 27.954 |
| ABU-1 | 12.098 | 0.063 | -0.001 | 7.557 |
| ABU-2 | 11.235 | 0.094 | -0.007 | 0.172 |
| ABU-10 | N/A | 0.003 | -0.002 | -0.423 |
| ABU-14 | N/A | N/A | N/A | -0.421 |
| ABU-4 | N/A | 0.005 | -0.002 | -0.387 |
| ABU-5 | N/A | -0.002 | -0.004 | -0.383 |

**Table 2.3** Summary of PLAAC results from the ABU family of proteins (Note that some proteins do not have values attributed to them – these represent outputs for which no domain of minimal core length could be identified)

Of particular interest are those proteins that score highly across multiple readouts – for example, whilst ABU-2 scores highly in prion propensity (0.094), it scores much lower in terms of CORE score (11.235) compared to other members of this family. Hence, to maximise our chances at identifying a genuine prion-like domain-containing protein, we correlated scores between different algorithms to highlight proteins that scored consistently well across each of these programs (Figure 2.3D-G), including yeast and human proteins in this analysis for comparison.

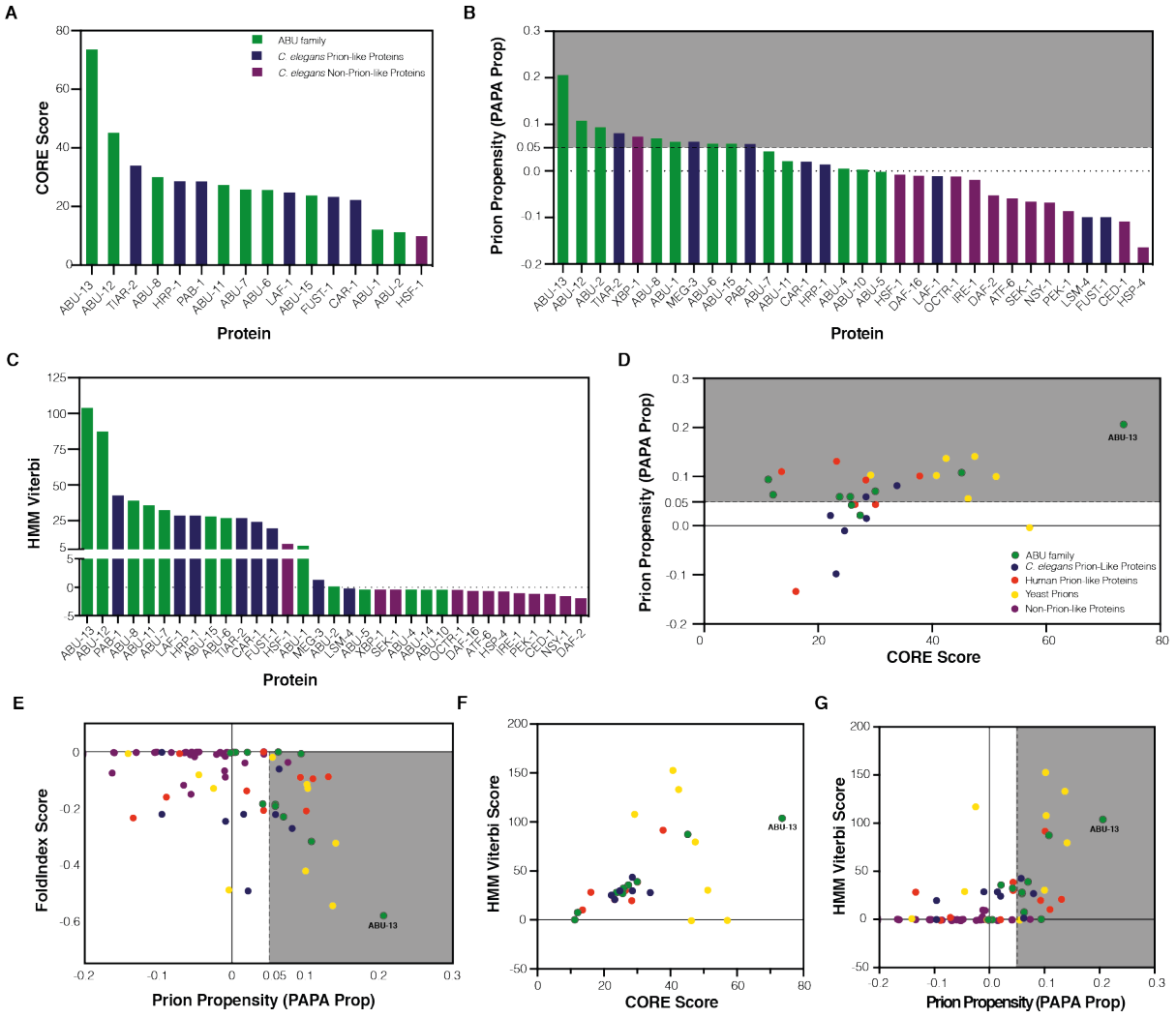
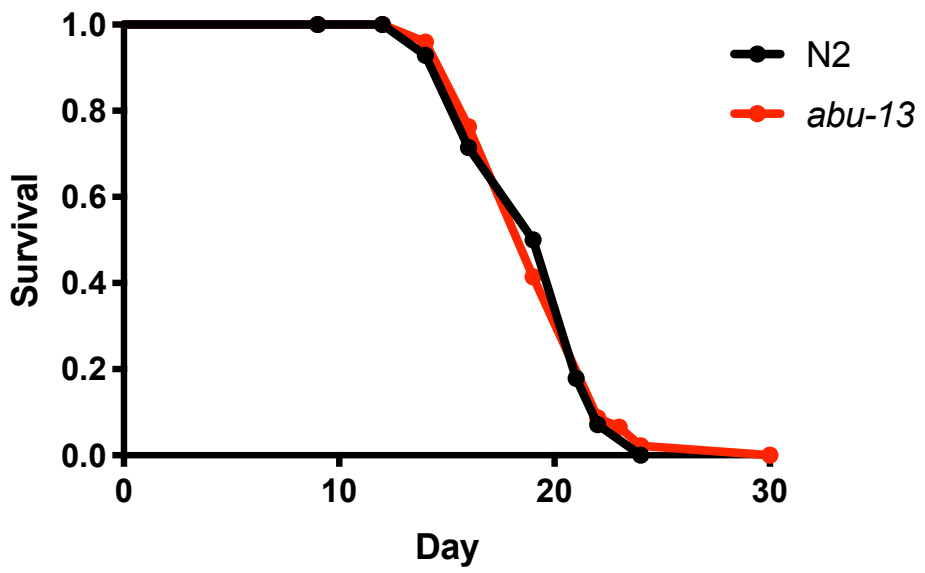



**Figure 2.3** Comparison of nematode PLAAC scores (A-C) Comparisons of ABU family PLAAC results with *C. elegans* specific prions and non-prions, including (A) CORE score [One way ANOVA, no significant difference]; (B) Prion Propensity [One way ANOVA, ABU family v *C. elegans* Non-Prion, $p$=0.0005]; (C) Viterbi-parsed HMM scores [One way ANOVA, ABU family v *C. elegans* Non-Prion, $p$=0.0128] (D) Correlation between CORE score and Prion Propensity [Pearson correlation, r=0.4408; $p$=0.0167; grey = significant PAPA value]; (E) Correlation between Prion Propensity and FoldIndex [Pearson correlation, r=-0.4976, $p$=<0.0001; grey = significant PAPA value]; (F) Correlation between CORE score and the Viterbi-parsed HMM scores [Pearson correlation, r=0.4634, $p$=0.0113]; (G) Correlation between Prion Propensity and the Viterbi-parsed HMM scores [Pearson correlation, r=0.5610, $p$=<0.0001; grey = significant PAPA value]

Within these correlations, species-specific clustering of prion scores becomes more evident with yeast prions consistently grouping together in regions highly predictive of prionogenic activity, whilst human and nematode scores tend to cluster at lower prediction scores. In keeping with some of the individual algorithms, the ABU family in general clusters with the prion-like proteins, rather than with the non-prion-like proteins. Consistently, ABU-13 outperforms other prion-like proteins, including the yeast prion clusters, with the high scoring nature of this protein manifestly clear from these plots. As a result, we decided to investigate and characterise ABU-13, as the most promising candidate from our computational predictions.

### *2.2.5 ABU-13 plays a non-redundant role in both innate immune and ER stress responses*

A




B

| | Replicate 1 | Replicate 2 | Replicate 3 |
|---|---|---|---|
| *N2* | 20 | 23 | 23 |
| *abu-13(tm6224)* | 19 | 23 | 23 |
| *p-value* | 0.2643 | 0.5307 | 0.0004 |

**Figure 2.4** Knockout of *abu-13* does not consistently alter lifespan of animals (A) Kaplan-Meier survival curves of *abu-13(tm6224)* and N2 animals assessed at 20°C; (B) Table indicating median lifespans (in days) of each replicate, along with the p-values of these replicates, with significance calculated by Mantel-Cox log rank test

As a role for the *abu* family in cellular stress responses has been implicated previously[13,14], we sought to determine whether ABU-13 plays a non-redundant role in protection upon exposure to a number of environmental challenges. Using animals with a deletion mutation of *abu-13* (allele *tm6224,* truncated at residue 177, compared to the full length of 493 amino acids) obtained from the Japanese Knockout Consortium, we assessed responses of these knockout animals to ER stress, pathogenic exposure, and heat shock. Stress response pathways and lifespan are intimately linked, with components of a variety of these pathways being implicated in the modulation of lifespan. Constitutive splicing of the $UPR^{ER}$ transcription factor *xbp-1* in neurons, for example, results in lifespan extension, whilst *xbp-1* mutant animals have a reduced lifespan[15].

To determine whether *abu-13* knockouts have a similarly shortened lifespan, we performed lifespan analyses on these animals, ultimately determining that they have a similar lifespan to wild type animals. This suggests that ABU-13 does not play a fundamental role in the maintenance of longevity, thus uncoupling any stress resistance observed in these animals from longevity (Figure 2.4A-B). Whilst similar effects have been observed before, this decoupling is unusual, potentially highlighting the redundancy of this pathway with other stress resistance pathways in lifespan maintenance.

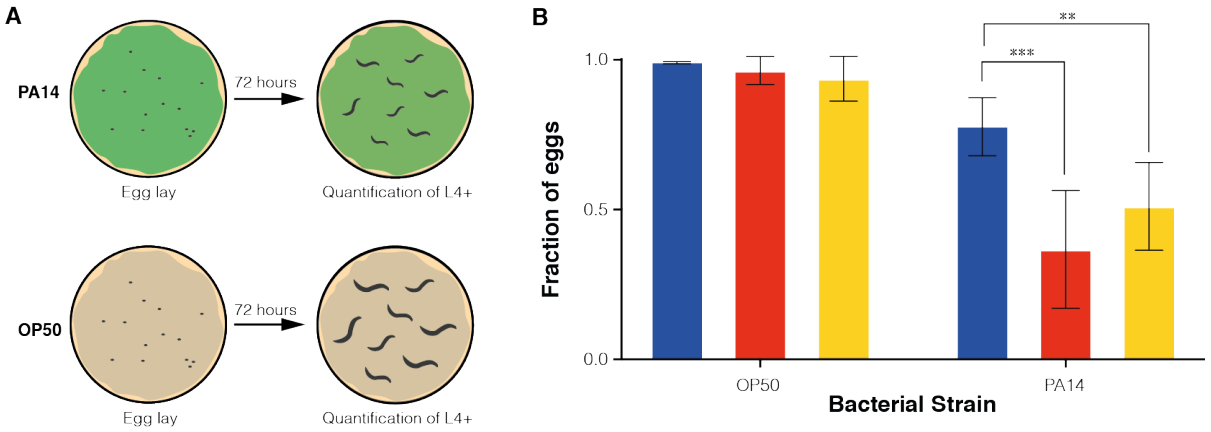



**Figure 2.5** ABU-13 plays a non-redundant role in the innate immune response (A) Schematic of the developmental assay used, with animals grown on either an OP50 control or pathogenic PA14; (B) Fraction of eggs developing to L4 after 72 hours when fed with either OP50 or PA14 in N2 (blue), *abu-13(tm6224)* (red) and *xbp-1(zc12)* (yellow) animals (Two way ANOVA, N2 v *abu-13(tm6224)* on PA14 – $p$=0.0002, ***; N2 v *xbp-1(zc12)* – $p$=0.0091, **)

We then assessed the role that ABU-13 plays in the innate immune response, by comparing the developmental times of N2 animals, *abu-13(tm6224)* animals, and *xbp-1* (allele *zc12*) mutant animals that have a point mutation in codon 34 resulting in premature termination of XBP-1 upstream of its functional domains, on either a non-pathogenic *E. coli* (OP50) bacterial lawn or a pathogenic *Pseudomonas aeruginosa* (PA14) lawn (Figure 2.5B). Bacterial avoidance is a strategy that nematodes use to improve survival when exposed to such immune threats; to eliminate the confounding effects of this behaviour on our results, we used a large bacterial lawn, ensuring the entire NGM plate was covered, removing the ability of animals to avoid PA14.

*Xbp-1(zc12)* animals are known to have an attenuated response to pathogens and, in keeping with this, show a significant reduction in developmental time when grown on PA14 in comparison to N2 animals, which only show a small defect in the fraction of eggs reaching the L4 postembryonic developmental stage after 72 hours. ABU-13 knockout animals also show reduced pathogen resistance, exhibiting a reduction in developmental time when compared to N2 (Figure 2.5B). This effect is driven by the pathogen itself, as these developmental differences are not observed when animals are grown on OP50.

To determine their sensitivity to ER stress, *abu-13(tm6224)* animals were subjected to a tunicamycin development assay, with plates supplemented with 0, 1.5 or 3µg/mL of tunicamycin (Figure 2.6A). Similar to the PA14 assay, the fraction of eggs reaching L4 after 72 hours was scored and compared to the fractions of both N2 and *xbp-1(zc12)* mutant animals reaching this stage. *Xbp-1(zc12)* mutant animals exhibit significantly reduced survival and development upon exposure to tunicamycin compared to wild type. As expected, in our assay N2 animals show minimal developmental defects when exposed to 1.5µg/mL tunicamycin, whilst *xbp-1(zc12)* animals have an extreme developmental delay, with only a very small percentage reaching L4 after 72 hours. ABU-13 knockout animals have a moderate developmental delay upon tunicamycin treatment when compared to N2. Similar results were

observed at 3µg/mL, but the trend is much less obvious as this concentration of tunicamycin is sufficient to cause a significant developmental delay even in the N2 control animals (Figure 2.6B). This suggests that ABU-13 does indeed play a role in protecting against the effects of ER stress, though to a lesser degree than XBP-1.

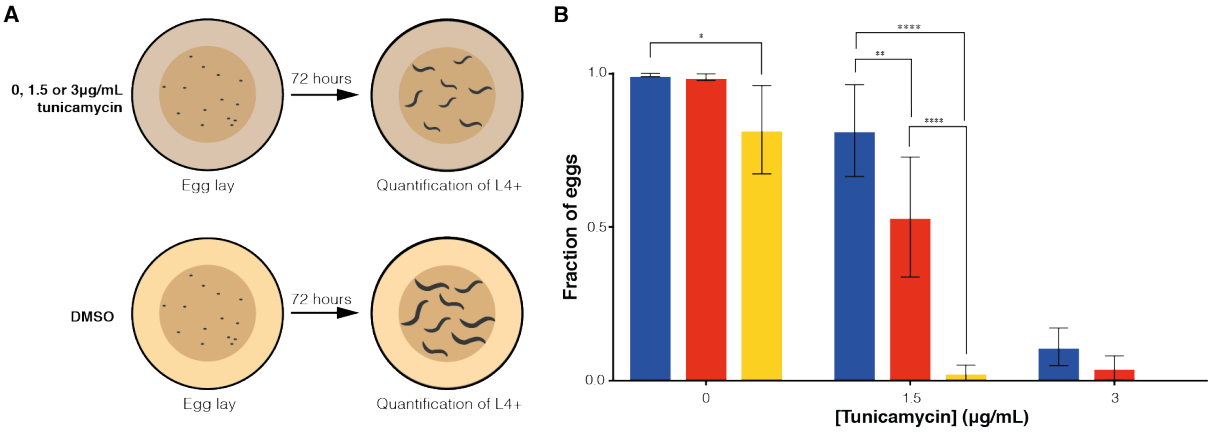



**Figure 2.6** ABU-13 plays a non-redundant role in ER stress responses (A) Schematic of the developmental assay used, with animals grown on either DMSO-supplemented plates or plates supplemented with varying concentrations of DMSO; (B) Fraction of eggs developing to L4 after 72 hours when grown on plates supplemented with either DMSO (represented at 0µg/mL), 1.5µg/mL or 3µg/mL of tunicamycin in N2 (blue), *abu-13(tm6224)* (red) and *xbp-1(zc12)* (yellow) animals (Two way ANOVA, N2 v AGD972 on 0µg/mL – $p$=0.0423; N2 v *abu-13(tm6224)* on 1.5µg/mL tunicamycin – $p$=0.0011, **; N2 v AGD972 on 1.5µg/mL tunicamycin – $p$= <0.0001, ****; *abu-13(tm6224)* v *xbp-1(zc12)* on 1.5µg/mL – $p$=<0.0001, ****)

We then examined resistance to a 2.5 hour 37°C heat shock treatment in N2, *abu-13(tm6224)* animals, and, as a positive control, animals with a substitution mutation in the heat shock factor *hsf*-1 (allele *sy441*) leading to a truncation mutant lacking its transactivation domain. There was, however, no significant difference between the heat shock survival of N2 and *abu-13(tm6224)*, whereas *hsf-1(sy441)* animals showed a significant reduction in survival when compared to either of these strains (Figure 2.7A-B). This suggests that ABU-13 is important in response to specific stresses, rather than to cellular stress in general.

### *2.2.6 Interplay between the canonical and non-canonical UPR$^{ER}$ pathway*

As the ABU family were identified as being upregulated in response to inhibition of the IRE-1/XBP-1 branch of the UPR$^{ER}$, we crossed *abu-13*(*tm6224)* animals into a strain containing a *hsp-4p::gfp* transgene. HSP-4 is the *C. elegans* ortholog of the mammalian ER stress sensor BiP, which is upregulated upon ER stress. Its induction depends entirely on the presence of IRE-1 and XBP-1[16] and this strain therefore acts as a readout for IRE-1/XBP-1 activation. Thus, using this transcriptional reporter, we were able to determine whether the activation of XBP-1 was changed in *abu-13* knockout animals, using GFP expression as a surrogate for activation.

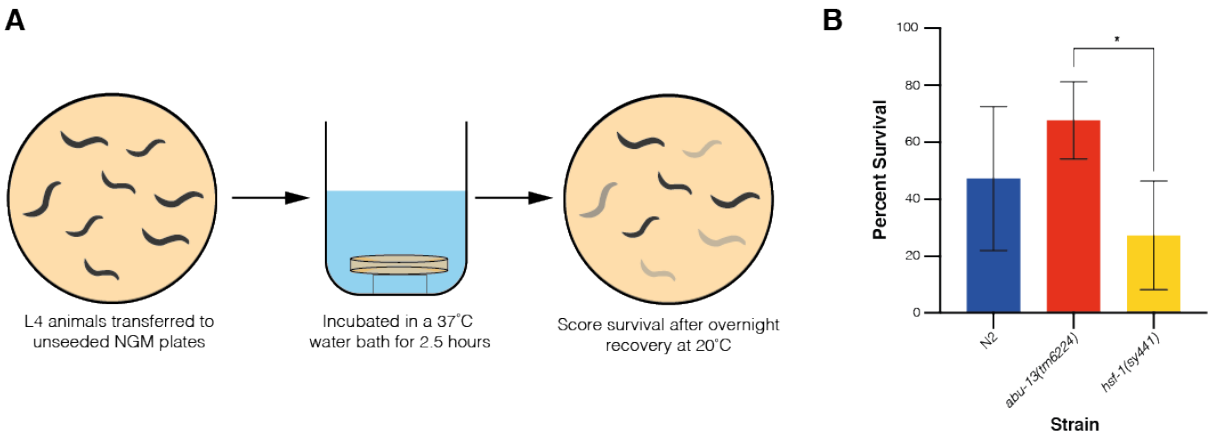



**Figure 2.7** ABU-13 does not play a non-redundant role in responses to heat shock (A) Schematic of the heat shock assay used, with survival of animals scored after a 2.5-hour heat shock treatment at 37°C; (B) Percent survival following a 2.5-hour heat shock at 37°C (One-way ANOVA, *abu-13(tm6224)* v *hsf-1(sy441)* – $p$=0.0439, *)

In animals expressing the *hsp-4p::gfp* transgene in a wild type background (SJ4005), GFP levels are low under basal conditions as the UPR is not activated. Upon tunicamycin treatment, however, the expression of GFP is induced, resulting in a large increase in observable fluorescence, demonstrating the effective activation of these downstream targets. In an *xbp-1* mutant background, however, this fluorescence cannot be seen either before or after tunicamycin treatment, demonstrating that these targets are not activated, as would be expected when this pathway is disrupted[15]. No significant effect on the activation of the *hsp-4p::gfp* transgene was observed in *abu-13* knockout animals, with similar levels of fluorescence recorded in these animals compared to the control SJ4005 animals. This suggests that ABU-13 does not influence the expression of XBP-1 targets upon ER stress, and that the phenotypic effects observed are a consequence of a different pathway.

In addition to this, as part of the innate immune system, proteostatic mechanisms are upregulated to mitigate the effects of the immune response itself. In keeping with this, existing evidence has demonstrated that *hsp-4p*::GFP expression is increased upon exposure to PA14. As we have identified a role for ABU-13 in both of these responses, we repeated this experiment in both the *hsp-4p::*GFP reporter control and those crossed into *abu-13(tm6224)* as well. We see a small increase in fluorescence in all animals following PA14 exposure, but no difference in reporter expression levels between these strains. This therefore further suggests that the role ABU-13 is playing in these responses is independent of the IRE-1/XBP-1 branch of the UPR$^{ER}$.

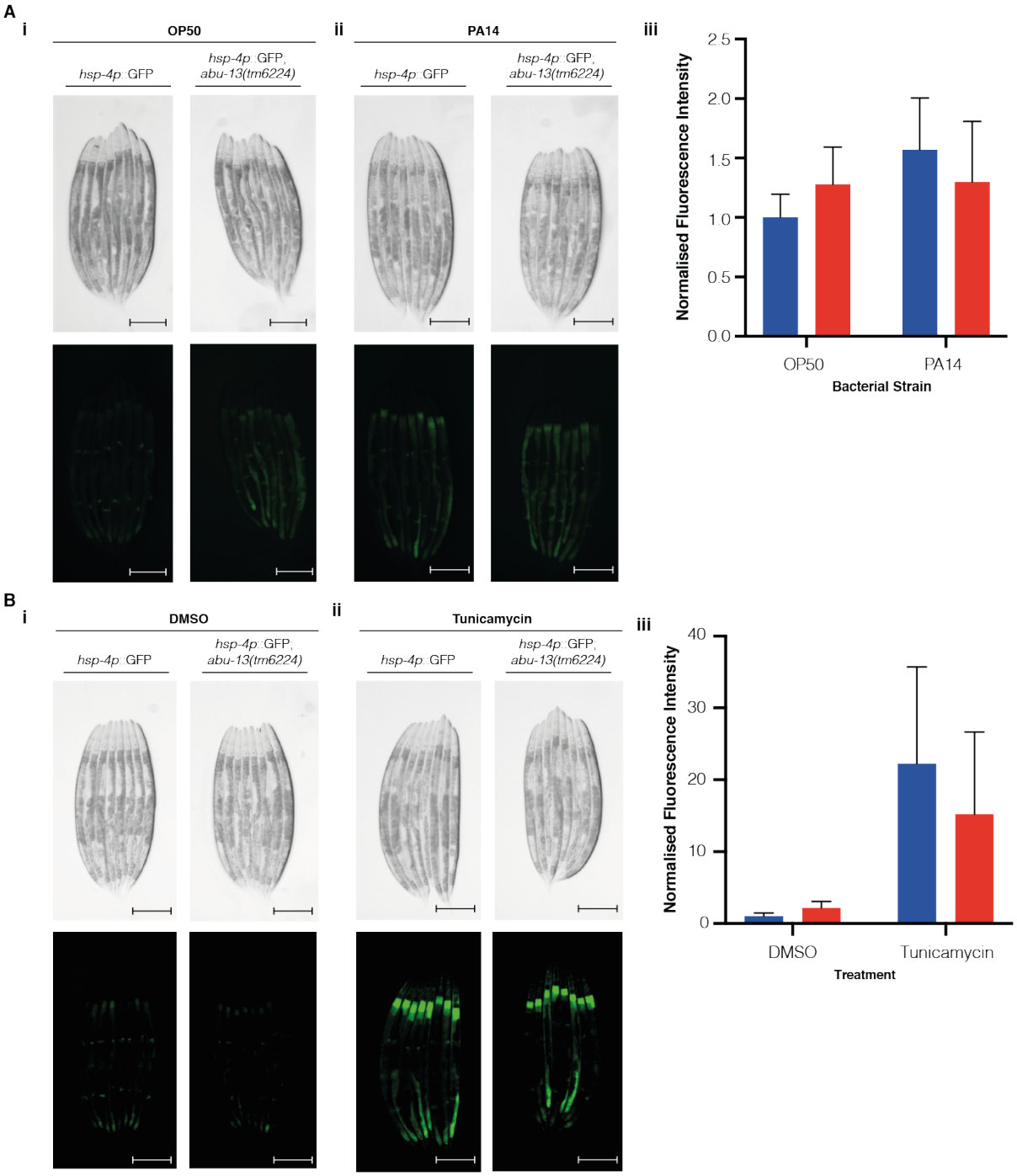



**Figure 2.8** ABU-13 does not influence the expression of *hsp-4*, a readout of IRE-1/XBP-1 activationMicrographs of *hsp-4p::*GFP (SJ4005) animals and *hsp-4p::*GFP;*abu-13(tm6224)* animals grown [i] OP50 for 6 hours; [ii] PA14 for 6 hours; [ii] Normalised fluorescence intensity values of the *hsp-4p::*GFP reporter after growth on OP50 or PA14 (*SJ4005 v abu-13(tm6224) on OP50, p = 0.8111, ns; SJ4005 v abu-13(tm6224) on PA14, p = 0.8226, ns; SJ4005 on OP50 v SJ4005 on PA14, p = 0.3290, ns; abu-13(tm6224) on OP50 v abu-13(tm6224) on PA14, p = 0.9998, ns)* (B) Micrographs of *hsp-4p::*GFP (SJ4005) animals and *hsp-4p::*GFP;*abu-13(tm6224)* animals treated with [i] DMSO for 4 hours; [ii] 25ng/µL tunicamycin for 4 hours [iii] Normalised fluorescence intensity values of the *hsp-4p::*GFP reporter after DMSO or 25ng/µL tunicamycin treatment (*SJ4005 v abu-13(tm6224) treated with DMSO, p = 0.9984, ns; SJ4005 v abu-13(tm6224) treated with tunicamycin, p = 0.7705, ns; SJ4005 treated with DMSO v SJ4005 treated with tunicamycin, p = 0.0730, ns; abu-13(tm6224) treated with DMSO v abu-13(tm6224) treated with tunicamycin, p = 0.3381, ns), Scale bars = 250µm*

## 2.3 Discussion

The primary determinant of a prion-like domain is amino acid composition, often in the form of Q/N-rich sequences[2,7]. In yeast, this is particularly the case, with many prions within this species shown to contain an enrichment of these residues; as these prions form the basis of many computational predictions, this has resulted in a bias, with algorithms preferentially identifying domains with these characteristics to the neglect of prion-like domains that utilise different mechanisms to confer their prionogenic abilities. This is best illustrated by both the prion protein itself, as well as the fungal prion HET-s, both of which have strong prion activity whilst lacking a Q/N-rich domain[17]. Structural studies have shown that these proteins form beta-helical and beta-solenoid structures, capable of acting as prionogenic scaffolds – this suggests that any secondary structure domain capable of driving the formation of such a tertiary structure should be classed as a prion domain.

This heterogeneity of prion domain composition is reflected in our PLAAC results, with many proteins known to possess prion-like characteristics scoring low in some of these algorithms. This suggests they do not conform to the standard Q/N-rich model for prion-like domains but rather they represent alternative mechanisms through which these characteristics can be conferred; RBM14 and DDX4, for example, are both proteins whose prion-like domains facilitate their involvement in the formation of the phase separated organelles paraspeckles and nuage, respectively[18,19]. Using these algorithms, however, they score low across multiple readouts, suggesting that they represent non-prion-like proteins – this is a clear demonstration of the omission of non-conforming prion domains from these computational results.

This may also explain the differences observed between the scores of prions in different species, specifically between yeast prions, and their human and nematode counterparts. If there are a higher proportion of prions that use Q/N-rich domains in yeast compared to those in human and nematode species, then the average score of known, characterised prion-like domain proteins will be lower in latter, thus biasing the output towards higher average scores for the yeast prion datasets. However, this does not exclude the possibility that the Q/N-rich prion-like domains identified in the PLAAC algorithm for *C. elegans* do indeed represent genuine novel prion-like proteins; instead, it highlights the systematic underrepresentation of non-conforming prion-like domains in these results.

In contrast to this omission, the false positive rate of many of these Q/N-rich domain prion predictions has been high with just 19 out of the 100 candidates screened in the initial validation of the Alberti algorithm possessing genuine prionogenic activity[8]. This has been mitigated by the development of algorithms such as that developed by Toombs et al.[9,12], which predicts domains on their basis of residue interactions. Thus, by using multiple computational predictions in conjunction with one another, we

were able to cross-validate and strengthen high-ranking candidates in our screen of the *C. elegans* proteome.

Amongst these highly ranked candidates, ABU-13 is one of the top hits across multiple outputs, scoring higher than well-characterised yeast prions by some readouts. This could be further validated moving forward using a Sup35 prion domain replacement assay, in which the prion-domain of Sup35 is swapped with that of the candidate prion. This chimeric protein is then tested for its ability to propagate the [PSI+] prion state – with successful propagation indicating the presence of a genuine prion-like domain. If the prion-like domain of ABU-13 is indeed capable of recapitulating these prionogenic effects then this would hugely strengthen the case for this protein acting as a prion in *C. elegans*.

Based on the persuasively high score of ABU-13 across multiple readouts, however, we moved forward with exploring the functions of this protein as a promising candidate for a novel prion-like protein. Surprisingly, we identified a non-redundant role for ABU-13 in the response to tunicamycin treatment in a wild type background; this suggests that this protein is involved in the ER stress response pathway even when the IRE-1/XBP-1 branch of the UPR is intact. This must be functioning separately from the canonical pathway as, when crossed into an *hsp-4* transcriptional reporter line, the expression of HSP-4 was unaffected by knockout of *abu-13*. This suggests that under wild type conditions ABU-13 plays a role in stimulating a stress response; upon inhibition or disruption to *xbp-1*, however, this protein is highly upregulated, compensating for the loss of *xbp-1*. Alternatively, this could represent differential developmental effects of ABU-13 and XBP-1 pathways. Both have been shown to be regulated downstream of the neuronal GPCR, OCTR-1, but these influences differ between larval stages and adulthood. No regulation of XBP-1 by OCTR-1 is observed during larval development - this ability is gained in adulthood, as demonstrated by the dependence on *xbp-1* for the enhanced survival of *octr-1* mutants upon PA14 in adulthood only.

ABU-13, meanwhile, is negatively regulated by OCTR-1 throughout life in a CED-1-dependent fashion. As the effect of *abu-13* knockout has been investigated in developmental assays, it remains possible that the fundamental, non-redundant role we have observed for ABU-13 in ER stress responses is only seen during development. Indeed, in data not shown, tunicamycin survival assays of *abu-13* knockout animals initiated during L4 had highly variable survival rates – this is possibly as a result of inconsistent L4 staging, with those closer to adulthood surviving less than their slightly younger counterparts - however, controlled survival assays of adult animals would be required in order to draw any conclusions. This common neuronal regulation of ABU-13 and XBP-1 does, however, suggest an interesting relationship between the canonical and non-canonical UPR pathways. Further work is needed to characterise and clarify the nature of this relationship.

Whilst the role of the ABU family has been identified in the context of a compensatory response to *xbp-1* inhibition, its potential roles in respect to the other branches of the UPR$^{ER}$ have been neglected. To investigate this, we would perform stress assays on ATF-6 and PEK-1 mutants, monitoring any changes in response when these animals are crossed into *abu-13* knockouts. This would be in addition to the work performed with *xbp-1;abu-13* double mutants - however, as these genes are both found on chromosome III, this would require editing of the endogenous locus via CRISPR to achieve this.

In addition to this, the *abu-13* knockout animal strain used (*tm6224*) contains a truncated form of the protein. Whilst this does abolish protein functioning entirely, it is not possible to determine whether the functional effects observed are a consequence of the prion-like domain specifically. To assess this, we will produce animals with deletion mutations of the entire predicted prion-domain, leaving the rest of the protein intact, and repeat the aforementioned phenotypic assays to determine whether these effects are still observed.

Comparison of the ABU family of proteins suggests that this family is predisposed to prion-like features, with many of the proteins in this family scoring relatively high for prionogenicity. This suggests that low complexity might aid their involvement in stress response pathways – how this is brought about, however, is yet to be fully understood. There are, however, a number of intrinsically disordered proteins that have already been implicated in stress responses – for example, LSM-1 is involved in the formation of stress responsive phase separated organelles known as P-bodies[20], whilst TIA-1 is involved with stress granule formation[21]. This potentially points towards a role for ABU-13 in the formation of phase separated organelles in response to stress. *In vivo* work will help to investigate this hypothesis further.

In addition to this, it is also possible that the intrinsic disorder of the ABU family of proteins acts to 'mimic' misfolded protein events within the ER. In the canonical UPR$^{ER}$, the chaperone HSP-4 (BiP in humans) is expressed in the endoplasmic reticulum and, under basal conditions, binds to the stress sensor IRE-1, maintaining this transmembrane protein in a monomeric, inactive state. When there is an accumulation of misfolded or unfolded proteins in the ER, however, HSP-4 preferentially binds to these proteins instead – this derepresses IRE-1, triggering its dimerization and subsequent activation, ultimately leading to a downstream stress response[16,22,23]. If the ABU family were acting to molecularly 'mimic' misfolded protein events, this could stimulate an enhanced activation of the canonical pathway by sequestering HSP-4 away from IRE-1, leading to an increased activation of this stress response. If this were the case, then the increased expression of ABU proteins observed in *xbp-1* mutant animals would represent a mechanism aimed at driving the increased activity of the non-*xbp-1*-related functions of IRE-1, as well as increasing the activation of the PEK-1 pathway, which also involves the chaperone HSP-4. Hence the ABU family could be functioning to upregulate compensatory UPR pathways, rather than stimulating a novel pathway entirely. There is, however, no observable change in the *hsp-4::*GFP reporter when combined with the *abu-13(tm6224)* mutants, as would be expected if this were the

mechanism of action – this, however, might be reflection of redundancy amongst the ABU family, with the phenotypic deficiencies observed in *abu-13(tm6224)* animals a consequence of the activity of functional domains outside of intrinsically disordered regions. On the whole, this hypothesis could be tested further by assessing the expression of downstream target genes of these compensatory pathways in *xbp-1;abu-13* double mutant backgrounds compared to *xbp-1* single mutants.

Alternatively, this prion-like characteristic of the ABU family might enable these proteins to bind to similarly unfolded regions found in misfolded proteins. This ability to has been observed in a number of prion-like proteins – often to a pathological end – however, in this case, this interaction might represent a physiological binding event. Akin to processes such as ubiquitination, in which proteins are targeted for degradation, binding of ABU proteins to misfolded proteins could act as a mechanism to identify and trigger their degradation. It would be interesting to investigate this further and identify whether there is any binding of the ABU proteins directly to misfolded proteins, and to compare the level of proteasomal activity in strains expression misfolding-prone proteins in both wild type and *abu* mutant backgrounds to determine whether such a binding event is used to stimulate degradation. It is important to note that, with the exception of ABU-13, the proteins within the ABU family all contain a transmembrane domain. This could point towards an alternative role – perhaps via recruitment of misfolded proteins to specific compartments, such as lysosomes, for degradation. In keeping with this, GFP-fused ABU-1 has been shown to localise in vesicular structures[13]. A better understanding of the cellular localisation of these proteins would improve our understanding of the role they might play.

Although a defined role for the ABU family remains unclear, the unique properties of ABU-13 suggest it plays a distinct role within these responses, potentially one that differs significantly from the rest of the family. As mentioned previously, this protein is by far the most likely candidate for a prion in the *C. elegans* proteome. In addition to this, the presence of a potential RNA binding domain in the N-terminal region of this protein is highly reminiscent of the domain structure observed in phase separating proteins. The combination of a prion-like domain and an RNA interacting motif has been shown to play fundamental roles in the process of liquid liquid phase transitions due to the high valency that such protein architecture bestows. In fact, disruption of the RNA binding ability of the prion-like domains of such proteins can abolish their ability to form such structures at all. Multivalency is a key feature of these proteins, as this ability to interact with numerous molecules at once drives the coalescence of multiple of cellular components, allowing distinct regions of high concentration to form. The presence of these domains within ABU-13, as well as the prediction of binding sites within the intrinsically disordered region and a zinc-binding site in the C terminus of the protein add weight to the hypothesis that ABU-13 is indeed acting as a scaffold for such transitions.

## 2.4 Material and Methods

*Maintenance of C. elegans*

Nematodes were maintained in accordance with standard protocols. Briefly, animals were grown on nematode growth media (NGM) plates seeded with the *E. coli* strain, OP50, at either 15°C for maintenance or at 20°C for experimental work. Prior to seeding, OP50 was cultured in Luria-broth overnight at 37°C, after which 100µL of bacteria was dispensed onto each plate and dried overnight at room temperature before use.

*Computational Predictions*

For predictions of prion-like domains, the PLAAC algorithm was used (http://plaac.wi.mit.edu/), using reference proteomes obtained from UniProt. All queries were programmed with the core length of 60 as standard. For the yeast proteome, the background frequency ($\alpha$) was set to 100%, whilst the human and *C. elegans* proteomes were set to 50%. Output from this computation also gave residue-by-residue charge values from which a charge distribution plot was constructed.

Domain conservation was determined using the NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), with settings maintained at default. ProinW was used to predict presence of nucleating amyloidogenic sequences (http://bioinf.uab.cat/prionw/). Signal sequences were identified using Signal-P5.0 (http://www.cbs.dtu.dk/services/SignalP/). The presence of zinc-binding residues was predicted by PredZinc (http://predzinc.bioshu.se/).

For determination of disorder, the IUPRED2A and ANCHOR2 algorithms were applied (https://iupred2a.elte.hu/), with scores from the original ANCHOR algorithm used for prediction of binding sites within the region of disorder.

*Lifespan Analysis*

Age-synchronised populations were grown at 20°C on OP50-seeded NGM plates and animals scored for viability every other day throughout adulthood (with the end of larval development corresponding with t = 0). To prevent offspring production and egg laying, animals were treated with 100µg/mL of FUDR at t = 0 and at day 5 of adulthood. Lifespan analyses were conducted using Prism 7 software, with P-values determined using the log-rank (Mantel-Cox) method.

*Development Assays*

For tunicamycin development assays, NGM plates were supplemented with 1.5µg/mL or 3µg/mL of tunicamycin (Sigma Aldrich), or an equivalent volume of DMSO as a control. Beyond this point, plates were seeded with OP50 as per the standard protocol. For PA14 development assays, bacteria was cultured overnight at 37°C after which high peptone (0.35%) NGM plates were seeded with 100µL of bacteria evenly spread across the agar using a 10µL disposable inoculating loop (Thermo Scientific). PA14-seeded plates were incubated at 37°C overnight and then left at room temperature for 8 hours before use.

Once ready for use, a timed egg lay was carried out on experimental plates to produce age-synchronised populations of animals. After this, plates were stored at 20°C for 72 hours before each plate was scored for developmental stage.

*Heat Shock Assay*

A timed egg lay was performed on seeded-NGM plates to age-synchronise populations. Once animals had reached the L4 stage of development, ~20 animals were transferred onto unseeded plates and wrapped in parafilm, ensuring they were fully watertight, before being placed into a 37°C water bath. After 2.5 hours, plates were removed and the survival of animals scored following an overnight period of recovery.

*Stress Reporter Treatment*

For tunicamycin imaging, age synchronised animals were collected at L4 and transferred into 1.5mL Eppendorf tubes containing either 25ng/µL of tunicamycin or an equivalent volume of DMSO. Animals were then incubated at room temperature on a rotating platform for 4 hours, before being transferred onto unseeded plates and immobilised with 50mM sodium azide.

For *Pseudomonas* imaging, age synchronised animals were collected at L4 and transferred onto plates seeded with either OP50 or PA14, prepared as described above. Animals were then left on their respective bacterial lawns at 20°C for 6 hours, before being transferred onto unseeded plates and immobilised with 50mM sodium azide.

Images were taken at 1X magnification using a Leica M205 FA microscope and LAS X software. Quantification of fluorescence was performed using ImageJ.

*Quantification of Fluorescence Imaging*

To determine the normalised fluorescence intensity values for each condition, the corrected total cell fluorescence (CTFC) for each individual animal was initially calculated. To do this, animals were

isolated within the image and measurements of integrated density and total area of the animal were taken, along with the average gray area of 3 different background regions. The CTFC was calculated as follows;

$$CTFC = Integrated\ Density - (Area\ of\ animal * Mean\ gray\ area)$$

An average CTFC value was determined for the control condition animals, and all CTFC values normalised to this average. For all these measurements, Prism 8 software was used for statistical analysis.

*Statistical Analysis*

All statistical analyses were performed using Prism 8 software. One- and two-way ANOVA adjusted with Tukey's multiple comparison correction was used to determine significance.

## 2.5 References


1. Li, L. & Lindquist, S. Creating a Protein-Based Element of Inheritance. *Science.* **287**, 661 – 664 (2000).
2. Ross, E. D., Baxa, U. & Wickner, R. B. Scrambled Prion Domains Form Prions and Amyloid. *Mol. Cell. Biol.* **24**, 7206 – 7213 (2004).
3. DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. A Critical Role for Amino-Terminal Glutamine/Asparagine Repeats in the Formation and Propagation of a Yeast Prion. *Cell* **93**, 1241–1252 (1998).
4. Michelitsch, M. D. & Weissman, J. S. A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci.* **97**, 11910–11915 (2000).
5. Maurer-Stroh, S. *et al.* Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* **7**, 237–242 (2010).
6. Gonzalez Nelson, A. C. *et al.* Increasing Prion Propensity by Hydrophobic Insertion. *PLoS One* **9**, e89286 (2014).
7. Michelitsch, M. D. & Weissman, J. S. 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**, (2000).
8. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* **137**, 146–158 (2009).
9. Toombs, J. A., McCarty, B. R. & Ross, E. D. Compositional Determinants of Prion Formation in Yeast. *Mol. Cell. Biol.* **30**, 319 – 332 (2010).
10. Toombs, J. A. *et al.* De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci.* **109**, 6519–6524 (2012).
11. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).
12. Toombs, J. A. *et al.* De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci.* **109**, 6519–6524 (2012).
13. Urano, F. *et al.* A survival pathway for *Caenorhabditis elegans* with a blocked unfolded protein response. *J. Cell Biol.* **158**, 639 – 646 (2002).
14. Sun, J., Singh, V., Kajino-Sakamoto, R. & Aballay, A. Neuronal GPCR Controls Innate Immunity by Regulating Noncanonical Unfolded Protein Response Genes. *Science.* **332**, 729 – 732 (2011).
15. Taylor, R. C. & Dillin, A. XBP-1 Is a Cell-Nonautonomous Regulator of Stress Resistance and Longevity. *Cell* **153**, 1435–1447 (2013).
16. Calfon, M. *et al.* IRE1 couples endoplasmic reticulum load to secretory capacity by processing

the XBP-1 mRNA. *Nature* **415**, 92–96 (2002).

17. Coustou, V., Deleu, C., Saupe, S. & Begueret, J. The protein product of the het-s heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proc. Natl. Acad. Sci.* **94**, 9773 – 9778 (1997).
18. Hennig, S. *et al.* Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **210**, 529–539 (2015).
19. Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **57**, 936–947 (2015).
20. Gallo, C. M., Munro, E., Rasoloson, D., Merritt, C. & Seydoux, G. Processing bodies and germ granules are distinct RNA granules that interact in *C. elegans* embryos. *Dev. Biol.* **323**, 76–87 (2008).
21. Gilks, N. *et al.* Stress Granule Assembly Is Mediated by Prion-like Aggregation of TIA-1. *Mol. Biol. Cell* **15**, 5383–5398 (2004).
22. Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P. & Ron, D. Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat. Cell Biol.* **2**, 326–332 (2000).
23. Amin-Wetzel, N. *et al.* A J-Protein Co-chaperone Recruits BiP to Monomerize IRE1 and Repress the Unfolded Protein Response. *Cell* **171**, 1625-1637.e13 (2017).

# Chapter 3

## *In vivo* and *in vitro* characterisation of ABU-13

## 3.0 Abstract

Throughout biology, prion-like proteins are routinely overrepresented in pathology, with their propensity for aggregation promoting the formation of aberrant fibrillar species. More recently, however, evidence has emerged for a physiological role of these proteins, with many involved in the process of liquid-liquid phase separation. To determine whether ABU-13 acts as a scaffold for the formation of these phase-separated compartments, we produced both fluorescently labelled overexpression and endogenously-modified ABU-13 lines, allowing us to monitor the activity of our candidate prion-like protein *in vivo*. In line with our hypothesis, we do indeed see puncta formation throughout the body of the animal, reminiscent of these membraneless organelles, suggesting that ABU-13 is indeed involved with formation of these transient structures. To determine whether these puncta were dependent on prionic mechanisms, we removed the prion domain from both of these lines, finding conflicting results, potentially suggesting that, whilst ABU-13 can act as a scaffold protein, there are other scaffold proteins expressed that play a compensatory role in these transitions in prion domain mutant animals.

To investigate the biophysical properties of these puncta, we subjected them to FRAP to determine their mobility, finding that the larger puncta found within the overexpression line were largely immobile, whilst those within the endogenous line showed a broad range of mobility. Interesting, we observe a positive correlation between size and mobility, with larger puncta in the endogenously tagged line showing increased mobility compared to the smaller puncta. We propose that this is a consequence of multiphase coexistence, with smaller, low mobility structures representing core regions containing constitutive components of these compartment, whilst larger structures represent multiple core regions surrounded by a high mobility shell rich with more transient, highly turned over components.

As we had already identified non-redundant roles for ABU-13 in both ER stress and innate immune responses, we next sought to understand how these puncta related to function. Whilst exposure of these animals to ER stress and heat shock conditions resulted in no observable change in puncta formation, we did see a highly significant change in puncta characteristics after exposure to the pathogenic bacteria *Pseudomonas aeruginosa*, with fewer yet larger puncta seen following invasion, as well as an upregulation of diffuse ABU-13 expression in the pharynx. This therefore led us to hypothesise that ABU-13 was acting via different mechanisms in response to these environmental stresses.

To aid our understanding of how these compartments may be functioning, we investigated where ABU-13 was expressed, with tissue predictions suggesting many hypodermal tissues. Indeed, an ABU-13 transcriptional reporter line does appear to show expression in tissues including the hypodermis, seams cells, amphid sheath glia and the rectal epithelial cells. We then crossed our fluorescently tagged ABU-

13 lines into a number of tissue marker strains, confirming expression in the seam cells, and potentially in the hypodermis, whilst excluding expression from a range of other tissues.

There are a number of known stress-related phase separated organelles in *C. elegans*. To determine whether ABU-13 was localised to these, we crossed our tagged lines into stress granule and P-body marker strains, showing no colocalisation before and after the application of stress. This led us to propose that ABU-13 puncta may represent a novel form of phase separated compartment.

To understand the mechanics of these phase separation properties without the confounding influence of the cellular environment, we moved into an *in vitro* system to determine whether ABU-13 was capable of driving phase transitions itself. Applying a range of buffer conditions to these purified proteins demonstrated an ability to reversibly form these puncta *in vitro*, thus supporting our previous hypothesis that ABU-13 can act as a scaffold protein for the formation of these structures.

On the whole, we demonstrate that ABU-13 is involved in the formation of a dynamically responsive, novel stress-related phase separated compartment in *C. elegans*.

## 3.1 Introduction

### *3.1.1 Phase separation*

Traditionally, cellular function is thought of as being compartmentalised within membrane-bound organelles[1]. This allows protein and nucleic acid-based molecules to be retained and regulated spatially, as the properties of the phospholipid bilayer largely restrict unregulated movement between the internal and external environments of the organelle. Instead, protein-based transporters are often relied upon to facilitate this movement - a process that is both energetically costly and time-consuming. As such, these organelles are often not rapidly dynamic. Even a rapid process such as the release of a neurotransmitter is reliant upon a cascade of processes. Phase separation, on the other hand, allows for rapid and responsive organelle dynamics, with organelle specificity deriving from the specificity of protein and nucleic acid interactions, rather than from the selective import and export of components.

This act of coalescence is dependent upon the intrinsic biophysical properties of the constituent components themselves - often acting in response to changes in the cellular environment directly, or as a result of posttranslational modification[2], for example. This removes some of the rate-limiting steps that membrane-bound organelles face when responding to stimuli. This is particularly true of liquid-like assemblies; these are often extremely transient, forming readily for the duration of their response before dissolving. Whilst many phase separated organelles undergo transitions into liquid-like states, not all do - some are capable of forming more solid-like structures known as hydrogels[3,4]. These are less transient than their liquid-like counterparts, with less free diffusion of components within them, and less exchange of molecules between them and the freely-diffusing surrounding environment.

These liquid and hydrogel properties do not exist in a discrete system, but are rather part of a spectrum, with different phase separated organelles exhibiting a host of different properties. It has been hypothesised that this transition ability, when left unchecked, can result in pathological fibrillar states leading to the onset of disease[5].

There are a few mechanisms through which the biogenesis of these compartments is stimulated. There is some evidence of RNA forming structural scaffolds onto which protein components can interact with and recruit further components[6–8]. The major route thought to drive this formation, however, is the intrinsic propensity of certain proteins, known as scaffold proteins, to coalesce[9]. Amongst scaffold proteins, there is an overrepresentation of prion-like proteins as their low complexity, intrinsically disordered domains facilitate such aggregatory behaviour[10,11]. However, this is not the only requirement for scaffold proteins - in order to promote the accumulation of many clients, this scaffold must have a degree of multivalency[9,12]. Some scaffolds achieve this through repetitive elements that promote interactions with one type of partner multiple times, whilst others interact with many different partners.

Interactions with RNA, in particular, are hallmarks of phase separated organelles for a number of reasons – chiefly, many of these structures play roles in the regulation of RNA[13] but, on top of this, RNA-species within these ribonucleoprotein complexes can help to maintain solubility and avoid aberrant transitions into fibrils[14]. Thus, the presence of both a prion-like domain and RNA interacting motifs within a protein is often indicative of a role for that protein in phase separation.

### *3.1.2 CRISPR/Cas9*

In recent years, the ability to edit endogenous loci has improved in ease and efficiency. The advent of CRISPR technology has aided this process phenomenally. First identified as part of the bacterial adaptive immune system, this system involves the formation of ribonucleoprotein complexes capable of targeting specific DNA sequences and inducing double stranded breaks 3-4 nucleotides upstream of a protospacer adjacent motif (PAM), where the endonuclease Cas9 binds[15].

In the bacterial immune system, nucleotide sequences from invading DNA and RNA species are incorporated endogenously into clustered regularly interspaced short palindrome repeats (CRISPR) within the bacterial genome. These foreign incorporated sequences can then be expressed as CRISPR RNA (crRNA) capable of interacting with the Cas9 protein and guiding it to and cleaving DNA and RNA species upon re-invasion. Thus, these guide RNA sequences effectively allow the activity of Cas9 to be reprogrammable for different genetic loci[16–19]. For genome editing, this has been exploited to induce double strand breaks at specific target sites. Unlike the bacterial immune system, however, the addition of a repair template can then be used to edit these specific-loci via homologous recombination.

In *C. elegans*, CRISPR technology continues to be optimised. Akin to the insertion of extrachromosomal array lines, these endogenous edits can be achieved via microinjection - however the efficiency differs between approaches. For small edits and deletions, Paix et al. (2015) designed a basic protocol using pre-annealed site-specific Cas9-crRNA complexes in conjunction with single stranded donor oligonucleotide (ssODN) repair templates. For larger insertions, however, they recommend using double stranded PCR templates flanked by double stranded homology arms[20].

Alternative methods for insertions have been developed that purportedly improve efficiency in comparison to this approach. One such method is that of nested CRISPR, developed by Vicencio et al. (2019), that aims to improve the efficiency of large insertions by splitting them up into two smaller inserts (Figure 3.1)[21]. Initially, the full-size insert is split into three fragments; a small N terminal fragment, a large central fragment and a small C terminal fragment. An ssODN coding for these two small fragments flanked by appropriate homology arms, purposefully excluding the large central fragment, is then used for the first insertion. A double stranded PCR repair template of the full-size insert is then used for the next insertion - unlike the previous protocol, however, this does not require the addition of homology arms as the N and C terminal fragments already inserted fulfil this role.

The efficiency differences observed when using single stranded versus double stranded repair templates led Dokshin et al. (2019) to develop a method using partially single stranded repair templates[22]. This involves single stranded overhangs corresponding to homology arms sequences flanking the main double stranded repair template. This improves the efficiency of using dsPCR templates alone.

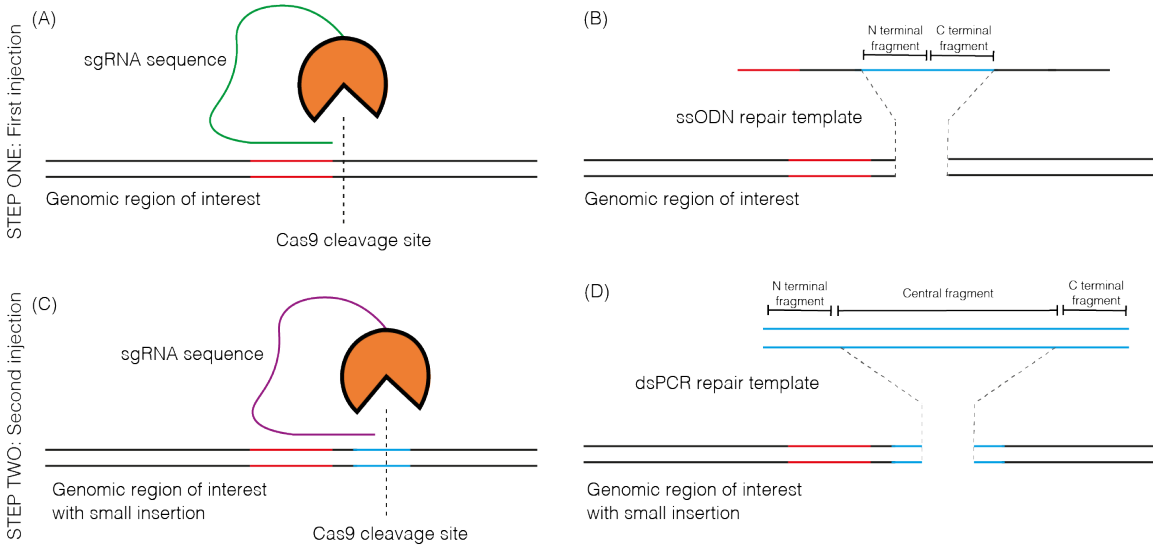



**Figure 3.1** A schematic outlining the multi-step process involved with Nested CRISPR

Whilst these methods rely on the injection of pre-prepared Cas9-crRNA RNP complexes, this can present some technical difficulties, such as a tendency for these complexes to aggregate and block the needle during microinjection. As such, other methods rely on the formation of these complexes *in vivo* using a plasmid-based system to express components such as the repair template, homology arms and guide RNA. SapTrap, as this method is known, eliminates the need for complex cloning using a single tube 'modular' cloning strategy, allowing for the interchange of various fluorescent tags, linkers and self-excising cassettes using the SapI enzyme (Figure 3.2) [23,24]. This enzyme results in a non-canonical digestion, with the resulting overhang found adjacent to the SapI recognition site, as opposed to the recognition and digestion sites being one and the same - as such, these overhangs can be made up of any combination of nucleotides. Exploiting this property of SapI allows components to be excised from PCR products and donor plasmids and inserted into a destination plasmid using a complementary combination of overhang sequences. This can then be injected into animals alongside a Cas9-expression vector or with the protein itself. Formation of the active RNP complex and the resulting edit then proceeds entirely *in vivo*. The inclusion of a self-excising cassette (SEC) allows for the positive selection of successfully edited animals - after which point, heat shock or hygromycin treatment can be used to excise this cassette, leaving just the insert of interest.

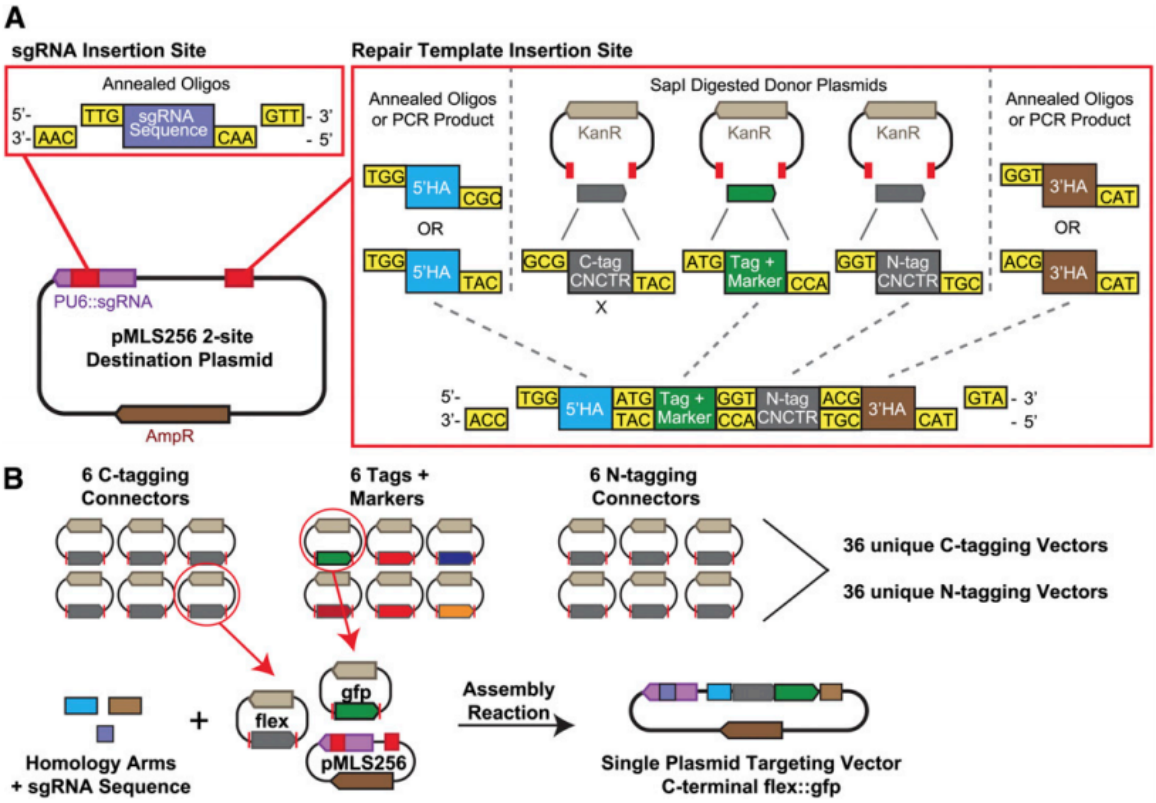



**Figure 3.2** Illustration of the modular nature of SapTrap, with exchangeable donor plasmids and overlapping SapI digestion overhangs (taken from *Schwartz and Jorgenson, 2016*)

In this chapter, we will develop a number of fluorescently labelled ABU-13 transgenic animals, optimising the CRISPR/Cas9 protocols outlined above to produce endogenously edited lines in addition to extrachromosomal array lines. Using these strains, we will then characterise the biophysical properties of ABU-13 *in vivo*, demonstrating that, not only can this protein form puncta through the animal, but that these puncta vary in their mobility, suggesting that they represent both liquid-like and solid-like structures. Further characterisation of these structures will demonstrate a broad expression pattern, with higher expression levels in the hypodermis, as well as a dynamic change in these puncta in response to pathogenic bacteria. Finally, we will characterise the properties of ABU-13 *in vitro*, demonstrating that it retains the ability to form puncta in a range of conditions, further supporting our hypothesis that ABU-13 is a phase separating scaffold protein.

## 3.2 Results

### *3.2.1 Construction of fluorescently tagged ABU-13 transgenic animals*

In order to determine the *in vivo* characteristics of ABU-13 and probe its potential prionogenicity, we produced transgenic lines expressing fluorescent proteins fused to ABU-13. Initially, we generated an extrachromosomal array line expressing tagRFP-fused ABU-13 under control of the *abu-13* promoter (Figure 3.3A) - we chose this fluorescent protein due to its monomeric nature and low propensity for aggregation. This is of particular importance in the context of prion-like proteins due to their inherent propensity for such aggregatory behaviour. Visualisation of these animals showed clear punctate structures throughout the body, with these puncta varying in size but showing little variation over time (Figure 3.3B). The presence of these puncta does not negatively influence the health of the animal, with no overt structural or developmental phenotypes observed, indicating that these structures are not having an obvious adverse effect on these nematodes, as might be expected of pathological aggregation.

One concern was that the distribution of ABU-13 in this system might be a result of overexpression. To circumvent the effects of overexpression, we sought to produce an endogenous insertion into the ABU-13 locus using CRISPR/Cas9 technology. Whilst endogenous insertions are possible in *C. elegans*, the efficiency of such techniques remains low and, as such, we exhausted a number of insertion strategies before achieving a successful edit (Table 3.1). Initially, we relied on a protein-based strategy, whereby a pre-incubated Cas9-crRNA complex is injected directly into the gonad of maturing animals alongside a double-stranded PCR template of the insertion coding sequence flanked by 35 nucleotide overhangs corresponding to the locus of interest. The efficiency of this technique proved to be low, with no successful edits, despite success being achieved with the co-CRISPR marker, with a single-stranded oligonucleotide (ssODN) corresponding to the *dpy-10* gene successfully editing the desired region and resulting in the appropriate roller phenotype for that small edit. This demonstrated that, whilst this method could produce small edits with a high efficiency, the efficiency for large insertions was significantly lower.

We next moved to a different strategy, using a plasmid-based system alongside the Cas9 protein, known as SapTrap. Unlike the previous technique, the guide RNA and the repair template are cloned into an insertion plasmid, as opposed to the former being assembled into a complex with Cas9 prior to injection and the latter being injected alongside it. This technique purportedly results in high efficiency insertions in a high throughput fashion, with a modular 'one-tube' cloning system allowing for a simplistic interchange of fusion proteins, linkers and insertion sites within the plasmid by exploiting the non-conventional endonuclease activity of the SapI enzyme. This, however, proved not to be the case in our hands, with no successful cloning events and thus we had no opportunity to test the efficiency of this system at endogenous insertions.

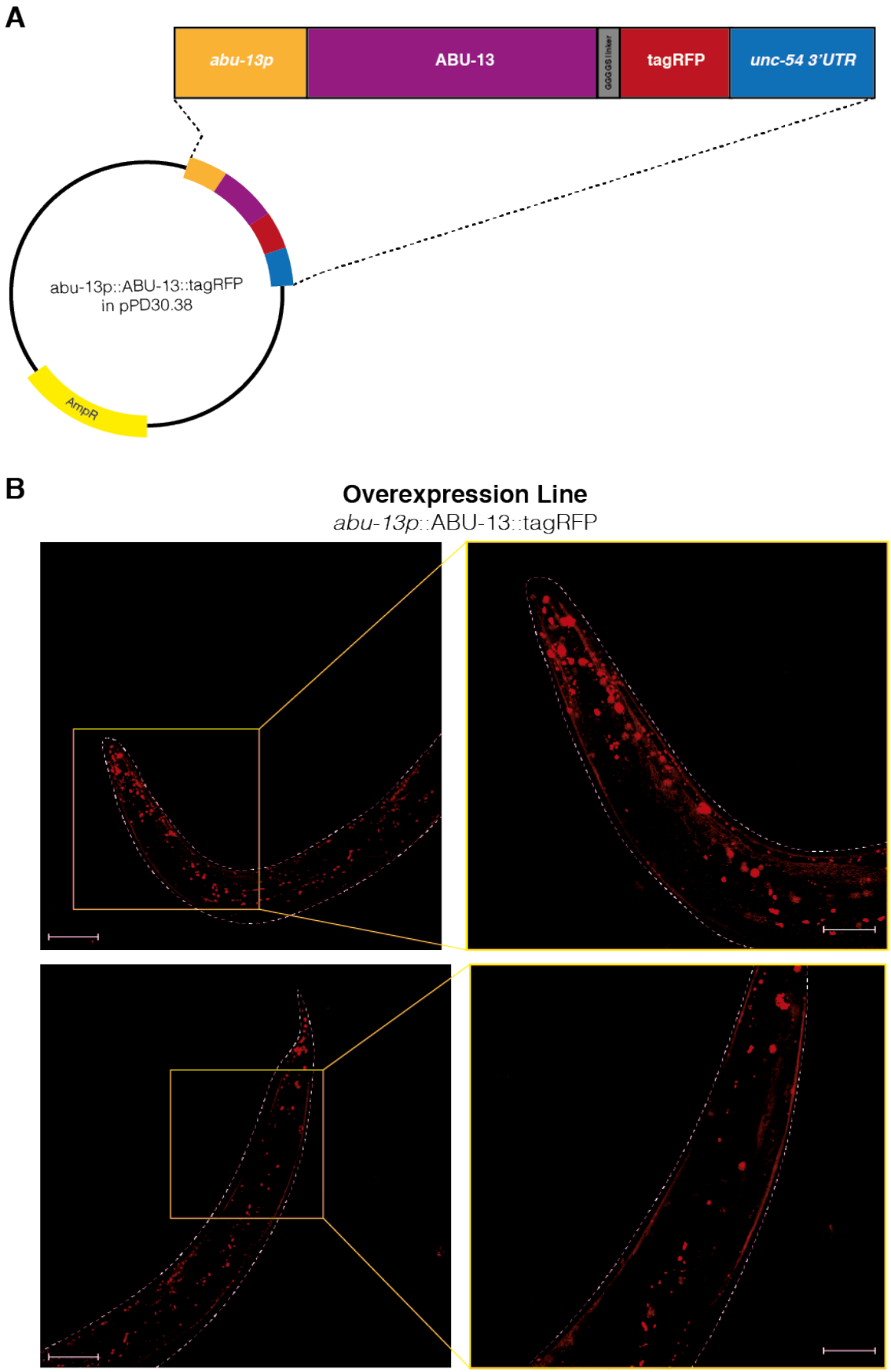



**Figure 3.3** Extrachromosomal array lines overexpressing tagRFP-fused ABU-13 show puncta throughout the animal (A) Schematic depicting the expression vector used to drive ABU-13::tagRFP overexpression (B) Confocal micrographs of ABU-13::tagRFP overexpression animals (*Scale bars, 100μm; 25μm*)

We next attempted a nested CRISPR approach to mitigate the highly significant decline in efficiency observed as the insertion size increases (Figure 3.1). Indeed, we had some success with this approach. In particular, we saw a high insertion efficiency of the initial ssODN repair template sequence, which primes the locus for subsequent full insertion. We did then successfully edit the locus in this second insertion step, albeit with low efficiency. This was possibly due to the large insertion size of this fragment compared to the first. The insertion produced, however, contained many errors that rendered the fluorescent protein non-functional. Many of these errors clustered around the site of Cas9 cleavage, and illustrated the importance of editing the PAM site within the repair template to minimise the risk of recutting and subsequent unnecessary, error-prone recombination events. Although we did attempt to repair these edits, our efforts proved unsuccessful, with more errors arising within the region of interest the more we edited. Despite further injections, no successful edits were achieved, demonstrating that, whilst this approach was indeed more successful than the initial strategy, the efficiency was still low.

| Method | Reference | Details |
|---|---|---|
| Small ssODN or double-stranded PCR repair template | Paix et al. (2015) | dsPCR template approach unsuccess for the endogenous insertion of tagRFP<br><br>ssODN repair successful for PrLD deletion |
| Nested CRISPR | Vicencio et al. (2019) | Initial insertion successful with a high efficiency<br><br>Second with dsPCR repair template inefficient, with any successful inserts containing numerous errors |
| SapTrap | Schwartz & Jorgensen (2016)<br>Dickinson et al. (2018) | No success with initial cloning step |
| Partially single-stranded repair template | Dokshin et al. (2019) | Full insert success, however small errors around the Cas9 cleavage site, repaired with a second ssODN edit<br><br>Adapted protocol to use a co-CRISPR strategy with *dpy-10* rather than a co-injection marker approach |

**Table 3.1** A summary of the CRISPR approaches used for endogenous locus editing

We next moved on to a more recently published methodology. This strategy was similar to those above, but instead of using double stranded PCR templates with double stranded homology arms, it used a double stranded repair template with single-stranded homology arms. Similarly to the other approaches taken, we initially saw little success - this led us to re-evaluate our strategy as a whole - specifically, the fluorescent protein tag we were using.

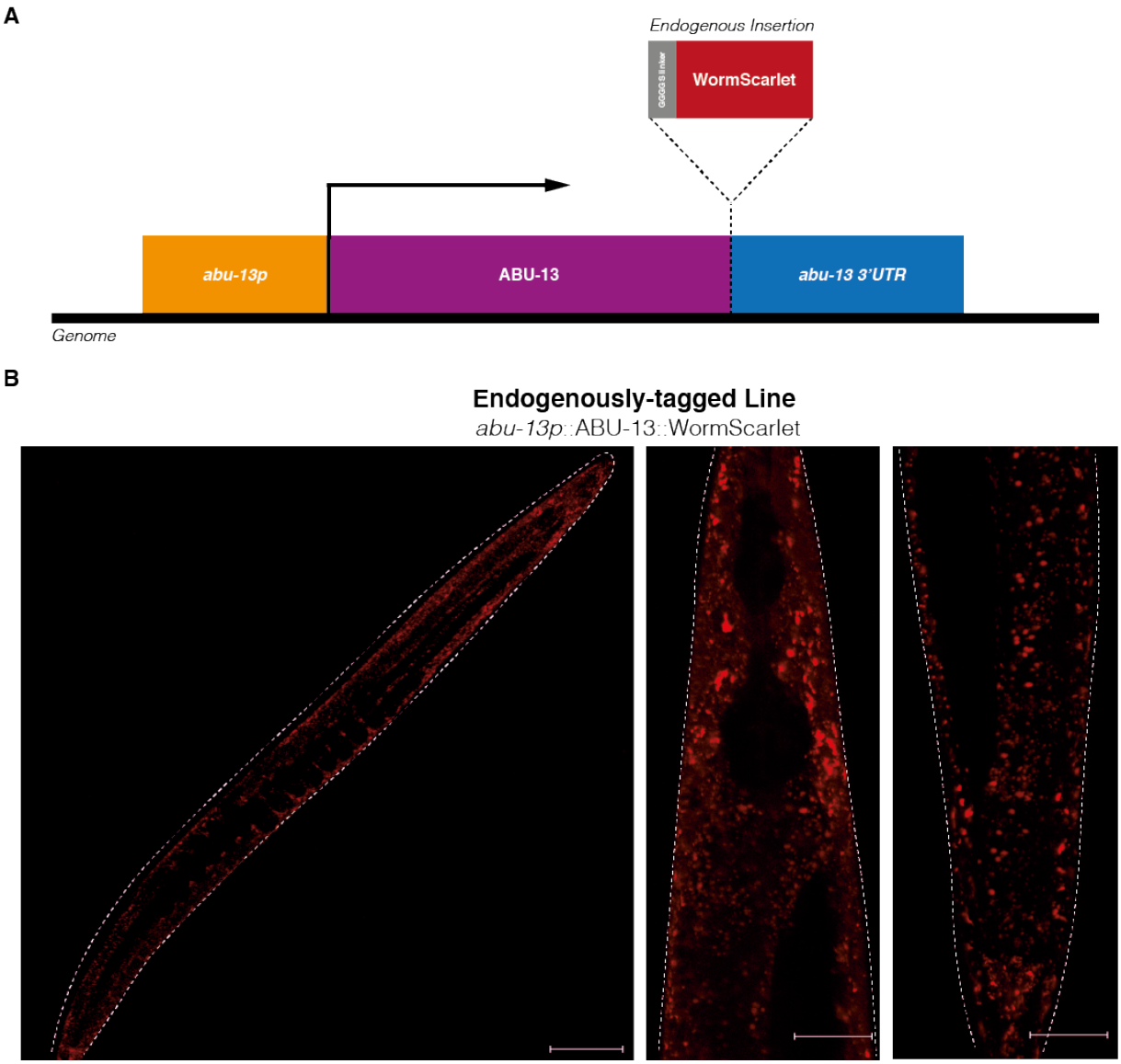



**Figure 3.4** CRISPR-edited endogenously-tagged ABU-13::WormScarlet lines show puncta throughout the animal (A) Schematic depicting the CRISPR-mediated insertion used for this ABU-13::WormScarlet endogenously-edited line (B) Confocal micrographs of ABU-13::WormScarlet endogenously-tagged animals (*Scale bars, 100μm; 25μm*)

As with the extrachromosomal array line, we initially sought to introduce tagRFP endogenously. However, the coding sequence for this protein is ~ 900 base pairs long. As previously mentioned, the larger the insertion, the more difficult it becomes to insert it. As such, we moved to using a WormScarlet tag instead that, as well as being much smaller (~700bp instead), also has the advantage of being much brighter than tagRFP - which is particularly useful if ABU-13 is expressed at low levels. Indeed, after switching to this new fluorescent protein, we saw a higher success rate than with previous approaches, although the error rate remained high. To counteract this, we performed an additional editing step to repair the errors - a step that ultimately proved successful. Thus, with a WormScarlet endogenously tagged ABU-13 line, we were able to proceed and determine whether similar puncta formation was observed in these animals to that which we had seen in the overexpression line.

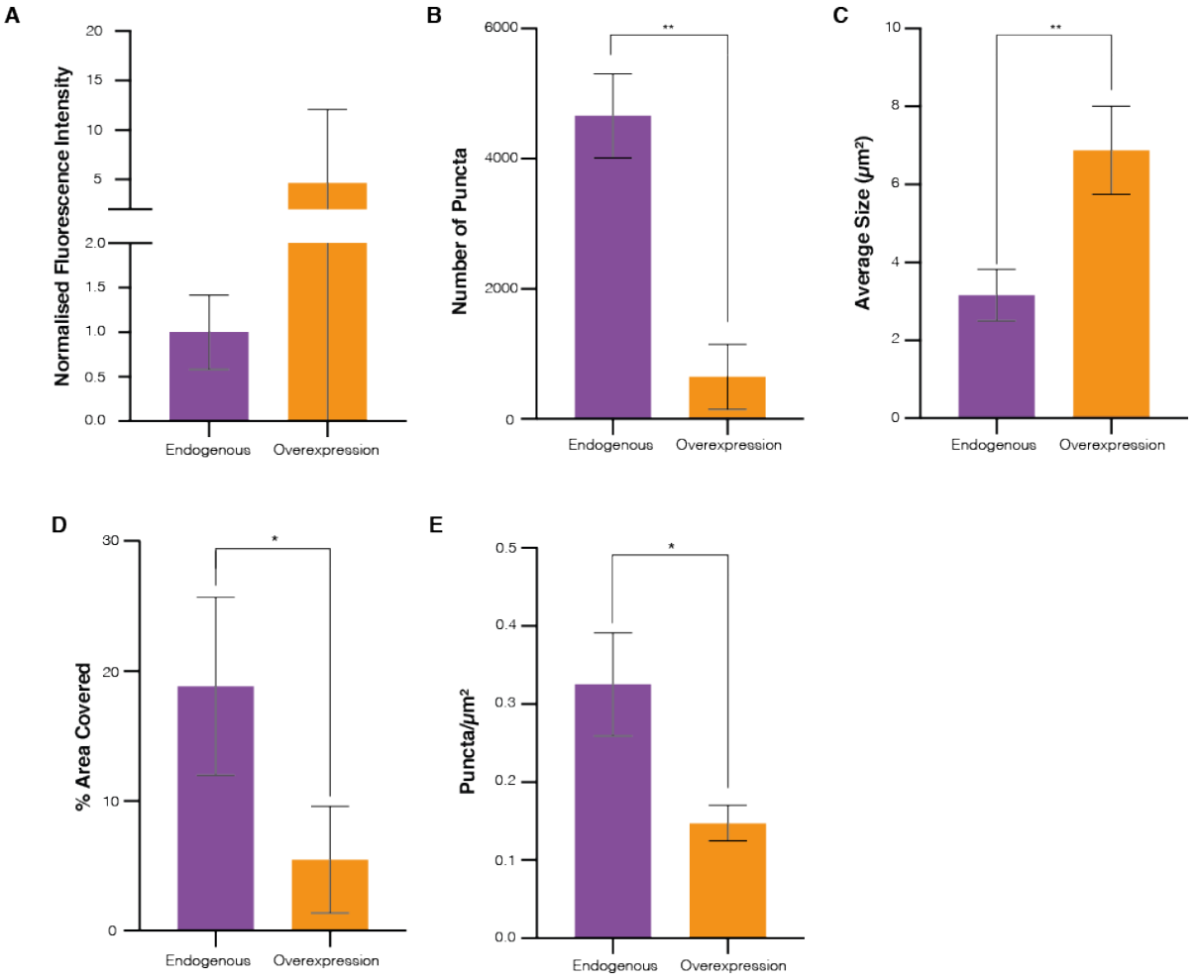



**Figure 3.5** ABU-13 puncta in overexpression lines are larger and less numerous than in the endogenously tagged lines (A-E) Comparisons between overexpression and endogenous strain for (A) Normalised fluorescence intensity value (Unpaired t test, p =0.4365; ns); (B) Number of puncta (Unpaired t test, p = 0.0010; **); (C) Average size (*Unpaired t test, p = 0.0079;* **); (D) % area covered (*Unpaired t test, p =0.0444;* *); (E) Puncta/μm$^2$ (*Unpaired t test, p = 0.0118;* *)

Whilst these animals also contained punctate structures, these puncta were different from those observed in the ABU-13::tagRFP extrachromosomal array line (Figure 3.4A,B). The puncta in animals expressing endogenously-tagged ABU-13 were expressed more broadly than in the overexpression line - not necessarily a surprise, as the expression of the extrachromosomal array can vary across generations, and the overexpression construct may lack endogenous expression elements. In addition to this, the puncta themselves were also different in both number and size - the endogenous line contained significantly more puncta than the overexpression line, with these puncta generally being smaller in size but more plentiful in number (Figure 3.5A-E). This demonstrates that, whilst the overexpression does not itself drive puncta formation, it does alter the characteristics of these puncta. Overall, the presence of a prion-like domain and a potential RNA binding domain alongside the inclusion of this protein into punctate assemblies is reminiscent of proteins that form phase separated organelles.

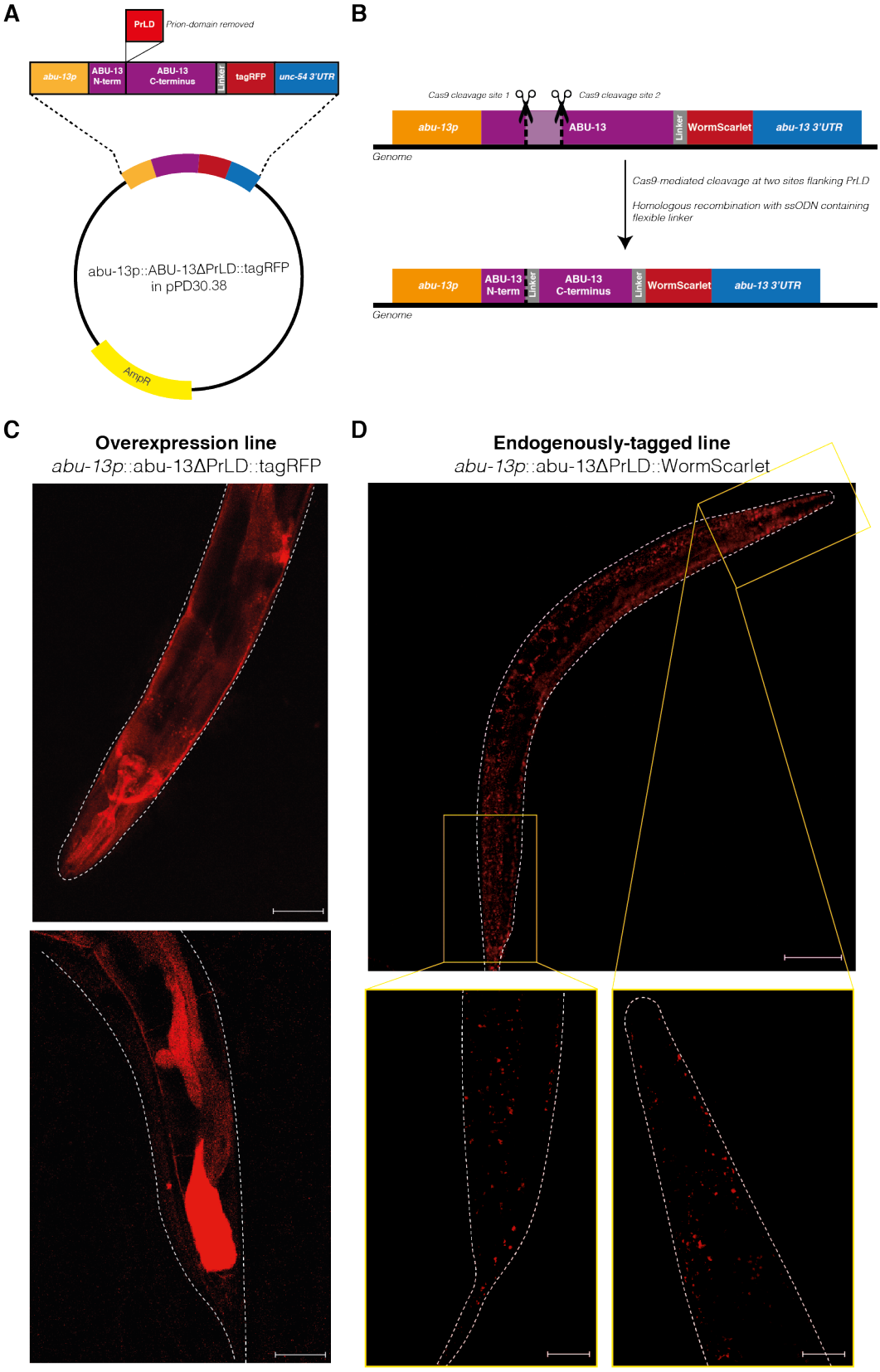



**Figure 3.6** ABU-13ΔPrLD mutants show different distributions in the overexpression versus endogenously-tagged line (A) Schematic depicting the modified expression vector used to generate the extrachromosomal array overexpression line (B) Experimental outline of the CRISPR/Cas9 edited ABU-13 locus to delete the prion domain from the genome (C-D) Confocal micrographs depicting (C) ABU-13ΔPrLD::tagRFP overexpression animals (*Scale bar, 50μm*); and (D) ABU-13ΔPrLD::WormScarlet endogenously tagged animals (*Scale bars, 100μm; 25μm*)

### *3.2.2 Removal of the prion-like domain abolishes the ability of ABU-13 to form puncta in the overexpression line but not the endogenously-tagged line*

We next investigated the influence that the prion-like domain has on the formation of these punctate structures by removing the domain from both the overexpression and endogenously-tagged lines (Figures 3.6A-B). For the former, we produced another extrachromosomal array line, the injection plasmid used for which lacked the PrLD. For the latter, however, we used a CRISPR/Cas9 strategy outlined in Paix et al. (2015); this involved using two sgRNA sequences to target sites either side of the PrLD for cleavage. Once cleaved, this region was repaired using an ssODN template containing a GGGGS linker flanked by homology arms corresponding to the sequences adjacent to the removed prion-domain sequence – but not the PrLD itself, thus resulting in ABU-13 with the prion-domain removed and the N and C termini either side of it connected via a flexible linker. This is particularly useful when assessing the involvement of a protein in the formation of phase separated structures as prion domain-containing proteins often act as scaffold proteins for these transitions and, as such, if ABU-13 were indeed a scaffold protein, then we would expect the removal of this domain to prevent such transitions.

To our surprise, the removal of the PrLD resulted in different effects in these two strains; the ability to form puncta was entirely abolished in the overexpression animals but was unaffected in the endogenously tagged animals. In the overexpression line, ABU-13 was diffuse across a range of tissue types, including the hypodermis, pharynx and intestine (Figure 3.6C). The endogenously tagged strain, however, showed a very similar punctate structure to that observed in the full-length variant (Figure 3.6D). This difference may result as a consequence of a linker placed between the N and C terminal regions of ABU-13 in the endogenous line that is not present in the overexpression line (Figures 3.6A-B). Without this linker, the functionality of the regions flanking the prion-like domain may be compromised. This potentially provides an interesting insight into the biogenesis of these puncta, as it may suggests that, whilst the prion-like domain is not necessary for puncta formation, either or both of the N and C termini play a role in either the formation or the localisation of ABU-13 to these puncta.

### *3.2.3 Increased mobility of ABU-13 puncta in the endogenously-tagged compared to the overexpression strain*

Due to the large observable differences between the puncta found within the endogenous and overexpression lines, we next sought to identify whether these puncta displayed different biophysical properties. In particular, we used fluorescence recovery after photobleaching (FRAP) to determine the mobility of ABU-13 within these puncta. In ABU-13::WormScarlet animals, we observed a highly

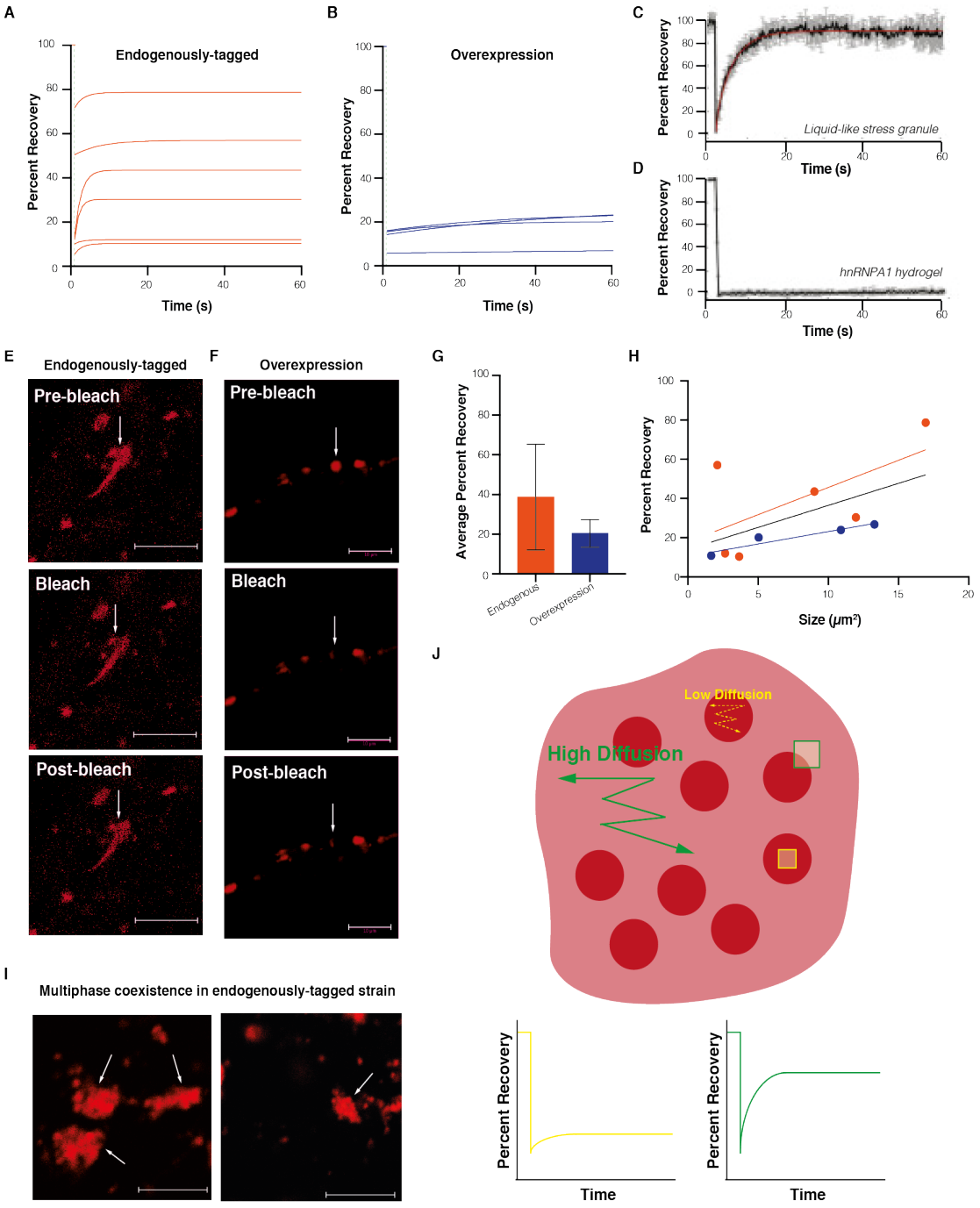



**Figure 3.7** ABU-13 puncta have highly variable mobility (A) FRAP recovery curves for individual puncta in endogenously-tagged ABU-13::WormScarlet animals (green dashed line = photobleaching event); (B) FRAP recovery curves for individual puncta in ABU-13::tagRFP overexpression animals (green dashed line = photobleaching event); (C) Representative recovery plot of liquid-like stress granules (adapted from Molliex et al., (2015)); (D) Representative recovery plot of hnRNPA1 hydrogels (adapted from Molliex et al., (2015)); (E) Confocal micrographs depicting the fluorescence recovery of an WormScarlet-tagged ABU-13 puncta from endogenously tagged animals (*Scale bar, 10μm*); (F) Confocal micrographs depicting the fluorescence recovery of a tagRFP-tagged ABU-13 puncta from overexpression animals (*Scale bar, 10μm*); (G) Average FRAP recovery comparison between endogenously tagged and overexpression animals (*Unpaired t test, p = 0.2247; ns*); (H) Correlation between puncta size and percentage FRAP recovery between endogenously tagged animals (*orange; r = 0.6119; p=0.1967*), overexpression animals (blue; r =0.9519; p=0.0481) and all puncta combined (*black; r=0.5446; p =0.1036*); (I) Representative confocal micrographs depicting ABU-13 puncta clustering together in a multiphase structure (*Scale bar, 10μm*)); (H) Schematic proposing multiphase coexistence in ABU-13 puncta, with recovery dynamics illustrated below

variable rate of recovery between puncta, varying between puncta with mobile fractions of just 10.34%, and others with mobile fractions of 78.69% (Figure 3.7A,E). On the other hand, the puncta found within ABU-13::tagRFP overexpression animals showed consistently low mobility, suggesting that they are

largely immobile (Figure 3.7 B,F). Higher recovery is indicative of more liquid-like structures, with similar mobility observed in dynamic phase-separated compartments such as stress granules (Figure 3.7C). Lower recovery, however, is associated with low mobility structures such as hydrogels (Figure 3.7D). This suggests that the puncta within the endogenously-tagged strain exhibit a mixture of liquid- and hydrogel-like properties, whilst all of those in the overexpression strain adopt more hydrogel-like structures. The average mobility of the puncta between these strains is not significantly different (Figure 3.7G), however, this is more a reflection of the high variation of mobility within the endogenously tagged animals - a larger sample size would likely clarify this difference.

There was, however, a significant relationship between puncta size and mobility with larger puncta exhibiting a greater mobility (Figure 3.7H). This was contrary to expectations, as larger puncta are traditionally associated with uncontrolled solid-like aggregation and fibrillisation – a hallmark of pathology. Further imaging revealed that many of these puncta cluster together, forming small subpopulations of assemblies within one region (Figure 3.7I). The individual puncta within these clusterings appear less mobile, whilst the FRAP of the general cluster shows more mobility. This suggests that these clusters are made up of a number of low mobility 'core' structures surrounded by a more mobile, diffuse 'shell'. As such, the 'core' regions of these puncta are reminiscent of the structures observed in the overexpression system. It is possible that these puncta represent hydrogel formations - such phase transition states show more solid-like properties than traditional liquid-like assemblies whilst retaining physiological function, as opposed to pathological dysfunction.

### *3.2.4 PA14 exposure alters ABU-13 puncta dynamics, whilst tunicamycin treatment and heat shock do not*

To understand the relationship between these ABU-13 puncta and the phenotypic effects of *abu-13* knockout, we subjected endogenously tagged ABU-13::WormScarlet animals to a number of stress related stimuli. Initially, we treated animals with 25ng/mL of tunicamycin for 4 hours before imaging. Compared to DMSO-treated control animals, there were no observable differences in puncta properties. with the same number, size, and coverage of puncta observed between these conditions (Figure 3.8A-G). This suggests that the role of ABU-13 in ER stress responses, as identified in *abu-13* knockout animals upon tunicamycin treatment, does not involve gross changes in puncta dynamics, such as increased formation or dissolution, or changes in component recruitment to these puncta.

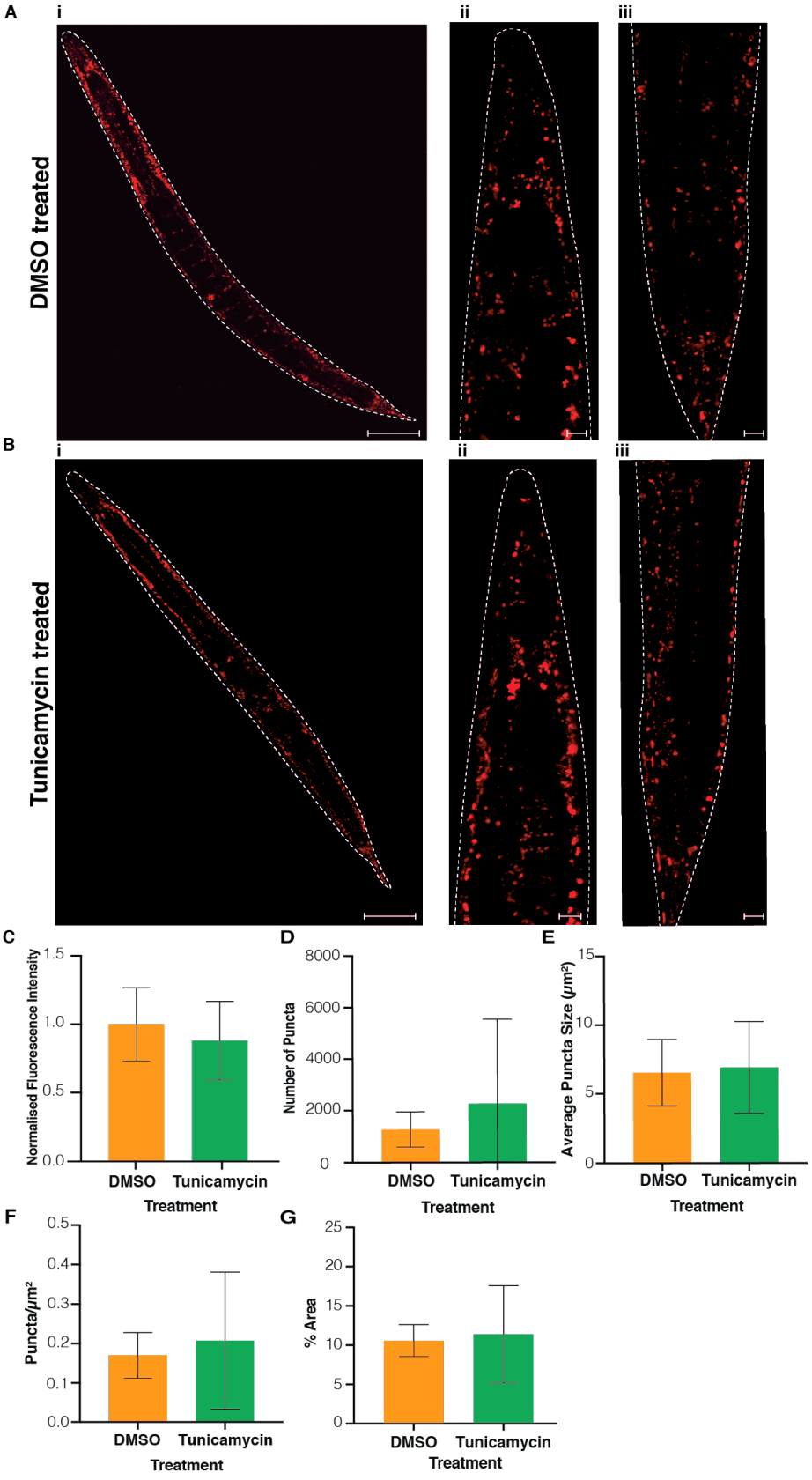



**Figure 3.8** ABU-13 puncta do not change upon tunicamycin treatment (A) Confocal micrographs of endogenously tagged ABU-13::WormScarlet animals treated with DMSO (*Scale bars, 100µm; 25µm*); (B) Confocal micrographs of endogenously tagged ABU-13::WormScarlet animals treated with 25ng/µL tunicamycin (*Scale bars, 100µm; 25µm*); (C-G) Comparisons between DMSO and tunicamycin treated animals for (C) Normalised fluorescence intensity value (*Unpaired t test, p =0.5145; ns*); (D) Number of puncta (*Unpaired t test, p = 0.5210; ns*); (E) Average puncta size (*Unpaired t test, p = 0.8517; ns*); (F) Puncta/µm$^2$ (*Unpaired t test, p = 0.6623; ns*); (G) % area covered (*Unpaired t test, p =0.7875; ns*)

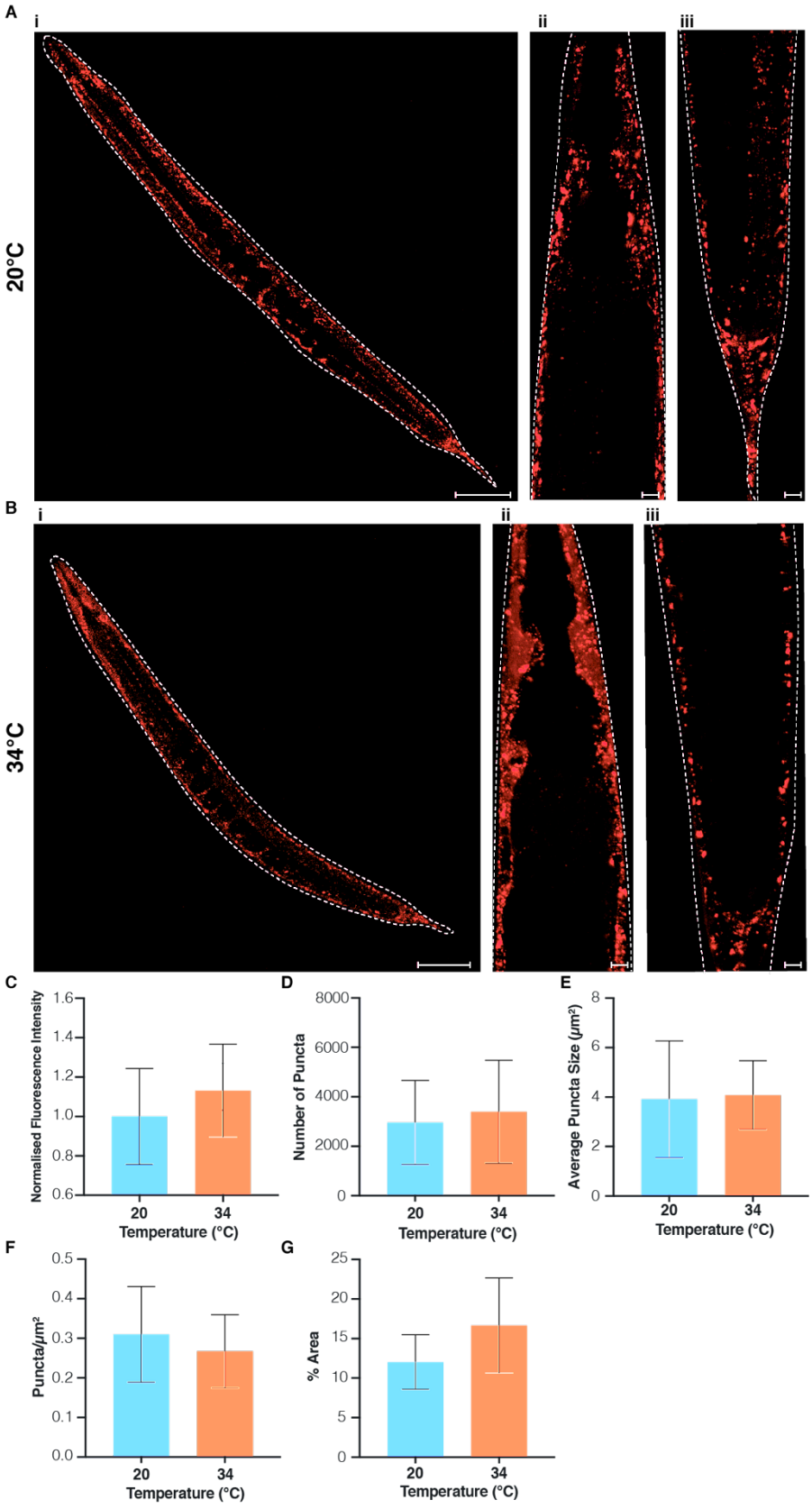



**Figure 3.9** ABU-13 puncta do not change upon heat shock (A) Confocal micrographs of endogenously tagged ABU-13::WormScarlet animals grown at 20°C (*Scale bars, 100µm; 25µm*); (B) Confocal micrographs of endogenously tagged ABU-13::WormScarlet animals subjected to a 6 hour 34°C heat shock (*Scale bars, 100µm; 25µm*); (C-G) Comparisons between 20°C and 34°C treated animals for (C) Normalised fluorescence intensity value (*Unpaired t test, p =0.4699; ns*); (D) Number of puncta (*Unpaired t test, p = 0.7575; ns*); (E) Average puncta size (*Unpaired t test, p = 0.9072; ns*); (F) Puncta/µm$^2$ (*Unpaired t test, p = 0.5951; ns*); (G) % area covered (*Unpaired t test, p =0.2318; ns*)

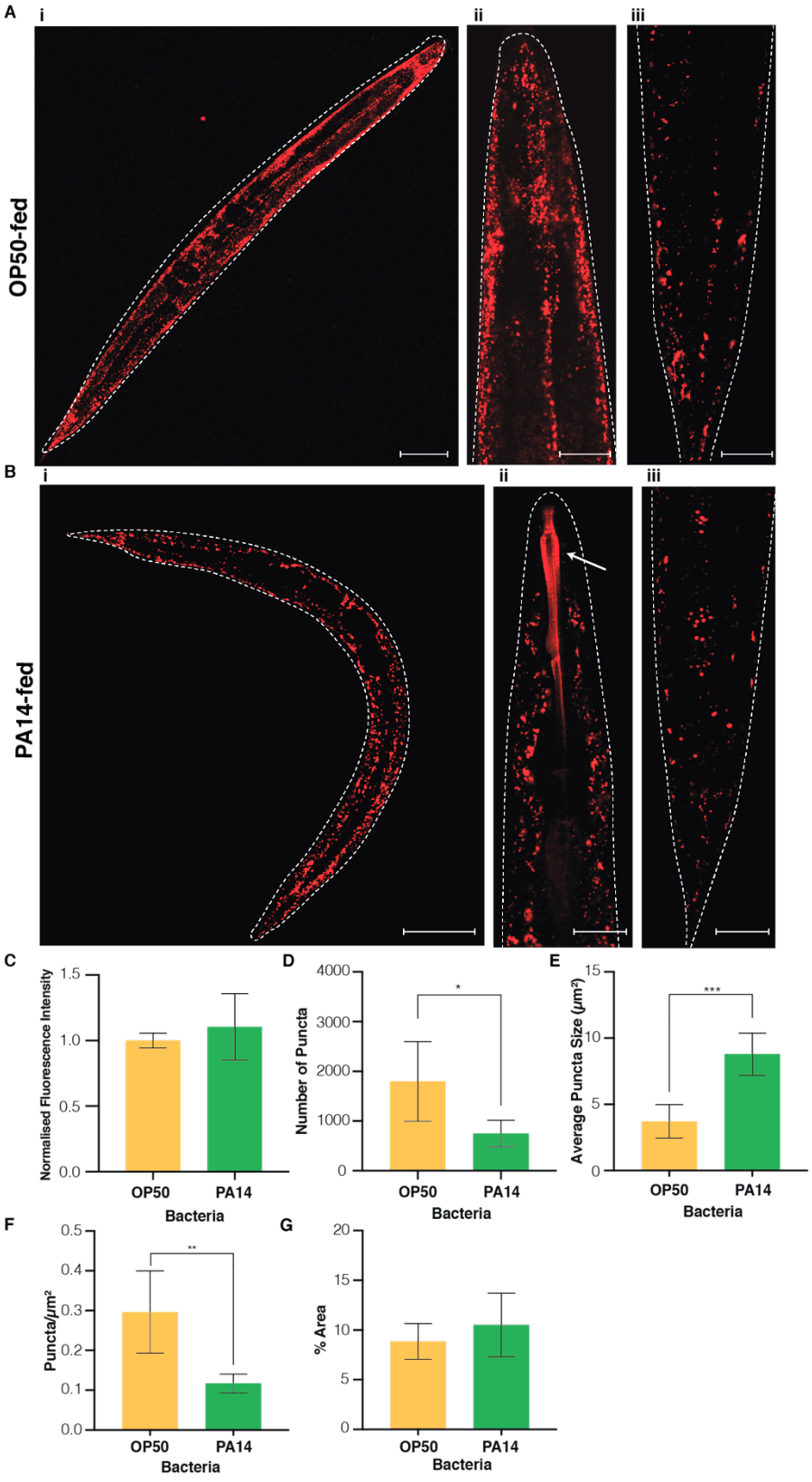



**Figure 3.10** ABU-13 puncta change following exposure to *Pseudomonas aeruginosa* (PA14) (A) Confocal micrographs of endogenously tagged ABU-13::WormScarlet animals exposed to OP50 (*Scale bars, 100µm; 25µm*); (B) Confocal micrographs of endogenously tagged ABU-13::WormScarlet animals exposed to PA14. The arrow indicates pharyngeal expression (*Scale bars, 100µm; 25µm*); (C-G) Comparisons between animals exposed to OP50 and PA14 for (C) Normalised fluorescence intensity value (*Unpaired t test, p =0.3911; ns*); (D) Number of puncta (*Unpaired t test, p = 0.0240;* *); (E) Average puncta size (*Unpaired t test, p = 0.0005;* ***); (F) Puncta/µm$^2$ (*Unpaired t test, p = 0.0053;* **); (G) % area covered (*Unpaired t test, p =0.3420; ns*)

Although no phenotypic effect was identified for knockout animals following heat stress, it is possible that there is redundancy compensating for loss of ABU-13 in this context, especially considering the unknown roles of many of the other ABU proteins. As such, we subjected ABU-13 endogenously tagged animals to a 34°C heat shock for 6 hours before imaging (Figure 3.9A-B). Similar to those treated with tunicamycin, we saw no gross changes to the ABU-13 puncta (Figure 3.9C-G), suggesting that there are no changes in ABU-13 dynamics upon heat stress. Paired with the lack of phenotypic readout for heat shock responses, this suggests that ABU-13 does not play a role in cellular responses to heat shock.

Finally, we subjected animals to PA14 for 6 hours to identify any changes in ABU-13 dynamics upon exposure to a pathogenic bacteria. Compared to animals treated with the non-pathogenic *E. coli* strain, OP50, there were significant differences in puncta dynamics (Figure 3.10A-B). In animals treated with PA14, there was a dramatic reduction in puncta number, with a concurrent increase in the size of puncta (Figure 3.10C-G). This indicates that ABU-13 puncta are either fusing, or that some are being dissolved whilst others increase their recruitment allowing them to increase in size. In addition to this, we observed a significant increase in the diffuse expression of ABU-13 in the pharynx, potentially in the structural marginal cells (Figure 3.10B). This is interesting for a number of reasons: firstly, this expression is induced at the primary interface between the external environment and the animal, suggesting that the induction is driven as a direct consequence of pathogenic attack in the pharynx. Secondly, unlike in other regions, there is no obvious ABU-13 puncta formation - this potentially points towards different roles of ABU-13 in the pharynx compared to other tissues.

On the whole, we have identified large structural changes to ABU-13 puncta in response to pathogenic insult but not in response to tunicamycin treatment. This suggests that these two stress-response pathways may depend upon ABU-13 via different mechanisms.

### *3.2.5 ABU-13 is expressed in tissues including the hypodermis, seam cells & amphid sheath glia*

To predict the expression pattern of *abu-13*, we utilised an online tissue expression prediction algorithm that assigned specific prediction scores across all tissue types within the animal. This algorithm, developed by Kaletsky et al. (2008), combines datasets from a wide range of microarray and RNA-seq experiments to produce comprehensive tissue expression prediction scores. For ABU-13, the highest scoring tissue types in this output include many epidermal-derived tissues such as the hypodermis – in particular the hyp-7 syncytium, arcade cells, seam cells and the amphid sheath cells (Figure 3.11A-B). Due to the punctate nature of ABU-13 *in vivo*, it is difficult to determine tissue expression using the ABU-13::tagRFP fusion protein. To aid this process, we therefore constructed a transcriptional reporter line expressing tagRFP under the *abu-13p* promoter (Figure 3.11C). This reporter line appeared to support some of these tissue predictions, with tagRFP visible within seam cells, the hypodermis, in the amphid sheath cell and within the rectal epithelium. It is important to note, however, that the *abu-13* coding region contains large intronic sequences - this is often indicative of regulatory elements within

introns that influence expression of the gene itself. Thus, without these regulatory introns, the full endogenous expression pattern of *abu-13* may not be fully realised using this reporter strain.

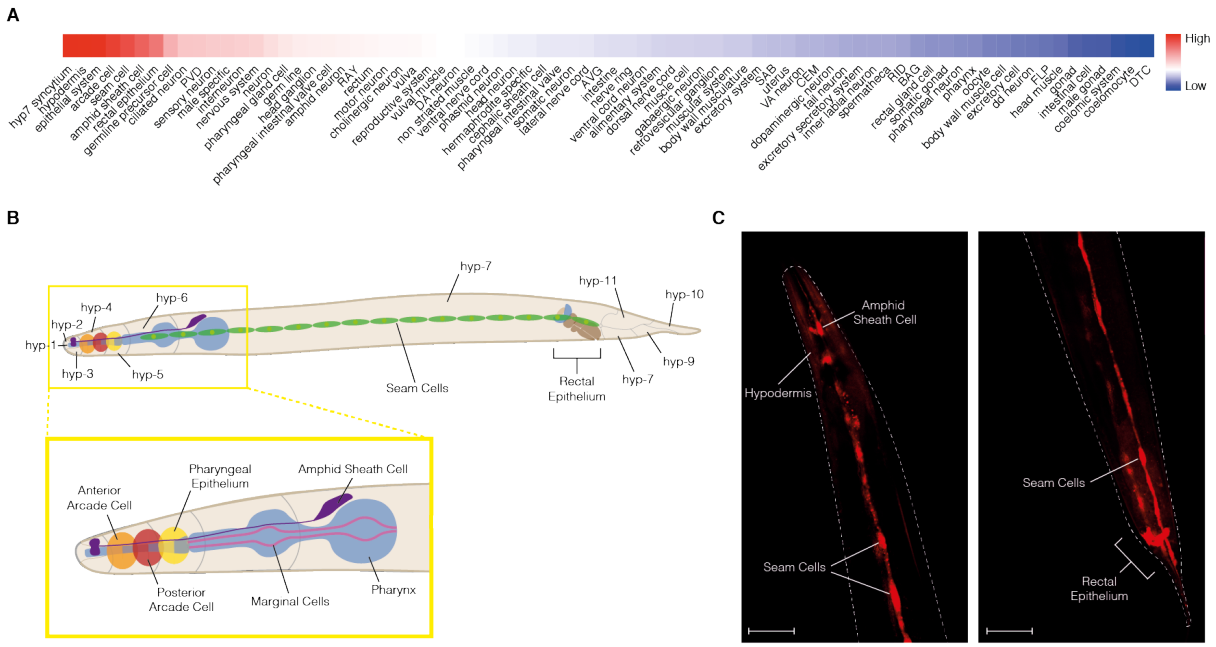



**Figure 3.11** ABU-13 is expressed widely in epidermally-derived tissues (A) A heat map of ABU-13 expression prediction scores; (B) A schematic highlighting tissues of interest within the adult hermaphrodite; (C) Confocal micrographs of *abu-13p*::tagRFP transcriptional reporter animals, highlighting expression in a number of tissues including the hypodermis and seam cells (*Scale bar = 50μm*)

To further determine the expression pattern of ABU-13, we initially crossed the fluorescently tagged ABU-13 overexpression animals into a number of strains expressing fluorescent markers for specific tissues, to identify any colocalization between these markers. As ABU-13 puncta appear to be broadly visible across the animal, we performed a number of crosses, including with markers for tissue types in which ABU-13 expression was not predicted computationally.

We first sought to determine whether ABU-13 puncta are present in neuronal cell bodies (Figure 3.12A) by crossing ABU-13::tagRFP animals into an *unc-119p*::GFP neuronal marker strain. Whilst the majority of puncta did not show any colocalization with *unc-119*p::GFP, there were some regions of overlap. This is particularly the case in the head region, where there is a high density of neuronal cell bodies. As a result, it is difficult to fully spatially resolve the colocalization between these two fluorescent proteins in this region. However, there does appear to be some proximal localisation observed around the amphid neurons, with several puncta surrounding this structure – this may represent expression of ABU-13 in the AMsh glia that support the neuron, in keeping with our previous computational predictions. In addition to this, there is potentially some colocalization along the ventral cord.

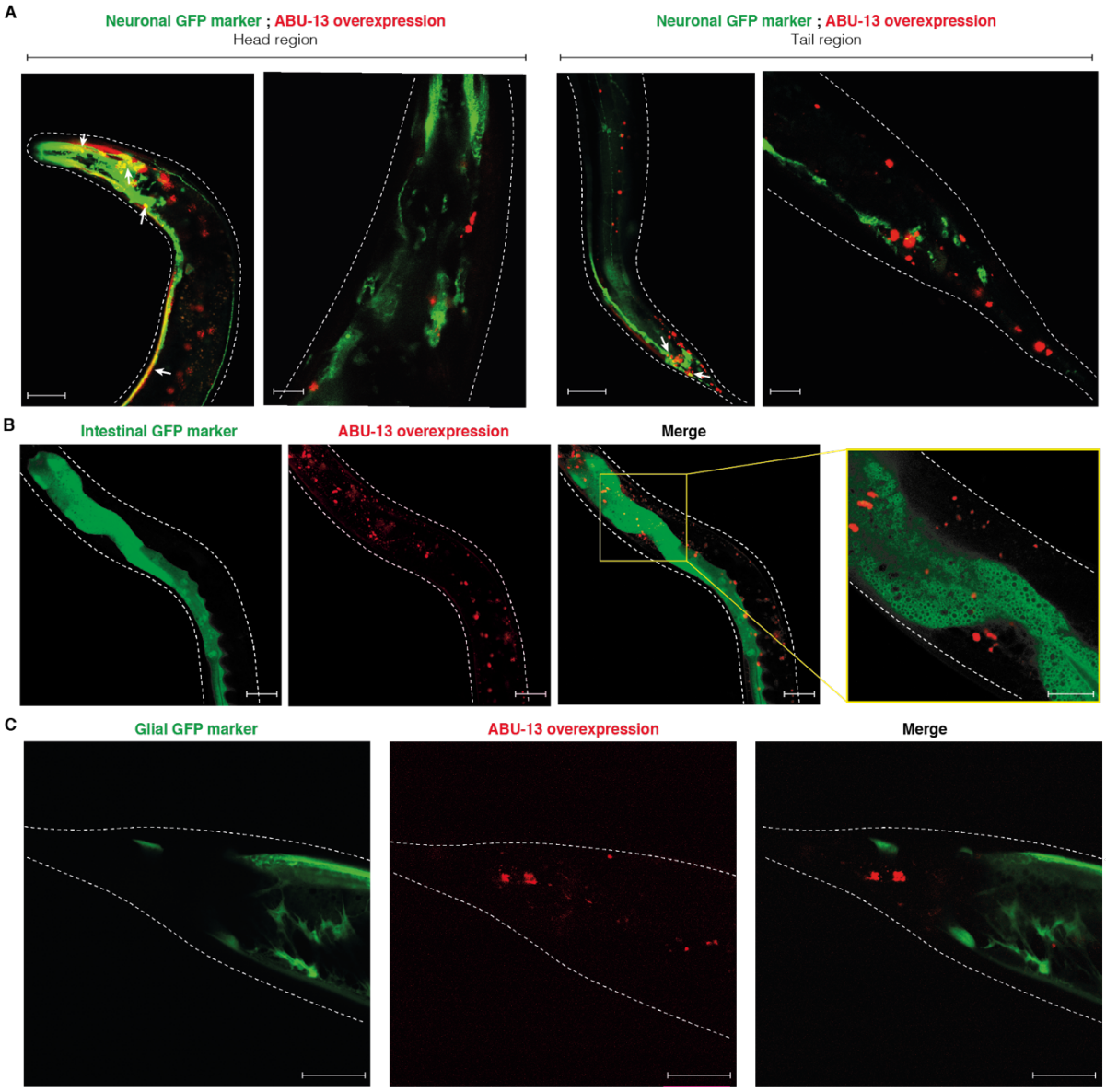



**Figure 3.12** ABU-13::tagRFP colocalises with amphid neurons in overexpression line (A) Confocal micrographs of ABU-13::tagRFP overexpression animals crossed into (A) a neuronal marker strain (*unc-119*p::GFP), *Scale bars = 50μm*; (B) an intestinal marker strain (*ges-1p*::GFP), *Scale bars = 50μm, 25μm*; (C) a glial marker strain (*pitx-1*::GFP), *Scale bars = 50μm, 25μm*

We next crossed the ABU-13::tagRFP overexpression animals into *ges-1p*::GFP animals enabled us to identify any intestinal expression of ABU-13 (Figure 3.12B). Although there does appear to be a small degree of overlap, this is not consistent among animals and it is difficult to rule out large puncta from other planes bleeding through into the imaging plane of interest, giving a false impression of intestinal localisation. One point to note is the discrepancy between the prion-domain mutant overexpression animals and the transcriptional reporter line, the latter of which shows strong ABU-13 expression in the anterior intestine whilst the former shows none. As neither of these lines truly recapitulate the endogenous expression of ABU-13, with both strains lacking some if not all intronic sequences that regulate *abu-13*, repeating this imaging in the endogenously-tagged animals will likely be informative

To investigate the expression of ABU-13 within glia, we constructed transgenic animals expressing GFP under the *itx-1* promoter (Figure 3.12C). We found that, in this strain, posterior glia were well labelled

and anterior glial less so - this is likely a consequence of poor penetrance, and further work is needed to improve this expression. In the tail region, however, where glial cells were strongly labelled, we observed no colocalization between the ABU-13::tagRFP puncta and glia. Unfortunately, this analysis did not allow us to determine whether ABU-13 is indeed found within AMsh glia.

It is important to recognise, however, that this imaging was performed in tagRFP extrachromosomal array lines, and thus expression of ABU-13 may be more variable and less complete than in the endogenously-tagged ABU-13::WormScarlet animals. As such, we have repeated some of these crosses and imaging in the endogenously-tagged strains. However, due to time constraints, we were unable to image the ABU-13::WormScarlet animals in all of these tissue marker strains.

We initially repeated this imaging with ABU-13::WormScarlet animals crossed into the *unc-119p::*GFP neuronal marker strain (Figure 3.13A). As with the overexpression line, we did see colocalization between some neuronal structures and ABU-13::WormScarlet. This is notably the case in the head region where a substantial colocalization is identified with the amphid neurons. These neurons are involved in processing sensory outputs from the environment, and this localisation is in keeping with the role of ABU-13 in responding to external stress from the surrounding environment. That being said, it is curious that we see overlap with the amphid neurons specifically as we have previously found evidence of *abu-13* expression in the amphid sheath glia. By nature, these glia surround the amphid neurons; we therefore may not be observing direct colocalization with the neuron itself, but rather with the sheath glia supporting it. Beyond this, we do not observe any further neuronal colocalization, apart from a small amount around the vulva that appears to represent juxtapositioned structures, as opposed to a direct overlap. Thus, this supports a role of ABU-13 in the amphid sheath neurons.

We then crossed endogenously-tagged ABU-13::WormScarlet animals into an SCMp::GFP seam cell marker strain (Figure 3.13B). This seam cell marker is particularly well expressed during larval development, so we chose to image animals during the L3 developmental stage, slightly earlier than the other localisation analyses. These animals displayed a clear colocalization with the seam cells both dorsally and laterally, in keeping with our results using the transcriptional reporter line. This supports a role for ABU-13 within the seam cells as well as pointing towards a role for this protein in the hypodermal system in general, as these cells are closely anatomically linked to the hyp-7 syncytium, as well as sharing a great deal of functionality with the hypodermis. We were, however constrained by the quality of markers available for a tissue marker cross of the hypodermis; while we did make several attempts to produce our own transgenic GFP marker strain for the hypodermis, under the *dpy-7* promoter, these animals were difficult to maintain with many exhibiting anatomical defects, likely as a consequence of GFP overexpression. Similar issues were encountered with other hypodermal marker strains used, of which there is limited availability with GFP markers compared to red fluorophores - these animals tended to be difficult to maintain due to the low expression of the extrachromosomal array,

and attempts at crossing proved unsuccessful. As such, we were unable to perform a successful cross with a general marker for the hypodermis.

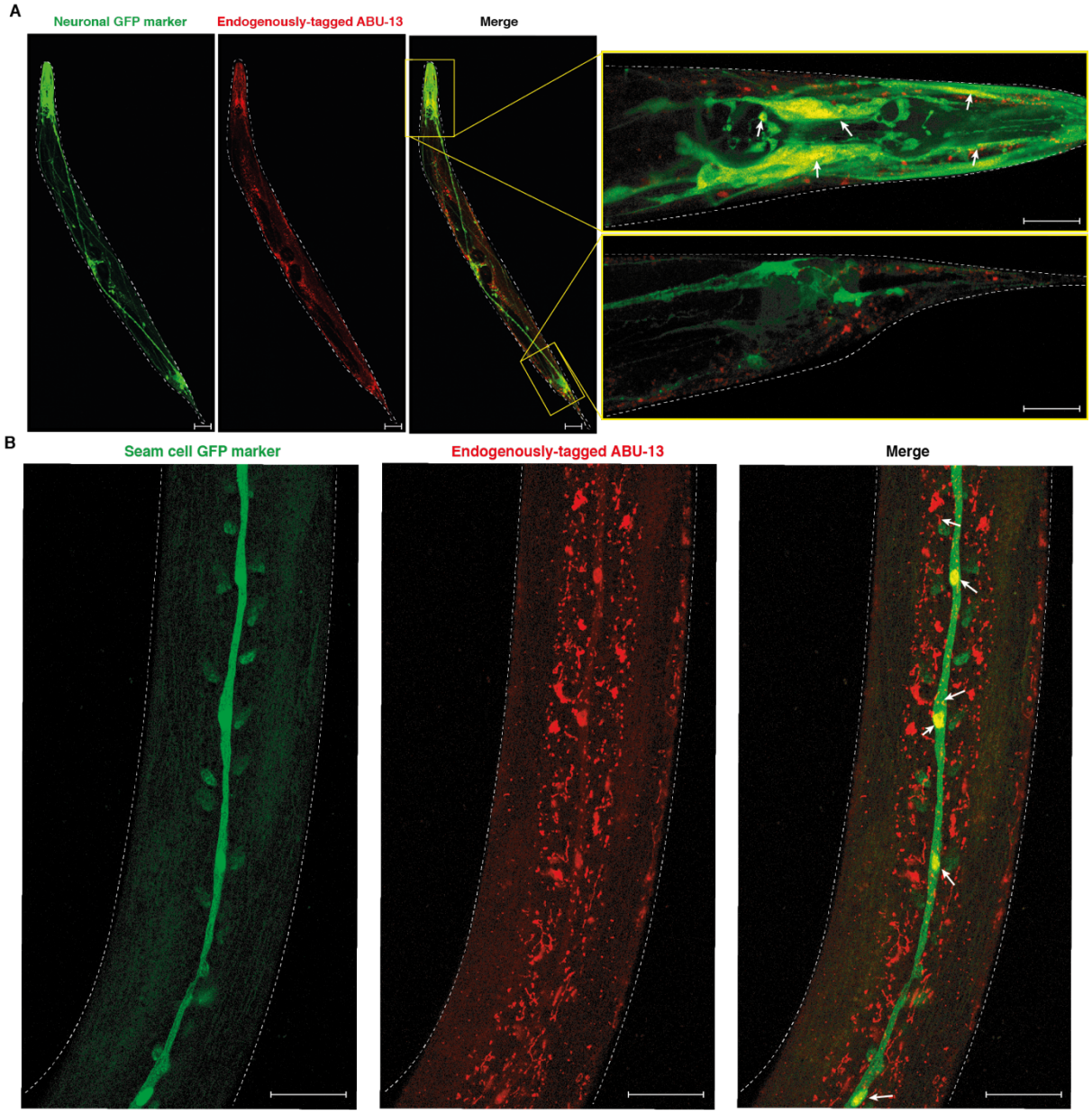



**Figure 3.13** ABU-13::WormScarlet colocalises with amphid neurons and seam cells in endogenously-tagged animals (A) Confocal micrographs of endogenously-tagged ABU-13::WormScarlet animals crossed into (A) a neuronal marker strain (*unc-119*p::GFP), *Scale bars = 50µm, 25µm*; (B) a seam cell marker strain (*SCMp*::GFP), *Scale bars = 25µm*

### *3.2.6 ABU-13 puncta are juxtanuclear*

We next sought to determine a subcellular localisation for ABU-13. As ABU-13 forms punctate structures that might be consistent with a nuclear localisation, we stained animals with the nuclear dye DAPI to determine whether there was colocalization of nuclei with tagged ABU-13 (Figure 3.14). This was not the case, with no direct colocalization between DAPI-stained nuclei and the ABU-13 puncta in ABU-13::tagRFP overexpression animals. We did, however, notice that these puncta appeared juxtaposed with nuclei – that is, many were found adjacent to or surrounding nuclei, suggesting that

they represent juxta-nuclear bodies. One possibility is that this represents an endoplasmic reticulum localisation, which may support a role for ABU-13 in protection from ER stress.

As we have already established, the puncta found in the overexpression strain are significantly larger on the whole than those found in the endogenously tagged strain. To address this, we attempted to stain endogenously-tagged animals with DAPI to determine whether this juxtaposition was maintained consistently across tissues. However, we encountered issues with this staining as the freeze-crack protocol used for DAPI-staining caused ABU-13 puncta to dissolve during the staining process. This dissolution of puncta after death possibly suggests that puncta maintenance is ATP-dependent in the endogenously-tagged line – this may not be the case in the overexpression line due to their potential propensity for aggregation. Alternative approaches to nuclear staining will be explored in the future to get around this issue and determine the relationship between ABU-13 and nuclei.

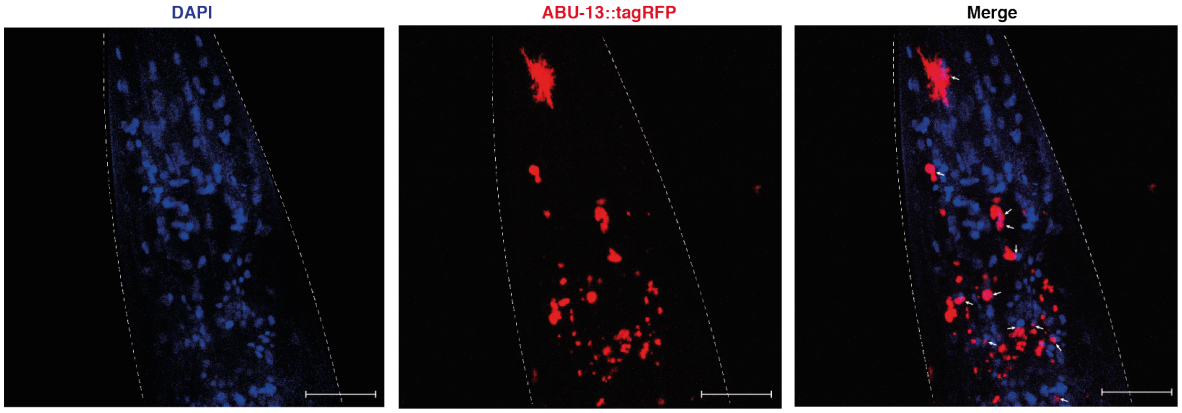



**Figure 3.14** ABU-13 is found in juxtanuclear positions, but shows no nuclear localisation (A) Confocal micrographs of ABU-13::tagRFP overexpression animals stained with DAPI to mark nuclei (*Scale bar = 25μm*)

To further explore the potential localisation of ABU-13 to the endoplasmic reticulum, we attempted staining of overexpression animals with an ER-sensitive dye, ER-Tracker™, following a similar freeze-crack protocol. This staining, however, showed a high degree of background and non-specific staining, with no ER structures immediately visible during imaging. As such, we deemed ER-Tracker (Blue-White)™ unsuitable for these animals. We then moved to using an ER-localised fluorescent marker; whilst most of these use red fluorophores, we found a limited number expressing GFP-tagged ER proteins as extrachromosomal array lines. The penetrance of these strains was low, and the markers themselves difficult to monitor and subsequently maintain. As such, no successful crosses could be performed. Moving forward, we will work to express an endogenous ER-resident GFP marker to eliminate issues with low penetrance and to facilitate successful crosses with ABU-13::WormScarlet animals.

### *3.2.7 ABU-13 does not colocalize with known stress-related membraneless organelles*

Following our hypothesis that ABU-13 is a scaffold protein involved with the formation of a phase separated compartment, we next explored whether this protein was a component of known stress-related membraneless organelles. Within *C. elegans*, the main stress-related compartments of this nature are stress granules and Processing bodies. Whilst P-bodies are constitutively present, they are also stress-responsive, with increased formation upon the application of a stressful stimuli. Stress granules, on the other hand, do not form under basal conditions and are instead triggered upon such a stimulus. To identify whether ABU-13 was a constituent of either of these organelles, we imaged ABU-13::tagRFP overexpression animals crossed into *ife-2p*::IFE-2::GFP animals, marking stress granules, or *dcap-1*::DCAP-1::GFP animals, marking P-bodies, before and after the application of a stress.

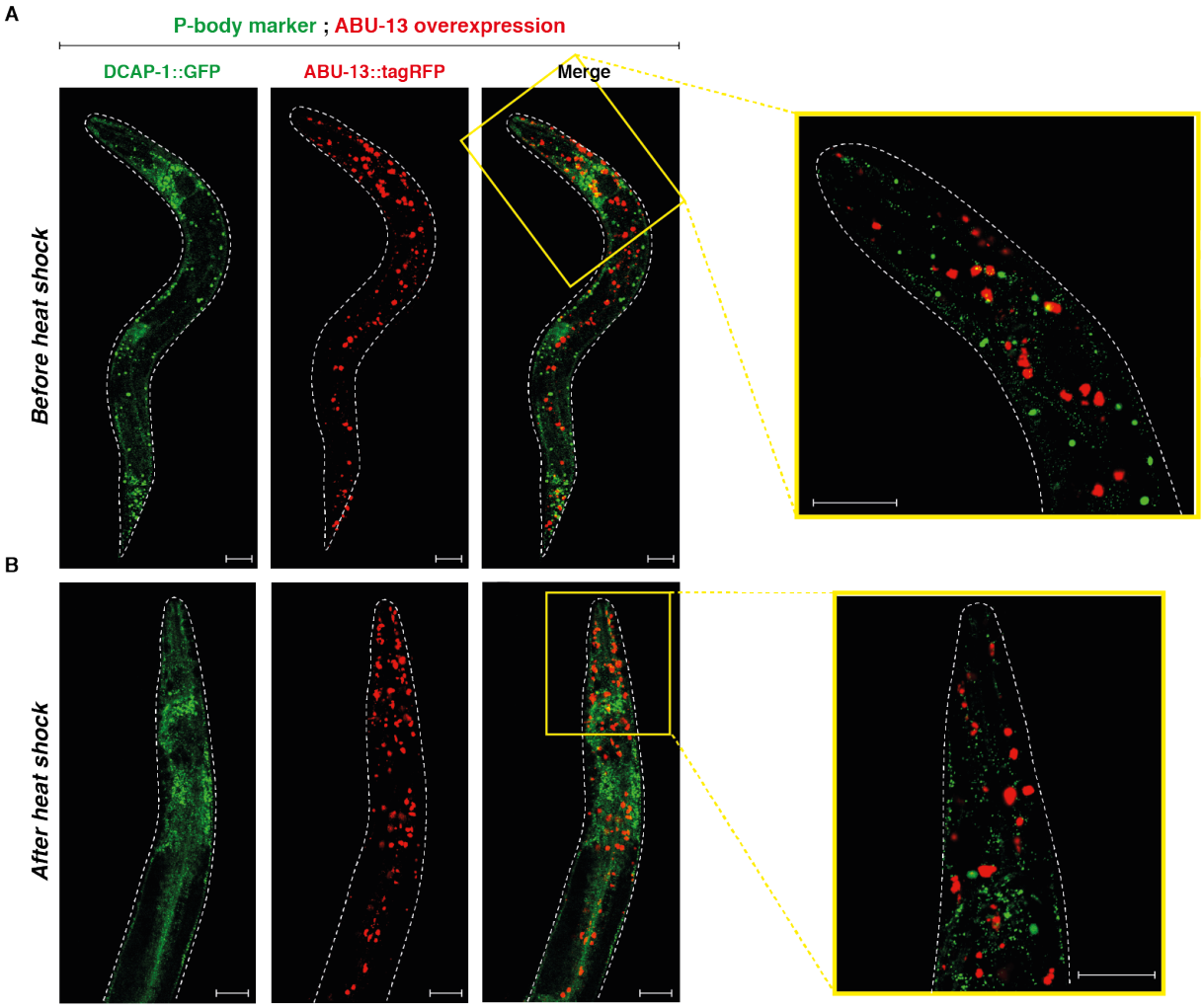



**Figure 3.15** ABU-13::tagRFP does not colocalise with P-bodies before or after heat shock in overexpression animals (A) Confocal micrographs of ABU-13::tagRFP animals crossed into a P-body marker strain (*dcap-1p*::DCAP-1::GFP) before heat shock, *Scale bars = 25µm* ; (B) Confocal micrographs of ABU-13::tagRFP animals crossed into a P-body marker strain (*dcap-1p*::DCAP-1::GFP) after heat shock, *Scale bars = 25µm*

Heat shock is a traditional mechanism through which these stress-related organelles are stimulated, thus we initially subjected these crossed animals to a 34°C heat shock for 6 hours before imaging. In animals expressing a P-body marker, large P-bodies are evident prior to the heat shock (Figure 3.15A) - these constitute the constitutively-active P-body population that are present in basal conditions. Following this heat shock, there is a large increase in the formation of small P-bodies - these represent the stress-responsive population (Figure 3.15B). Both before and after application of a heat stress stimuli, the

ABU-13 puncta are larger than the concurrent P-bodies with no colocalization observed between these structures. This suggests that within the overexpression model at least, these structures are unrelated.

We then repeated this heat shock treatment in animals with the stress granule marker. As expected for these animals, there were initially very minimal GFP-positive puncta as these components are diffuse under basal conditions (Figure 3.16A). However, following heat shock treatment, there was a large coalescence of GFP into small puncta, indicative of stress granule formation (Figure 3.16B); these were present ubiquitously throughout the animal. However, as with the P-bodies, these stress granules are much smaller than the co-expressed ABU-13::tagRFP puncta and no colocalization is observed. As such, this also suggests that these two membraneless organelles do not overlap and supports the notion that ABU-13 puncta represent a novel species of stress-responsive phase-separated compartment in *C. elegans*.

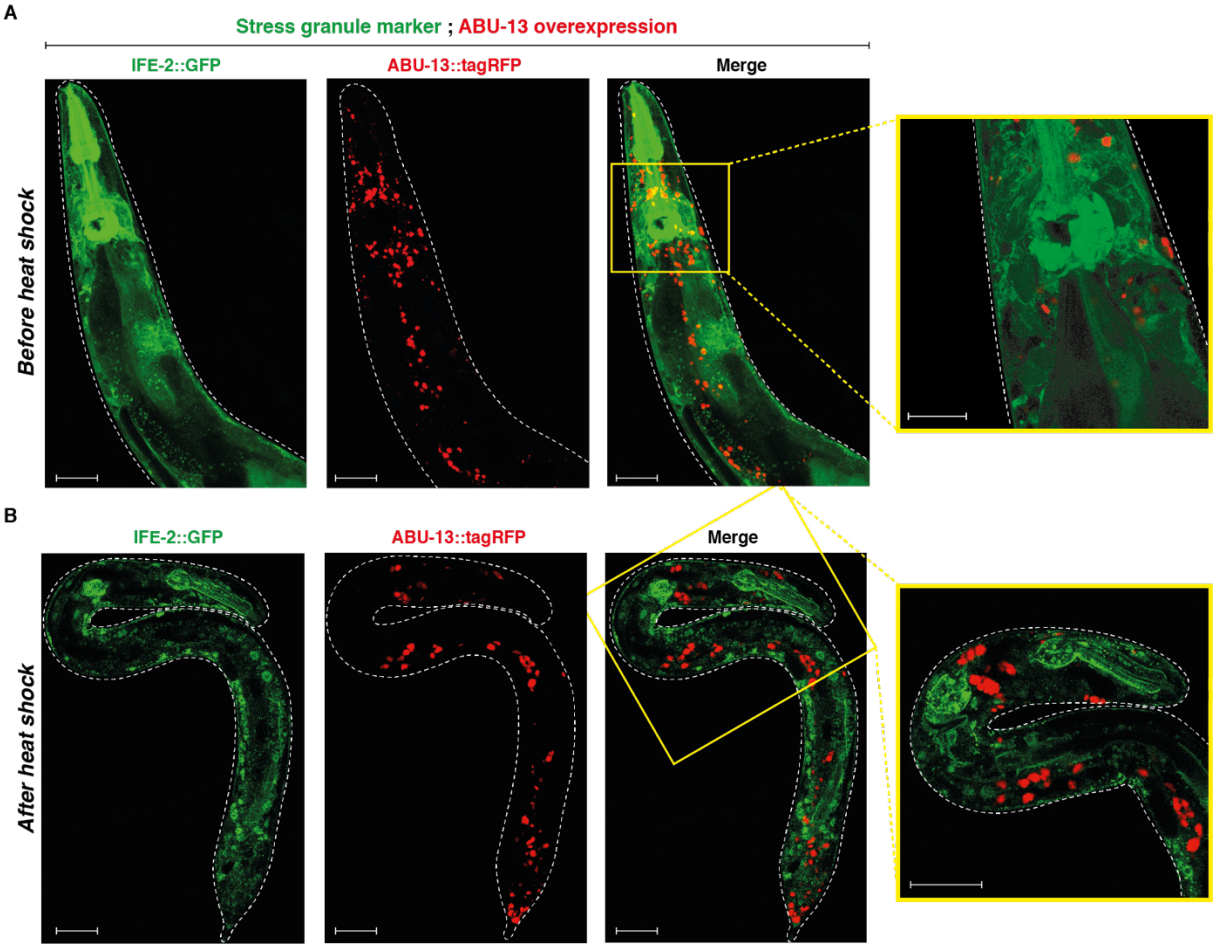



**Figure 3.16** ABU-13::tagRFP does not colocalise with stress granules before or after heat shock in overexpression animals (A) Confocal micrographs of ABU-13::tagRFP animals crossed into a stress granule marker strain (*ife-2p*::IFE-2::GFP) before heat shock, *Scale bars = 25μm* ; (B) Confocal micrographs of ABU-13::tagRFP animals crossed into a stress granule marker strain (*ife-2p*::IFE-2::GFP) after heat shock, *Scale bars = 25μm*

There are, however, limitations in the use of the overexpression animals. As we have already established, the puncta formed within these animals are significantly larger than those observed in the endogenously tagged animals. As such, we moved towards using these endogenously tagged animals to assess the potential interplay between these stress-related organelles.

Using these animals, we initially determined the effect of tunicamycin treatment. This was of particular interest as, unlike the PA14-treated animals, no observable changes to ABU-13 puncta were observed in these animals, and thus we hypothesised that the effect on ER stress resistance may be conferred via interactions with other membraneless organelles. Endogenously tagged animals crossed into the P-body marker strain showed a much more similar size distribution of P-body and ABU-13 puncta than in the overexpression model, as expected (Figure 3.17A-B). Despite this, however, there does not appear to be substantial colocalization between these puncta in either the DMSO-treated control animals nor the tunicamycin treated animals. It does appear, however, that some of the P-body puncta overlap with the larger ABU-13 puncta, potentially representing P-bodies engulfed by ABU-13 puncta or those in close physical proximity with them. This could be clarified using FRET to determine the proximity of these compartments, or using superresolution microscopy to improve spatial resolution.

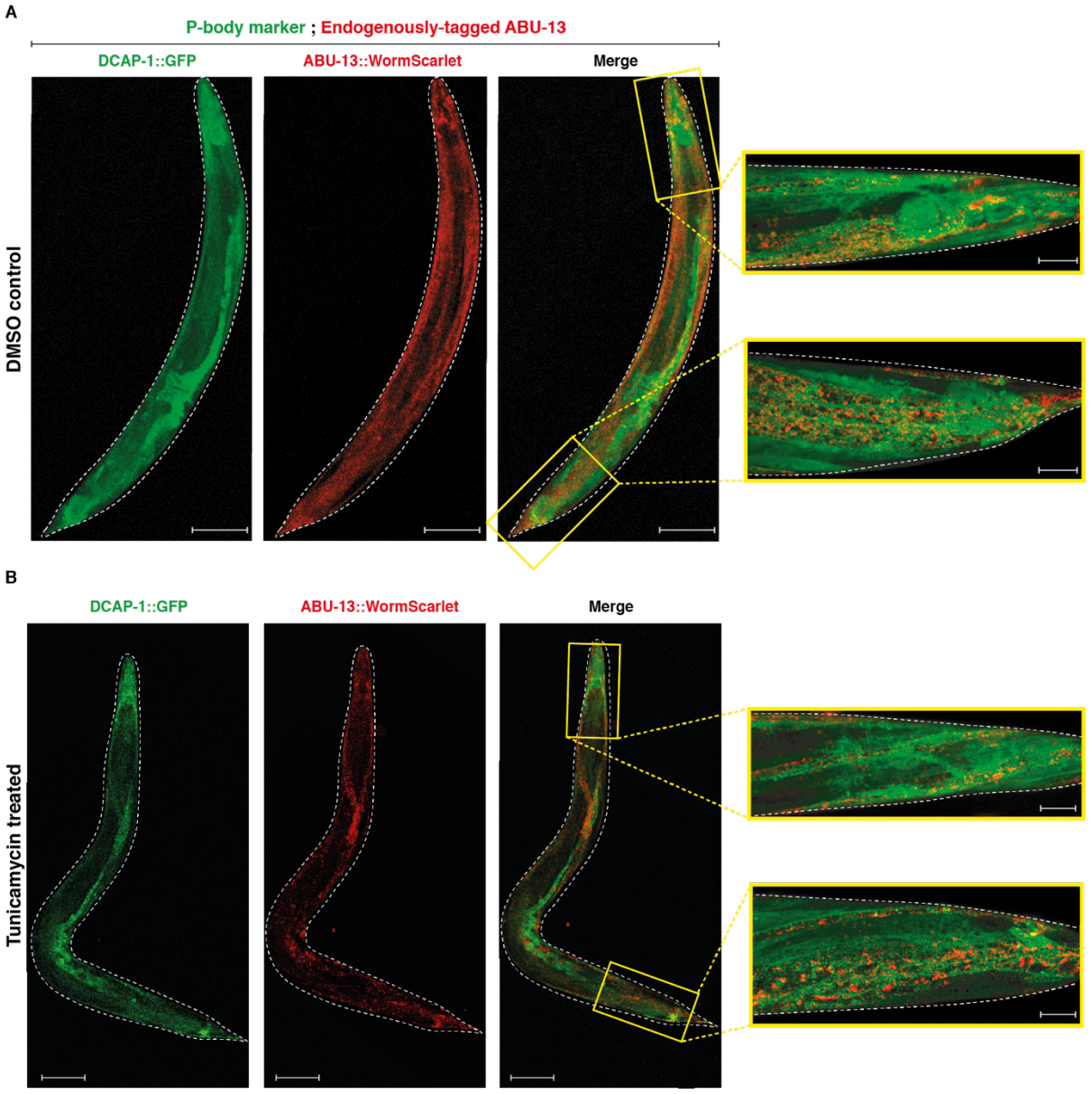



**Figure 3.17** ABU-13::WormScarlet does not substantially colocalise with P-bodies before or after 25ng/µL tunicamycin treatment in endogenously-tagged animals (A) Confocal micrographs of ABU-13::WormScarlet animals crossed into a P-body marker strain (*dcap-1p*::DCAP-1::GFP) after DMSO treatment, *Scale bars, 25µm* ; (B) Confocal micrographs of ABU-13::WormScarlet animals crossed into a P-body marker strain (*dcap-1p*::DCAP-1::GFP) after tunicamycin treatment, Scale bars = 25µm

Further work is needed to determine whether any interaction is observed between stress granules and the endogenously tagged ABU-13 puncta. In addition to this, assessing these interactions after treatment with a range of stress-stimuli may shed light on differential interactions in different conditions. As we have already established for ABU-13 and as has already been shown for stress granules and P-bodies, these stress-related organelles respond in a stimulus-specific fashion - as such, any potential inter-organelle interaction may also be stimulus-specific; if this were indeed the case, it could help point towards molecular mechanisms conserved between these organelles.

### *3.2.8 Purification of ABU-13 in preparation for in vitro phase separation assays*

To understand the dynamics of puncta formation and test our hypothesis that ABU-13 is involved in the formation of phase separated structures, we next sought to purify both ABU-13 and ABU13ΔPrLD and perform *in vitro* phase separation assays. However, substantial challenges were met during this purification process (Figure 3.18A). We began using a BL21 bacterial expression system, with the exonic coding region for *abu-13* and *abu-13ΔPrLD* inserted in frame with GFP within a pET expression vector under control of a T7 promoter. Following induction, we saw no GFP signal within these bacteria. To determine whether this was as a consequence of low expression, we continued with the purification and, after SDS-PAGE, saw no band corresponding to the size of ABU-13. This issue persisted across numerous attempts.

As bacteria lack some of the protein folding machinery compared to eukaryotes, we moved into an insect cell expression system, using *Drosophila* S2 cells, expressing *abu-13* and *abu-13ΔPrLD* within a pMT vector, fused to GFP. Both sets of transformed cell cultures showed poor growth and could not be taken further for protein purification.

As we knew that ABU-13 was expressed consistently within *C. elegans*, we attempted to purify directly from populations of worms in which ABU-13 was endogenously tagged. For this, we inserted a 3xFLAG tag into the C-terminus of *abu-13* as to reduce the confounding effects that a larger fluorescent tag may have on the phase transition behaviour of ABU-13 *in vitro*. As such, we then attempted to purify both a FLAG-tagged and a WormScarlet-tagged form of ABU-13. We grew these animals in liquid culture before lysis and pull-down of the protein using FLAG or WormScarlet-specific beads before eluting the bound protein. However, upon visualisation of the purified protein via Western blot, we saw no bands corresponding to ABU-13, demonstrating that this purification had failed.

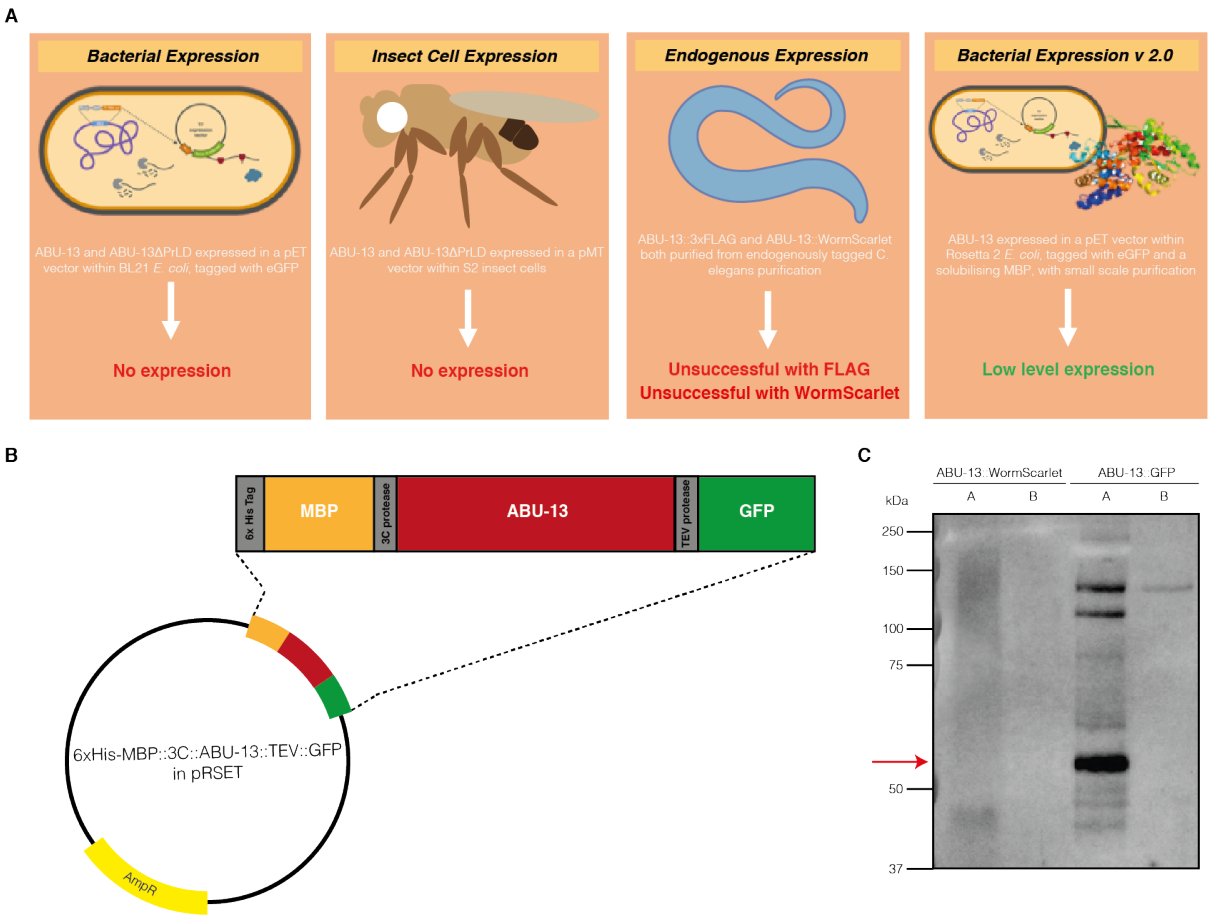



**Figure 3.18** Outline of ABU-13 protein purification work (A) Schematic detailing the various approaches taken to purify ABU-13 in preparation for in vitro work ; (B) A plasmid map of the pRSET vector used for successful purification of GFP-tagged ABU-13 in which a MBP tag was used to solubilise the protein; (C) Western blot showing multiple attempts at protein purification from ABU-13::WormScarlet expressing *C. elegans*, or ABU-13::GFP expressing Rosetta 2 bacteria, probed with anti-GFP antibody (A = first attempt; B = second attempt; red arrow = correct ABU-13::GFP band)

As it seemed likely that the prion-like nature of ABU-13 was causing issues with this purification, we moved to a protocol optimised for prion-like proteins. This involved the expression of ABU-13 fused to a 3C protease-cleavable maltose-binding protein tag, the addition of which aided the solubilisation of this protein in bacteria (Figure 3.18B). We then expressed this construct in Rosetta 2 bacteria, optimised for the expression of eukaryotic proteins. Following induction with IPTG and bacterial collection, the expression of GFP within bacteria expressing the control GFP-only construct was clear - however, the expression of ABU-13::GFP was not visible. We continued with this purification, and confirmed the presence of GFP-tagged ABU-13 in the final eluate by Western blot (Figure 3.18C).

During this process, however, the aggregation-prone nature of ABU-13 became evident as it underwent substantial aggregation upon cleavage of the solubilising MBP tag. If the purified protein was too concentrated as this cleavage event occurred, we observed significant formation of GFP-positive aggregated structures visible to the naked eye.

### *3.2.9 ABU-13 forms reversible puncta in vitro*

The presence of a prion-like domain and an RNA binding motif are indicative not only of components involved with phase separation compartments, but proteins with these characteristics often act as scaffolds, driving formation of these structures. Whilst *in vivo* data with prion-domain mutant animals is inconclusive, the ability of ABU-13 to phase separate *in vitro* would support a role for this protein in puncta formation. As such, using purified ABU-13, we conducted a series of *in vitro* phase separation assays to determine whether this protein possessed the intrinsic ability to nucleate a phase transition.

One of the triggers of phase separation is an increase in molecular crowding conditions akin to those observed within the cytoplasm, which increases the propensity for phase separating proteins to coalesce. This can be mimicked *in vitro* using a molecular crowding agent such as dextran or polyethylene glycol (PEG). As would be expected for phase separating proteins, the formation of GFP-positive ABU-13 puncta increases as the molecular crowding conditions are increased. This is not seen in the GFP control suggesting that this ability is intrinsic to ABU-13 itself rather than being driven by the GFP tag (Figure 3.19A).

Recently, it has been demonstrated that the chemical 1,6-hexanediol is capable of disrupting the weak interactions that hold phase separated compartments together. We tested this in our set up, imaging ABU-13 at either 50mg/mL or 100mg/mL PEG, with and without treatment with 5% 1,6-hexanediol Figure 3.19B). Whilst little difference is seen before and after treatment in the low molecular crowding conditions, there is a difference in the high molecular crowding condition sample. This suggests that some, but not all, puncta within these samples represent liquid-like phase separated compartments. It is possible that the remaining puncta possess more solid-like characteristics and cannot be dissolved with this treatment - whether this is a hydrogel or an aggregate is unclear, however, this could support previous evidence showing multiple species of ABU-13 puncta; one with more liquid-like properties and another with more solid-like properties.

As mentioned previously, phase separating proteins are exquisitely sensitive to concentration - a property that has been illustrated in both our overexpression line and in the propensity of ABU-13 to aggregate during the purification process. Within a small concentration range, however, phase separating proteins are able to reversibly transition from diffuse to punctate and back again as the concentration is increased and then subsequently decreased. To determine whether ABU-13 is capable of this, we compared puncta formation in samples diluted from 100nM ABU-13 in 100mg/mL PEG to 5nM in 5mg/mL to a baseline state of 5nM ABU-13 in 5mg/mL that had never been further concentrated (Figure 3.19C). In keeping with our hypothesis, we indeed observed a reversibility of puncta formation. As with the 1,6-hexanediol treatment, a small fraction of puncta remained in some of the imaging fields of view, potentially representing more solid-like structures; however, the majority of puncta were dissolved.

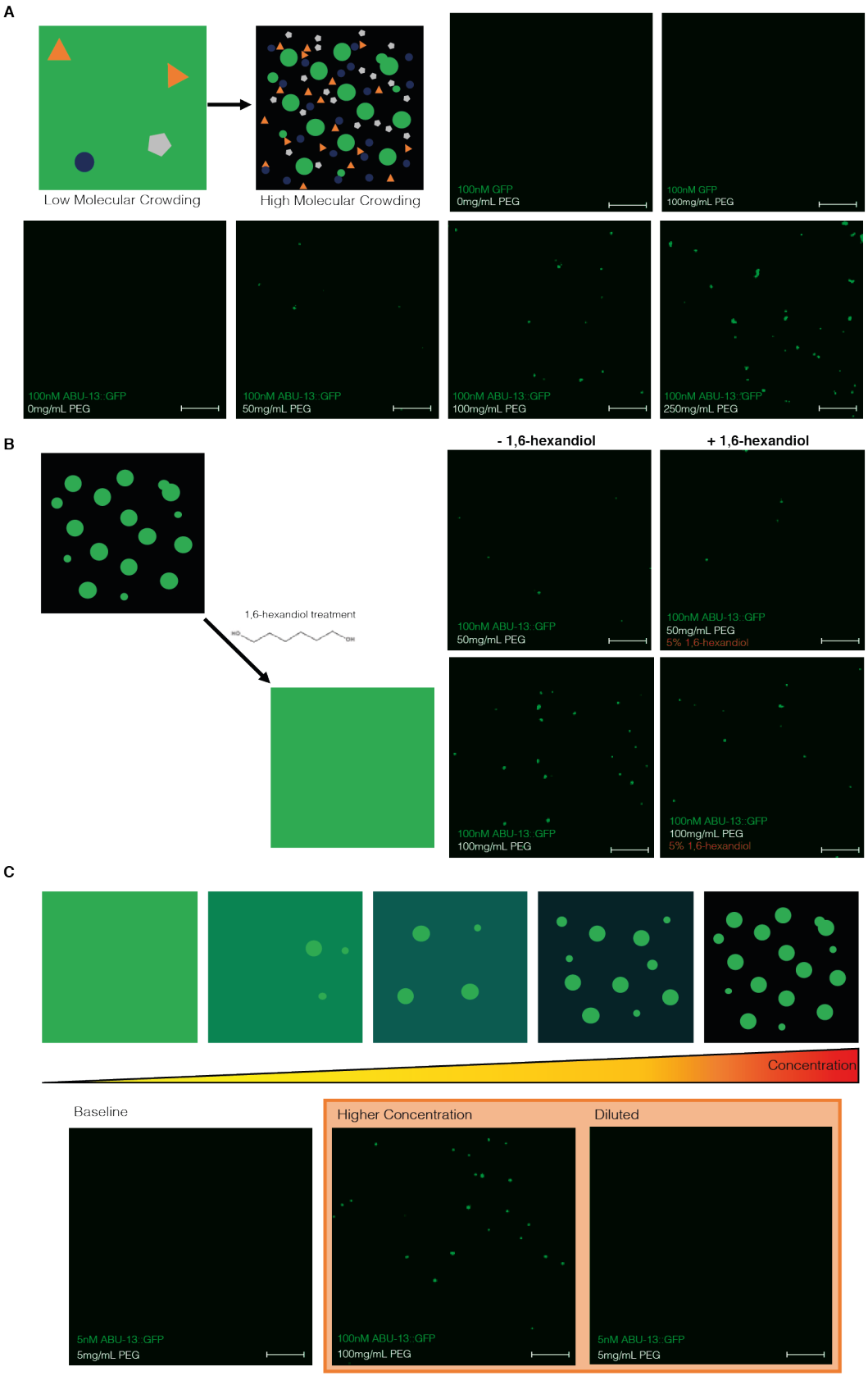



**Figure 3.19** ABU-13 is able to form puncta *in vitro*, reminiscent of phase separated structures (A) Confocal micrographs of purified ABU-13::GFP or GFP alone treated in varying molecular crowding conditions (*Scale bar = 100μm*) ; (B) Confocal micrographs of purified ABU-13::GFP in either 50mg/mL of PEG or 100mg/mL PEG, before and after 5% 1,6-hexanediol treatment (*Scale bar = 100μm*); (C) Confocal micrographs of purified ABU-13::GFP before and after dilution, showing the basal levels of puncta formation, and then the levels of puncta formation following dilution from a higher concentration (*Scale bar =100μm*)

As we were only able to purify ABU-13 on a small scale, we could not further validate the conditions in which ABU-13 transitions - for example, phase separation also differs depending on factors such as salt concentration and temperature. However, further purifications have been performed, allowing us to construct a full phase diagram in the future.

On the whole, the evidence thus far supports the role of ABU-13 in nucleating the formation of phase separated compartments, suggesting that it may indeed represent a scaffold protein, rather than a client protein. Further work to determine whether puncta formation is abolished upon removal of the prion domain would demonstrate this more conclusively.

## 3.3 Discussion

### *3.3.1 Phase separation properties of ABU-13*

Supporting our previous hypothesis that, due to its prion-like domain and predicted RNA binding properties, ABU-13 was involved in phase transitions, we observe punctate structures throughout the animal. Phase separated organelles often play a role in the regulation of RNA, which, as this protein has been implicated in ER stress responses[25,26], could be a mechanism through which ABU-13 confers stress resistance. For example, ABU-13 may sequester specific RNA species upon stress, preventing their translation - or similarly, it may interact with other RNA species and promote translation. Understanding the protein and RNA interactors of ABU-13, particularly those localising to these biomolecular condensates, would help us to determine the molecular mechanisms operating within these puncta and establish exactly how this stress resistance might be conferred.

Not all the components of a phase separating compartment have the intrinsic ability to drive their formation, however. Instead, many proteins are recruited by components already found within these structures - these are known as client proteins, whilst the proteins driving phase separation are known as scaffolds. A general role for prion-like proteins as scaffolds for the nucleation of these compartments has been proposed - for example, TIA-1 and TDP-43[27,28], and Lsm4[29] are all proteins containing a PrLD that are involved in the formation of stress granules and P-bodies respectively. As such, we might expect the removal of the prion-like domain of ABU-13 to abolish its ability to nucleate puncta formation. Whilst we do observe this in the overexpression model, we still see puncta formation in the endogenous line. We do not believe that this difference can be accounted for by differences in expression level between the overexpression line and the endogenously-tagged line; whilst this could be an influencing factor, we observe a much greater expression of ABU-13::tagRFP in the overexpression model by several orders of magnitude compared to the endogenously-tagged line. As we see diffuse ABU-13 when there is higher expression, it is unlikely that the punctate structures in the endogenously-tagged line are driven by protein concentration. Instead, we hypothesise that this difference is a consequence of a flexible linker added between the N and C terminal regions flanking the removed prion-like domain in the endogenous line that is absent in the extrachromosomal array of the overexpression line. It is likely that, without this linker, the function of the remaining protein of ABU-13 is compromised. This is, however, insightful nonetheless as it suggests that localisation to ABU-13 puncta relies on regions outside of the prion-like domain.

This can be interpreted in a number of ways; either ABU-13 is a client protein and incorporation into pre-existing condensates requires multivalent interaction with non-prionogenic domains, or there is redundancy amongst scaffold proteins and even without the prion-like domain, ABU-13 can continue to be incorporated due to its other interactions. This latter interpretation has been observed in a number

of organelles, with removal of single scaffold proteins not abolishing puncta formation[30,31]. Whilst it is common for phase-separated organelles to contain multiple scaffold proteins, many of which contain prion-like domains, it would be surprising to find a prion-like protein within these organelles that acted as a client protein. This predisposes us towards the interpretation that ABU-13 is acting as a scaffold, however more work is required to firmly come to this conclusion. For example, identifying whether transgenic lines expressing only the prion-domain can form puncta *in vivo* would help to clarify this further - though it is likely that these puncta may form non-specific aggregate-like structures in these conditions, as the multivalent interactions of these proteins play an important role in modulating the activity of these proteins. If this were indeed the case, alternative approaches to test this hypothesis *in vivo* could be employed - for example, identification of protein interactors would be useful in determining the molecular makeup of these puncta. This would allow puncta formation to be monitored by labelling another component - thus can be used to show whether there is still puncta formation in ABU-13ΔPrLD overexpression animals, even if ABU-13 is diffuse.

A role of ABU-13 as a scaffold protein is further supported by our *in vitro* work with purified GFP-tagged ABU-13. The formation of GFP-positive puncta demonstrates that coalescence is an inherent biophysical ability of ABU-13, whilst the reversibility of this formation indicates that these structures are not solid aggregates but rather dynamic assemblies truly reminiscent of phase separated organelles. Moving forward, purification of ABU-13ΔPrLD would allow us to determine more confidently whether this ability to form puncta is indeed reliant on the prion-like domain, with this *in vitro* setting removing any confounding interactions between existing puncta, and the N and C termini *in vivo*. It is also possible that, even if loss of the prion-domain does not abolish puncta formation, that these mutants have altered puncta dynamics and function – thus examining the functional consequences of prion-domain mutants would be of interest moving forward.

### *3.3.2 Multiphase coexistence of ABU-13*

Traditionally, phase separation involves a transition from a diffuse to a liquid-like state, with many membraneless organelles exhibiting rapid recovery dynamics following a photobleaching event. The recovery rate of ABU-13 was variable, with larger puncta showing a more rapid recovery than smaller puncta. This variability suggests that some puncta are more dynamic than others. Further investigation of these differences demonstrated to us that these larger puncta were comprised of a clustering of smaller puncta. These smaller sub-punctae had lower mobility, whilst the area bounding them was more mobile. This is reminiscent of core-shell architecture seen in a number of phase-separating organelles, with core components forming a stable inner region, and more transient interactions occurring in the periphery[32,33].

Within these internal clusterings, very little fusion occurs, demonstrating that these phases can coexist whilst remaining immiscible. This is very similar to the behaviour of nucleoli, in which multiple phases

exist in a tripartite organisation, with differential surface tensions between compartments preventing their fusion[32,34]. Core-shell architecture has also been demonstrated in organelles considered incredibly liquid-like, including stress granules[35]. This therefore led us to suggest that ABU-13 puncta have variably mobility, with small core regions exhibiting hydrogel-like properties and representing a core set of compartment components, whilst the shell is more liquid-like and represents more transient interactions. It is important to note at this point that scaffold and client proteins can be found within either or both of the core and shell regions – the differential properties observed within these regions are determined by the intermolecular interactions in these different environments, rather than entirely different protein content. The presence of a scaffold in the core, for example, does not preclude it from being in the shell region as well.

There are fascinating implications of this supramolecular organisation for stress responses - for example, it is possible that within a stress response, the core components of these assemblies do not change, but those within the outer layer do. If this were the case, the core components would represent those that are necessary for puncta functionality, whilst the outer components acting as dynamic 'adaptors', modulating the activity of the puncta in response to environmental stimuli. Determining both the mobility of these puncta in various physiological stress conditions, and identifying potential core versus shell proteins in these contexts would be useful for gaining a deeper understanding of puncta dynamics and function.

### *3.3.3 Role in ER stress responses*

Despite identifying a role for ABU-13 in the response to tunicamycin treatment, we did not see any observable effects on puncta dynamics upon ER stress. This suggests that the role of this protein in ER stress responses does not involve large structural changes to these assemblies, as is observed in some stress-responsive organelles. Instead, we hypothesise that these responses are facilitated via the differential recruitment of components to these structures. This would allow the activity of these constitutively active puncta to be modulated dynamically, with client proteins recruiting RNA and protein specific to the stress response.

At the moment, it is unclear whether the role that ABU-13 plays in ER stress responses is dependent on puncta formation. While we know that puncta formation still occurs in endogenously-edited prion-domain mutants, it is unclear whether these animals exhibit the same phenotypic deficits seen in *abu-13* knockout animals. Similarly, we do not know whether these *abu-13(tm6224)* knockout animals, which express a truncated form of the protein, are capable of forming puncta *in vivo* or not. Repeating these phenotypic assays in *abu-13ΔPrLD* mutants would clarify whether or not these animals retain their ability to facilitate an ER stress response. If they do not exhibit the same phenotypic effects, this suggests that their presence within puncta is sufficient in enabling their function. Thus, identifying

alternative ways to prevent association of ABU-13 with puncta (i.e. by mutating RNA binding domains) would be necessary to further understand the link between the function of ABU-13 and its inclusion in these compartments.

As many phase separated organelles are involved in the regulation of RNA, it is possible that ABU-13 exerts its action via a similar mechanism. As such, determining the RNA species within these puncta, and specifically those that ABU-13 interacts with, will help to elucidate whether this protein is involved with the regulation of RNA. This hypothesis could be further tested by identifying and subsequently mutating any RNA binding motifs within ABU-13. As with the prion-like domain mutants, it would be interesting to determine whether the phenotypic deficits observed in *abu-13(tm6224)* animals are recapitulated when these RNA binding motifs are mutated. It is, however, possible that the two compensate for one another – as such, if no phenotypic effect is observed in either of these mutants, it might be the case that mutants for both the RNA binding domain and the prion-domain are necessary for actualisation of these deficits.

In addition to this, to further explore the role that ABU-13 puncta play in the ER stress response, we are in the process of producing ABU-13::WormScarlet animals with an *xbp-1(zc12)* mutation, thus replicating the conditions in which *abu-13* was first identified as a modulator of the non-canonical unfolded protein response. As these genes both reside on chromosome III, however, we are working on a CRISPR-derived mutant to monitor this effect.

### *3.3.4 Role in the immune response*

The changes observed in ABU-13 puncta upon exposure to *Pseudomonas aeruginosa* indicate significant changes in their dynamics. The reduction in puncta number paired with the increase in size suggests one of two possibilities – that either puncta are fusing and thus overcoming their immiscible nature, or that some puncta are being dissolved whilst others are increasing their recruitment. The switch from a state of immiscibility to fusion would indicate that there is an unknown factor modulating the free mixing of these puncta in their basal state. There is evidence that both RNA and ATP can buffer these condensates, preventing their solid-like transition[14,36] - thus this switch may be a reflection of increased ATP interaction or RNA recruitment that in turn overcomes the immiscible nature of these puncta and ultimately facilitates their fusion.

Beyond these puncta, we also observe an increase in diffuse expression of ABU-13 within a sub-region of the pharynx, possibly within the marginal cells. These cells line the pharynx in three segments, surrounded by pharyngeal muscle. Little is known about their function, however, they do appear to possess the ability to scavenge cholesterol and small hydrophobic molecules, known as wactives, from bacteria being fed upon in a sphingomyelin-dependent fashion[37]. This scavenging behaviour is

particularly interesting in the context of ABU-13, which is known to be activated downstream of the scavenger receptor, CED-1[38,39]. Whilst there is no evidence of *ced-1* expression in these marginal cells, it is possible that ABU-13 is responsive to scavenger processes within this tissue via another mechanism. Although no such function has been identified as of yet, as the marginal cells are amongst the first cells to contact potentially pathogenic ingested material, it is logical for these to play a role in immune surveillance, activating an early innate immune response. Perhaps, for example, these cells are capable of scavenging bacterial components that can act as antigens and adjuvants, stimulating this immune response. For such a mechanism to be useful, however, these marginal cells would need to signal to other tissues.

Interestingly, within these cells, ABU-13 is diffuse. This was surprising to us as we had expected ABU-13, as a potential scaffold protein, to nucleate puncta formation when expressed. The lack of coalescence observed could point towards a different cellular environment within these cells that increases the critical concentration required for a phase transition. This is possible, for example, by modulating the buffering capacity of the aqueous environment by increasing salt concentration or, as previously mentioned, increasing RNA concentration. Alternatively, this could support a role for ABU-13 as a client protein, rather than a scaffold, with the lack of condensation resulting from no expression of a nucleating scaffold protein. Whilst we cannot fully exclude this possibility, given our *in vitro* evidence, this explanation seems unlikely.

Overall, we propose that pharyngeal ABU-13 expressed within marginal cells functions to process scavenged bacterial antigens, potentially stimulating a downstream immune response. To fully investigate this hypothesis further, a greater understanding of the physiology of marginal cells in the context of the immune system is required, and then ultimately how ABU-13 might fit into this.

### *3.3.5 Long-term immune protection?*

With many environmental stresses, an initial exposure to a threat can have a hormetic effect and result in a bolstering of future responses against the future stress. This has been observed in nematodes following heat shock[40,41], for example, with improved survival of animals subjected to an initial stress earlier in life compared to those experiencing a stress for the first time. This points towards an ability of stress response pathways to 'prime' themselves for recurrent stress. Within the immune system, there are a number of mechanisms through which such priming can be achieved - in higher organisms, for example, the adaptive immune system can facilitate this.

Whilst *C. elegans* lacks the specialised, mobile immune cells required for an adaptive immune system, there are examples of primitive responses in invertebrates that play a comparable role. Although discussed in another context, the bacterial CRISPR/Cas9 system is an excellent example of this, using

Cas9-crRNA ribonucleoprotein complexes to stimulate a form of adaptive immune response in bacteria against specific invading DNA and RNA species[17,19]. This, however, relies on insertion of foreign DNA elements into CRISPR motif regions within the bacterial genome[16]. Whilst ABU-13 does not play a homologous role within *C. elegans*, as demonstrated by its lack of nuclear localisation, it is possible that the constitutive presence of ABU-13 structures may allow for priming of the immune response, protecting animals against reinfection later. This constitutive presence within animals enables more long-term interactions to be forged, unlike those that can be made within the limited lifespan of more transient phase separated structures. As such, responses of these puncta to bacterial invasion could prompt a form of 'molecular memory' that is propagated beyond the immune response itself. There is evidence for similar long-lived effects in other membraneless organelles as well - for example, the prion-like scaffold protein CPEB3 is capable of forming functional aggregate-like structures whose long-term persistence within synapses is involved in the facilitation of a fear conditioned memory[42–44]. Thus, it is possible that retention of immune system components within these assemblies post-infection may act as a 'molecular imprint' of previous exposure, and thus protect animals against reinfection.

This hypothesis, however, is contingent on a number of unknown factors that require future experimental work. Firstly, it would be interesting to determine whether the changes in puncta dynamics observed in response to PA14 treatment are maintained post-infection - if so, this would suggest a persistence of this response. In addition to this, determining the molecular interactions that ABU-13 makes within these puncta in different stress conditions - particularly post PA14 exposure - would allow us to determine how these puncta fit into the *C. elegans* immune system in general. It would be particularly interesting to identify if any immune system components are retained within these puncta post-exposure and, if this were the case, whether they might be involved in priming the animal for a future response. Similarly, in the spirit of a truly 'proto'-adaptive immune response, it is possible that, in keeping with the role that these puncta play in sequestering components from the environment, bacterial components may also be sequestered within these puncta, allowing for them to be more readily identified upon re-exposure. This underscores the pertinence of performing co-immunoprecipitation studies with ABU-13 to identify protein interactors in different environmental contexts, as this could unveil potential novel mechanisms of stress response.

### *3.3.6 Role of the epithelial system in stress responses*

A number of epithelial-derived cell types were predicted as potential tissues for *abu-13* expression, including the hypodermis, the amphid sheath (AMsh) glia and the arcade cells. The arcade cells, of which there are 9, form two separated epithelial syncytia and contain very little endoplasmic reticulum[45]. ABU-13 expression in these cells was not observed in our transcriptional reporter line, so we did not explore the role that arcade cells might play in ABU-13 function any further. We did, however, see localisation of ABU-13 around amphid neurons in the neuronal marker strain, as well structures

reminiscent of the AMsh glia within the transcriptional reporter, suggesting that we may indeed see ABU-13 expression within these sheath cells. These cells have been primarily linked with maintenance of the amphid neurons, with ablation of the AMsh glia causing morphological abnormalities in the dendritic tips of these neurons and preventing the detection of the dauer pheromone[46–48]. Although these functions do not seem related to functions we have identified for ABU-13, recent evidence has proposed a broad role for the cephalic sheath glia in coordinating global responses to ER stress[49]; thus, it is possible that the AMsh glia also play unidentified roles in systemic signalling of stress - a mechanism through which ABU-13 might act within these cells.

The widespread expression of ABU-13 within what appears to be the hypodermis and seam cells is of particular interest in the context of immune responses. These tissues possess a barrier function, protecting animals from the invasion of potentially harmful pathogens. This protection derives in part from the formation of the cuticle, which acts as a physical barrier to the outside world protecting against a range of environmental threats including bacterial infection and osmotic shock. Alongside other specialised epithelial cells, the hypodermis and seam cells secrete cuticle components including collagen, structural glycoproteins and many insoluble proteins, allowing for the formation of an exoskeleton-like structure to form[50]. Whilst some of the ABU-13 expressed does appear to be located towards the body surface, a lot is expressed throughout the syncytial hypodermis suggesting that, even though this prion-like protein may have a propensity for insolubility, its main function is not as a part of the cuticle. This does not, however, exclude a role for these puncta in the processing of cuticular components.

Beyond a role in the production and maintenance of the cuticle, these tissues play a role in immune signalling of invasive pathogens. This is particularly the case for pathogens that adhere to the cuticle surface, leading to the downstream activation of an innate immune response[51,52]. However, the primary route of *Pseudomonas* infection is via gut invasion[53] and thus, this suggests a cell non-autonomous mechanism is involved in the response of hypodermal ABU-13 puncta to PA14 exposure. Such a mechanism would allow other tissues to pre-emptively prepare for bacterial invasion, thus improving the chances of survival. Pre-emptive immune responses such as this have been identified in *C. elegans* previously - for example, upon cuticular injury, antimicrobial peptide expression is increased in anticipation of any opportunistic pathogenic invasion[54,55]. Hence, moving forward, it would be of interest to determine whether there is indeed a signalling pathway mediating the link between bacterial invasion in the gut and the structural changes observed in ABU-13 puncta, or whether there is a more direct mode of activation.

Overall, we propose that ABU-13 represents a scaffold protein capable of driving the assembly of multiphase biomolecular condensates capable of facilitating both ER stress and immune responses. The former may involve the context-dependent association of client molecules with the phase-separated

body, whilst the latter involves large structural changes to the overall puncta. Further work to determine the molecular interactions within these puncta in different environmental conditions will elucidate the mechanisms underlying these functions.

## 3.4 Material and Methods

*Construction of transgenic strains expressing extrachromosomal arrays*

To construct the pRT11[*abu-13p::abu-13::tagRFP*] plasmid, animals were first lysed in nematode lysis buffer (1X Platinum™ *Taq* High Fidelity DNA Polymerase (Invitrogen); 200μg/mL Proteinase K (Promega); $MgCl_2$ (New England Biolabs)), before amplification of the *abu-13p::abu-13* locus from the isolated genomic DNA. This PCR amplification was performed in 5 fragments, with each fragment flanked by overhangs complementary to the adjacent region.

To prepare these fragments for insertion into the expression vector pPD30.38, the plasmid was digested with HpaI and EcoRV following an hour-long incubation at 37°C. These fragments and the digested plasmid were then run on a 1% low-melting point agarose gel at 90V before excision and purification using a QiaQUICK Gel Purification Kit (Qiagen). The purified components were then combined, with 50ng of vector added along with a two-fold molar excess of each fragment. 10μL of Gibson Assembly Master Mix (NEB) was then added to this mix, and the total volume brought up to 20μL with the addition of deionised water. The sample was then incubated in a thermocycler for 60 minutes at 50°C before electroporation into DH5α electrocompetent cells. These transformed cells were plated onto ampicillin-containing TYE plates and incubated overnight at 37°C. Colonies were then singled out and grown in 2XTY for 8 hours before being spun down and collected. Recombinant plasmids were then purified from bacteria using a QIAprep Spin Miniprep Kit (Qiagen) and sequenced to confirm identity.

Following this, the full length *abu-13p::abu-13* transcript was amplified from this plasmid and *tagRFP* amplified from pDD286. The expression vector pPD30.38 was digested using HpaI and SalI. The fragments and vector were ligated and the subsequent recombinant plasmid purified as per the protocol outlined above.

For construction of the *abu-13p::tagRFP* transcriptional reporter plasmid (pRT31[*abu-13p:: tagRFP::unc-54 3'UTR]*), *abu-13p* and *tagRFP* were both amplified from pRT11 and inserted into pPD30.38 via Gibson assembly following digestion of the plasmid with HpaI and SalI.

For construction of pRT26, the plasmid containing *abu-13* without the prion domain (pRT26[*abu-13p::abu-13ΔPrLD::tagRFP::unc-54 3'UTR]*), the regions either side of the prion-like domain were amplified from pRT11 before insertion into a HpaI/SalI digested pPD30.38.

For construction of the pRT27[*pitx-1::gfp::unc-54 3'UTR*], the glial reporter line plasmid, the glial-specific promoter *itx-1* was amplified from nematode genomic DNA whilst *C. elegans* codon-optimised

*gfp* was amplified from pPD95.77. These fragments were then cloned into HpaI/EcoRV-digested pPD30.38 using Gibson Assembly, as outlined above.

These plasmids were then injected into L4 animals mounted on 2% agarose pads at 100ng/μL using a Femtojet® 4x (Eppendorf) and InjectMan® 4 (Eppendorf). Nematodes were recovered in M9 media and transferred to OP50-seeded NGM plates for recovery. F1 progeny were then screened for transgenic animals - these animals were then singled out onto individual plates and propagated on for several generations before use to maximise the penetrance of the extrachromosomal array.

*SapTrap cloning for endogenous insertion*

To produce the insertion plasmid for the SapTrap protocol, both 5' and 3' homology arms flanking the insert region were amplified in preparation, with the former amplified from pRT11 and the latter amplified following a nested PCR directly from genomic DNA. These homology arms were 57 bps in length, with an additional 3 bps added adjacent to the SapI recognition site, corresponding to the specific overhang sequences exposed following digestion. In addition to this, 10μM of complementary single-stranded oligonucleotides (ssODN) coding for the sgRNA sequence were mixed in 1X oligo annealing buffer (from a 1.25X buffer: 300uL of 1M HEPES [pH 7.5]; 1mL of 1M potassium acetate; 6.7mL of $H_2O$) and annealed following incubation of the mixture at 95°C in a thermocycler for 5 minutes. The annealed oligos were then cooled at room temperature for 10 minutes before diluting to 150nM in TE buffer. Note that within the annealed sgRNA duplex, each strand contained a similar 3 bp overhang to allow ligation into the destination vector later on.

Following this, a DNA mix was prepared of all components in preparation for cloning into a destination vector (pPDD379). This included 300fmol of the annealed sgRNA duplex, 100fmol of the 5' and 3' homology arms, 100fmol of the mKate2-containing plasmid pDD375, 100fmol of the self-excising cassette-containing plasmid pDD363, 100fmol of the C-terminal linker-containing plasmid pMLS287 and 100fmol of pDD379. The reaction mix was then made up of 1μL of DNA mix, 1μL of SapTrap reaction buffer (for 6X SapTrap reaction buffer: 24μL of 10X T4 DNA ligase buffer (NEB); 1uL of 1M potassium acetate; 15uL of $H_2O$), 1μL of SapTrap enzyme mix (for 6X SapTrap enzyme mix: 28μL of SapI (NEB); 6μL T4 polynucleotide kinase (NEB); 6μL 400U/uL T4 DNA ligase (NEB)) and 3μL of nuclease free water. This mixture was then incubated at room temperature overnight before bacterial transformation, performed via electroporation, as outlined previously.

*Protein-based CRISPR*

Adapted protocol described in Paix et al. (2015)[20]. Briefly, the repair template was prepared by amplifying *tagRFP* from pDD286, flanked by 35 bp homology arms on either side of the transcript. The

PCR product was subsequently gel purified. Before preparation of the injection mix, the crRNA was designed using the CCTop algorithm (https://crispr.cos.uni-heidelberg.de/), which predicts off and on target scores for that specific crRNA. The crRNA itself made up of a 20 nucleotide crRNA sequence corresponding to the region immediately upstream of a PAM motif (NGG) - the site of Cas9 cleavage - along with a universal sequence that allows the crRNA to anneal with a tracrRNA to make the full sgRNA. The tracrRNA interacts with Cas9 itself, facilitating the formation of ribonucleoprotein complexes that target specific genomic loci. Hence, the 1uL of 8ug/µL target crRNA (IDT), 0.4µL of the co-CRISPR *dpy-10* crRNA (IDT) and 5µL of 4ug/uL tracrRNA (IDT) were incubated together for 5 minutes at 95°C in the thermocycler, followed by a gradient cooling to 4°C.

After this, 10µL of 5µg/µL TrueCut™ Cas9 Protein v2 (Invitrogen) was added to the mix and incubated at 37°C for a further 10 minutes before leaving the RNP mix at room temperature for 10 minutes. Once cooled, 0.55µL of 500ng/µL *dpy-10* ssODN repair template (IDT) and up to 500ng/µL of the PCR repair template was added, along with nuclease free water to bring the volume of the injection mix up to 20µL, if necessary. The mix was then spun down at 13,000rpm for 10 minutes and the injection mixed used to inject early adult animals.

F1 progeny displaying the dumpy phenotype were single plated and, upon egg laying, subjected to PCR genotyping to identify any correct insertions.

*Nested CRISPR*

This protocol is adapted from Vicencio et al. (2019)[21] and shares many similarities with the Paix et al. (2015) protocol described above. It differs fundamentally its two-step nature, however, with an initial small insertion followed by a second larger insertion. The preparation of the sgRNA-Cas9 complex is the same as outlined previously. However, when adding the repair templates, rather than using a PCR repair template the initial insertion uses a 200 nucleotide 4nmol ultramer ssODN repair template (IDT) instead. This ssODN consists of a 45 bp 5' homology arm with the PAM from this first insertion mutated, a N terminal fragment from tagRFP, a C terminal fragment, and then a 45 bp 3' homology arm. Mutating the initial PAM site prevents recutting. The constituent components of the initial injection mix differ from that used previously;

| Component | Initial Concentration | Final Concentration | Volume (µL) |
|---|---|---|---|
| TrueCut™ Cas9 v2 | 10000ng/µL | 250ng/µL | 0.75 |
| tracrRNA | 320µM | 6.40µM | 0.60 |
| *dpy-10* crRNA | 50µM | 1.25µM | 0.75 |

| | | | |
|---|---|---|---|
| *abu-13* crRNA | 100μM | 5μM | 1.5 |
| *dpy-10* ssODN | 32.7μM | 0.92μM | 0.84 |
| *tagRFP* ssODN | 100μM | 2.20μM | 0.66 |
| Nuclease free water | | | 24.9 |
| Total volume | | | 30 |

The injection mix was prepped as before, and early adults singled out if they displayed the dumpy phenotype. Following egg lay, these animals were then lysed and genotyped via PCR, with gel electrophoresis using a 2% agarose gel used to identify any insertions. Non-dumpy offspring from homozygous animals were then singled out, and the genotyping repeated until homozygous animals identified. From this point, the second insertion step could proceed. Unlike the first insertion, a PCR product of tagRFP was used as the repair template; this repair template did not require the addition of *abu-13*-specific homology arms as the N and C terminal fragments inserted in the first step act as homology arms in this second step. The injection mix was prepared as before, using the following concentrations;

| Component | Initial Concentration | Final Concentration | Volume (μL) |
|---|---|---|---|
| TrueCut™ Cas9 v2 | 10000ng/μL | 250ng/μL | 0.75 |
| tracrRNA | 320μM | 10.25μM | 0.96 |
| *dpy-10* crRNA | 100μM | 2μM | 0.6 |
| *tagRFP* crRNA | 100μM | 8μM | 2.4 |
| *dpy-10* ssODN | 32.7μM | 0.92μM | 0.84 |
| *tagRFP* PCR | 1000ng/μL* | 570ng/μL | 17.1* |
| Nuclease free water | | | 7.35 |
| Total volume | | | 30 |

*Volume of tagRFP PCR product added depended on the concentration of the PCR product after gel excision and purification.

As with the initial step, following injection, F1 animals with the dumpy phenotype were singled out and genotyped following egg lay, with correct genotyping bands purified and sent for sequencing for further validation (Eurofins Genomics).

*Long, partially single-stranded DNA-based CRISPR*

This protocol utilises a dual PCR amplification process to produce double stranded products of differential sizes. Either tagRFP or WormScarlet was amplified from pDD286 or pSEM87 respectively, producing one set of PCR products the length of the fluorescent protein coding sequence itself and another using ultramer primers made up of 120 bp homology arms and 20 bps of the fluorophore sequence (IDT). Due to their length, a 1μM working concentration of each primer was used, diluted 1:100 from a 100μM stock solution, along with 20ng/μL of plasmid to optimise this amplification.

These products were purified using the QIAquick PCR Purification Kit (Qiagen), combined at a 1:1 ratio and incubated in a thermocycler at 95°C for 2 minutes to melt the double stranded DNA, before a gradient cooling to 4°C to allow reannealing of single strands from different PCR products. This resulted in a population of double stranded DNA flanked by single-stranded ends corresponding to the 120 bp 5' and 3' homology arms.

The injection mix was then prepared following the same protocol as the second insertion stop of the nested CRISPR protocol outlined above. Note that in the original Dokshin et al. (2019) protocol, rather than using a co-CRISPR marker, a co-injection marker was used - instead of this, we used the *dpy-10* co-CRISPR marker as before.

Due to the high error rate observed during this process, we proceeded to correct a small error in positively edited animals that resulted in a lack of fluorescent protein expression using the protocol for the first insertion step described in the nested CRISPR protocol.

*FRAP*

Animals were mounted on 2% agarose pads and imaged under a Zeiss 710 confocal microscope at 40X or 63X. To minimise imaging time and allow for maximal temporal resolution, the imaging window was restricted to small areas within the field of view. An initial region of interest (ROI) was highlighted corresponding to the area to be bleached, along with two further ROIs corresponding both a reference and a background region. Upon imaging, the first ROI was bleached 5 times consecutively using the 514nm and 561nm lasers set to 100% power following an initial imaging step to visualise the pre-bleach fluorescence intensity. A series of imaging steps were then repeated post-bleaching to monitor the changes in fluorescence of these regions over time.

For FRAP analysis, the background intensity was subtracted from both the fluorescence intensity values of the bleached and the reference regions. The bleached region was then normalised to the reference region, counteracting the effects of general bleaching during the imaging process. These values were then fitted to the following exponential recovery curve using Fiji (ImageJ) software;

$$y = a * \left(\frac{1}{e^{-b*x}}\right) + c$$

Within this exponential recovery curve, $a + c$ corresponds to the mobile fraction of the bleached region. This curve was then plotted using Prism 8 software.

*Stress Imaging*

For tunicamycin and *Pseudomonas* exposure, animals were treated as outlined in Section 2.4. For heat shock, however, age synchronized animals were collected at L4 and transferred onto unseeded plates. These plates were then placed in a 20˚C or 34˚C incubator for 6 hours prior to imaging.

Following induction of stress, animals were mounted on 2% agarose pads and imaged using a Zeiss 710 Confocal microscope, with either a 10X, 20X, 40X Oil, or 63X Oil objective. Images were then analysed using Fiji (ImageJ) software. To determine the normalised fluorescence intensity values for each condition, we first determined the corrected total cell fluorescence (CTFC) for each individual animal. To do this, animals were isolated within the image and measurements of integrated density and total area of the animal were taken, along with the average gray area of 3 different background regions. The CTFC was calculated as follows;

$$CTFC = Integrated\ Density - (Area\ of\ animal * Mean\ gray\ area)$$

An average CTFC value was determined for the control condition animals, and all CTFC values normalised to this average.

For puncta measurements, images were initially thresholded using Autothreshold settings. Whole animals were then isolated in the image and the Analyse Particles function used to measure the properties of these puncta. For all these measurements, Prism 8 software was used for statistical analysis.

*Expression prediction software*

To predict the expression pattern of *abu-13*, an online 'tissue-ome' database was used (http://worm.princeton.edu/), developed by Kaletsky et al. (2018). These prediction values were then inputted into Prism 8 software and a heat map generated.

*Freeze-crack and DAPI staining*

Prior to preparation of nematodes, a flat metal plate was cooled on a bed of dry ice to ensure ice-cold. Day 1 animals were then collected in M9 and washed 3 times to remove all bacteria, aspirating and

discarding the M9 after each wash. Animals were then resuspended in dd$H_2O$ and 10 - 20µL transferred onto polylysine coated slides, with individual animals being 'drawn out' using a pipette tip as to separate out individual nematodes. A coverslip was then added to the slide before placing it onto the pre-cooled metal plate for 2 hours. The coverslip was then carefully removed from the slide. Acetone was then added and incubated with nematode samples in a humid chamber for 30 minutes, after which, the slide was gently washed with PBS twice, making sure to remove as much liquid as possible without disrupting the prepared animals on the slide. DAPI was then added to the sample and left to stain animals for a further 30 minutes in a humid chamber. After a final set of PBS washes, a coverslip was secured to the slide and animals were imaged using a Zeiss 710 Confocal microscope.

*Protein purification from BL21 E. coli*

In preparation for protein expression, we amplified out the exonic sequences of *abu-13* and *abu-13ΔPrLD* from pRT11 and pRT26 respectively and, using Gibson assembly, cloned these into pPD30.38. For bacterial purification, we then amplified out the full open reading frames for these genes and, using FseI and AscI digestion sites, cloned them sequence into an adapted pET 28b expression vector. The resulting recombinant expression vectors (pRT28 and pRT29 respectively) were then transformed into BL21(DE3) *E. coli*, adding 1µL of plasmid to 50µL of bacteria. This mixture was placed on ice for 30 minutes before heat shock was applied at 42C for 10 seconds. The sample was then immediately placed back onto ice for 5 minutes before recovering bacteria in 950µL of SOC media. Following incubation of this mixture at 37C for 60 minutes in a shaking incubator, the liquid culture was spread on kanamycin selection plates and left at 37C overnight.

The resulting colonies were then scraped into 60mL of 2XTY and grown overnight to prepare a pre-culture. This pre-culture was added to 3L of 2XTY and grown in a shaking incubator set to 37°C until OD 0.8. Following a period of cooling at 4°C, 0.5µM of IPTG was added to induce expression of *abu-13* and incubated at 20°C overnight in a shaking incubator. Bacteria was then pelleted down at 6000rpm for 20 minutes at 4°C, supernatant discarded, and the remaining pellet resuspended in 30mL PBS. After spinning down at 3900rpm for 30 minutes at 4°C and discarding the supernatant once more, this washed pellet was then resuspended in 200mL of lysis buffer (20mM HEPES; 150mM KCl; 5% glycerol; 20mM imidazole, pH 7.7; 1% Triton X-100; 5mM $MgCl_2$; 10ug/mL DNase I; 1xcOmplete protease inhibitor cocktail (Roche); 1mM PMSF). The bacteria was then sonicated for 5 minutes, using a programme set to 5 seconds on/5 seconds off at a 40% amplitude. This sonicated bacteria was then pelleted at 16,000rpm for 30 minutes at 4°C to clear any remaining debris. The clarified supernatant was then collected.

For every 3mL of Ni-NTA agarose resin (Qiagen), 25mL of lysis buffer was added to equilibrate the beads, before spinning them down at 1000rpm. The buffer was then removed, replaced with the bacterial

lysate and left rotating overnight at 4°C. After this point, the incubated beads were left to settle for 30 minutes, the supernatant discarded before resuspending and washing in 1x extract buffer (for 10xXB: 1M KCl; 1mM $CaCl_2$; 10mM $MgCl_2$; 50mM sucrose; 100mM HEPES, pH 7.8) supplemented with 20mM imidazole, 4 times. 1mM DTT and 0.5mM EDTA was then added along with 1:50 v/v of TEV protease (NEB) and rocked overnight at 4°C. The sample was then dialysed twice with XB before incubating with 20mM imidazole for 2 hours.

The beads were then loaded into an Econo-Pac Chromatography Column (Biorad) pre-equilibrated with lysis buffer and the flow through collected before concentration of the sample to 1mL using an Amicon® Ultra-15 Centrifugal Filter Unit (Millipore). This concentrated sample was then ultracentrifuged at 100,000g for 5 minutes and the supernatant collected before being flask frozen and stored at -80°C.

*Protein purification from S2 cells*

In preparation for expression in S2 cells, the *abu-13* and *abu-13ΔPrLD* open reading frame sequences were amplified out of pRT28 and pRT29, and cloned into an adapted *gfp*-containing pMT expression vector gifted by Joseph Watson (Derivery lab, MRC LMB) following FseI and AscI digestion. The resulting recombinant plasmids (pRT29 and pRT30 respectively) were transfected into S2 cells using Lipofectamine® 3000 as follows. S2 cells were initially seeded and grown to between 70 and 90% confluence. Lipofectamine® 3000 Reagent (Invitrogen) was then diluted in Opti-MEM® Medium (Invitrogen), whilst 5μg of plasmid was diluted in Opti-MEM® Medium and P3000™ Reagent (Invitrogen) added. These two diluted reagents were then combined at a ratio of 1:1 and incubated for 5 minutes at 4°C before being added to cells. These cells were then incubated for 3-4 days at 37°C and then visualised to monitor growth.

*Protein purification from C. elegans*

For preparation of liquid culture, 3L of OP50 were grown in LB overnight at 37°C before being pelleted at 2500g for 10 minutes. This concentrated OP50 was then used to inoculate S-medium (1L S Basal [5.85g NaCl; 1g $K_2HPO_4$; 6 g $KH_2PO_4$; 1mL of 5mg/mL cholesterol; 999mL $H_2O$]; 10mL of 1M potassium citrate, pH 6 [for 1L: 20g citric acid monohydrate, 293.5g tri-potassium citrate monohydrate; 1L $H_2O$]; 10mL trace metals solution (for 1L: 1.86g disodium EDTA; 0.69g $FeSO_4$•7 $H_2O$; 0.2 g MnCl2•4 $H_2O$; 0.29 g $ZnSO_4$•7 $H_2O$; 0.025 g $CuSO_4$•5 $H_2O$; 1L $H_2O$]; 3mL 1M $CaCl_2$; 3 mL 1M $MgSO_4$). Several large plates of ABU-13::WormScarlet or ABU-13::3xFLAG animals recently cleared of bacteria were washed into this inoculated S-media and flasks transferred into a shaking incubator set at 20C.

These liquid cultures were monitored daily to ensure no bacterial growth or depletion of the food source. By the fourth or fifth day, when there were a significant number of adult animals for harvesting, these flasks were put on ice and cooled for 15 minutes allowing animals to settle. The liquid was then aspirated and animals washed in M9, before aspirating off the liquid once more and resuspending animals in RIPA buffer (10mM Tris-HCl, pH 7.5; 150mM NaCl; 0.5mM EDTA; 0.1% SDS; 1% Triton X-100; 1% deoxycholate; 1x cOmplete protease inhibitor cocktail (Roche); 1mM PMSF) before flash freezing in liquid nitrogen. These samples were then thawed on ice before being lysed by sonication. Lysis was then spun down at 4000g for 15 minutes and the clarified supernatant collected. The sample used with the RFP-Trap beads was then diluted in a dilution buffer (10mM Tris-HCl, pH 7.5; 150mM NaCl; 0.5nM EDTA; 1x cOmplete protease inhibitor cocktail (Roche); 1mM PMSF) at a ratio of 2:3, sample:buffer.

Prior to purification of the tagged protein, RFP-Trap® agarose beads (Chromotek) and ANTI-FLAG® M2 magnetic beads (Sigma) were vortexed to resuspend beads in their storage buffer. A 500µL bead slurry of each was then diluted in 10mL of dilution buffer. RFP-Trap beads were then spun down at 2500g for 2 minutes at 4°C, whilst anti-FLAG beads were separated using a magnetic separator before removing the supernatant and resuspending the beads in the same volume once again. This was repeated twice more. After the final wash, diluted lysate was added to the RFP-Trap beads, whilst undiluted lysate was added to the anti-FLAG beads. The lysate-bead mix was then left on a rotator overnight at 4°C.

Beads were then collected as described previously and supernatant discarded. After two further rounds of washing in dilution buffer, we aspirated off the supernatant and eluted the bound protein from the beads. For the RFP-Trap beads, we added 1mL of 0.2M glycine, pH 2.5 to the beads, mixing constantly for 30 seconds before immediately spinning down at 2500g for 2 minutes, collecting the supernatant and neutralizing it in 100µL of 1M Tris, pH 10.4. For the anti-FLAG beads, 1mL of 0.1 glycine, pH 3.0 was added and incubated with the beads for 10 minutes on a rotator before separating with a magnetic separator, collecting the supernatant and neutralizing with 200uL of 0.5M Tris-HCl, pH 7.4 with 1.5M NaCl. Collected eluate was then flash frozen and stored at -80°C until use.

*Protein purification for prion-like proteins*

Protocol adapted from Alberti et al. (2018)[56]. The 6xHis::MBP::3C::ABU-13::TEV::GFP expression vector was constructed in two stages; initially, the individual components were combined in pPD30.38 by Gibson Assembly. The MBP sequence was cloned out of pMal c5X gifted by Dr. Ketan Malhotra (Bertolotti lab, MRC LMB) with a 6xHis::3C protease site sequence added to the forward primer. The *abu-13* open reading frame was cloned from pRT28, with the addition of the 3C protease target site sequence at the 5' end whilst the *gfp* sequence was cloned out of an adapted pMT-GFP expression vector gifted by Joseph Watson (Derivery lab, MRC LMB), with the TEV-cleavage site and a flexible linker

sequence added to the forward primer. These were then assembled as previously described into pPD30.38 following a SalI/HpaI digestion.

Using this intermediate recombinant plasmid, either the full 6xHis::MBP::3C::ABU-13::TEV::GFP sequence or a GFP-only sequence were amplified and inserted into an adapted pRSET expression vector gifted by Joseph Watson (Derivery lab, MRC LMB), following digestion of all components with FseI and AscI, via T4 DNA ligase-mediated ligation overnight at 15C before being transformed into DH5α electrocompetent cells and prepped for use using methods previously outlined. Once prepared, 1uL of purified plasmid was added to 50μL of Rosetta™ 2(DE3)pLysS Singles™ Competent Cells (Novagen) and incubated on ice for 5 minutes. The mix was then heated for exactly 30 seconds at 42°C in a stationary thermoshaker and transferred back onto ice immediately for 2 minutes before being suspended in 950μL of SOC media and incubated at 37C for 1 hour in a shaking incubator. This liquid culture was then spread onto ampicillin-containing TYE plates and incubated overnight at 37°C.

Colonies from this plate were then scraped into 60mL of 2XTY and grown overnight at 37°C. 20mL of this pre-culture was then added to each litre of 2XTY (3 litres made overall), and grown in a shaking incubator until OD 0.8. The culture was then cooled at 4°C before the addition of 0.5μM of IPTG. These induced cultures were then grown in the shaking incubator overnight at 20C. The next day, bacteria was spun down at 4500rpm for 20 minutes at 4°C, the supernatant removed and the bacterial pellet resuspended in 30mL PBS. These resuspending bacterial pellets were then spun down once again, this time at 3900rpm for 30 minutes.

These washes pellets were then resuspended in lysis buffer (50mM Tris-HCl, pH 7.4; 1M KCl; 5% glycerol; 10mM imidazole; 1x cOmplete protease inhibitor cocktail (Roche); 1mM PMSF) and lysed by sonication. The lysis was then cleared of debris following centrifugation at 4500rpm for 20 minutes at 4°C and the supernatant incubated with pre-equilibrated Ni-NTA agarose resin (Qiagen) overnight. Agarose beads were collected via centrifugation at 1500rpm for 2 minutes and washed three times in lysis buffer, with the beads spun down once more after each wash. After the final wash, the beads were resuspended once more and added to pre-equilibrated Econo-Pac Chromatography Columns (Biorad). The beads were then washed once more with two columns worth of lysis buffer. Bound protein was then eluted from the beads using elution buffer (50nM Tris-HCl, pH 7.4; 1M KCl; 5% glycerol; 500nM imidazole).

This eluted protein was then incubated for 1 hour at room temperature with PreScission protease (Promega) at a ratio of 1:50 to cleave off the 6xHis-MBP tag. Buffer exchange and concentration of the purified protein samples was then performed using Amicon® Ultra-4 Centrifugal Filter Units (Millipore), subjecting the sample to an initial spin at 4500rpm for 15 minutes, and then three additional spins following addition of storage buffer to the concentrated sample (50mM Tris-HCl, pH 7.4; 500mM

KCl; 1mM DTT; 5% glycerol). These samples were then aliquoted, frozen in liquid nitrogen and stored at -80°C until use.

*SDS-PAGE and Western blot*

Initially, 10uL of 4x LDS Sample Buffer (Invitrogen) was added 30µL of purified protein sample and incubated for 10 minutes at 95°C. Samples were not standardised prior to electrophoresis. These heated samples were then loaded in duplicate onto a NuPage 4-12% Bis-Tris precast gel (Invitrogen) alongside a Precision Plus Protein Dual Color Standard (Biorad) protein ladder and run at 200V in 1x MOPS-SDS running buffer for 30 minutes.

Once run, the gel was removed from the plastic casing. The gel was then split in half, with the first set of samples being fixed in 40%EtOH/10% acetic acid for 15 minutes, rinsed in water and stained overnight in QC Colloidal Coomassie Stain (Biorad). This gel was then rinsed in water once more and imaged using a Chemidoc Touch Imaging System (Biorad), with the settings set to detect Coomassie.

The second set of samples was then subjected to a dry transfer onto an iBlot™ Transfer Stack PVDF membrane (Invitrogen) using iBlot® Dry Blotting System (Invitrogen). This membrane was then blocked for 2 hours in 5% bovine serum albumin dissolved in PBS-T, before the addition of either an RFP antibody (6G6, Chromotek) against WormScarlet at a 1:1000 concentration, or an anti-GFP antibody (ab290, Abcam) at a 1:10000 concentration. This membrane was then incubated at 4°C overnight before being washed 3 times in PBS-T and 5% BSA solution containing secondary antibody added. For the RFP antibody-probed membrane, 1:5000 of anti-mouse secondary antibody, whilst 1:10000 of anti-Rabbit HRP-conjugated secondary antibody (W401B, Promega) was used against the anti-GFP antibody. After a 2 hour incubation at room temperature, membranes were washed 3 times in PBS-T and then a final time in PBS. Imaging was then performed using a Chemidoc Touch Imaging System (Biorad), with the settings set to detect Chemiluminescence, following visualisation of HRP-conjugated antibodies with Pierce™ ECL Western Blotting Substrate (Thermo Scientific).

*In vitro phase separation assays*

The concentration of purified GFP and GFP-fused ABU-13 was calculated using an Eppendorf Biospectrometer® set to 488nm before standardising samples with storage buffer. Prior to imaging, airtight chambers were prepared using glass coverslips passivated with PLL-g-PEG (20K) at 1mg/mL in buffer containing 10mM Tris-HCl, pH 7.5 and 150mM KCl for 1 hour at room temperature. Experimental conditions were tested in a standard phase separation buffer (50mg/mL 20K PEG (Sigma); 25nM Tris-HCl, pH 7.4; 150mM KCl), with any deviation from this standard specified. 5µL of sample was loaded onto pre-prepared slides and imaged using a custom spinning disc confocal microscope.

## 3.5 References


1. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
2. Mitrea, D. M. *et al.* Structural polymorphism in the N-terminal oligomerization domain of NPM1. *Proc. Natl. Acad. Sci.* **111**, 4466 – 4471 (2014).
3. Kato, M. *et al.* Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767 (2012).
4. Hennig, S. *et al.* Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **210**, 529–539 (2015).
5. Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).
6. Ferrandon, D., Koch, I., Westhof, E. & Nüsslein–Volhard, C. RNA–RNA interaction is required for the formation of specific bicoid mRNA 3′ UTR–STAUFEN ribonucleoprotein particles. *EMBO J.* **16**, 1751–1758 (1997).
7. Fox, A. H., Nakagawa, S., Hirose, T. & Bond, C. S. Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends Biochem. Sci.* **43**, 124–135 (2018).
8. Van Treeck, B. *et al.* RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc. Natl. Acad. Sci.* **115**, 2734 – 2739 (2018).
9. Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016).
10. Andrusiak, M. G. *et al.* Inhibition of Axon Regeneration by Liquid-like TIAR-2 Granules. *Neuron* **104**, 290-304.e8 (2019).
11. Chakravarty, A. K. & Jarosz, D. F. More than Just a Phase: Prions at the Crossroads of Epigenetic Inheritance and Evolutionary Change. *J. Mol. Biol.* **430**, 4607–4618 (2018).
12. Mitrea, D. M. *et al.* Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *Elife* **5**, e13571 (2016).
13. Garre, E., Pelechano, V., Sánchez del Pino, M., Alepuz, P. & Sunnerhagen, P. The Lsm1-7/Pat1 complex binds to stress-activated mRNAs and modulates the response to hyperosmotic shock. *PLOS Genet.* **14**, e1007563 (2018).
14. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science.* **360**, 918 – 921 (2018).
15. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
16. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
17. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial

Immunity. *Science*. **337**, 816 – 821 (2012).

18. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci*. **109**, E2579 LP-E2586 (2012).
19. Garneau, J. E. *et al*. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
20. Paix, A., Folkmann, A., Rasoloson, D. & Seydoux, G. High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* **201**, 47 – 54 (2015).
21. Vicencio, J., Martínez-Fernández, C., Serrat, X. & Cerón, J. Efficient Generation of Endogenous Fluorescent Reporters by Nested CRISPR in *Caenorhabditis elegans*. *Genetics* **211**, 1143 – 1154 (2019).
22. Dokshin, G. A., Ghanta, K. S., Piscopo, K. M. & Mello, C. C. Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in *Caenorhabditis elegans*. *Genetics* **210**, 781 – 787 (2018).
23. Schwartz, M. L. & Jorgensen, E. M. SapTrap, a Toolkit for High-Throughput CRISPR/Cas9 Gene Modification in *Caenorhabditis elegans*. *Genetics* **202**, 1277 – 1288 (2016).
24. Dickinson, D. J., Slabodnick, M. M., Chen, A. H. & Goldstein, B. SapTrap assembly of repair templates for Cas9-triggered homologous recombination with a self-excising cassette. *microPublication Biology* (2018). doi:10.17912/W2KT0N
25. Urano, F. *et al*. A survival pathway for *Caenorhabditis elegans* with a blocked unfolded protein response. *J. Cell Biol*. **158**, 639 – 646 (2002).
26. Sun, J., Liu, Y. & Aballay, A. Organismal regulation of XBP-1-mediated unfolded protein response during development and immune activation. *EMBO Rep*. **13**, 855–860 (2012).
27. Gilks, N. *et al*. Stress Granule Assembly Is Mediated by Prion-like Aggregation of TIA-1. *Mol. Biol. Cell* **15**, 5383–5398 (2004).
28. March, Z. M., King, O. D. & Shorter, J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res*. **1647**, 9–18 (2016).
29. Decker, C. J., Teixeira, D. & Parker, R. Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae . *J. Cell Biol*. **179**, 437–449 (2007).
30. Teixeira, D. & Parker, R. Analysis of P-Body Assembly in Saccharomyces cerevisiae. *Mol. Biol. Cell* **18**, 2274–2287 (2007).
31. Rao, B. S. & Parker, R. Numerous interactions act redundantly to assemble a tunable size of P bodies in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci*. **114**, E9569 LP-E9578 (2017).
32. Feric, M. *et al*. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).

33. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science*. **357**, eaaf4382 (2017).
34. Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proc. Natl. Acad. Sci.* **108**, 4334 – 4339 (2011).
35. Jain, S. *et al*. ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487–498 (2016).
36. Patel, A. *et al*. ATP as a biological hydrotrope. *Science*. **356**, 753 – 756 (2017).
37. Kamal, M. *et al*. The marginal cells of the *Caenorhabditis elegans* pharynx scavenge cholesterol and other hydrophobic small molecules. *Nat. Commun*. **10**, 3938 (2019).
38. Haskins, K. A., Russell, J. F., Gaddis, N., Dressman, H. K. & Aballay, A. Unfolded Protein Response Genes Regulated by CED-1 Are Required for *Caenorhabditis elegans* Innate Immunity. *Dev. Cell* **15**, 87–97 (2008).
39. Lamitina, T. & Cherry, S. Dangerous Liaisons: The Apoptotic Engulfment Receptor CED-1 Links Innate Immunity to the Unfolded Protein Response. *Dev. Cell* **15**, 3–4 (2008).
40. Kumsta, C., Chang, J. T., Schmalz, J. & Hansen, M. Hormetic heat stress and HSF-1 induce autophagy to improve survival and proteostasis in *C. elegans*. *Nat. Commun*. **8**, 14337 (2017).
41. Olsen, A., Vantipalli, M. C. & Lithgow, G. J. Lifespan extension of *Caenorhabditis elegans* following repeated mild hormetic heat treatments. *Biogerontology* **7**, 221 (2006).
42. Si, K. *et al*. A Neuronal Isoform of CPEB Regulates Local Protein Synthesis and Stabilizes Synapse-Specific Long-Term Facilitation in Aplysia. *Cell* **115**, 893–904 (2003).
43. Drisaldi, B. *et al*. SUMOylation Is an Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. *Cell Rep*. **11**, 1694–1702 (2015).
44. Fioriti, L. *et al*. The Persistence of Hippocampal-Based Memory Requires Protein Synthesis Mediated by the Prion-like Protein CPEB3. *Neuron* **86**, 1433–1448 (2015).
45. Wright, K. A. & Thomson, J. N. The buccal capsule of *Caenorhabditis elegans* (Nematoda: Rhabditoidea): an ultrastructural study. *Can. J. Zool*. **59**, 1952–1961 (1981).
46. Vowels, J. J. & Thomas, J. H. Multiple chemosensory defects in *daf-11* and *daf-21* mutants of *Caenorhabditis elegans*. *Genetics* **138**, 303 – 316 (1994).
47. Perens, E. A. & Shaham, S. *C. elegans daf-6* Encodes a Patched-Related Protein Required for Lumen Formation. *Dev. Cell* **8**, 893–906 (2005).
48. Procko, C., Lu, Y. & Shaham, S. Glia delimit shape changes of sensory neuron receptive endings in *C. elegans*. *Development* **138**, 1371 – 1381 (2011).
49. Frakes, A. E. *et al*. Four glial cells regulate ER stress resistance and longevity via neuropeptide signaling in *C. elegans*. *Science*. **367**, 436 – 440 (2020).
50. Taffoni, C. & Pujol, N. Mechanisms of innate immunity in *C. elegans* epidermis. *Tissue Barriers* **3**, e1078432 (2015).
51. Hodgkin, J., Kuwabara, P. E. & Corneliussen, B. A novel bacterial pathogen, Microbacterium

nematophilum, induces morphological change in the nematode *C. elegans*. *Curr. Biol.* **10**, 1615–1618 (2000).

52. Muir, R. E. & Tan, M.-W. Virulence of *Leucobacter chromiireducens subsp. solipictus* to *Caenorhabditis elegans*: Characterization of a Novel Host-Pathogen Interaction. *Appl. Environ. Microbiol.* **74**, 4185 – 4198 (2008).
53. Tan, M.-W., Mahajan-Miklos, S. & Ausubel, F. M. Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *Proc. Natl. Acad. Sci.* **96**, 715 – 720 (1999).
54. Pujol, N. *et al*. Distinct Innate Immune Responses to Infection and Wounding in the *C. elegans* Epidermis. *Curr. Biol.* **18**, 481–489 (2008).
55. Tong, A. *et al*. Negative regulation of *Caenorhabditis elegans* epidermal damage responses by death-associated protein kinase. *Proc. Natl. Acad. Sci.* **106**, 1457 – 1461 (2009).
56. Alberti, S. *et al*. A User's Guide for Phase Separation Assays with Purified Proteins. *J. Mol. Biol.* **430**, 4806–4820 (2018).

# Chapter 4

## Understanding the RNA and protein interactions of ABU-13

## 4.0 Abstract

The formation of phase-separated structures is reliant on the multivalent interactions of a number of scaffold proteins, facilitating the coalescence of a wide range of protein and RNA components. These structures compartmentalise cellular function akin to a membrane-bound compartment, albeit lacking the membrane. Often these structures are involved in RNA processing of some kind – whether through the sequestration of relevant RNA machinery, preventing translation or other related processes, or via active RNA processes such as mRNA decapping.

Thus far, we have demonstrated that ABU-13 is capable of forming punctate structures that resemble those observed within phase-separated organelles. As such, gaining an insight into the nature of protein and RNA interactions within these structures is likely to aid our mechanistic understanding of their role within *C. elegans*. Although the ability to bind RNA has been predicted by domain conservation predictions in the N terminus of the protein, this ability has not been verified experimentally and thus, we initially sought to determine whether ABU-13 was an RNA binding protein. Using an adapted cross-linking and immunoprecipitation protocol, we show that this is indeed the case, further supporting the involvement of ABU-13 within phase-separated structures.

We next investigated the protein-protein interactions of ABU-13 by co-immunoprecipitating ABU-13::tagRFP and performing mass spectrometry on the pulled down proteins. In keeping with our hypothesis that this protein plays a role in ER stress responses and RNP complex formation, we see an overrepresentation of gene ontology terms relating to these processes.

As we hypothesised that ABU-13 is involved in the formation of hydrogel-like structures, we performed a biotinylated isoxazole precipitation of hydrogel-forming proteins to determine if ABU-13 is found amongst these proteins. This was not the case, and ABU-13 was not identified within this precipitate. This dataset, however, provides a novel insight into hydrogel-forming proteins in *C. elegans*, highlighting potential candidates for further investigation.

## 4.1 Introduction

### *4.1.1 RNA Binding Domains*

Multivalency is a powerful driver of phase separation due to the inherent propensity of multivalent proteins to interact with a variety of other cellular components, including proteins and RNA. This facilitates the coalescence of numerous components into a restricted cellular space, ultimately driving this process of phase transition. Amongst these phase separating proteins, there is an enrichment of both proteins with prion-like domains and those with RNA binding domains. Whilst the former is vital for driving self-assembly of these proteins due to their intrinsic proclivity for aggregation-like behaviour[1], the presence of RNA binding can act as a modulator of these processes, with evidence suggesting that RNA can act to buffer these transitions, keeping them in a more liquid-like state and preventing their transition to solid-like pathological aggregates[2].

Beyond this, however, these RNA associations often play important functions in the regulation of RNA – from sequestration to degradation – and thus understanding the specificity of RNA binding under different physiological circumstances can be a powerful tool for determining molecular function of these biological condensates. Furthermore, determining the sites of interaction with specific RNAs can point towards different types of regulation; for example, the RNA binding protein TIS11B is involved in the formation of TIS granules that form a meshwork around the endoplasmic reticulum. Within a subcompartment of this membraneless compartment, known as the TIGER domain, TIS11B facilitates the differential regulation of mRNA transcripts based on their 3'UTRs[3]. Thus, understanding both the identity of the RNA involved and the site of interaction can be useful for determining the molecular mechanisms at work within these organelles.

### *4.1.2 Hydrogels*

Physiological hydrogels allow phase-separated structures to stably persist within a cellular environment, as opposed to the often-transient formations observed with liquid-liquid phase transitions. Many of these structures are involved in multiphase systems, such as the nucleolus, where a tripartite structure of phases facilitates a sequential processing of ribosomal components. Similarly, biphasic structures are commonplace amongst some of the more transient compartments too – for example, stress granules contain a core-shell architecture whereby the outer layer has more liquid-like properties and the inner layer has more solid-like properties. This allows for retention of vital components of these condensates, whilst the outer layer allows for weaker associations of more transient components to occur.

It is possible to isolate proteins with a propensity for forming hydrogels by selectively precipitating them out from lysate using the chemical biotinylated isoxazole (b-isox) (Figure 4.1). This has allowed the sequence determinants of such transitions to be explored. Notably, low complexity domains were shown to be both necessary and sufficient for this precipitation. This has been illustrated using the stress granule-associated protein, TIA1; evidence from mice has shown that this protein is precipitated from a range of cell and tissue types following homogenisation and b-isox treatment. This precipitation is abolished, however, if the low complexity domain is removed – this is not the case when the RNA recognition motifs are removed, with precipitation occurring as before in this context. Despite this, RNA binding proteins are highly enriched amongst this dataset, with a highly significant representation of RNA granule proteins found within these precipitated protein groups in comparison to the much lower frequency of such proteins within the mouse proteome in general[4].

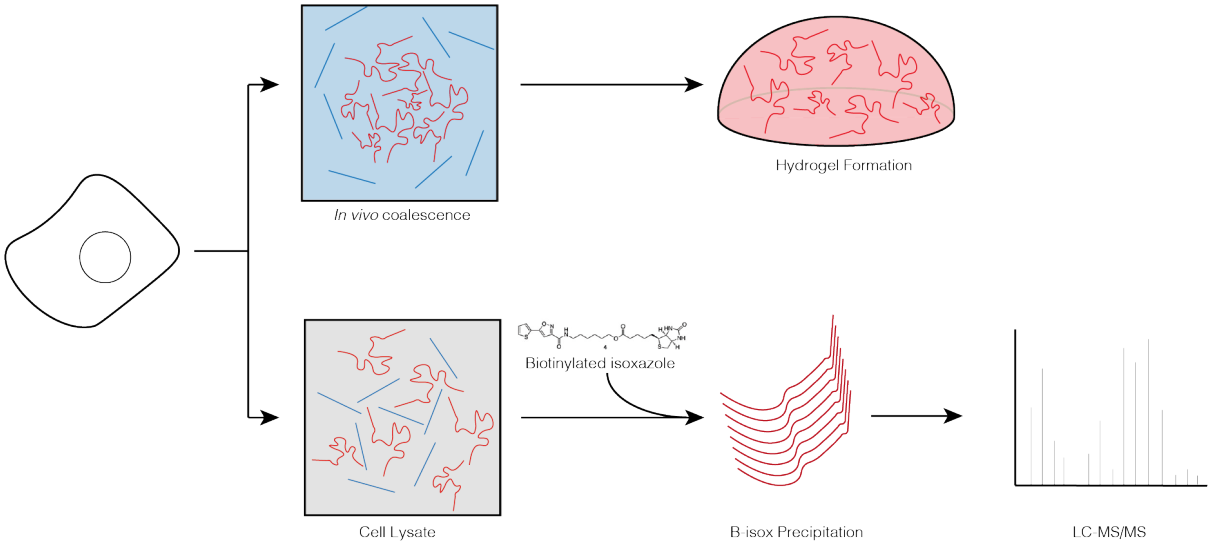



**Figure 4.1** Schematic depicting the experimental process allowing biotinylated isoxazole precipitation of hydrogel forming proteins

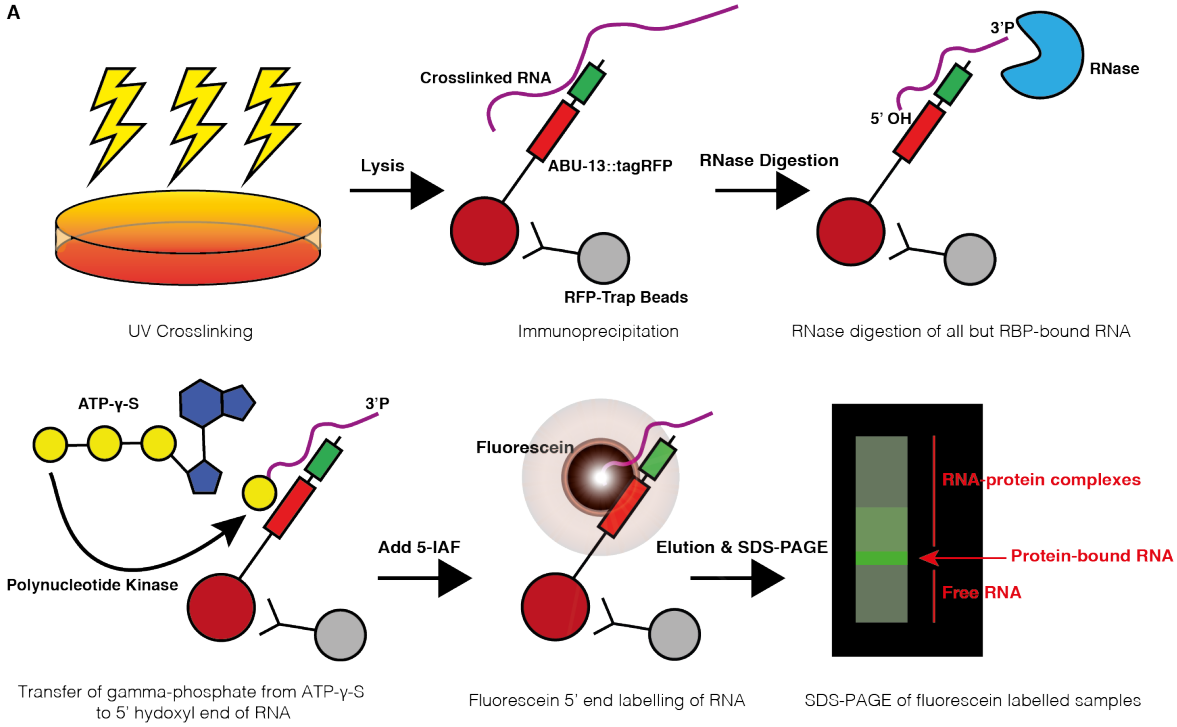
Within these low complexity domains, there is also a correlation between hydrogel formation and the presence of [G/S]Y[G/S] motifs. This has been shown *in vitro* by the abolition of hydrogen formation of purified FUS and hnRNPA2 LC domains upon mutation of these tyrosine residues. These *in vitro* hydrogel structures exhibit diffraction rings of 4.6-4.7Å and 10Å, suggesting they form cross-β structures, similar to those seen in a range of prion-like and amyloid-like protein fibrils. However, unlike pathogenic amyloid fibrils, these structures are readily disassembled and display a more dynamic nature, capable of heterotypic incorporation of other RNA binding proteins into pre-existing polymers. Similar findings have also been shown for the prion-like domain of RBM14, involved in paraspeckle formation, which can form an amyloidogenic hydrogel as well[5]. As with FUS and hnRNPA2, these hydrogels show diffraction reflections at around 4.6Å and 10Å, highly reminiscent of the amyloid structures previously discussed. Thus, it is clear that low complexity domains play a role in driving the transition of RNA binding proteins into hydrogel states.

In this chapter we will demonstrate that ABU-13 is indeed an RNA binding protein, capable of interacting with RNA whilst the protein is in its full-length state – but not when the prion-like domain is removed. This further supports the role of this protein in phase transitions, as RNA binding proteins are overrepresented in such processes. Beyond RNA, we next sought to identify the protein interaction partners of ABU-13, showing an enrichment of other RNA interacting proteins immunoprecipitated alongside ABU-13, suggesting a role for this protein in the regulation of RNA. In addition to this, a number of glycosylation-related proteins were pulled down as well, pointing towards a further role for ABU-13 in the folding quality control of glycosylated proteins. To investigate the hypothesis that ABU-13 represents a species of hydrogel-forming proteins, we took an unbiased approach, precipitating hydrogel-forming protein en mass from wild type animals using b-isox. We did not find ABU-13 amongst these precipitated proteins suggesting that it may not in fact drive the formation of a hydrogel-like structure *in vivo*. This dataset, however, represents a novel characterisation of the hydrogel proteome of *C. elegans*, with a number of interesting candidate proteins identified that warrant further investigation going forward.

## 4.2 Results

### *4.2.1 ABU-13 is an RNA binding protein*

To test whether ABU-13 represents a species of RNA binding protein, as predicted by conserved domain searches, we adapted protocols for crosslinking and immunoprecipitation (CLIP)[6], and nucleic acid end-labelling[7], allowing us to pull-down ABU-13 following UV crosslinking of the protein to any RNA interactors *in vivo* (Figure 4.2A).



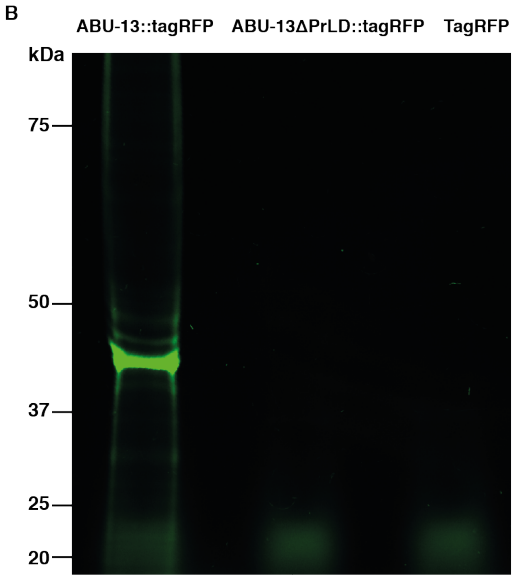




**Figure 4.2** ABU-13 is capable of binding RNA, (A) a schematic illustrating the adapted CLIP protocol used to determine whether ABU-13 is an RNA binding protein; (B) SDS-PAGE of ABU-13::tagRFP, ABU-13 Δ::tagRFP and tagRFP alone, following cross linking and immunoprecipitation

This RNA could then be chemically altered, using polynucleotide kinase to facilitate the transfer of a gamma phosphate group from ATP-γ-S to the 5'-hydroxyl end of any bound RNA, priming this 5' end for conjugation with 5-iodacetamidofluorescein (5-IAF). SDS-PAGE of these immunoprecipitated samples allowed any RNA-bound protein complexes to be directly visualised due to the fluorescent properties of this compound. Using extrachromosomal array lines, ABU-13::tagRFP, ABU-13ΔPrLD::tagRFP and tagRFP alone were tested for their abilities to bind RNA. Out of these three samples, only the full length ABU-13 co-precipitated with RNA, suggesting that ABU-13 is indeed capable of binding RNA. ABU-13ΔPrLD::tagRFP and tagRFP, on the other hand, showed no ability to bind RNA (Figure 4.2B). This suggests that the RNA binding abilities observed by the ABU-13::tagRFP animals were not a consequence of the fluorescent tag and may also be dependent upon the prion-like domain.

These RNA binding results could not, however, be tested in our endogenous expression systems due to the inability of the anti-FLAG or anti-RFP immunoprecipitation beads to successfully pull down sufficient ABU-13::3xFLAG or ABU-13::WormScarlet.

### *4.2.2 ABU-13 interacts with a range of proteins*

To determine the protein interactions of ABU-13, we co-immunoprecipitated tagRFP-fused ABU-13 using RFP-Trap beads (Chromotek) and performed mass spectrometry to identity any proteins pulled down alongside this protein, compared to those in a tagRFP control. Whilst there were many proteins identified in both samples, we applied thresholds of detection to eliminate background proteins that represented non-specific hits shared between both conditions. This involved filtering out all the proteins that showed a less than 50% increase between tagRFP and ABU-13::tagRFP. Within this filtered dataset, a number of proteins relating to particular processes were significantly overrepresented (Figure 4.3). Many of these were processes involved in ER-localised processes, RNA processing and translation; this is in keeping with our previous findings that ABU-13 contains a signal sequence for ER targeting and plays a non-redundant role in ER homeostasis. In addition to this, many terms relating to occyte development were significantly overrepresented – whilst these could represent a role for ABU-13 in the development of eggs, they may also suggest a slight difference in the developmental stage at which nematode populations were harvested for the co-IP. The harvesting of animals at Day 1 of adulthood could explain this result – at this stage, animals are still undergoing significant oogenesis and thus any mismatch - even if minor - in the ages between the nematodes of each sample group could result in a significant overrepresentation of genes involved with these responses.

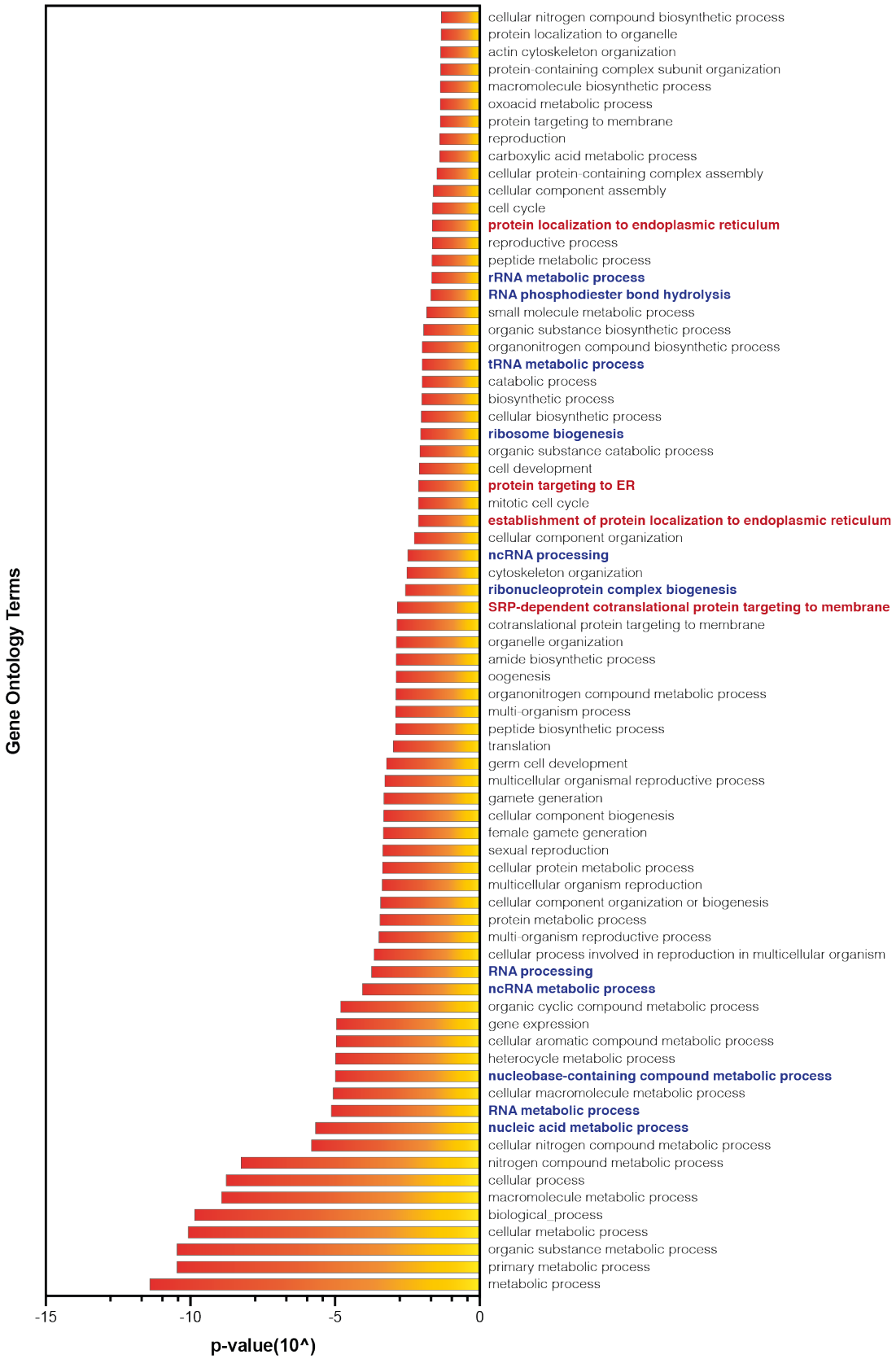



**Figure 4.3** Gene Ontology terms for proteins identified as overrepresented within the mass spectrometry results of a tagRFP-fused ABU-13 pull-down compared to a tagRFP control, as determined using PANTHER Overrepresentation Test, with proteins ranked by significance (Fisher test, with FDR adjustments to correct for multiple comparisons); red = terms relating to the ER and stress responses; blue = RNA processing terms.

### *4.2.3 ABU-13 interacts with stress related proteins, including those of the IRE-1 branch of the UPR*

Following our initial results that ABU-13 plays a non-redundant role in the ER stress response, as demonstrated by the developmental deficits observed in knockout animals exposed to tunicamycin, we sought to determine if any known ER stress response proteins were identified as interacting with ABU-13. This was indeed the case, with many proteins related to stress responses shown to be overrepresented (Table 4.1). The presence of ER resident proteins, alongside those involved with ER stress responses, suggests that ABU-13 is indeed localised to this compartment.

Components of the ERAD pathway were also pulled down, including UBXN-4, which positively regulates RNA Polymerase II transcription as part of the ER stress and ERAD responses. This is particularly interesting due to the homology of the N-terminal region of ABU-13 with RNA Polymerase II – if this homology region shares structural similarity to RNA Polymerase II, UBXN-4 may also be involved in its regulation too. In addition to this, a number of interactors associated with the IRE-1 branch of the unfolded protein response were identified; these represent promising candidates when considering the role of ABU-13 in the non-canonical UPR pathway following inhibition of *xbp-1*. UGGT-1, for example, is an UDP-glucose glycoprotein glucosyltransferase found in the lumen of the ER. As well as playing a role in the IRE-1-mediated UPR, this protein also functions in larval development. Among the ER stress responsive proteins identified, a number of IRE-1 pathway-related proteins were also pulled down – these include the GTP binding protein, TAG-335, and the ribosomal binding protein, K12H4.2.

| | Mapped ID | Gene Name | PANTHER Protein Class |
|---|---|---|---|
| Protein targeting to the ER | *trap-3* | TRanslocon-Associated Protein | - |
| | *K12H4.4* | Probable signal peptidase complex subunit 3 | protease |
| | *asna-1* | ATPase asna-1 | nucleotide phosphatase |
| | | | transporter |
| | *ZK512.4* | Signal recognition particle 9 kDa protein | RNA binding protein |
| | | | receptor |
| ERAD Pathway | *mans-3* | alpha-1,2-Mannosidase | chaperone |
| | *F44B9.5* | Ancient ubiquitous protein 1 homolog | acyltransferase |
| | *erl-1* | ERLin (ER lipid raft associated protein) homolog | - |
| | *ubxn-4* | UBX domain-containing protein 4 | - |
| | *uggt-1* | UDP-Glucose Glycoprotein glucosylTransferase | glycosyltransferase |
| Response to ER stress | *ubc-13* | Ubiquitin-conjugating enzyme E2 13 | - |
| | *tag-335* | Mannose-1-phosphate guanyltransferase beta | - |
| | *Y54G2A* | 4D656 | membrane traffic protein |
| | | | transporter |
| | *K12H4.4* | Probable signal peptidase complex subunit 3 | protease |
| | *mans-3* | alpha-1,2-Mannosidase | chaperone |
| | *F44B9.5* | Ancient ubiquitous protein 1 homolog | acyltransferase |

| | | | |
|---|---|---|---|
| | *K12H4.2* | Uncharacterized protein K12H4.2 | - |
| | *erl-1* | ERLin (ER lipid raft associated protein) homolog | - |
| | *gale-1* | UDP-glucose 4-epimerase | - |
| | *Y54G2A.23* | MANF/CDNF-like protein | - |
| | *ubxn-4* | UBX domain-containing protein 4 | - |
| | *uggt-1* | UDP-Glucose Glycoprotein glucosylTransferase | glycosyltransferase |
| Cellular Response to Stress | *ubc-13* | Ubiquitin-conjugating enzyme E2 13 | - |
| | *his-3* | Histone H2A | - |
| | *tag-335* | Mannose-1-phosphate guanyltransferase beta | - |
| | *Y54G2A* | 4D656 | membrane traffic protein |
| | | | transporter |
| | *him-1* | Structural maintenance of chromosomes protein 1 | - |
| | *rpa-2* | Replication Protein A homolog | - |
| | *K12H4.4* | Probable signal peptidase complex subunit 3 | protease |
| | *mans-3* | alpha-1,2-Mannosidase | chaperone |
| | *F44B9.5* | Ancient ubiquitous protein 1 homolog | acyltransferase |
| | *smk-1* | SMEK (Dictyostelium Suppressor of MEK null) homolog | - |
| | *K12H4.2* | Uncharacterized protein K12H4.2 | - |
| | *erl-1* | ERLin (ER lipid raft associated protein) homolog | - |
| | *gale-1* | UDP-glucose 4-epimerase | - |
| | *haf-1* | HAlF transporter (PGP related) | cysteine protease |
| | | | serine protease |
| | *flr-4* | Serine/threonine-protein kinase flr-4 | - |
| | *Y54G2A.23* | MANF/CDNF-like protein | - |
| | *ubxn-4* | UBX domain-containing protein 4 | - |
| | *Y54G2A.50* | Uncharacterized protein | DNA helicase |
| | *uggt-1* | UDP-Glucose Glycoprotein glucosylTransferase | glycosyltransferase |
| | *ragc-1* | RAs-related GTP binding protein C homolog | small GTPase |
| | *smc-6* | Structural maintenance of chromosomes protein 6 homolog smc-6 | - |

**Table 4.1** Summary of ER and stress related gene ontology terms corresponding to the proteins significantly overrepresented amongst ABU-13::tagRFP samples following mass spectrometry of co-immunoprecipitated interaction partners

### *4.2.4 ABU-13 interacts with proteins involved with RNA processing*

In addition to the ER-related terms, there is also an overrepresentation of terms relating to RNA processing (Table 4.2) – and within these, of particular interest, are terms relating to ribonucleoprotein complex assembly. As we have previously established, ABU-13 forms punctate structures within nematodes reminiscent of phase separated organelles – as such, the presence of other RNP complex components is consistent with the hypothesis that ABU-13 forms phase separated structures. Amongst these potential interactors are a number of translation initiation factors and ribosome biogenesis factors – this suggests that that ABU-13 may play a role in sequestering translation machinery to alleviate protein folding load on the ER.

| | Mapped ID | Gene Name | PANTHER Protein Class |
|---|---|---|---|
| Ribonucleoprotein complex assembly | *eif-2d* | Eukaryotic Initiation Factor | receptor |
| | *eif-3.E* | Eukaryotic translation initiation factor 3 subunit E | translation initiation factor |
| | *eif-3.C* | Eukaryotic translation initiation factor 3 subunit C | translation initiation factor |
| | *pptr-2* | Serine/threonine protein phosphatase 2A regulatory subunit | protein phosphatase |
| | *C43E11.9* | 60S ribosome subunit biogenesis protein NIP7 homolog | - |
| | *ZC434.4* | Ribosomal RNA-processing protein 7 homolog | ribosomal protein |
| | *lpd-6* | LiPid Depleted | - |
| | *Y54E10A.10* | Ribosome production factor 2 homolog | nuclease |
| | *snr-4* | Probable small nuclear ribonucleoprotein Sm D2 | mRNA processing factor |
| | *K12H4.3* | Ribosome biogenesis protein BRX1 homolog | ribosomal protein |
| RNA Processing | *elpc-1* | Elongator complex protein 1 | kinase activator |
| | *teg-4* | Tumorous Enhancer of Glp-1(Gf) | damaged DNA-binding protein |
| | | | mRNA polyadenylation factor |
| | | | mRNA splicing factor |
| | *riok-1* | Serine/threonine-protein kinase RIO1 | - |
| | *emb-4* | Uncharacterized protein | DNA helicase |
| | | | RNA helicase |
| | | | endoribonuclease |
| | | | hydrolase |
| | *snu-23* | Putative zinc finger matrin-type protein snu-23 | - |
| | *ssb-1* | Sjogren Syndrome antigen B homolog | ribonucleoprotein |
| | *rbd-1* | RBD (RNA binding domain) protein | mRNA splicing factor |
| | *Y54G2A.12* | Uncharacterized protein | mRNA splicing factor |
| | *B0024.11* | Putative pseudouridine synthase B0024.11 | - |
| | *ZC434.4* | Ribosomal RNA-processing protein 7 homolog | ribosomal protein |
| | *dcr-1* | Endoribonuclease dcr-1 | endodeoxyribonuclease |
| | *lsm-1* | LSM Sm-like protein | mRNA splicing factor |
| | | | nuclease |
| | *Y54E10A.10* | Ribosome production factor 2 homolog | nuclease |
| | *ZK686.2* | Putative ATP-dependent RNA helicase ZK686.2 | - |
| | *snr-4* | Probable small nuclear ribonucleoprotein Sm D2 | mRNA processing factor |
| | *Y54G2A.75* | Uncharacterized protein | - |
| | *B0511.7* | Uncharacterized protein | - |
| | *B0511.6* | RNA helicase | - |
| | *rps-7* | 40S ribosomal protein S7 | ribosomal protein |

**Table 4.2** Summary of RNA-related gene ontology terms corresponding to the proteins significantly overrepresented amongst ABU-13::tagRFP samples following mass spectrometry of co-immunoprecipitated interaction partners

Beyond these RNP components, there are also a number of mRNA splicing factors that coIP alongside ABU-13. This includes splicing factors involved in nematode development such as RBD-1, which coincidentally is also found within the phase-separated structure, the nucleolus, once more hinting at the potential involvement of ABU-13 in phase transitions. This is further supported by the presence of LSM-1 in that dataset – this protein, involved in RNA capping and degradation, has also been shown to localise in P-bodies and stress granules. Overall, these overrepresented RNA-related terms are in keeping with

our hypothesis that ABU-13 is involved with multivalent interactions with other RNA interacting proteins, potentially through the coalescence of phase-separating components. It is important, however, to recognize the possibility that some of these co-immunoprecipitated proteins represents non-physiological interactions occurring post-lysis. Further validation and characterisation of these candidate interactors would clarify this further.

### *4.2.5 Identification of hydrogel forming proteins in C. elegans*

As our *in vivo* work suggested involvement of ABU-13 in both liquid-like and hydrogel-like structures, we decided to take an unbiased approach to identifying hydrogel-forming proteins in wild type animals. This involved selective precipitation of these proteins using biotinylated isoxazole (b-isox), before subjecting the precipitate to mass spectrometry. Although this approach has been undertaken in a number of species-derived cells and tissue homogenates, including *Drosophila* S2 cells, mouse NIH 3T3 embryonic stem cells, as well as mouse brain and testis homogenate[4], this approach has never been performed in *C. elegans*. As we hypothesised that ABU-13 was capable of forming a hydrogel-like structure, extracting this information from N2 animals would allow us to determine this in the absence of any genetic modification or protein tagging, ensuring that all protein components were in their native conformation.

Unlike the original b-isox precipitation study[4], we did not identify the same 106 'core' proteins – however, this may reflect species-specific differences between proteins involved in the formation of hydrogels. On the other hand, however, the proteins precipitated from the nematode lysate showed a strong overrepresentation of RNA binding terms when compared to the DMSO control treated animals (Figure 4.4A), a finding that is in keeping with these previous experimental datasets. RNA binding proteins are highly enriched amongst proteins that facilitate phase separation, thus their enrichment within proteins involved in solid-like phase separation makes sense. This population likely represents a subset of scaffold proteins that possess a propensity for pathological transition – indeed, within our dataset, TDP-1 is identified, the *C. elegans* homolog of the ALS-associated protein TDP-43. However, among these hydrogel-forming proteins precipitated by b-isox, ABU-13 was not identified, suggesting that ABU-13 may not act as a hydrogel-forming protein.

Despite the absence of ABU-13, this dataset of b-isox precipitated hydrogel-forming proteins from *C. elegans* is novel and as such we explored these results further. Initially we compared the 'prion-like' abilities of these proteins, plotting Prion Propensity (PAPAprop) against the output values from the Viterbi-parsed HMM (Figure 4.4B). The majority of the results cluster in low scoring regions, indicating that they do not represent prion-like proteins as predicted by these algorithms. There are, however, a few that score higher, indicating that these may indeed represent protein with prionogenic abilities – these include MDT-15, MUT-16, GLD-2, PBRM-1, AMA-1, DUR-1, HUM-1, MSI-1, SMG-6 and

PUF-6. These proteins may therefore be hydrogel-forming scaffold proteins that utilise prion-like domains to drive their phase transitions, although further work is needed to verify this.

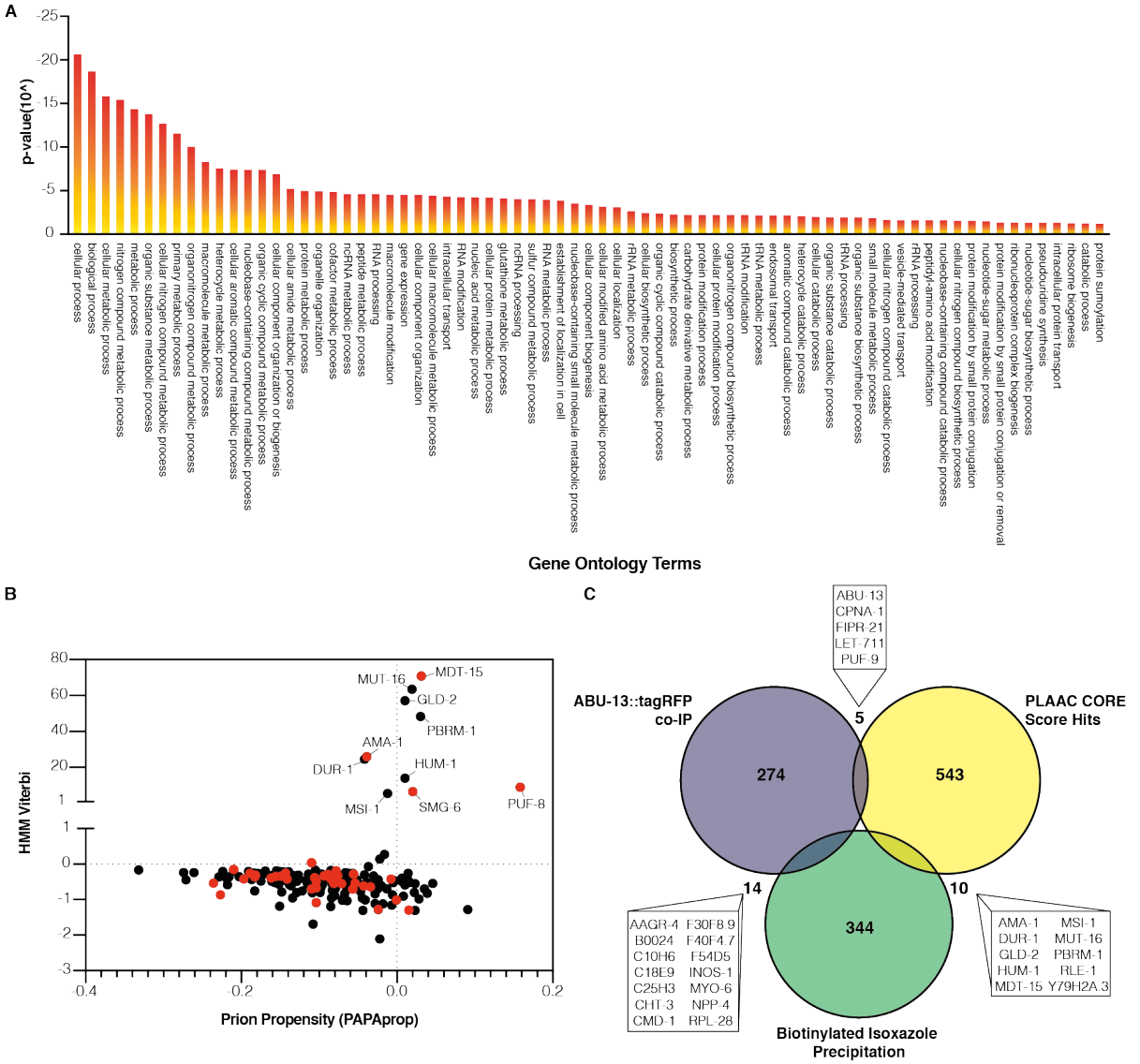



**Figure 4.4** Summary of results from biotinylated isoxazole precipitation of N2 animals showing (A) Gene Ontology terms for proteins identified as being overrepresented within the mass spectrometry results of b-isox precipitated samples compared to DMSO control samples, as determined using PANTHER Overrepresentation Test, with proteins ranked by significance (Fisher test, with FDR adjustments to correct for multiple comparisons); (B) Prion-like domain scores for b-isox samples, black = non-RNA associated proteins, red = RNA-related protein (*Unpaired t-test, p = 0.5234, no significant difference*); (C) Venn diagram comparing the proteins identified following co-immunoprecipitation of ABU-13::tagRFP (purple), with CORE scores within the PLAAC algorithm output (yellow), and with those proteins precipitated following b-isox treatment (green)

As prionogenic proteins tend to contain both a prion-like domain and RNA interacting motifs, we next sought to determine whether the RNA binding proteins overrepresented within this dataset were more likely to be predicted prion-like proteins. Whilst a limited number of these did crop up in the highest scoring candidates within this dataset (MDT-15, AMA-1, SMG-6 and PUF-6), the majority clustered alongside the other low scoring candidates and demonstrated no significant difference in the prion-like prediction scores from the non-RBPs amongst this set of proteins ($p$ = 0.5234). This suggests that hydrogel forming proteins within *C. elegans* do not rely on the presence of a prion-like domain to facilitate their liquid-to-solid transitions.

We then went on to compare the proteins identified from a number of these approaches to see if there was any overlap that might prove physiologically promising (Figure 4.4C). Specifically we compared; (1) co-immunoprecipitation of ABU-13::tagRFP; (2) b-isox precipitation; and (3) PLAAC algorithm results. Whilst there were no proteins shared between all three of these datasets, there are a number that were shared between individual categories. For example, there are a number of RNA-related proteins shared between the b-isox dataset and the PLAAC scores, including AMA-1 and GLD-2, involved with RNA polymerisation and polyadenylation respectively. In addition to this, there are also a number of proteins involved that are flagged as common between the ABU-13 pull-down and the b-isox precipitate, suggesting that these might represent hydrogel-forming proteins associated with ABU-13 puncta – these include AAGR-4, CHT-3, INOS-1 and NPP-4 amongst others. NPP-4, for example, is particularly interesting as this is the ortholog of a human nucleoporin that, as previously discussed, is involved in the formation of physiological hydrogels. Finally, we also compared the ABU-13 co-immunoprecipitation results to the PLAAC scores, identifying a small number of proteins, some of which exhibited promising functionality – PUF-9, for example, has been predicted to possess mRNA 3'UTR binding activity[8], reminiscent of TIS11B in TIGER domains. LET-711, on the other hand, has been implicated in larval development and the regulation of gene expression through its involvement in P-body activity, via recruitment of LSM-1 and LSM-3[9] – this is particularly interesting considering the presence of LSM-1 within these pulled down proteins as well, implicating a role for ABU-13 with P-body components. Overall, there do appear to be candidate proteins shared between these groups that represent interesting avenues for future investigation, with some of these candidates pointing towards potential pathways in which ABU-13 might be involved.

## 4.3 Discussion

### *4.3.1 Protein-RNA Interactions*

The ability of ABU-13 to bind RNA validates the prediction of nucleic acid interaction domains in the N terminus of the protein, although the exact sites of interaction are yet to be determined. This is in keeping with our hypothesis that ABU-13 is a prion-like protein involved in phase transitions, as the combination of PrLDs and RNA binding domains are highly represented amongst proteins involved with this process. It is, however, particularly interesting to note the lack of RNA binding ability of ABU-13ΔPrLD. There are a number of possible explanations for this result. Most likely, removal of this domain disrupts proper folding of the rest of the protein, ultimately preventing RNA binding due to a general structural perturbation. Alternatively this could indicate that the prion-like domain is necessary for RNA binding either directly, via interactions with residues within this intrinsically disordered region, or indirectly as a consequence of the reliance of the overexpressing animals on the prion-domain for incorporation into phase separated compartments. For example, it is possible that ABU-13 does not bind RNA in its diffuse state, as is observed in ABU-13ΔPrLD::tagRFP animals, and instead requires recruitment to phase separated structures before this can occur. If this were the case, this would suggest that target RNAs are 'shepherded' to ABU-13, rather than this protein acting as a recruiter of RNA itself. This would indicate that ABU-13 is acting to stabilise these structures via a multivalent array of interactions, or that it is potentially playing a functional role with its RNA interactions. Optimisation of endogenous ABU-13 pull-downs would aid this understanding – this is particularly the case considering the different behaviours of the ABU-13ΔPrLD species observed between our two models. If failure to localise ABU-13 within these phase separated compartments is indeed why the ABU-13ΔPrLD::tagRFP animals do not exhibit any RNA binding activity following CLIP, then, for this theory to be true, we would hypothesise that the endogenous ABU-13ΔPrLD::WormScarlet animals would demonstrate RNA-binding ability due to their localisation within these structures.

Differential binding of RNA species under specific physiological conditions could be a mechanism through which ABU-13 exerts its functional effects. The optimisation of the pull down technique using endogenously tagged ABU-13 would also enable us to identify which RNAs are bound under a range of conditions – i.e. in basal conditions, under ER stress and after exposure to an immune threat. If these were indeed variable, the identity of these RNA species would help to elucidate the role of ABU-13 and the puncta in which it is found.

### *4.3.2 Protein-Protein Interactions*

The use of overexpression models in co-immunoprecipitation studies presents some challenges in extracting biologically relevant information. This is especially the case in light of recent evidence

demonstrating that phase separated organelles adopt different morphologies when their scaffold components are overexpressed compared to expression at endogenous levels[3]. However, as we have shown with our *in vivo* endogenous tagging, ABU-13 does appear to be upregulated in situations of stress – thus an overexpression system may be recapitulating some of the interactions involved in such responses. This is supported by the identification of a number of components relevant to the pathways that ABU-13 has been implicated in. Some of the most interesting hits resulting from these mass spectrometry results are those implicated in the IRE-1 branch of the $UPR^{ER}$, due to the role of the ABU family of proteins in responding to a blockage in the IRE-1/XBP-1 branch - these candidates potentially represent physiologically-relevant interactors.

One of these candidates of interest is the UDP-glucose glycoprotein glucosyltransferase, UGGT-1; this protein has been identified as an ER stress responsive protein, upregulated following tunicamycin treatment in an IRE-1-dependent fashion, and is thought to play a role in monitoring the folding quality of glycoproteins[10]. Although GFP-fused UGGT-1 reporter lines only show expression within the nervous system, it has been noted that, due to the lack of introns and 3'UTR sequence within the construct used to produce these transgenic animals, this may not be truly representative of its true expression pattern; this is particularly the case as no changes in the amount or expression pattern of this GFP reporter was observed in these animals upon the induction of ER stress – a result that is in contradiction to qPCR data that shows a significant increase in *uggt-1* expression in these conditions. Thus, it is possible that *uggt-1* is indeed expressed beyond the nervous system, particularly following ER stress, although this has yet to be determined. As such, this does not exclude the possibility that UGGT-1 is indeed a physiological interactor of ABU-13 in extra-neuronal tissues.

The presence of an XBP-1 binding site in the promoter of *uggt-1* suggests its expression is negatively correlated with *abu-13* expression – i.e. inhibited XBP-1 activity would reduce *uggt-1* expression but increase *abu-13* expression. This negative correlation between the expression patterns of these two genes, paired with their potential protein-protein interaction, could point towards a mechanism of regulation, with UGGT-1 binding to and negatively regulating the activity of ABU-13. If ABU-13 is involved with a positive feedback loop to increase its own expression, it is possible that UGGT-1 prevents this, acting as a downstream readout of XBP-1 activity and thus inhibiting the further expression of ABU-13. Conversely, when XBP-1 is inhibited, *uggt-1* would not be expressed, thus disinhibiting this feedback loop and leading to an increase in *abu-13* expression. Crucially, this hypothesis neglects the requirement for an ER stress signal – thus, there must be other stress-responsive modulators that are involved independently of the XBP-1 pathway, perhaps via the secondary non-*xbp-1* pathway driven by IRE-1, potentially acting as enhancers of gene expression. This hypothesis could be tested by constitutive expression of *uggt-1* in an *xbp-1* mutant background and monitoring *abu-13* transcript levels following tunicamycin treatment; if this regulation were indeed occurring, there would be no observed increase in *abu-13* expression compared to the *xbp-1* mutant animals alone.

Following evidence that ABU-13 is indeed an RNA binding protein, its role in RNA processing and in ribonucleoprotein complexes make an attractive hypothesis. In support of this, we have also observed a significant overrepresentation of RNA-related terms within these mass spectrometry results. The presence of other RNP components, in particular, may help point towards a molecular understanding of what ABU-13 is doing, if it is indeed involved in the formation of such complexes. Within these components includes a number of translation initiation factors, including EIF-2D, EIF-3.E and EIF-3.C, and ribosome biogenesis factors, such as K12H4.3 and Y54E10A.10 – sequestering of such components can reduce translational rates during periods of proteostatic stress, alleviating protein folding burden and helping restore protein homeostasis to the organism. LSM-1 and LET-711 also represent promising RNP-related proteins pulled down alongside ABU-13 – these proteins are involved in Processing-body activity in the development embryo, with LET-711 recruiting the decapping proteins LSM-1 and LSM-3 to these condensates within somatic blastomeres, resulting in the degradation of maternally derived mRNA. As we have previously shown, P-bodies are constitutively active, stress-responsive organelles within *C. elegans* – however, we see no association between ABU-13 puncta and P-bodies *in vivo*. It is, however, possible that components of these organelles are shared, even if the organelles themselves do not overlap, potentially representing some common functionality.

In addition to this, candidates within this dataset also hint at a mechanistic regulation of any potential phase separation that might occur in our system. PPTR-2, for example, is a protein phosphatase previously identified as having a redundant role in the regulation of P granule formation, dephosphorylating MEG proteins and facilitating their involvement in P granule biogenesis. An initial screen of its redundant partner, PPTR-1 identified interactions with MEG-1, MEG-3 and MEG-4 – a finding that was demonstrated in PPTR-2 as well, with these MEG proteins being pulled down alongside PPTR-2[11]. As PPTR-1 is not immunoprecipitated alongside ABU-13, it is possible that PPTR-2 is playing a non-redundant role in this system, regulating ABU-13 to modulate dynamics of its puncta. Identification of a physiologically relevant phosphorylation site would elucidate the role that this posttranslational modification might have on ABU-13. Thus far, we have been unable to successfully detect ABU-13 after endogenous tagging with a 3xFLAG motif or WormScarlet tag, for reasons that remain elusive. Optimisation of this protocol would allow us to initially determine whether ABU-13 is indeed phosphorylated, and if so, under what conditions. From here, it would then be possible to determine whether *pptr-2* knockouts prevented this dephosphorylation and, alongside imaging data, demonstrate whether this resulted in changes in puncta dynamics. If this were indeed the case, identification of the relevant phosphorylation site could then provide direct evidence of this regulation using phospho-mimetic and phospho-deficient mutants of ABU-13.

Identification of protein-protein interactions within an endogenous expression system would be useful to delineate which, if any, of these hits represent proteins of physiological interest. As briefly mentioned,

the overexpression of ABU-13 may result in protein interactions that are normally observed upon stress, when *abu-13* is physiologically upregulated. As with the RNA interaction work, optimisation of the endogenous pull down will help determine the protein interactome of ABU-13 under different conditions, helping us to further characterise its mechanism of action.

### *4.3.3 The hydrogel-forming proteome of C. elegans*

The low mobility nature of some ABU-13 puncta indicated to us that this protein might be involved in the formation of hydrogel-like structures – particularly considering the presence of a prion-like domain within the centre of this protein, often indicative of proteins involved in the biogenesis of such puncta. However, the absence of this protein amongst the biotinylated isoxazole precipitated proteins calls this hypothesis into question. There are a number of possible explanations for these contradictory results; firstly, we cannot rule out the possibility that our precipitation protocol was not entirely successful. This is especially relevant considering the absence of the majority of shared 'core' proteins precipitated following b-isox treatment in previous datasets – these proteins represent many recognisable phase-separation related proteins including hnRNPA1 and TIA-1. While this absence may reflect evolved differences in the propensity of species-specific proteins for solid transitions, repetition of this precipitation would help to clarify these dataset differences, although we have been restricted in this pursuit by the very limited supply of biotinylated isoxazole.

There are, of course, a multitude of potential physiological explanations for the lack of ABU-13 within this dataset as well. The tagging of ABU-13, for example, may drive the protein towards more solid-like transitions; we did attempt to mitigate this effect using a tagRFP fluorescent protein, which has been specifically developed to be monomeric as to prevent fluorophore aggregation. However, we were unable to use this tag for our endogenous tagging system and thus the solid-like formations observed in the ABU-13::WormScarlet animals may result from the tag rather than the protein itself. Meanwhile, in the ABU-13::tagRFP animals, we may be observing an overexpression effect resulting in punctae appearing to be more solid than they would be under physiological conditions. As has been previously discussed, prion-like proteins involved in phase transitions are acutely sensitive to protein concentration and thus any abnormal expression of these components could drive these structures into non-physiological states.

Another explanation for these contradictory results could be that ABU-13 is not the only protein component within these compartments capable of driving phase separation, with other proteins driving the liquid to solid transition. Hence, rather than relying on a single protein component to drive synthesis of these structures, a number of scaffold proteins instead work in concert to modulate the dynamic nature of these puncta – so, whilst ABU-13 may be involved in more liquid-like phase separation, others may be responsible for the transition into a hydrogel structure.

Besides the absence of ABU-13, the low scoring nature of the majority of pulled down proteins when subjected to prion-prediction algorithms is particularly interesting. It might be expected, due to their roles in facilitating phase transitions, that these proteins would represent candidate prionogenic species. However, according to these algorithms, this is not the case. Whilst this may be a consequence of an underrepresentation of hydrogel-forming proteins – for example, as already mentioned, many of the 'core' proteins precipitated by b-isox in mammalian cells are not seen in this dataset – it may also be indicative of a different mechanism through which hydrogel-forming proteins phase separate. Unlike proteins involved in liquid-liquid phase transitions, these proteins may rely on different domain structures to drive solid transitions. Further computational analysis of liquid-like versus solid-like scaffold proteins would be useful in elucidating these differences. For example, as has already been shown in previous work, the number of hydrophobic residues within prion domains is tightly controlled to increase the propensity for coalescence and phase separation without prompting aberrant fibrillisation. Thus, if these hydrogel-forming proteins do contain a 'prion-like domain', it is possible that they possess different residue compositions to drive the biogenesis of more solid-like structures.

Alternatively, it is possible that hydrogel proteins are not necessarily themselves the nucleating, scaffold proteins that drive initial assembly of these phase-separated condensates. Instead, these organelles may require sequential progression through different phase transitioned states, from diffuse to liquid to hydrogel. As such, there is not necessarily a need for a prion-like domain within these proteins. They do, however, clearly play roles more fundamental to these assemblies than other client proteins, capable of fine-tuning and modulating the dynamic natures of these structures. As such, they may represent an intermediate between scaffold and client proteins, a species of protein that we have termed transitional proteins. Whilst these proteins do not necessarily represent fundamental core components that drive initial phase separation events, they do push the organelle towards more solid-like states. Crucially, these proteins enact this transition as a consequence of their secondary structure, rather than through biologically active processes that regulate other protein components within these condensates. Whilst this might not be the case for all hydrogel-forming proteins, we propose this sequential phase transition model as a potential mechanism through which b-isox precipitated proteins in *C. elegans* facilitate their intrinsic hydrogel forming abilities, as opposed to a direct action of these proteins to drive hydrogel formation from diffuse states. Further work to determine the modulatory activity of these proteins is necessary to understand the exact influence they have on existing or *de novo* phase-separated structures. Such a model would fit with the absence of ABU-13 from this precipitation data – rather than driving hydrogel formation itself, ABU-13 may instead be involved in the biogenesis of these biological condensates as a scaffold protein for liquid-like phase transitions. Once in these liquid-like states, a further 'transitional' component may be responsible for the switch into a hydrogel state later on.

To determine what the identity of this transitional component might be, we compared the proteins precipitated by biotinylated isoxazole treatment to those pulled down with ABU-13::tagRFP. Of the 14 crossover candidates, there are a number that warrant further investigation. NPP-4 is one of these candidates of interest; this protein is the ortholog of the human Nucleoporin 62[12]. As we have previously described, nucleoporins are involved in the formation of physiological hydrogel structures, with NPP-4 and its homologs interacting with other nucleoporins to drive the formation of nuclear pore complexes. This is supported by evidence that, in *Xenopus*, Nup62 contains an FG-rich domain that, when purified, is capable of rapidly forming hydrogel structures *in vitro*[13]. These structures show a low permeability, with movement of molecules through this hydrogel showing a highly selectivity. In addition to an FG-rich repeat region, this protein has also been shown to contain a serine-threonine-rich linker region that acts as a site of glycosylation, with the addition of N-acetylglucosamine residues modulating the activity of this protein[14–16]. This modification weakens inter-FG-repeat connections, reducing the selectivity of this barrier and facilitating the entry of larger nuclear transport receptors[13]. The regulation of Nucleoporin 62 via glycosylation is particularly interesting in the context of ABU-13, which has been shown to co-immunoprecipitate alongside the glycosylation-related protein UGGT-1 and a glucosidase II alpha subunit ortholog, AAGR-4, both of which are involved with the folding of nascent glycoproteins. This points towards a general role for ABU-13 in the regulation of glycosylation. In the context of NPP-4, it is possible that ABU-13 may be acting in concert with glycosylation components and subsequently with NPP-4 to regulate the movement of specific molecules through the nuclear pore complex, helping to alleviate the protein-folding burden upon stress via this regulation.

Adding weight to the involvement of ABU-13 in glycosylation is the presence of CHT-3 within this limited set of crossover proteins. CHT-3 is the ortholog of the human proteins CHI3L1, a secreted glycoprotein containing a chitinase-like domain, and the glycoprotein OVGP1, involved in reproductive health. This is particularly interesting contextually as *CHI3L1* expression is increased during inflammation, suggesting a role for this protein in immune responses. In keeping with this, the presence of a chitinase-like domain[17] points towards an interaction with chitin-containing structures, of which the fungal cell wall is a prime candidate. It is important to note, however, that no active chitinolytic behaviour has been identified from this protein – whether this is also the case for CHT-3 remains unclear. The co-immunoprecipitation of CHT-3 alongside ABU-13 could suggest a mechanism through which ABU-13 is involved in the innate immune response, perhaps via the regulated glycosylation of immune proteins upon stress.

An alternative mechanism through which ABU-13 may be exerting its stress-responsive effects through CHT-3 is via its ability to modulate the PI3K-AKT pathway. Although not studied in nematodes, increased expression of CHI3L1 has been shown to increase the kinase activity of PI3K following gamma-irradiation in U87 cells, resulting in the increased phosphorylation of AKT and subsequent increased inhibition of FOXO-mediated transcription[18]. In *C. elegans*, the FOXO homolog DAF-16 has

been shown to restore ER proteostasis in an IRE-1/XBP-1-independent manner, with reduced insulin/IGF-1 signalling rescuing the effects of *xbp-1* and *ire-1* mutations following tunicamycin treatment. This is facilitated by the improved degradation of misfolded lumenal protein following their EDEM-1-mediated release from the ER folding cycle and subsequent export into the cytosol for degradation via autophagy[19]. This could provide a mechanistic pathway through which ABU-13 and the non-canonical UPR pathway acts upon inhibition of the canonical UPR – using the DAF-2/DAF-16 IIS pathway to drive alternative degradation systems. However, such a strong restoration of proteostasis is not observed in *xbp-1* mutants, despite components of the non-canonical pathway being upregulated in these circumstances. It may be that the influence of ABU-13 on this DAF-16 pathway is minimal in comparison to the experimental paradigm utilising *daf-2* mutants to mimic reduced IIS – an intervention that represents a much larger perturbation and elicits a much greater downstream effect than an endogenous pathway might induce physiologically. Either way, this branch warrants more exploration to determine if this is indeed mechanism of action for ABU-13 – crossing these animals into *cht-3* mutants, for example, could be useful for determining whether such mutants display similar phenotypes to the *abu-13* knockout animals.

There are, however, a number of proteins identified from this precipitation that do have predicted prion-like properties. Whilst these are not likely to be involved with ABU-13-related processes, they do represent an interesting group of proteins whose intrinsic hydrogel-forming ability intrinsic may arise as a consequence of a prion-like domain. Within this list of candidate proteins, certain functions are overrepresented – particularly those relating to the regulation of RNA, whether through RNA polymerase activity, such as with AMA-1, or with the regulation of mRNA post-transcriptionally – MSI-1, for example, regulates mRNA within the cytoplasm. Of these RNA regulating proteins, GLD-2 is of particularly note due to its association with the developmental phase-separated organelle, P granules. This protein contains a cytoplasmic poly(A) polymerase (PAP) motif, enabling it to polyadenylate mRNA within the cytoplasm, increasing its expression. Unlike conventional PAP motifs, GLD-2 does not contain an RNA recognition motif but rather interacts with the RNA binding protein GLD-3, which shepherds appropriate RNA to it[20]. This mechanism is akin to that which we proposed for ABU-13, whereby removal of the prion-like domain abolished RNA binding ability, suggesting that it was only able to bind RNA when within localised within puncta – it is possible that GLD-2 is acting in a similar way, with RNA being shuttled to it by another component.

Although not pulled down with ABU-13, the hydrogel-forming protein candidate RLE-1 stands out amongst this list due to its role in protein homeostasis and the effect that this protein has on organismal longevity. RNAi against this E3 ubiquitin ligase in metastable mutants that exhibit paralysis phenotypes due to a loss of proteostasis and subsequent protein aggregation results in a reduction in this paralysis, suggesting a role for RLE-1 in protein homeostasis[21]. Beyond this, RLE-1 functions as a modulator of lifespan, catalysing the ubiquitination of DAF-16 and triggering its degradation. Consequently,

downstream transcriptional pathways cannot be expressed, ultimately reducing lifespan. The identification of RLE-1 as a potential hydrogel-forming protein is particularly interesting in the context of its involvement in the ubiquitination-proteosomal system, with recent evidence suggesting a role for phase separation in these processes[22]. Whilst the UPS components within this study were shown to form liquid-like condensates, it is possible that RLE-1 represents a component of this system involved in less transient, more solid-like assemblies that are constitutively present within *C. elegans*. Further work to understand the influence of RLE-1 on hydrogel formation *in vivo* would be required to determine this, and whether or not this propensity plays a role in its function as an E3 ubiquitin ligase.

Overall, this enrichment of nucleotide binding proteins within CORE-scoring biotinylated isoxazole precipitated proteins is further validation of the role that many phase separating organelles and their constituent proteins have in the regulation of RNA.

## 4.4 Material and Methods

*Crosslinking and Immunoprecipitation (CLIP)*

Nematodes were grown to day 1 of adulthood before transfer onto unseeded NGM plates ready for UV irradiation using a UV Stratalinker 2400 (254nm), set to 300μJ(x100). Following irradiation, populations were harvested in M9 and pelleted down, aspirating off the supernatant and resuspending animals in RIPA buffer before transferring to Precellys tubes and flash freezing samples in liquid nitrogen. After thawing, animals were lysed at 4°C using a Precellys 24 (Bertin Instruments) programmed for 3 spins of 15 seconds, with a 30 second interval between each spin. To clear any debris, samples were then spun down at 8000g for 15 minutes at 4°C and transferred to fresh precooled 1.5mL Eppendorf tubes.

RNase I (Ambion, AM2295) was prepared in a 1/500 dilution, after which 10μL of diluted RNase I was added to the lysate alongside 2μL Turbo DNase. Samples were incubated for 3 minutes at 37°C, shaking at 1100rpm before being transferred immediately to ice. A further spin was then performed at 22,000g for 20 minutes at 4°C to clear the lysate once more, and subsequently the supernatant collected.

In preparation for the immunoprecipitation step, RFP-Trap Agarose beads (Chromotek, rta-20) were vortexed and 2μL of beads mixed with 500μL of cooled dilution buffer (10mM Tris-HCl, pH 7.5; 150mM NaCl; 0.5mM EDTA; 1x cOmplete protease inhibitor cocktail (Roche); 1mM PMSF). This bead mix was then centrifuged at 2500g for 2 minutes at 4°C and the supernatant discarded before resuspending in 500μL of dilution buffer once more and repeating this step a further two times. Once the supernatant had been removed a final time, the prepared lysate was added and the sample incubated with the beads for 1 hour at 4°C. Beads were then collected by centrifugation at 2500g for 2 minutes at 4°C and the supernatant discarded. 500μL of dilution buffer added and spun down using the same conditions a further two times to wash the beads.

Once the supernatant had been discarded, beads were resuspended in 20μL of PNK mix (15μL water; 4μL 5x PNK buffer, pH 6.5 [350mM Tris-HCl; 50mM $MgCl_2$; 25mM dithiothreitol]; 0.5μL PNK enzyme (NEB); 0.5μL RNasin (Promega)) and incubated for 20 minutes at 37°C. After this, 500μL of high-salt buffer (50mM Tris-HCl, pH 7.4; 1M NaCl; 1mM EDTA; 1% NP-40; 0.1% SDS; 0.5% sodium deoxycholate) was added before collecting the beads via centrifugation at 2500g for 2 minutes at 4°C and resuspending in high-salt buffer and spinning down once more. These wash steps were then repeated twice more with wash buffer (20mM Tris-HCl, pH 7.4; 10mM $MgCl_2$; 0.2% Tween-20) before removing the supernatant and adding 8μL of hot PNK mix (0.4μL PNK; 0.8μL ATPγS; 0.8μL 10x PNK buffer (NEB); 6μL water) and incubating overnight at 37°C.

Immediately prior to use, prepare a 10mM stock solution of 5-iodacetamidofluorescein (Thermo Scientific) in DMSO. Spin down beads and remove the hot PNK mix before resuspending beads in 50µL of fluorescein mix (7.5µL of 10mM 5-IAF; 42.5µL of 2mM HEPES, pH 7.4). Mix gently at room temperature for 2 – 3 hours in the dark before centrifuging the beads and discarding the supernatant. Beads were then resuspending in 20µL of 1x NuPage LDS Sample Buffer (Invitrogen) and incubated for 10 minutes at 70°C. Beads were then spun down a final time and the supernatant loaded onto a NuPage 4-12% Bis-Tris precast gel (Invitrogen) and run at 200V in 1X MOPS running buffer alongside a Precision Plus Protein Dual Color Standard (Biorad) protein ladder. Once the gel had run, it was removed from its casing and rinsed in PBS before being imaged using a Chemidoc Touch Imaging System (Biorad), with the settings set to detect fluorescein.

*Co-Immunoprecipitation*

Day 1 adult animals were lysed in RIPA using a Preceylls 24, as previously described. Lysates were spun down at 8000g for 15 minutes at 4°C and cleared supernatant transferred into pre-cooled 1.5mL Eppendorf tubes. As with CLIP, 25 µL of RFP-Trap Agarose beads were prepared after vortexing in 500 µL of ice-cold dilution buffer. Following three wash steps, the details of which are detailed above, lysate was added to the beads and incubated overnight at 4°C, after which, three further washes were performed. After the final wash, beads were spun down at 2500g a final time and the supernatant discarded before resuspending the washed beads in 100 µL of 2x LDS Sample Buffer and incubating the mix for 10 minutes at 95C. Following this boiling step, beads were collected by centrifugation at 2500x g for 2 minutes and the supernatant loaded onto a NuPage 4-12% Bis-Tris precast gel (Invitrogen) and run as above.

Once run, the gel was removed from the plastic casing and fixed in 40%EtOH/10% acetic acid for 15 minutes before rinsing in water and staining overnight in QC Colloidal Coomassie Stain (Biorad). The gel was then rinsed in water once more and imaged using a Chemidoc Touch Imaging System (Biorad), with the settings set to detect Coomassie. The full gel lane was then fragmented and excised from the gel, and prepped for mass spectrometry.

*Biotinylated isoxazole precipitation*

Day 1 adult animals were homogenised in lysis buffer (20mM Tris buffer; 150mM NaCl; 5mM $MgCl_2$; 20mM β-mercaptoethanol; 0.5% NP-40; 10% glycerol; 1x cOmplete protease inhibitor cocktail (Roche); 1:100 RNasin ribonuclease inhibitor (Promega); 2mM vanadyl ribonucleoside complex (NEB); 0.1mM PMSF), using the Preceylls 24 system as described previously. Cleared lysates were then divided in two – one to be used as the experimental sample, the other as the control. Biotinylated isoxazole (Sigma-

Aldrich) was prepared as a 10mM stock solution in DMSO and a 1/100 dilution added to one of the lysates, resulting in a final b-isox concentration of 10uM. An equivalent volume of DMSO was added to the control sample. Both were then tumbled at 4°C for 90 minutes before spinning down at 10,000x g for 10 minutes at 4°C. Supernatant from both samples was discarded and pellets washed twice in lysis buffer before suspending in 2x LDS Sample Buffer and loading on to a NuPage 4-12% Bis-Tris precast gel (Invitrogen). Once run, the gel was then stained and prepared for mass spectrometry, as previously described.

*Mass Spectrometry*

Samples underwent tandem mass spectrometry following proteolysis of excised gel fragments. The resulting peptide fragments were subjected to LC-MS/MS using a Velos Orbitrap mass spectrometer (Thermo Scientific). Protein identity was attributed to peptides using the MASCOT database (Matrix Science) and results collated using Scaffold4 (Proteome Software).

Results from mass spectrometry were compared between experimental and control conditions (i.e. ABU-13::tagRFP versus tagRFP only; B-isox versus DMSO). Experimental samples that demonstrated a greater than 50% increase in Exclusive Unique Peptide Count compared to their DMSO counterparts were taken forward for further analysis. Using the PANTHER Overrepresentation Test (http://www.pantherdb.org/), gene ontology terms relating to the proteins identified could be determined in comparison to a *C. elegans* reference genome.

## 4.5 References


1. March, Z. M., King, O. D. & Shorter, J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res*. **1647,** 9–18 (2016).
2. Maharana, S. *et al*. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*. **360,** 918 LP-921 (2018).
3. Ma, W. & Mayr, C. A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3′UTR-Mediated Protein-Protein Interactions. *Cell* **175,** 1492–1506.e19 (2018).
4. Kato, M. *et al*. Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149,** 753–767 (2012).
5. Hennig, S. *et al*. Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol*. **210,** 529–539 (2015).
6. Konig, J. *et al*. iCLIP - Transcriptome-wide Mapping of Protein-RNA Interactions with Individual Nucleotide Resolution. *JoVE* e2638 (2011). doi:doi:10.3791/2638
7. Zearfoss, N. R. & Ryder, S. P. in (ed. Conn, G. L.) 181–193 (Humana Press, 2012). doi:10.1007/978-1-62703-113-4_14
8. Nolde, M. J., Saka, N., Reinert, K. L. & Slack, F. J. The *Caenorhabditis elegans* pumilio homolog, *puf-9*, is required for the 3′UTR-mediated repression of the *let-7* microRNA target gene, hbl-1. *Dev. Biol*. **305,** 551–563 (2007).
9. Gallo, C. M., Munro, E., Rasoloson, D., Merritt, C. & Seydoux, G. Processing bodies and germ granules are distinct RNA granules that interact in *C. elegans* embryos. *Dev. Biol*. **323,** 76–87 (2008).
10. Buzzi, L. I., Simonetta, S. H., Parodi, A. J. & Castro, O. A. The Two *Caenorhabditis elegans* UDP-Glucose:Glycoprotein Glucosyltransferase Homologues Have Distinct Biological Functions. *PLoS One* **6,** e27025 (2011).
11. Wang, J. T. *et al*. Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in *C. elegans*. *Elife* **3,** e04591 (2014).
12. Schetter, A., Askjaer, P., Piano, F., Mattaj, I. & Kemphues, K. Nucleoporins NPP-1, NPP-3, NPP-4, NPP-11 and NPP-13 are required for proper spindle orientation in *C. elegans*. *Dev. Biol*. **289,** 360–371 (2006).
13. Labokha, A. A. *et al*. Systematic analysis of barrier-forming FG hydrogels from *Xenopus* nuclear pore complexes. *EMBO J*. **32,** 204–218 (2013).
14. Starr, C. M., D'Onofrio, M., Park, M. K. & Hanover, J. A. Primary sequence and heterologous expression of nuclear pore glycoprotein p62. *J. Cell Biol*. **110,** 1861–1871 (1990).
15. Lubas, W. A., Smith, M., Starr, C. M. & Hanover, J. A. Analysis of nuclear pore protein p62 glycosylation. *Biochemistry* **34,** 1686–1694 (1995).
16. Davis, L. I. & Blobel, G. Nuclear pore complex contains a family of glycoproteins that

includes p62: glycosylation through a previously unidentified cellular pathway. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 7552–7556 (1987).

17. BE, H., White C FAU - Recklies, A. D. & AD, R. Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family. PG - 25803-10.
18. Francescone, R. A. *et al.* Role of YKL-40 in the angiogenesis, radioresistance, and progression of glioblastoma. *J. Biol. Chem.* **286,** 15332–15343 (2011).
19. Safra, M. *et al.* The FOXO Transcription Factor DAF-16 Bypasses ire-1 Requirement to Promote Endoplasmic Reticulum Homeostasis. *Cell Metab.* **20,** 870–881 (2014).
20. Wang, L., Eckmann, C. R., Kadyk, L. C., Wickens, M. & Kimble, J. A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. *Nature* **419,** 312–316 (2002).
21. Alavez, S., Vantipalli, M. C., Zucker, D. J. S., Klang, I. M. & Lithgow, G. J. Amyloid-binding compounds maintain protein homeostasis during ageing and extend lifespan. *Nature* **472,** 226–229 (2011).
22. Yasuda, S. *et al.* Stress- and ubiquitylation-dependent phase separation of the proteasome. *Nature* **578,** 296–300 (2020).

# Chapter 5

## Conclusions

## 5.0 Conclusions

Within this thesis, I have established that the *C. elegans* protein ABU-13 is a prion-like domain-containing component of a phase separated compartment that plays important roles in immune responses and ER stress resistance (Figure 5.1).

After subjecting the nematode proteome to computational analysis to identify prionogenic candidates[1–4], we cross-validated multiple algorithms and found that ABU-13, a previously uncharacterised protein, scored consistently high across computations. The ABU family in general exhibit a higher prion propensity than the average non-prion-like protein within these animals, however, on the whole, ABU-13 ranks the highest amongst both this family and the rest of the proteome, even outperforming other well-characterised prion-like proteins already identified in *C. elegans*.

This family has previously been shown to be upregulated within innate immune responses and in the non-canonical unfolded protein response of the endoplasmic reticulum, however little has been known about the function of the ABU proteins individually[5–7]. Using a series of stress resistance assays, we compared *abu-13(tm6224)* animals to N2 animals, identifying developmental delays in *abu-13(tm6224)* mutants when treated with the ER stress-inducing agent tunicamycin, and when exposed to the pathogenic bacteria *Pseudomonas aeruginosa*.

To gain mechanistic insight into how ABU-13 might be involved with such functions, we proceeded to produce transgenic strains expressing fluorescently-tagged forms of this protein. Initially this involved using extrachromosomal array lines, however, after a process of optimisation, we were successful in establishing a line of endogenously-tagged animals using a CRISPR/Cas9-mediated genome insertion as well. *In vivo* visualisation of this protein demonstrated a localisation to discrete punctate structures throughout the animals. Along with the presence of a prion-like domain and a potential RNA binding domain, as evidenced by conserved domain predictions that suggest nucleotide binding activity in the N-terminus, we hypothesise that these puncta represent phase-separated compartments.

The puncta within the endogenously-tagged line are smaller and more numerous than those in the overexpression line. FRAP analyses from both of these strains demonstrate that the ABU-13 puncta show a greater variation in mobility within endogenously-tagged animals than within overexpression line animals, with the latter exhibiting consistently low FRAP recovery. Whilst a high mobility is indicative of more diffuse, liquid-like assemblies, lower mobility is associated with more solid, hydrogel-like formations. As such, we proposed that the ABU-13 puncta within the endogenously-tagged strain exist along a spectrum of phase-transitioned states.

Akin to other biomolecular condensates, these structures may display multiphase coexistence, with hydrogel forming puncta clustering within larger, less viscous assemblies[8–10]. If this were the case, little fusion of these smaller hydrogel-like structures would occur in a basal state. Instead, these more immiscible solid states likely represent a coalescence of 'core' proteins vital for the functionality of the organelle. Meanwhile, the surrounding liquid-like environment potentially represents 'adaptor' components that are interchanged depending on the cellular context, thus allowing for the function of these constitutively active puncta to be dynamically modulated. As ABU-13 is not precipitated following treatment with biotinylated isoxazole, we propose that it does not represent the 'scaffold' of such hydrogel formations. The role of this protein in nucleating liquid-liquid phase transitions, however, is unclear, with *in vitro* phase separation assays demonstrating an inherent biophysical ability to phase separate, whilst *in vivo* data from endogenously-edited prion-domain mutants showing retention of ABU-13 within punctate compartments. Taking this into consideration we propose that ABU-13 plays a redundant role as a scaffold protein for liquid-liquid phase separation, whilst a secondary 'transitional' co-scaffold protein interacting with ABU-13 puncta is responsible for the liquid to hydrogel phase transition.

To assess the involvement of these puncta in the functions we identified for ABU-13, we subjected endogenously-tagged animals to a range of environmental stresses. In response to ER stress, ABU-13 puncta exhibit no significant change, with similar puncta number, size and distribution observed between DMSO-treated controls and tunicamycin-treated animals. We hypothesise that these puncta are dynamically exchanging components with the diffuse environment, bringing together stress-specific clients to facilitate a downstream sub-branch of the ER stress response. This exchange results in no observable net gain or loss of puncta, suggesting that many of the components required are already located within these structures under basal conditions. Although initially identified as representing a non-canonical unfolded protein response of the endoplasmic reticulum[6,7], our evidence points towards ABU-13 playing a role in the ER stress response when the canonical pathway is operational. Moving forward, it would be interesting to investigate whether the response of ABU-13 in *xbp-1* mutant animals represents an upregulation of the response seen in wild type conditions, or whether ABU-13 is involved in other mechanisms to compensate for the loss of XBP-1.

The response of ABU-13 puncta to pathogenic bacteria, however, is different, with significant growth and fusion of puncta following exposure to *P. aeruginosa*. We propose that, in response to pharyngeal and gut invasion of PA14, cell non-autonomous stress signals coordinate a global immune response, priming the organism to deal with an ongoing pathogenic attack as it spreads. Within the pharynx, we observe an upregulation of diffuse ABU-13 upon exposure to this bacteria, potentially within the marginal cells – however, these cells do not have any well-characterised global signalling capacity. We do, however, see potential expression within the amphid sheath glia – whilst the role of expression of ABU-13 in this support cell is unclear, it is possible it is involved in a form of stress signalling in

coordination with the amphid neurons, which are involved in processing sensory information from the environment.

Beyond the pharynx, it is also possible that other sites of bacterial invasion may be involved. The intestine, for example, has been shown to modulate metabolism systemically[11], as well as recent evidence demonstrating a role for this tissue in the transmission of ER stress signals between cells[12]. Thus, we hypothesise that PA14 invasion of the intestine may stimulate a global stress state, leading to downstream activation of ABU-13 puncta. This activation might involve a change in transition state of these compartments, with immiscible structures adopting less viscous properties, allowing for their subsequent fusion. This likely results from recruitment of client molecules that buffer these assemblies, such as RNA or hydrotropic ATP[13,14].

Although a number of phase-separated compartments have been linked to stress responses previously, we do not see any colocalization of ABU-13 with stress granules nor P-bodies before or after exposure to an environmental stress, suggesting that these ABU-13 puncta may represent novel phase-separated organelles.

In both of the relevant stress response conditions, recruitment of client molecules to ABU-13 compartments is likely a mechanism through which these puncta can facilitate their downstream effects on stress responses. This is observed in many phase separated organelles, with recruitment of particular protein and nucleic acid components having functional consequences. These functional consequences often involve the regulation of RNA – for example, through the benign sequestration of specific RNA species or through the active processing of these transcripts. In keeping with this, we have determined that ABU-13 is indeed an RNA binding protein. As we have already postulated, the involvement of RNA in these assemblies is potentially vital for preventing aberrant transitions into pathological fibrils due to the aggregation-prone behaviour of prion-like proteins[13,15]. However, beyond this, it is likely that the RNA species recruited to these ABU-13 puncta are involved in the downstream stress response in some capacity.

As such, further work is necessary to determine the protein-protein and protein-RNA interactions with ABU-13 in different cellular stress conditions. This will shed light on the molecular mechanisms at hand and help elucidate the pathways that might govern the function of these puncta. Preliminary coimmunoprecipitation studies of ABU-13 from the overexpression model identified a number of proteins involved in RNA regulation, supporting this hypothesis. Beyond this, a number of components pulled down are involved in the quality control of glycosylated proteins, suggesting that ABU-13 may play a role in monitoring the folding of glycoproteins. This has important implications for ABU-13-mediated stress responses – for example, glycoproteins are a vital component of the cuticle, maintenance of which is important for protecting the animal against pathogenic invasion. Performing these studies in

endogenously tagged animals in these stress contexts will further clarify how some of these components relate to the downstream stress response. This is similar for RNA interactions; whilst protein interactions may help to identify the function of these puncta – i.e. in mRNA decapping, for example – understanding the RNA species that are processed will allow us to gain molecular details of the stress pathway involved.

Prion-like proteins are proving to play functional roles in the biology of phase separation with their multivalent, intrinsically disordered nature providing the ideal biophysical conditions for these biomolecular assemblies to form[16,17]. Whilst many of these organelles represent transient structures responding dynamically to specific stimuli, others are more persistent. These persistent structures have been implicated in epigenetic regulation, with the heritability of some prion species suggesting that these compartments may be able to pass on information transcellularly and transgenerationally[18–20]. Whilst this is most clearly evident in yeast species due to their asexual reproduction by budding, similar effects have also been observed in eukaryote organisms too, with phase separated organelles being implicated in transgenerational epigenetic inheritance[21].

It would be interesting moving forward to determine whether the persistence of ABU-13 puncta enables the transmittance of stress-related information either via a 'molecular memory' mechanism within an organism, or transgenerationally, priming offspring for adverse environmental conditions.

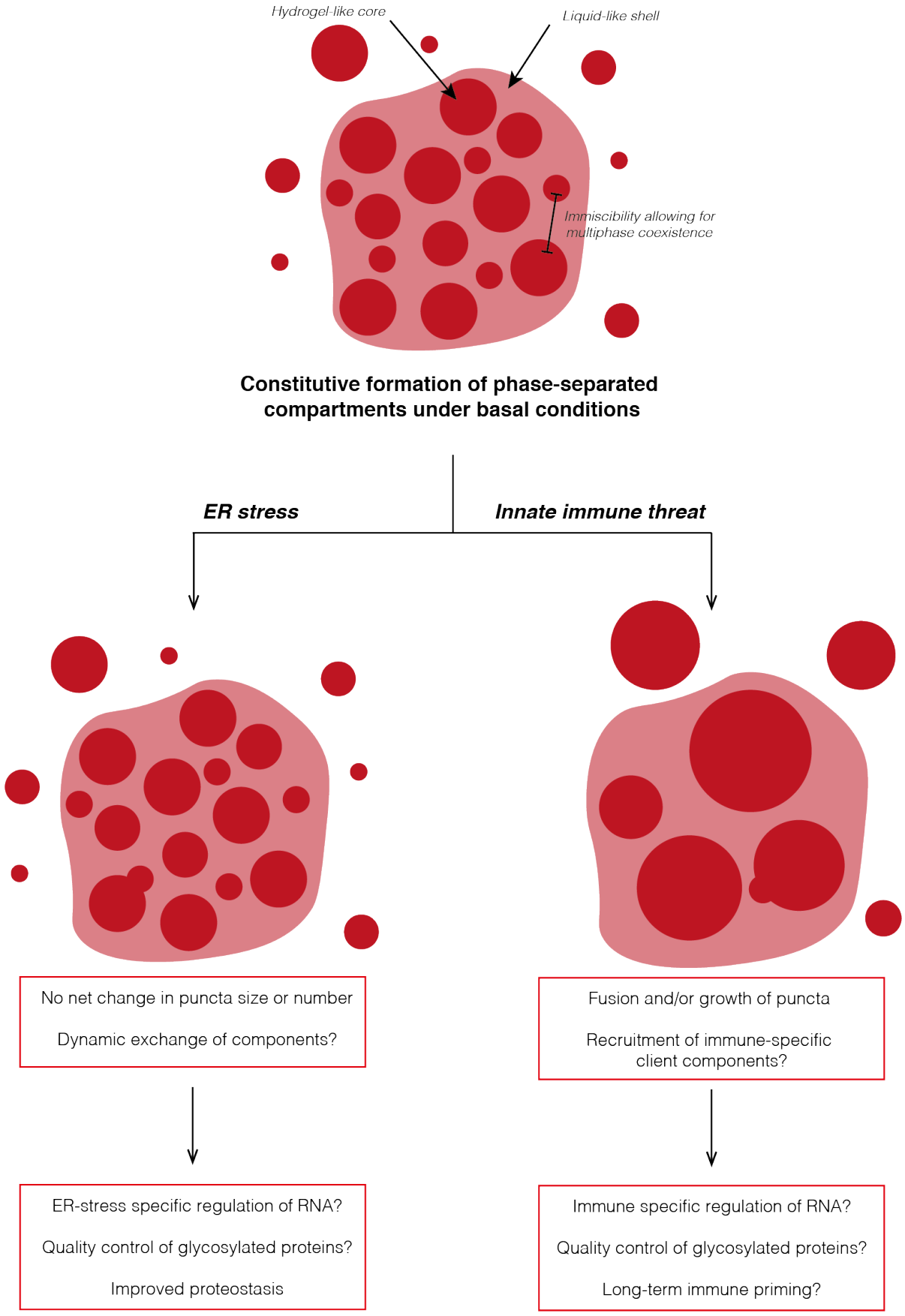



**Figure 5.5** A schematic summarising the role of ABU-13, showing the responses of ABU-13 puncta (red) in response to ER stress and innate immune activation

## 5.1 References


1. Toombs, J. A., McCarty, B. R. & Ross, E. D. Compositional Determinants of Prion Formation in Yeast. *Mol. Cell. Biol.* **30**, 319 – 332 (2010).
2. Toombs, J. A. *et al.* De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci.* **109**, 6519–6524 (2012).
3. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* **137**, 146–158 (2009).
4. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).
5. Urano, F. *et al.* A survival pathway for *Caenorhabditis elegans* with a blocked unfolded protein response. *J. Cell Biol.* **158**, 639 – 646 (2002).
6. Sun, J., Singh, V., Kajino-Sakamoto, R. & Aballay, A. Neuronal GPCR Controls Innate Immunity by Regulating Noncanonical Unfolded Protein Response Genes. *Science.* **332**, 729 – 732 (2011).
7. Haskins, K. A., Russell, J. F., Gaddis, N., Dressman, H. K. & Aballay, A. Unfolded Protein Response Genes Regulated by CED-1 Are Required for *Caenorhabditis elegans* Innate Immunity. *Dev. Cell* **15**, 87–97 (2008).
8. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).
9. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science.* **357**, eaaf4382 (2017).
10. Jain, S. *et al.* ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487–498 (2016).
11. Imanikia, S., Sheng, M., Castro, C., Griffin, J. L. & Taylor, R. C. XBP-1 Remodels Lipid Metabolism to Extend Longevity. *Cell Rep.* **28**, 581-589.e4 (2019).
12. Imanikia, S., Özbey, N. P., Krueger, C., Casanueva, M. O. & Taylor, R. C. Neuronal XBP-1 Activates Intestinal Lysosomes to Improve Proteostasis in *C. elegans*. *Curr. Biol.* (2019). doi:10.1016/J.CUB.2019.06.031
13. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science.* **360**, 918 – 921 (2018).
14. Patel, A. *et al.* ATP as a biological hydrotrope. *Science.* **356**, 753 – 756 (2017).
15. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
16. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
17. Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of

Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).

18. March, Z. M., King, O. D. & Shorter, J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* **1647**, 9–18 (2016).
19. Chakravarty, A. K. & Jarosz, D. F. More than Just a Phase: Prions at the Crossroads of Epigenetic Inheritance and Evolutionary Change. *J. Mol. Biol.* **430**, 4607–4618 (2018).
20. Halfmann, R. & Lindquist, S. Epigenetics in the Extreme: Prions and the Inheritance of Environmentally Acquired Traits. *Science.* **330**, 629 (2010).
21. Wan, G. *et al.* Spatiotemporal regulation of liquid-like condensates in epigenetic inheritance. *Nature* **557**, 679–683 (2018).

# Chapter 6

# Appendices

## 6.0 Experimental Contributions

In Chapter 2, the lifespan analysis performed on *abu-13(tm6224)* animals was performed and statistical analysed by Dr. Rebecca Taylor

In Chapter 3, the bacterial purification of ABU-13 from BL21 *E. coli* and from *Drosophila* S2 cells was performed in conjunction with Joseph Watson.

In Chapter 3, imaging chambers for *in vitro* phase separation assays were prepared by and imaging performed in conjunction with Joseph Watson.

In Chapter 4, mass spectrometry was performed by the Mass Spectrometry facility at the MRC LMB.

## 6.1 Supplemental Tables

***Summary of strains***

| Strain | Genotype | Comment |
|---|---|---|
| N2 | Wild type, Bristol strain | Obtained from CGC |
| RCT214 | *abu-13(tm6224)*III | Obtained from Japanese Knockout Consortium; outcrossed to N2 6x |
| SJ4005 | zcls4*(hsp-4::GFP)*V | Obtained from CGC |
| RCT215 | zcls4*(hsp-4::GFP)*V; *abu-13(tm6224)*III | SJ4005 crossed into RCT214 |
| AGD972 | *xbp-1(zc12)*III; zcls4*(hsp-4::GFP*)V | SJ17 further outcrossed to N2 4x |
| PS3551 | *hsf-1(sy441)*I | |
| RCT96 | rmsEx12*[abu-13p::abu-13::tagRFP::unc-54 3'UTR]* | |
| RCT200 | *abu-13(rms12[abu-13::WormScarlet])*III | |
| RCT201 | *abu-13(rms-13[*rms12])*III | Prion-like domain deletion |
| RCT93 | rmsEx11[*abu-13p::tagRFP::unc-54 3'UTR*] | |
| RCT37 | edIs6*[unc-119p::gfp + rol-6(su1006)]*IV | DP132 (CGC) outcrossed to N2 6x |
| RCT216 | edIs6*[unc-119p::gfp + rol-6(su1006)]*IV; rmsEx12*[abu-13p::abu-13::tagRFP::unc-54 3'UTR]* | RCT37 crossed into RCT96 |
| RCT217 | edIs6*[unc-119p::gfp + rol-6(su1006)]*IV; *abu-13(rms12[abu-13::WormScarlet])*III | RCT37 crossed into RCT200 |
| RCT51 | zcIs18*[ges-1p::gfp]* | SJ4144 (CGC) outcrossed to N2 6x |
| RCT218 | zcIs18*[ges-1p::gfp]; abu-13(rms12[abu-13::WormScarlet])*III | RCT51 crossed into RCT200 |
| RCT142 | rmsEx13*[pitx-1::gfp::unc-54 3'UTR]* | Glial marker strain |
| RCT219 | rmsEx13*[pitx-1::gfp::unc-54 3'UTR]; abu-13(rms12[abu-13::WormScarlet])*III | RCT142 crossed into RCT200 |
| HZ470 | wIs51[*SCMp::gfp + unc-119(+)*] | Seam cell marker strain |
| RCT220 | wIs51[*SCMp::gfp + unc-119(+)*]; *abu-13(rms12[abu-13::WormScarlet])*III | HZ470 crossed into RCT200 |
| BRF70 | synEx115*[pife-2::ife-2::gfp::unc-54 3'UTR; rol-6(su1006)]* | Gifted by the Syntichaki lab |
| RCT221 | synEx115*[pife-2::ife-2::gfp::unc-54 3'UTR; rol-6(su1006)];* rmsEx12*[abu-13p::abu-13::tagRFP::unc-54 3'UTR]* | BRF70 crossed into RCT96 |
| BRF261 | synEx19[*pdcap-1::dcap-1::gfp::dcap-1 3'UTR; rol-6(1006)*] | Gifted by the Syntichaki lab |
| RCT222 | synEx19[*pdcap-1::dcap-1::gfp::dcap-1 3'UTR; rol-6(1006)*]; rmsEx12*[abu-13p::abu-13::tagRFP::unc-54 3'UTR]* | BRF261 crossed into RCT96 |
| RCT223 | synEx19[*pdcap-1::dcap-1::gfp::dcap-1 3'UTR; rol-6(1006)*]; *abu-13(rms12[abu-13::WormScarlet])*III | BRF261 crossed into RCT200 |

**Supplemental Table 1** Summary of *C. elegans* strains used throughout this thesis

***Summary of CRISPR reagents***

| Aim | Method | Components | Sequence |
|---|---|---|---|
| WormScarlet insertion into C-terminus | Partially single-stranded DNA repair template (Dokshin et al., 2019) | sgRNA | GAGTTACAGCTCCAGACTCT |
| | | WormScarlet FP sequence for dsODN | AAGCTTGTTCCAGCAAGAATTACTA TTCAAGAAGCTCGTCAAGCTGAGA GTGGAGCTGAAAAGTGCAGACAAG TAATCAGAATTGTGAAGCAGCGCC TAGAGTCTGGAGCTGTAACTCTGGG TGGTGGTGGTAGTG**TCAGCAAGGG AGAGGCAGT** |
| | | WormScarlet RP sequence for dsODN | acaaaaatataaataaattatttattattttattgcacaccaa aattccaaaaataaataaatttcaaattaggcgcctatttac acaataatctttccaattttacttctcacat**TTACTTG TAGAGCTCGTCCATTC** |
| Repair of WormScarlet insertion | ssODN repair template (Paix et al., 2015) | sgRNA | CTACCATAGGCACCTCTGGG |
| | | ssODN repair template | GCAAGAATTACTATTCAAGAAGCT CGTCAAGCTGAGAGTGGAGCTGAA AAGTGCAGACAAGTAATCAGAATT GTGAAGCAGCGCTACCATAGGCAC CTCTGGGTGGTGGTGGTAGT**GTCA GCAAGGGAGAGGCAGTTATCAAG GAGTTCATGCGTTTCAAGGTCCA CATGGAGGGATCCATGAACGGAC ACGAGTTCGAG** |
| 3xFLAG insertion into C-terminus | ssODN repair template (Paix et al., 2015) | sgRNA | GAGTTACAGCTCCAGACTCT |
| | | ssODN repair template | CTGAAAAGTGCAGACAAGTAATCA GAATTGTGAAGCAGCGACTAGAGT CTGGAGCTGTAACTCTG**GACTACA AGGACCACGACGGTGACTACAAG GACCACGACATCGACTACAAGGA CGACGACGACAAG**TAAatgtgagaagta aaattggaaagattattgtgtaaataggcgcctaatttgaaa tttattta |
| Deletion of proin-like domain | ssODN repair template (Paix et al., 2015) | sgRNA #1 | GTTGTGGAACAAATGCGTATGGG |
| | | sgRNA #2 | AGATTGCACTGGCTGTTGGC |
| | | ssODN repair template | TGTATAGATTAAAAGGTCTTCATGA AATTTAACATATTTTGTTTCAGTCC ATACGCATTT**GGTGGTGGTGGTAG T**TCTTTCGAGTCACATACCCAAATC GTGCAACCAGTGAGAGCTGTTCAA GCTAATGCTCAA |
| Identification of positively edited animals | *dpy-10* co-CRISPR (Paix et al., 2015) | sgRNA | GCTACCATAGGCACCACGAG |
| | | ssODN repair template | CACTTGAACTTCAATACGGCAAGAT GAGAATGACTGGAAACCGTACCGC ATGCGGTGCCTATGGTAGCGGAGC TT CACATGGCTTCAGACCAACAGCCT AT |

**Supplementary Table 2** CRISPR reagents used for successful editing events (underlined = flexible linker sequence; bold = insert sequence; uppercase = coding sequences; lowercase = non-coding sequences)

***Summary of genotyping primers***

| Target | Type of Primer | Sequence |
|---|---|---|
| C-terminal tagging of *abu-13* | Forward | GAAGAGGGAGAATGCCGTGC |
| | Reverse | GAGATGCCAGAATCATTCCTTG |
| Prion-like domain deletion | Forward 1 | TGTTTCAGCCCATACGCATTT |
| | Forward 2 | TATAGCAGTAGTTGACATACCCTAGGATACCATG |
| | Reverse | TACCTCAGCATCAAGAGAAGC |
| *abu-13(tm6224)* mutation | Forward | TATAGCAGTAGTTGACATACCCTAGGATACCATG |
| | Reverse | TTACAGAGTTACAGCTCCAGACTCTAGGC |
| *xbp-1(zc12)* mutation | Forward | CCCTCACTTTCCTCATCACAAC |
| | Reverse | TTTGTTCGAGGTGTCCATCTTC |

**Supplementary Table 3** Summary of genotyping primers used throughout this thesis

***Summary of proteins co-immunoprecipitated with ABU-13::tagRFP***

| Gene ID | Details |
|---|---|
| *eef-1B* | Probable elongation factor 1-beta/1-delta 2 OS=Caenorhabditis elegans GN=eef-1B.2 PE=1 SV=4 |
| *act-4* | Actin-4 OS=Caenorhabditis elegans GN=act-4 PE=3 SV=2 |
| *CELE_R08E5* | R08E5.3 OS=Caenorhabditis elegans GN=CELE_R08E5.3 PE=1 SV=1 |
| *C46G7* | C46G7.2 OS=Caenorhabditis elegans GN=C46G7.2 PE=1 SV=1 |
| *abu-13* | ABU-13 OS=Caenorhabditis elegans GN=abu-13 PE=1 SV=3 |
| *sodh-1* | Alcohol dehydrogenase 1 OS=Caenorhabditis elegans GN=sodh-1 PE=2 SV=2 |
| *dnj-13* | DNJ-13 OS=Caenorhabditis elegans GN=dnj-13 PE=1 SV=1 |
| *CELE_F53F4* | F53F4.11 OS=Caenorhabditis elegans GN=CELE_F53F4.11 PE=1 SV=2 |
| *CELE_F31D4* | F31D4.9 OS=Caenorhabditis elegans GN=CELE_F31D4.9 PE=1 SV=2 |
| *rpia-1* | Probable-ribose 5-phosphate isomerase OS=Caenorhabditis elegans GN=rpia-1 PE=3 SV=1 |
| *elpc-1* | Elongator complex protein 1 OS=Caenorhabditis elegans GN=elpc-1 PE=1 SV=2 |
| *B0564* | B0564.7, isoform a OS=Caenorhabditis elegans GN=B0564.7 PE=1 SV=2 |
| *C43E11* | 60S ribosome subunit biogenesis protein NIP7 homolog OS=Caenorhabditis elegans GN=C43E11.9 PE=1 SV=1 |
| *cal-5* | CAL-5 OS=Caenorhabditis elegans GN=cal-5 PE=1 SV=2 |
| *CELE_F18C12* | F18C12.3 OS=Caenorhabditis elegans GN=CELE_F18C12.3 PE=1 SV=1 |
| *C36E6* | C36E6.1, isoform b OS=Caenorhabditis elegans GN=C36E6.1 PE=1 SV=1 |
| *asp-13* | ASP-13 OS=Caenorhabditis elegans GN=asp-13 PE=1 SV=1 |
| *CELE_Y92H12A* | Y92H12A.5, isoform a OS=Caenorhabditis elegans GN=CELE_Y92H12A.5 PE=1 SV=1 |
| *him-1* | Structural maintenance of chromosomes protein 1 OS=Caenorhabditis elegans GN=him-1 PE=1 SV=4 |
| *unc-96* | UNC-96, isoform b OS=Caenorhabditis elegans GN=unc-96 PE=1 SV=1 |
| *CELE_F55B11* | F55B11.2 OS=Caenorhabditis elegans GN=CELE_F55B11.2 PE=1 SV=2 |
| *fipr-21* | FIPR-21 OS=Caenorhabditis elegans GN=fipr-21 PE=1 SV=1 |
| *CELE_F17E9* | F17E9.5 OS=Caenorhabditis elegans GN=CELE_F17E9.5 PE=1 SV=2 |
| *ucr-2* | UCR-2.1, isoform a OS=Caenorhabditis elegans GN=ucr-2.1 PE=1 SV=1 |
| *B0511* | B0511.6 OS=Caenorhabditis elegans GN=B0511.6 PE=1 SV=2 |
| *cgr-1* | CGR-1 OS=Caenorhabditis elegans GN=cgr-1 PE=1 SV=3 |
| *CELE_ZC477* | ZC477.3, isoform a OS=Caenorhabditis elegans GN=CELE_ZC477.3 PE=1 SV=1 |
| *tag-163* | TAG-163, isoform a OS=Caenorhabditis elegans GN=tag-163 PE=1 SV=1 |
| *B0416* | B0416.5, isoform b OS=Caenorhabditis elegans GN=B0416.5 PE=1 SV=1 |
| *rmd-6* | RMD-6 OS=Caenorhabditis elegans GN=rmd-6 PE=4 SV=1 |
| *emb-4* | EMB-4, isoform a OS=Caenorhabditis elegans GN=emb-4 PE=1 SV=4 |
| *CELE_T04C9* | T04C9.1, isoform a OS=Caenorhabditis elegans GN=CELE_T04C9.1 PE=1 SV=1 |
| *C49C3* | C49C3.9 OS=Caenorhabditis elegans GN=C49C3.9 PE=1 SV=1 |
| *Y54E10A* | Ribosome production factor 2 homolog OS=Caenorhabditis elegans GN=Y54E10A.10 PE=3 SV=1 |
| *nlp-24* | NLP-24 OS=Caenorhabditis elegans GN=nlp-24 PE=1 SV=2 |

| | |
|---|---|
| *gly-4* | Polypeptide N-acetylgalactosaminyltransferase 4 OS=Caenorhabditis elegans GN=gly-4 PE=2 SV=2 |
| *CELE_Y37D8A* | Y37D8A.19 OS=Caenorhabditis elegans GN=CELE_Y37D8A.19 PE=1 SV=1 |
| *CELE_Y43D4A* | Y43D4A.5, isoform c OS=Caenorhabditis elegans GN=CELE_Y43D4A.5 PE=4 SV=1 |
| *CELE_F40F4* | F40F4.7 OS=Caenorhabditis elegans GN=CELE_F40F4.7 PE=1 SV=4 |
| *pph-4* | Serine/threonine-protein phosphatase 4 catalytic subunit 1 OS=Caenorhabditis elegans GN=pph-4.1 PE=1 SV=1 |
| *F40F8* | UMP-CMP kinase 2 OS=Caenorhabditis elegans GN=F40F8.1 PE=1 SV=1 |
| *F44E2* | Uncharacterized protein F44E2.8 OS=Caenorhabditis elegans GN=F44E2.8 PE=4 SV=2 |
| *F14E5* | Golgi apparatus protein 1 homolog OS=Caenorhabditis elegans GN=F14E5.2 PE=1 SV=1 |
| *gei-15* | GEI-15 OS=Caenorhabditis elegans GN=gei-15 PE=1 SV=1 |
| *cpna-1* | CPNA-1, isoform b OS=Caenorhabditis elegans GN=cpna-1 PE=1 SV=1 |
| *pch-2* | Putative pachytene checkpoint protein 2 OS=Caenorhabditis elegans GN=pch-2 PE=1 SV=1 |
| *ugt-50* | Putative UDP-glucuronosyltransferase ugt-50 OS=Caenorhabditis elegans GN=ugt-50 PE=1 SV=2 |
| *npp-1* | NPP-1, isoform a OS=Caenorhabditis elegans GN=npp-1 PE=1 SV=3 |
| *mppb-1* | MPPB-1 OS=Caenorhabditis elegans GN=mppb-1 PE=1 SV=2 |
| *soc-2* | Leucine-rich repeat protein soc-2 OS=Caenorhabditis elegans GN=soc-2 PE=1 SV=3 |
| *rsu-1* | RSU-1 OS=Caenorhabditis elegans GN=rsu-1 PE=1 SV=1 |
| *trpp-8* | TRPP-8, isoform a OS=Caenorhabditis elegans GN=trpp-8 PE=1 SV=1 |
| *ZK177* | Uncharacterized protein ZK177.8 OS=Caenorhabditis elegans GN=ZK177.8 PE=4 SV=2 |
| *pptr-2* | PPTR-2, isoform c OS=Caenorhabditis elegans GN=pptr-2 PE=1 SV=1 |
| *vps-45* | VPS-45, isoform a OS=Caenorhabditis elegans GN=vps-45 PE=1 SV=1 |
| *CELE_F30F8* | F30F8.9, isoform b OS=Caenorhabditis elegans GN=CELE_F30F8.9 PE=1 SV=1 |
| *CELE_W04A4* | W04A4.5 OS=Caenorhabditis elegans GN=CELE_W04A4.5 PE=1 SV=2 |
| *clec-153* | CLEC-153 OS=Caenorhabditis elegans GN=clec-153 PE=1 SV=1 |
| *symk-1* | SYMK-1 OS=Caenorhabditis elegans GN=symk-1 PE=1 SV=2 |
| *CELE_Y56A3A* | Y56A3A.7, isoform a OS=Caenorhabditis elegans GN=CELE_Y56A3A.7 PE=1 SV=2 |
| *dap-3* | DAP-3 OS=Caenorhabditis elegans GN=dap-3 PE=1 SV=1 |
| *CELE_W02H5* | W02H5.8 OS=Caenorhabditis elegans GN=CELE_W02H5.8 PE=1 SV=1 |
| *CELE_T23E7* | T23E7.2, isoform a OS=Caenorhabditis elegans GN=CELE_T23E7.2 PE=1 SV=2 |
| *mans-3* | alpha-1,2-Mannosidase OS=Caenorhabditis elegans GN=mans-3 PE=1 SV=2 |
| *hpo-5* | HPO-5 OS=Caenorhabditis elegans GN=hpo-5 PE=1 SV=2 |
| *CELE_F53C11* | F53C11.1, isoform b OS=Caenorhabditis elegans GN=CELE_F53C11.1 PE=1 SV=1 |
| *C44E4* | C44E4.4 OS=Caenorhabditis elegans GN=C44E4.4 PE=1 SV=1 |
| *scav-4* | SCAV-4 OS=Caenorhabditis elegans GN=scav-4 PE=1 SV=2 |
| *C34C6* | C34C6.4 OS=Caenorhabditis elegans GN=C34C6.4 PE=1 SV=1 |
| *mrp-5* | MRP-5, isoform b OS=Caenorhabditis elegans GN=mrp-5 PE=1 SV=1 |
| *ugt-58* | Putative UDP-glucuronosyltransferase ugt-58 OS=Caenorhabditis elegans GN=ugt-58 PE=3 SV=2 |

| | |
|---|---|
| *lgc-28* | LGC-28 OS=Caenorhabditis elegans GN=lgc-28 PE=1 SV=1 |
| *CELE_M05D6* | M05D6.6 OS=Caenorhabditis elegans GN=CELE_M05D6.6 PE=1 SV=1 |
| *magu-4* | MAGU-4, isoform a OS=Caenorhabditis elegans GN=magu-4 PE=1 SV=1 |
| *ani-2* | Anillin-like protein 2 OS=Caenorhabditis elegans GN=ani-2 PE=2 SV=1 |
| *CELE_Y43F4B* | Y43F4B.5, isoform a OS=Caenorhabditis elegans GN=CELE_Y43F4B.5 PE=1 SV=1 |
| *CELE_ZC416* | ZC416.6, isoform a OS=Caenorhabditis elegans GN=CELE_ZC416.6 PE=1 SV=2 |
| *F13H6* | Esterase CM06B1 OS=Caenorhabditis elegans GN=F13H6.3 PE=3 SV=3 |
| *erfa-3* | ERFA-3, isoform a OS=Caenorhabditis elegans GN=erfa-3 PE=1 SV=2 |
| *iff-2* | Eukaryotic translation initiation factor 5A-2 OS=Caenorhabditis elegans GN=iff-2 PE=2 SV=1 |
| *arx-2* | Actin-related protein 2 OS=Caenorhabditis elegans GN=arx-2 PE=3 SV=1 |
| *CELE_T07D3* | T07D3.9, isoform a OS=Caenorhabditis elegans GN=CELE_T07D3.9 PE=1 SV=3 |
| *fli-1* | Protein flightless-1 homolog OS=Caenorhabditis elegans GN=fli-1 PE=2 SV=2 |
| *CELE_F55A4* | F55A4.7 OS=Caenorhabditis elegans GN=CELE_F55A4.7 PE=1 SV=4 |
| *CELE_T02H6* | T02H6.1, isoform b OS=Caenorhabditis elegans GN=CELE_T02H6.1 PE=1 SV=1 |
| *pmp-1* | PMP-1 OS=Caenorhabditis elegans GN=pmp-1 PE=1 SV=1 |
| *CELE_F25G6* | F25G6.9 OS=Caenorhabditis elegans GN=CELE_F25G6.9 PE=1 SV=2 |
| *CELE_T28D6* | T28D6.6, isoform a OS=Caenorhabditis elegans GN=CELE_T28D6.6 PE=1 SV=1 |
| *CELE_T08H10* | T08H10.1 OS=Caenorhabditis elegans GN=CELE_T08H10.1 PE=1 SV=2 |
| *hpo-34* | HPO-34, isoform a OS=Caenorhabditis elegans GN=hpo-34 PE=1 SV=3 |
| *CELE_M01E11* | M01E11.2 OS=Caenorhabditis elegans GN=CELE_M01E11.2 PE=1 SV=1 |
| *ugt-26* | UDP-glucuronosyltransferase OS=Caenorhabditis elegans GN=ugt-26 PE=1 SV=1 |
| *aagr-3* | AAGR-3, isoform a OS=Caenorhabditis elegans GN=aagr-3 PE=1 SV=1 |
| *ZK512* | Probable ATP-dependent RNA helicase DDX55 homolog OS=Caenorhabditis elegans GN=ZK512.2 PE=3 SV=2 |
| *asp-3* | Aspartic protease 3 OS=Caenorhabditis elegans GN=asp-3 PE=1 SV=3 |
| *CELE_M4* | M4.1, isoform b OS=Caenorhabditis elegans GN=CELE_M4.1 PE=1 SV=1 |
| *Y54G2A* | MANF/CDNF-like protein OS=Caenorhabditis elegans GN=Y54G2A.23 PE=3 SV=2 |
| *metr-1* | Probable methionine synthase OS=Caenorhabditis elegans GN=metr-1 PE=3 SV=1 |
| *acdh-11* | ACDH-11, isoform b OS=Caenorhabditis elegans GN=acdh-11 PE=1 SV=1 |
| *sac-1* | SAC-1 OS=Caenorhabditis elegans GN=sac-1 PE=1 SV=1 |
| *T26C12* | Acetolactate synthase-like protein OS=Caenorhabditis elegans GN=T26C12.1 PE=3 SV=2 |
| *puf-9* | PUF-9 OS=Caenorhabditis elegans GN=puf-9 PE=1 SV=2 |
| *C42C1* | C42C1.11, isoform c OS=Caenorhabditis elegans GN=C42C1.11 PE=1 SV=1 |
| *CELE_F42C5* | F42C5.9 OS=Caenorhabditis elegans GN=CELE_F42C5.9 PE=1 SV=1 |
| *alh-7* | ALH-7, isoform a OS=Caenorhabditis elegans GN=alh-7 PE=1 SV=2 |
| *CELE_R09H10* | R09H10.5 OS=Caenorhabditis elegans GN=CELE_R09H10.5 PE=1 SV=2 |
| *CELE_R13H4* | R13H4.2, isoform a OS=Caenorhabditis elegans GN=CELE_R13H4.2 PE=1 SV=1 |

| | |
|---|---|
| *E04D5* | Eukaryotic translation initiation factor 2A OS=Caenorhabditis elegans GN=E04D5.1 PE=3 SV=2 |
| *CELE_F21D5* | F21D5.7 OS=Caenorhabditis elegans GN=CELE_F21D5.7 PE=1 SV=1 |
| *hum-9* | HUM-9 OS=Caenorhabditis elegans GN=hum-9 PE=1 SV=2 |
| *tni-4* | Troponin I 4 OS=Caenorhabditis elegans GN=tni-4 PE=2 SV=2 |
| *ani-1* | Anillin-like protein 1 OS=Caenorhabditis elegans GN=ani-1 PE=2 SV=2 |
| *vha-19* | VHA-19 OS=Caenorhabditis elegans GN=vha-19 PE=1 SV=1 |
| *aagr-4* | AAGR-4 OS=Caenorhabditis elegans GN=aagr-4 PE=1 SV=3 |
| *CELE_Y71H10B* | Y71H10B.1, isoform c OS=Caenorhabditis elegans GN=CELE_Y71H10B.1 PE=1 SV=1 |
| *CELE_F32D8* | F32D8.5, isoform a OS=Caenorhabditis elegans GN=CELE_F32D8.5 PE=1 SV=1 |
| *art-1* | Probable very-long-chain enoyl-CoA reductase art-1 OS=Caenorhabditis elegans GN=art-1 PE=3 SV=1 |
| *mop-25* | MO25-like protein 2 OS=Caenorhabditis elegans GN=mop-25.2 PE=3 SV=1 |
| *abce-1* | ABCE-1 OS=Caenorhabditis elegans GN=abce-1 PE=1 SV=1 |
| *kin-3* | Casein kinase II subunit alpha OS=Caenorhabditis elegans GN=kin-3 PE=1 SV=1 |
| *acdh-7* | ACDH-7 OS=Caenorhabditis elegans GN=acdh-7 PE=1 SV=2 |
| *CELE_ZK1058* | ZK1058.9, isoform b OS=Caenorhabditis elegans GN=CELE_ZK1058.9 PE=1 SV=1 |
| *pgam-5* | Serine/threonine-protein phosphatase Pgam5, mitochondrial OS=Caenorhabditis elegans GN=pgam-5 PE=3 SV=2 |
| *clec-56* | CLEC-56 OS=Caenorhabditis elegans GN=clec-56 PE=1 SV=3 |
| *gsk-3* | Glycogen synthase kinase-3 OS=Caenorhabditis elegans GN=gsk-3 PE=1 SV=1 |
| *ula-1* | NEDD8-activating enzyme E1 regulatory subunit OS=Caenorhabditis elegans GN=ula-1 PE=3 SV=2 |
| *CELE_ZC247* | ZC247.1 OS=Caenorhabditis elegans GN=CELE_ZC247.1 PE=1 SV=1 |
| *H24K24* | Alcohol dehydrogenase class-3 OS=Caenorhabditis elegans GN=H24K24.3 PE=2 SV=1 |
| *trap-3* | TRAP-3 OS=Caenorhabditis elegans GN=trap-3 PE=1 SV=1 |
| *nex-3* | Annexin OS=Caenorhabditis elegans GN=nex-3 PE=1 SV=1 |
| *air-1* | AIR-1 OS=Caenorhabditis elegans GN=air-1 PE=1 SV=1 |
| *ttr-25* | TTR-25, isoform a OS=Caenorhabditis elegans GN=ttr-25 PE=1 SV=1 |
| *acdh-3* | ACDH-3 OS=Caenorhabditis elegans GN=acdh-3 PE=1 SV=1 |
| *chup-1* | CHUP-1, isoform b OS=Caenorhabditis elegans GN=chup-1 PE=1 SV=1 |
| *gna-2* | GNA-2 OS=Caenorhabditis elegans GN=gna-2 PE=1 SV=1 |
| *CELE_F19B6* | Uridine kinase OS=Caenorhabditis elegans GN=CELE_F19B6.1 PE=1 SV=1 |
| *agef-1* | AGEF-1, isoform a OS=Caenorhabditis elegans GN=agef-1 PE=1 SV=1 |
| *CELE_R09B3* | R09B3.3 OS=Caenorhabditis elegans GN=CELE_R09B3.3 PE=1 SV=1 |
| *CELE_T16G1* | T16G1.9 OS=Caenorhabditis elegans GN=CELE_T16G1.9 PE=1 SV=1 |
| *grld-1* | GRLD-1, isoform a OS=Caenorhabditis elegans GN=grld-1 PE=1 SV=1 |
| *rme-2* | RME-2, isoform a OS=Caenorhabditis elegans GN=rme-2 PE=1 SV=3 |
| *CELE_T03G6* | T03G6.3 OS=Caenorhabditis elegans GN=CELE_T03G6.3 PE=1 SV=2 |
| *snr-4* | Probable small nuclear ribonucleoprotein Sm D2 OS=Caenorhabditis elegans GN=snr-4 PE=1 SV=1 |
| *CELE_Y82E9BR* | Y82E9BR.14, isoform b OS=Caenorhabditis elegans GN=CELE_Y82E9BR.14 PE=1 SV=1 |

| | |
|---|---|
| *let-711* | LET-711 OS=Caenorhabditis elegans GN=let-711 PE=1 SV=3 |
| *faah-1* | Fatty acid amide hydrolase 1 OS=Caenorhabditis elegans GN=faah-1 PE=2 SV=1 |
| *rpb-2* | DNA-directed RNA polymerase II subunit RPB2 OS=Caenorhabditis elegans GN=rpb-2 PE=2 SV=2 |
| *lpd-6* | LPD-6 OS=Caenorhabditis elegans GN=lpd-6 PE=1 SV=2 |
| *F44B9* | Probable replication factor C subunit 5 OS=Caenorhabditis elegans GN=F44B9.8 PE=3 SV=3 |
| *ZK686* | Probable dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 3 OS=Caenorhabditis elegans GN=ZK686.3 PE=3 SV=2 |
| *haf-3* | HAF-3 OS=Caenorhabditis elegans GN=haf-3 PE=1 SV=2 |
| *riok-1* | Serine/threonine-protein kinase RIO1 OS=Caenorhabditis elegans GN=riok-1 PE=1 SV=5 |
| *odr-8* | Ufm1-specific protease OS=Caenorhabditis elegans GN=odr-8 PE=1 SV=1 |
| *CELE_R74* | R74.8, isoform a OS=Caenorhabditis elegans GN=CELE_R74.8 PE=1 SV=2 |
| *mtm-1* | MTM-1 OS=Caenorhabditis elegans GN=mtm-1 PE=1 SV=2 |
| *pfd-6* | Probable prefoldin subunit 6 OS=Caenorhabditis elegans GN=pfd-6 PE=2 SV=1 |
| *iffb-1* | Eukaryotic translation initiation factor eIF5B OS=Caenorhabditis elegans GN=iffb-1 PE=1 SV=1 |
| *CELE_JC8* | JC8.2 OS=Caenorhabditis elegans GN=CELE_JC8.2 PE=1 SV=2 |
| *kars-1* | Lysine--tRNA ligase OS=Caenorhabditis elegans GN=kars-1 PE=1 SV=1 |
| *pat-10* | PAT-10 OS=Caenorhabditis elegans GN=pat-10 PE=1 SV=1 |
| *rpl-10a* | 60S ribosomal protein L10a OS=Caenorhabditis elegans GN=rpl-10a PE=3 SV=1 |
| *CELE_Y71H2AM* | Y71H2AM.5 OS=Caenorhabditis elegans GN=CELE_Y71H2AM.5 PE=1 SV=1 |
| *unc-89* | Muscle M-line assembly protein unc-89 OS=Caenorhabditis elegans GN=unc-89 PE=1 SV=3 |
| *F37C4* | Protein F37C4.5 OS=Caenorhabditis elegans GN=F37C4.5 PE=1 SV=3 |
| *CELE_ZK418* | ZK418.9, isoform b OS=Caenorhabditis elegans GN=CELE_ZK418.9 PE=1 SV=1 |
| *pas-4* | Proteasome subunit alpha type-7 OS=Caenorhabditis elegans GN=pas-4 PE=1 SV=1 |
| *smk-1* | SMK-1, isoform a OS=Caenorhabditis elegans GN=smk-1 PE=1 SV=1 |
| *inos-1* | INOS-1 OS=Caenorhabditis elegans GN=inos-1 PE=1 SV=1 |
| *cdc-37* | CDC-37, isoform b OS=Caenorhabditis elegans GN=cdc-37 PE=1 SV=1 |
| *skpo-1* | Peroxidase skpo-1 OS=Caenorhabditis elegans GN=skpo-1 PE=2 SV=1 |
| *plst-1* | PLST-1, isoform a OS=Caenorhabditis elegans GN=plst-1 PE=1 SV=1 |
| *C31B8* | C31B8.8 OS=Caenorhabditis elegans GN=C31B8.8 PE=1 SV=1 |
| *fasn-1* | FASN-1 OS=Caenorhabditis elegans GN=fasn-1 PE=1 SV=2 |
| *B0024* | Putative pseudouridine synthase B0024.11 OS=Caenorhabditis elegans GN=B0024.11 PE=3 SV=1 |
| *ifb-2* | Intermediate filament protein ifb-2 OS=Caenorhabditis elegans GN=ifb-2 PE=1 SV=1 |
| *sur-6* | Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B OS=Caenorhabditis elegans GN=sur-6 PE=1 SV=1 |
| *lin-40* | LIN-40, isoform a OS=Caenorhabditis elegans GN=lin-40 PE=1 SV=2 |
| *nrs-1* | Asparagine--tRNA ligase, cytoplasmic OS=Caenorhabditis elegans GN=nrs-1 PE=3 SV=1 |
| *gpd-1* | Glyceraldehyde-3-phosphate dehydrogenase 1 OS=Caenorhabditis elegans GN=gpd-1 PE=3 SV=1 |

| | |
|---|---|
| *ttr-24* | TTR-24, isoform a OS=Caenorhabditis elegans GN=ttr-24 PE=1 SV=1 |
| *eif-3* | Eukaryotic translation initiation factor 3 subunit C OS=Caenorhabditis elegans GN=eif-3.C PE=3 SV=2 |
| *let-721* | Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial OS=Caenorhabditis elegans GN=let-721 PE=3 SV=2 |
| *myo-6* | MYO-6, isoform a OS=Caenorhabditis elegans GN=myo-6 PE=1 SV=2 |
| *tpp-2* | Tripeptidyl-peptidase 2 OS=Caenorhabditis elegans GN=tpp-2 PE=2 SV=1 |
| *hipr-1* | Huntington interacting protein related 1 OS=Caenorhabditis elegans GN=hipr-1 PE=3 SV=3 |
| *lrs-1* | Leucine--tRNA ligase OS=Caenorhabditis elegans GN=lrs-1 PE=3 SV=2 |
| *zyg-9* | Zygote defective protein 9 OS=Caenorhabditis elegans GN=zyg-9 PE=1 SV=1 |
| *lfi-1* | LFI-1, isoform a OS=Caenorhabditis elegans GN=lfi-1 PE=1 SV=1 |
| *ZC434* | Probable arginine kinase ZC434.8 OS=Caenorhabditis elegans GN=ZC434.8 PE=3 SV=1 |
| *vha-8* | VHA-8 OS=Caenorhabditis elegans GN=vha-8 PE=1 SV=1 |
| *CELE_Y57G11C* | Y57G11C.33, isoform b OS=Caenorhabditis elegans GN=CELE_Y57G11C.33 PE=1 SV=1 |
| *ret-1* | Reticulon-like protein OS=Caenorhabditis elegans GN=ret-1 PE=1 SV=1 |
| *cmd-1* | Calmodulin OS=Caenorhabditis elegans GN=cmd-1 PE=1 SV=3 |
| *vha-16* | VHA-16 OS=Caenorhabditis elegans GN=vha-16 PE=1 SV=1 |
| *kin-19* | Casein kinase I isoform alpha OS=Caenorhabditis elegans GN=kin-19 PE=3 SV=1 |
| *iff-1* | Eukaryotic translation initiation factor 5A OS=Caenorhabditis elegans GN=iff-1 PE=1 SV=1 |
| *CELE_F57F5* | F57F5.1 OS=Caenorhabditis elegans GN=CELE_F57F5.1 PE=1 SV=2 |
| *mlc-2* | Myosin regulatory light chain 2 OS=Caenorhabditis elegans GN=mlc-2 PE=3 SV=1 |
| *sel-9* | Suppressor/enhancer of lin-12 protein 9 OS=Caenorhabditis elegans GN=sel-9 PE=1 SV=1 |
| *mccc-1* | MCCC-1 OS=Caenorhabditis elegans GN=mccc-1 PE=1 SV=2 |
| *fbp-1* | FBP-1, isoform a OS=Caenorhabditis elegans GN=fbp-1 PE=1 SV=1 |
| *ard-1* | ARD-1, isoform a OS=Caenorhabditis elegans GN=ard-1 PE=1 SV=1 |
| *rps-30* | RPS-30 OS=Caenorhabditis elegans GN=rps-30 PE=1 SV=1 |
| *nuo-1* | NUO-1 OS=Caenorhabditis elegans GN=nuo-1 PE=1 SV=1 |
| *pdhb-1* | Pyruvate dehydrogenase E1 component subunit beta, mitochondrial OS=Caenorhabditis elegans GN=pdhb-1 PE=1 SV=2 |
| *CELE_W06A7* | W06A7.2, isoform a OS=Caenorhabditis elegans GN=CELE_W06A7.2 PE=1 SV=1 |
| *CELE_Y18D10A* | Y18D10A.11 OS=Caenorhabditis elegans GN=CELE_Y18D10A.11 PE=1 SV=1 |
| *rpl-28* | 60S ribosomal protein L28 OS=Caenorhabditis elegans GN=rpl-28 PE=1 SV=3 |
| *CELE_F42G8* | F42G8.10 OS=Caenorhabditis elegans GN=CELE_F42G8.10 PE=1 SV=1 |
| *ugt-19* | UGT-19 OS=Caenorhabditis elegans GN=ugt-19 PE=1 SV=2 |
| *asna-1* | ATPase asna-1 OS=Caenorhabditis elegans GN=asna-1 PE=1 SV=1 |
| *CELE_F45H10* | F45H10.3 OS=Caenorhabditis elegans GN=CELE_F45H10.3 PE=1 SV=1 |
| *asns-2* | ASNS-2, isoform a OS=Caenorhabditis elegans GN=asns-2 PE=1 SV=1 |
| *CELE_M153* | M153.1 OS=Caenorhabditis elegans GN=CELE_M153.1 PE=1 SV=1 |
| *hpo-18* | HPO-18 OS=Caenorhabditis elegans GN=hpo-18 PE=1 SV=1 |

| | |
|---|---|
| *ubxn-4* | UBX domain-containing protein 4 OS=Caenorhabditis elegans GN=ubxn-4 PE=3 SV=1 |
| *uggt-1* | UGGT-1 OS=Caenorhabditis elegans GN=uggt-1 PE=1 SV=1 |
| *rho-1* | Ras-like GTP-binding protein rhoA OS=Caenorhabditis elegans GN=rho-1 PE=1 SV=1 |
| *rpa-2* | 60S acidic ribosomal protein P2 OS=Caenorhabditis elegans GN=rpa-2 PE=3 SV=2 |
| *CELE_ZK1055* | ZK1055.7 OS=Caenorhabditis elegans GN=CELE_ZK1055.7 PE=1 SV=1 |
| *CELE_F36F2* | F36F2.1, isoform a OS=Caenorhabditis elegans GN=CELE_F36F2.1 PE=1 SV=1 |
| *set-27* | SET-27 OS=Caenorhabditis elegans GN=set-27 PE=1 SV=2 |
| *gale-1* | GALE-1, isoform a OS=Caenorhabditis elegans GN=gale-1 PE=1 SV=1 |
| *pfd-5* | Probable prefoldin subunit 5 OS=Caenorhabditis elegans GN=pfd-5 PE=2 SV=1 |
| *acdh-10* | Probable medium-chain specific acyl-CoA dehydrogenase 10, mitochondrial OS=Caenorhabditis elegans GN=acdh-10 PE=2 SV=1 |
| *C30H6* | C30H6.7, isoform a OS=Caenorhabditis elegans GN=C30H6.7 PE=1 SV=1 |
| *dhs-30* | DHS-30 OS=Caenorhabditis elegans GN=dhs-30 PE=1 SV=1 |
| *cpna-2* | Copine family protein 2 OS=Caenorhabditis elegans GN=cpna-2 PE=2 SV=4 |
| *aqp-8* | AQP-8, isoform b OS=Caenorhabditis elegans GN=aqp-8 PE=3 SV=1 |
| *C01H6* | C01H6.2 OS=Caenorhabditis elegans GN=C01H6.2 PE=1 SV=2 |
| *puf-5* | Pumilio domain-containing protein 5 OS=Caenorhabditis elegans GN=puf-5 PE=2 SV=1 |
| *mrpl-12* | MRPL-12 OS=Caenorhabditis elegans GN=mrpl-12 PE=1 SV=1 |
| *phy-2* | Prolyl 4-hydroxylase subunit alpha-2 OS=Caenorhabditis elegans GN=phy-2 PE=1 SV=1 |
| *CELE_F49E2* | F49E2.2, isoform c OS=Caenorhabditis elegans GN=CELE_F49E2.2 PE=1 SV=1 |
| *ttll-5* | TTLL-5, isoform b OS=Caenorhabditis elegans GN=ttll-5 PE=1 SV=1 |
| *C30H6* | C30H6.7, isoform a OS=Caenorhabditis elegans GN=C30H6.7 PE=1 SV=1 |
| *CELE_DC2* | DC2.5 OS=Caenorhabditis elegans GN=CELE_DC2.5 PE=1 SV=4 |
| *spp-10* | SPP-10, isoform b OS=Caenorhabditis elegans GN=spp-10 PE=1 SV=3 |
| *CELE_F56A11* | F56A11.5, isoform a OS=Caenorhabditis elegans GN=CELE_F56A11.5 PE=1 SV=1 |
| *CELE_Y46E12BL* | Y46E12BL.2, isoform a OS=Caenorhabditis elegans GN=CELE_Y46E12BL.2 PE=1 SV=1 |
| *tbc-14* | TBC-14 OS=Caenorhabditis elegans GN=tbc-14 PE=1 SV=2 |
| *deps-1* | DEPS-1, isoform a OS=Caenorhabditis elegans GN=deps-1 PE=1 SV=1 |
| *lov-1* | Location of vulva defective 1 OS=Caenorhabditis elegans GN=lov-1 PE=1 SV=4 |
| *K12H4* | Probable signal peptidase complex subunit 3 OS=Caenorhabditis elegans GN=K12H4.4 PE=1 SV=1 |
| *CELE_T19D12* | T19D12.4, isoform a OS=Caenorhabditis elegans GN=CELE_T19D12.4 PE=1 SV=1 |
| *CELE_Y7A5A* | Y7A5A.1 OS=Caenorhabditis elegans GN=CELE_Y7A5A.1 PE=1 SV=1 |
| *qns-1* | QNS-1, isoform a OS=Caenorhabditis elegans GN=qns-1 PE=1 SV=1 |
| *tba-1* | TBA-1, isoform b OS=Caenorhabditis elegans GN=tba-1 PE=1 SV=1 |
| *CELE_K04F10* | K04F10.7 OS=Caenorhabditis elegans GN=CELE_K04F10.7 PE=1 SV=1 |
| *snb-1* | Synaptobrevin-1 OS=Caenorhabditis elegans GN=snb-1 PE=1 SV=1 |
| *CELE_F35D11* | F35D11.4 OS=Caenorhabditis elegans GN=CELE_F35D11.4 PE=1 SV=2 |

| | |
|---|---|
| *klp-18* | Kinesin-like protein OS=Caenorhabditis elegans GN=klp-18 PE=1 SV=1 |
| *C25H3* | C25H3.9, isoform b OS=Caenorhabditis elegans GN=C25H3.9 PE=1 SV=1 |
| *sptl-2* | Serine palmitoyltransferase 2 OS=Caenorhabditis elegans GN=sptl-2 PE=3 SV=1 |
| *C18E9* | C18E9.4 OS=Caenorhabditis elegans GN=C18E9.4 PE=1 SV=1 |
| *bag-1* | BAG family molecular chaperone regulator 1 OS=Caenorhabditis elegans GN=bag-1 PE=1 SV=1 |
| *mmaa-1* | Methylmalonic aciduria type A homolog, mitochondrial OS=Caenorhabditis elegans GN=mmaa-1 PE=3 SV=2 |
| *rbd-1* | RBD-1, isoform a OS=Caenorhabditis elegans GN=rbd-1 PE=1 SV=1 |
| *CELE_Y23H5B* | Y23H5B.5 OS=Caenorhabditis elegans GN=CELE_Y23H5B.5 PE=1 SV=2 |
| *CELE_F54E2* | F54E2.1 OS=Caenorhabditis elegans GN=CELE_F54E2.1 PE=1 SV=2 |
| *ragc-1* | RAGC-1, isoform a OS=Caenorhabditis elegans GN=ragc-1 PE=1 SV=1 |
| *eif-1* | EIF-1 OS=Caenorhabditis elegans GN=eif-1 PE=1 SV=1 |
| *mpst-1* | Putative thiosulfate sulfurtransferase mpst-1 OS=Caenorhabditis elegans GN=mpst-1 PE=3 SV=1 |
| *ddo-2* | D-aspartate oxidase 2 OS=Caenorhabditis elegans GN=ddo-2 PE=1 SV=1 |
| *CELE_T24C12* | T24C12.3 OS=Caenorhabditis elegans GN=CELE_T24C12.3 PE=1 SV=2 |
| *gpx-7* | Glutathione peroxidase OS=Caenorhabditis elegans GN=gpx-7 PE=1 SV=1 |
| *F09B12* | Putative phospholipase B-like 2 OS=Caenorhabditis elegans GN=F09B12.3 PE=1 SV=2 |
| *npp-4* | NPP-4 OS=Caenorhabditis elegans GN=npp-4 PE=1 SV=1 |
| *mrps-15* | 28S ribosomal protein S15, mitochondrial OS=Caenorhabditis elegans GN=mrps-15 PE=3 SV=1 |
| *ZK177* | Uncharacterized protein ZK177.8 OS=Caenorhabditis elegans GN=ZK177.8 PE=4 SV=2 |
| *CELE_ZK1307* | ZK1307.8 OS=Caenorhabditis elegans GN=CELE_ZK1307.8 PE=1 SV=1 |
| *dod-19* | DOD-19 OS=Caenorhabditis elegans GN=dod-19 PE=1 SV=2 |
| *mdt-9* | MDT-9, isoform b OS=Caenorhabditis elegans GN=mdt-9 PE=1 SV=1 |
| *F54D5* | Probable glutaryl-CoA dehydrogenase, mitochondrial OS=Caenorhabditis elegans GN=F54D5.7 PE=1 SV=1 |
| *unc-61* | UNC-61, isoform c OS=Caenorhabditis elegans GN=unc-61 PE=1 SV=2 |
| *aipl-1* | AIPL-1 OS=Caenorhabditis elegans GN=aipl-1 PE=1 SV=1 |
| *CELE_F28B3* | F28B3.10 OS=Caenorhabditis elegans GN=CELE_F28B3.10 PE=1 SV=1 |
| *cht-3* | Chitinase-like protein C25A8.4 OS=Caenorhabditis elegans GN=cht-3 PE=1 SV=3 |
| *moag-4* | MOAG-4 OS=Caenorhabditis elegans GN=moag-4 PE=1 SV=1 |
| *dad-1* | Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit dad-1 OS=Caenorhabditis elegans GN=dad-1 PE=3 SV=1 |
| *CELE_W04C9* | W04C9.4 OS=Caenorhabditis elegans GN=CELE_W04C9.4 PE=1 SV=1 |
| *pfd-4* | Probable prefoldin subunit 4 OS=Caenorhabditis elegans GN=pfd-4 PE=2 SV=1 |
| *CELE_R09E10* | R09E10.6 OS=Caenorhabditis elegans GN=CELE_R09E10.6 PE=1 SV=1 |
| *flu-2* | Kynureninase OS=Caenorhabditis elegans GN=flu-2 PE=2 SV=1 |
| *mei-1* | Meiotic spindle formation protein mei-1 OS=Caenorhabditis elegans GN=mei-1 PE=1 SV=1 |
| *ttr-14* | TTR-14 OS=Caenorhabditis elegans GN=ttr-14 PE=1 SV=1 |
| *ugt-16* | UDP-glucuronosyltransferase OS=Caenorhabditis elegans GN=ugt-16 PE=1 SV=1 |
| *pmp-2* | PMP-2 OS=Caenorhabditis elegans GN=pmp-2 PE=1 SV=1 |

| | |
|---|---|
| *teg-4* | TEG-4 OS=Caenorhabditis elegans GN=teg-4 PE=1 SV=1 |
| *CELE_D1086* | D1086.6 OS=Caenorhabditis elegans GN=CELE_D1086.6 PE=1 SV=5 |
| *CELE_ZK546* | ZK546.14, isoform b OS=Caenorhabditis elegans GN=CELE_ZK546.14 PE=1 SV=1 |
| *pafo-1* | RNA polymerase II-associated factor 1 homolog OS=Caenorhabditis elegans GN=pafo-1 PE=2 SV=2 |
| *tre-3* | Trehalase OS=Caenorhabditis elegans GN=tre-3 PE=1 SV=1 |
| *T23D8* | Protein LTV1 homolog OS=Caenorhabditis elegans GN=T23D8.3 PE=3 SV=1 |

**Supplemental Table 4** Summary of all proteins co-immunoprecipitated with ABU-13::tagRFP from overexpression animals that show a greater than 50% change in abundance compared to *abu-13p*::tagRFP animals (OS = Organism Name; GN = Gene Name; PE = Evidence of Protein Existence; SV = Sequence Version)

***Summary of proteins precipitated from N2 animals following biotinylated isoxazole treatment***

| Gene ID | Details |
|---|---|
| *ttc-7* | Y39A3CR.3, isoform a OS=Caenorhabditis elegans GN=ttc-7 PE=1 SV=1 |
| *CELE_K11D2* | K11D2.5 OS=Caenorhabditis elegans GN=CELE_K11D2.5 PE=4 SV=2 |
| *CELE_Y43F8B* | Y43F8B.1, isoform e OS=Caenorhabditis elegans GN=CELE_Y43F8B.1 PE=1 SV=1 |
| *C04F12* | C04F12.1 OS=Caenorhabditis elegans GN=C04F12.1 PE=1 SV=1 |
| *CELE_F07F6* | F07F6.8 OS=Caenorhabditis elegans GN=CELE_F07F6.8 PE=1 SV=1 |
| *calu-1* | CALU-1, isoform a OS=Caenorhabditis elegans GN=calu-1 PE=1 SV=1 |
| *jmjd-1* | Lysine-specific demethylase 7 homolog OS=Caenorhabditis elegans GN=jmjd-1.2 PE=1 SV=1 |
| *cdc-6* | CDC-6 OS=Caenorhabditis elegans GN=cdc-6 PE=1 SV=1 |
| *dhhc-8* | Palmitoyltransferase OS=Caenorhabditis elegans GN=dhhc-8 PE=3 SV=3 |
| *cdf-1* | Cation diffusion facilitator family protein 1 OS=Caenorhabditis elegans GN=cdf-1 PE=1 SV=1 |
| *Y53F4B* | Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1 OS=Caenorhabditis elegans GN=Y53F4B.13 PE=3 SV=3 |
| *CELE_W02D7* | W02D7.4 OS=Caenorhabditis elegans GN=CELE_W02D7.4 PE=1 SV=1 |
| *H03A11* | Extracellular serine/threonine protein kinase CeFam20 OS=Caenorhabditis elegans GN=H03A11.1 PE=1 SV=1 |
| *CELE_K08D9* | K08D9.4 OS=Caenorhabditis elegans GN=CELE_K08D9.4 PE=4 SV=1 |
| *ech-1* | T08B2.7, isoform a OS=Caenorhabditis elegans GN=ech-1.2 PE=1 SV=1 |
| *C49F5* | C49F5.6 OS=Caenorhabditis elegans GN=C49F5.6 PE=4 SV=1 |
| *nol-10* | NOL-10 OS=Caenorhabditis elegans GN=nol-10 PE=1 SV=1 |
| *sem-5* | Sex muscle abnormal protein 5 OS=Caenorhabditis elegans GN=sem-5 PE=1 SV=1 |
| *xpf-1* | XPF-1 OS=Caenorhabditis elegans GN=xpf-1 PE=1 SV=1 |
| *myo-6* | MYO-6, isoform a OS=Caenorhabditis elegans GN=myo-6 PE=1 SV=2 |
| *spon-1* | SPON-1 OS=Caenorhabditis elegans GN=spon-1 PE=1 SV=2 |
| *sqv-1* | SQV-1, isoform a OS=Caenorhabditis elegans GN=sqv-1 PE=1 SV=1 |
| *CELE_F26F12* | F26F12.3, isoform a OS=Caenorhabditis elegans GN=CELE_F26F12.3 PE=1 SV=1 |
| *CELE_F56D5* | F56D5.3 OS=Caenorhabditis elegans GN=CELE_F56D5.3 PE=1 SV=1 |
| *pid-1* | 21U-RNA biogenesis factor pid-1 OS=Caenorhabditis elegans GN=pid-1 PE=1 SV=1 |
| *parn-1* | PARN-1 OS=Caenorhabditis elegans GN=parn-1 PE=4 SV=2 |
| *C32E8* | C32E8.9 OS=Caenorhabditis elegans GN=C32E8.9 PE=1 SV=2 |
| *cids-1* | CID domain-containing protein 1 OS=Caenorhabditis elegans GN=cids-1 PE=3 SV=2 |
| *Y48A6B* | Putative H/ACA ribonucleoprotein complex subunit 2-like protein OS=Caenorhabditis elegans GN=Y48A6B.3 PE=3 SV=1 |
| *CELE_R10H10* | R10H10.7 OS=Caenorhabditis elegans GN=CELE_R10H10.7 PE=1 SV=5 |
| *T27F2* | Uncharacterized protein T27F2.1 OS=Caenorhabditis elegans GN=T27F2.1 PE=1 SV=1 |
| *dsbn-1* | Dysbindin protein homolog OS=Caenorhabditis elegans GN=dsbn-1 PE=1 SV=1 |
| *CELE_Y77E11A* | Y77E11A.7, isoform a OS=Caenorhabditis elegans GN=CELE_Y77E11A.7 PE=1 SV=1 |

| | |
|---|---|
| *CELE_ZC410* | ZC410.5, isoform a OS=Caenorhabditis elegans GN=CELE_ZC410.5 PE=1 SV=1 |
| *maoc-1* | MAOC-1 OS=Caenorhabditis elegans GN=maoc-1 PE=1 SV=1 |
| *gpa-12* | Guanine nucleotide-binding protein alpha-12 subunit OS=Caenorhabditis elegans GN=gpa-12 PE=1 SV=2 |
| *C25H3* | C25H3.9, isoform a OS=Caenorhabditis elegans GN=C25H3.9 PE=1 SV=1 |
| *aap-1* | AAP-1 OS=Caenorhabditis elegans GN=aap-1 PE=1 SV=1 |
| *inx-22* | Innexin OS=Caenorhabditis elegans GN=inx-22 PE=3 SV=2 |
| *CELE_Y73C8B* | Y73C8B.3 OS=Caenorhabditis elegans GN=CELE_Y73C8B.3 PE=1 SV=1 |
| *CELE_M18* | M18.3 OS=Caenorhabditis elegans GN=CELE_M18.3 PE=1 SV=1 |
| *CELE_F53H1* | F53H1.1, isoform a OS=Caenorhabditis elegans GN=CELE_F53H1.1 PE=1 SV=1 |
| *F42A8* | Uncharacterized protein F42A8.1 OS=Caenorhabditis elegans GN=F42A8.1 PE=4 SV=1 |
| *mls-1* | MLS-1 OS=Caenorhabditis elegans GN=mls-1 PE=4 SV=1 |
| *sds-22* | Protein phosphatase 1 regulatory subunit SDS22 homolog OS=Caenorhabditis elegans GN=sds-22 PE=3 SV=1 |
| *golg-4* | F59A2.6 OS=Caenorhabditis elegans GN=golg-4 PE=1 SV=1 |
| *C06A6* | Thioredoxin domain-containing protein C06A6.5 OS=Caenorhabditis elegans GN=C06A6.5 PE=1 SV=2 |
| *laat-1* | Lysosomal amino acid transporter 1 OS=Caenorhabditis elegans GN=laat-1 PE=1 SV=2 |
| *C50D2* | Probable ADP-dependent glucokinase OS=Caenorhabditis elegans GN=C50D2.7 PE=1 SV=1 |
| *CELE_Y54E5A* | Y54E5A.5 OS=Caenorhabditis elegans GN=CELE_Y54E5A.5 PE=1 SV=1 |
| *rba-1* | Probable histone-binding protein rba-1 OS=Caenorhabditis elegans GN=rba-1 PE=3 SV=1 |
| *mdt-15* | Mediator of RNA polymerase II transcription subunit 15 OS=Caenorhabditis elegans GN=mdt-15 PE=1 SV=3 |
| *C33A12* | Probable NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5 OS=Caenorhabditis elegans GN=C33A12.1 PE=3 SV=1 |
| *CELE_K01C8* | K01C8.1 OS=Caenorhabditis elegans GN=CELE_K01C8.1 PE=1 SV=2 |
| *CELE_ZK484* | ZK484.4, isoform b OS=Caenorhabditis elegans GN=CELE_ZK484.4 PE=1 SV=1 |
| *R107* | ATP-dependent (S)-NAD(P)H-hydrate dehydratase OS=Caenorhabditis elegans GN=R107.2 PE=3 SV=3 |
| *syx-18* | SYX-18 OS=Caenorhabditis elegans GN=syx-18 PE=1 SV=1 |
| *C05D11* | Uncharacterized protein C05D11.1 OS=Caenorhabditis elegans GN=C05D11.1 PE=4 SV=2 |
| *rabn-5* | RABN-5 OS=Caenorhabditis elegans GN=rabn-5 PE=1 SV=1 |
| *pssy-2* | PSSY-2 OS=Caenorhabditis elegans GN=pssy-2 PE=4 SV=4 |
| *ipla-7* | T04B2.5 OS=Caenorhabditis elegans GN=ipla-7 PE=1 SV=1 |
| *rle-1* | Regulation of longevity by E3 ubiquitin-protein ligase OS=Caenorhabditis elegans GN=rle-1 PE=1 SV=4 |
| *arl-5* | ADP-ribosylation factor-like protein 5 OS=Caenorhabditis elegans GN=arl-5 PE=3 SV=3 |
| *sars-2* | SARS-2, isoform a OS=Caenorhabditis elegans GN=sars-2 PE=1 SV=1 |
| *CELE_R09F10* | R09F10.8 OS=Caenorhabditis elegans GN=CELE_R09F10.8 PE=1 SV=3 |
| *scc-1* | Sister chromatid cohesion protein 1 OS=Caenorhabditis elegans GN=scc-1 PE=1 SV=1 |

| | |
|---|---|
| *tut-2* | Cytoplasmic tRNA 2-thiolation protein 2 OS=Caenorhabditis elegans GN=tut-2 PE=3 SV=2 |
| *rab-33* | Rab-33 (Fragment) OS=Caenorhabditis elegans GN=rab-33 PE=2 SV=1 |
| *exos-7* | EXOS-7 OS=Caenorhabditis elegans GN=exos-7 PE=1 SV=1 |
| *CELE_Y46G5A* | Y46G5A.34 OS=Caenorhabditis elegans GN=CELE_Y46G5A.34 PE=4 SV=1 |
| *taf-5* | TAF-5 OS=Caenorhabditis elegans GN=taf-5 PE=1 SV=1 |
| *ccnk-1* | CCNK-1 OS=Caenorhabditis elegans GN=ccnk-1 PE=1 SV=2 |
| *dur-1* | DUR-1, isoform k OS=Caenorhabditis elegans GN=dur-1 PE=1 SV=1 |
| *B0495* | Putative endoplasmic reticulum metallopeptidase 1-A OS=Caenorhabditis elegans GN=B0495.7 PE=1 SV=2 |
| *CELE_Y48G1C* | Y48G1C.6 OS=Caenorhabditis elegans GN=CELE_Y48G1C.6 PE=4 SV=1 |
| *CELE_Y54E5A* | Y54E5A.5 OS=Caenorhabditis elegans GN=CELE_Y54E5A.5 PE=1 SV=1 |
| *ced-6* | Cell death protein 6 OS=Caenorhabditis elegans GN=ced-6 PE=1 SV=1 |
| *cdr-1* | CDR-1 OS=Caenorhabditis elegans GN=cdr-1 PE=1 SV=1 |
| *CELE_Y63D3A* | Y63D3A.7 OS=Caenorhabditis elegans GN=CELE_Y63D3A.7 PE=1 SV=2 |
| *T03F6* | Probable glucosamine-6-phosphate isomerase OS=Caenorhabditis elegans GN=T03F6.3 PE=3 SV=1 |
| *gosr-2* | Probable Golgi SNAP receptor complex member 2 OS=Caenorhabditis elegans GN=gosr-2.1 PE=3 SV=1 |
| *grp-1* | GTP exchange factor for ARFs 1 OS=Caenorhabditis elegans GN=grp-1 PE=3 SV=2 |
| *CELE_W10C8* | W10C8.5 OS=Caenorhabditis elegans GN=CELE_W10C8.5 PE=1 SV=1 |
| *cdd-1* | CDD-1 OS=Caenorhabditis elegans GN=cdd-1 PE=1 SV=1 |
| *lact-1* | LACT-1 OS=Caenorhabditis elegans GN=lact-1 PE=4 SV=2 |
| *anmt-2* | ANMT-2 OS=Caenorhabditis elegans GN=anmt-2 PE=1 SV=2 |
| *C01H6* | C01H6.2 OS=Caenorhabditis elegans GN=C01H6.2 PE=1 SV=2 |
| *wdr-5* | WD repeat-containing protein wdr-5.1 OS=Caenorhabditis elegans GN=wdr-5.1 PE=1 SV=1 |
| *CELE_K04G7* | K04G7.1 OS=Caenorhabditis elegans GN=CELE_K04G7.1 PE=1 SV=1 |
| *CELE_W01B11* | W01B11.6, isoform a OS=Caenorhabditis elegans GN=CELE_W01B11.6 PE=1 SV=2 |
| *Y39B6A* | Uncharacterized protein Y39B6A.33 OS=Caenorhabditis elegans GN=Y39B6A.33 PE=3 SV=1 |
| *B0024* | B0024.10 OS=Caenorhabditis elegans GN=B0024.10 PE=1 SV=2 |
| *lgg-3* | Ubiquitin-like protein ATG12 OS=Caenorhabditis elegans GN=lgg-3 PE=3 SV=1 |
| *cyc-2* | Cytochrome c 2.1 OS=Caenorhabditis elegans GN=cyc-2.1 PE=1 SV=2 |
| *CELE_Y4C6B* | Y4C6B.7, isoform b OS=Caenorhabditis elegans GN=CELE_Y4C6B.7 PE=1 SV=1 |
| *CELE_F41D9* | F41D9.2, isoform b OS=Caenorhabditis elegans GN=CELE_F41D9.2 PE=1 SV=1 |
| *dve-1* | DVE-1, isoform a OS=Caenorhabditis elegans GN=dve-1 PE=1 SV=1 |
| *gst-21* | GST-21, isoform b OS=Caenorhabditis elegans GN=gst-21 PE=1 SV=1 |
| *M142* | Protein-lysine N-methyltransferase M142.8 OS=Caenorhabditis elegans GN=M142.8 PE=3 SV=1 |
| *CELE_K09E4* | K09E4.3, isoform a OS=Caenorhabditis elegans GN=CELE_K09E4.3 PE=1 SV=1 |
| *CELE_F47G3* | F47G3.3 OS=Caenorhabditis elegans GN=CELE_F47G3.3 PE=1 SV=1 |
| *C44B7* | Uncharacterized protein C44B7.5 OS=Caenorhabditis elegans GN=C44B7.5 PE=1 SV=1 |

| | |
|---|---|
| *gos-28* | Golgi SNAP receptor complex member 1 OS=Caenorhabditis elegans GN=gos-28 PE=2 SV=1 |
| *mif-1* | MIF-1 OS=Caenorhabditis elegans GN=mif-1 PE=1 SV=1 |
| *paqr-1* | Progestin and adipoQ receptor-like protein 1 OS=Caenorhabditis elegans GN=paqr-1 PE=3 SV=2 |
| *T19B4* | Adenine phosphoribosyltransferase OS=Caenorhabditis elegans GN=T19B4.3 PE=2 SV=1 |
| *CELE_F59E11* | F59E11.5 OS=Caenorhabditis elegans GN=CELE_F59E11.5 PE=4 SV=2 |
| *gly-13* | Putative alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase OS=Caenorhabditis elegans GN=gly-13 PE=2 SV=2 |
| *lron-10* | LRON-10 OS=Caenorhabditis elegans GN=lron-10 PE=1 SV=1 |
| *fbxa-156* | FBXA-156 OS=Caenorhabditis elegans GN=fbxa-156 PE=4 SV=2 |
| *pud-2* | PUD-2.1 OS=Caenorhabditis elegans GN=pud-2.1 PE=1 SV=1 |
| *C13F10* | C13F10.5 OS=Caenorhabditis elegans GN=C13F10.5 PE=4 SV=2 |
| *dgk-2* | Diacylglycerol kinase OS=Caenorhabditis elegans GN=dgk-2 PE=1 SV=1 |
| *ras-2* | GTP binding protein OS=Caenorhabditis elegans GN=ras-2 PE=1 SV=1 |
| *cept-1* | CEPT-1, isoform b OS=Caenorhabditis elegans GN=cept-1 PE=1 SV=1 |
| *ZK632* | Probable mannose-6-phosphate isomerase OS=Caenorhabditis elegans GN=ZK632.4 PE=3 SV=3 |
| *CELE_Y87G2A* | Y87G2A.19 OS=Caenorhabditis elegans GN=CELE_Y87G2A.19 PE=1 SV=1 |
| *puf-8* | PUF-8 OS=Caenorhabditis elegans GN=puf-8 PE=1 SV=3 |
| *CELE_E04D5* | E04D5.2 OS=Caenorhabditis elegans GN=CELE_E04D5.2 PE=4 SV=2 |
| *CELE_Y53C12A* | Protein YIPF OS=Caenorhabditis elegans GN=CELE_Y53C12A.3 PE=1 SV=2 |
| *CELE_Y119D3B* | Y119D3B.12, isoform a OS=Caenorhabditis elegans GN=CELE_Y119D3B.12 PE=1 SV=1 |
| *strd-1* | STE20-related kinase adapter protein strd-1 OS=Caenorhabditis elegans GN=strd-1 PE=1 SV=1 |
| *CELE_D1069* | D1069.3, isoform b OS=Caenorhabditis elegans GN=CELE_D1069.3 PE=1 SV=1 |
| *alh-3* | 10-formyltetrahydrofolate dehydrogenase OS=Caenorhabditis elegans GN=alh-3 PE=1 SV=1 |
| *hum-1* | HUM-1 OS=Caenorhabditis elegans GN=hum-1 PE=1 SV=1 |
| *iglr-2* | Immunoglobulin domain and leucine-rich repeat-containing protein 2 OS=Caenorhabditis elegans GN=iglr-2 PE=1 SV=2 |
| *C49A9* | C49A9.9, isoform a OS=Caenorhabditis elegans GN=C49A9.9 PE=1 SV=1 |
| *CELE_D1069* | D1069.3, isoform b OS=Caenorhabditis elegans GN=CELE_D1069.3 PE=1 SV=1 |
| *ugt-39* | UGT-39, isoform a OS=Caenorhabditis elegans GN=ugt-39 PE=1 SV=1 |
| *CELE_K02F6* | K02F6.7 OS=Caenorhabditis elegans GN=CELE_K02F6.7 PE=1 SV=2 |
| *let-413* | Protein lap1 OS=Caenorhabditis elegans GN=let-413 PE=1 SV=3 |
| *ugt-47* | Putative UDP-glucuronosyltransferase ugt-47 OS=Caenorhabditis elegans GN=ugt-47 PE=1 SV=2 |
| *npp-4* | NPP-4 OS=Caenorhabditis elegans GN=npp-4 PE=1 SV=1 |
| *aagr-4* | AAGR-4 OS=Caenorhabditis elegans GN=aagr-4 PE=1 SV=3 |
| *pdcd-2* | PDCD-2, isoform a OS=Caenorhabditis elegans GN=pdcd-2 PE=1 SV=2 |
| *CELE_Y37A1B* | Y37A1B.17, isoform a OS=Caenorhabditis elegans GN=CELE_Y37A1B.17 PE=1 SV=1 |
| *prp-3* | PRP-3 OS=Caenorhabditis elegans GN=prp-3 PE=1 SV=1 |

| | |
|---|---|
| *mrpl-1* | MRPL-1 OS=Caenorhabditis elegans GN=mrpl-1 PE=1 SV=1 |
| *syx-5* | Putative syntaxin-5 OS=Caenorhabditis elegans GN=syx-5 PE=3 SV=1 |
| *CELE_T14B4* | T14B4.1 OS=Caenorhabditis elegans GN=CELE_T14B4.1 PE=1 SV=3 |
| *wdr-48* | WD repeat-containing protein 48 homolog OS=Caenorhabditis elegans GN=wdr-48 PE=1 SV=2 |
| *CELE_F21H7* | Major sperm protein OS=Caenorhabditis elegans GN=CELE_F21H7.5 PE=1 SV=1 |
| *coq-3* | COQ-3, isoform a OS=Caenorhabditis elegans GN=coq-3 PE=1 SV=1 |
| *CELE_T26A5* | T26A5.6 OS=Caenorhabditis elegans GN=CELE_T26A5.6 PE=1 SV=3 |
| *C05C8* | KRR1 small subunit processome component OS=Caenorhabditis elegans GN=C05C8.2 PE=1 SV=1 |
| *rad-51* | RAD-51, isoform b OS=Caenorhabditis elegans GN=rad-51 PE=1 SV=1 |
| *smc-6* | SMC-6 OS=Caenorhabditis elegans GN=smc-6 PE=1 SV=1 |
| *tdp-1* | Tar DNA-binding protein homolog 1 OS=Caenorhabditis elegans GN=tdp-1 PE=1 SV=1 |
| *C13B4* | C13B4.1, isoform a OS=Caenorhabditis elegans GN=C13B4.1 PE=1 SV=3 |
| *CELE_F10G7* | F10G7.10, isoform a OS=Caenorhabditis elegans GN=CELE_F10G7.10 PE=1 SV=1 |
| *CELE_Y34B4A* | Y34B4A.9 OS=Caenorhabditis elegans GN=CELE_Y34B4A.9 PE=1 SV=1 |
| *gad-3* | GAD-3 OS=Caenorhabditis elegans GN=gad-3 PE=1 SV=3 |
| *arf-1* | ADP-ribosylation factor 1-like 2 OS=Caenorhabditis elegans GN=arf-1.2 PE=2 SV=2 |
| *cyp-14a5* | CYP-14A5 OS=Caenorhabditis elegans GN=cyp-14a5 PE=1 SV=1 |
| *kin-1* | cAMP-dependent protein kinase catalytic subunit OS=Caenorhabditis elegans GN=kin-1 PE=1 SV=3 |
| *ero-1* | ERO-1, isoform c OS=Caenorhabditis elegans GN=ero-1 PE=1 SV=1 |
| *inos-1* | INOS-1 OS=Caenorhabditis elegans GN=inos-1 PE=1 SV=1 |
| *CELE_Y51A2D* | Y51A2D.7, isoform b OS=Caenorhabditis elegans GN=CELE_Y51A2D.7 PE=1 SV=2 |
| *CELE_W04B5* | W04B5.5 OS=Caenorhabditis elegans GN=CELE_W04B5.5 PE=1 SV=1 |
| *CELE_F40F4* | F40F4.7 OS=Caenorhabditis elegans GN=CELE_F40F4.7 PE=1 SV=4 |
| *CELE_Y6B3B* | Y6B3B.9, isoform a OS=Caenorhabditis elegans GN=CELE_Y6B3B.9 PE=1 SV=2 |
| *rsp-2* | Probable splicing factor, arginine/serine-rich 2 OS=Caenorhabditis elegans GN=rsp-2 PE=3 SV=1 |
| *mrpl-35* | Probable 39S ribosomal protein L35, mitochondrial OS=Caenorhabditis elegans GN=mrpl-35 PE=3 SV=2 |
| *tra-3* | Calpain-5 OS=Caenorhabditis elegans GN=tra-3 PE=1 SV=1 |
| *mrrf-1* | Ribosome-recycling factor, mitochondrial OS=Caenorhabditis elegans GN=mrrf-1 PE=3 SV=2 |
| *CELE_F32H2* | F32H2.10 OS=Caenorhabditis elegans GN=CELE_F32H2.10 PE=4 SV=1 |
| *parg-1* | PME-3, isoform f OS=Caenorhabditis elegans GN=parg-1 PE=1 SV=1 |
| *acl-3* | ACL-3 OS=Caenorhabditis elegans GN=acl-3 PE=1 SV=2 |
| *W09C3* | Probable DNA-directed RNA polymerase III subunit RPC6 OS=Caenorhabditis elegans GN=W09C3.4 PE=3 SV=1 |
| *mdt-20* | Mediator of RNA polymerase II transcription subunit 20 OS=Caenorhabditis elegans GN=mdt-20 PE=3 SV=2 |
| *C34E10* | C34E10.10 OS=Caenorhabditis elegans GN=C34E10.10 PE=1 SV=1 |
| *CELE_F25E2* | F25E2.2, isoform a OS=Caenorhabditis elegans GN=CELE_F25E2.2 PE=1 SV=2 |

| | |
|---|---|
| *CELE_Y94H6A* | Y94H6A.10 OS=Caenorhabditis elegans GN=CELE_Y94H6A.10 PE=1 SV=1 |
| *C56C10* | C56C10.11 OS=Caenorhabditis elegans GN=C56C10.11 PE=1 SV=2 |
| *C05D12* | C05D12.3, isoform a OS=Caenorhabditis elegans GN=C05D12.3 PE=1 SV=1 |
| *clk-1* | 5-demethoxyubiquinone hydroxylase, mitochondrial OS=Caenorhabditis elegans GN=clk-1 PE=1 SV=1 |
| *nhr-35* | Nuclear hormone receptor family member nhr-35 OS=Caenorhabditis elegans GN=nhr-35 PE=2 SV=3 |
| *erl-1* | ERL-1 OS=Caenorhabditis elegans GN=erl-1 PE=1 SV=1 |
| *tag-257* | TAG-257 OS=Caenorhabditis elegans GN=tag-257 PE=1 SV=1 |
| *jac-1* | JAC-1, isoform e OS=Caenorhabditis elegans GN=jac-1 PE=1 SV=1 |
| *CELE_F58H1* | F58H1.8 OS=Caenorhabditis elegans GN=CELE_F58H1.8 PE=1 SV=1 |
| *trpp-4* | TRPP-4 OS=Caenorhabditis elegans GN=trpp-4 PE=1 SV=1 |
| *hat-1* | HAT-1 OS=Caenorhabditis elegans GN=hat-1 PE=1 SV=2 |
| *clec-78* | CLEC-78 OS=Caenorhabditis elegans GN=clec-78 PE=1 SV=2 |
| *CELE_Y7A9A* | Y7A9A.1 OS=Caenorhabditis elegans GN=CELE_Y7A9A.1 PE=1 SV=2 |
| *rae-1* | mRNA export factor rae-1 OS=Caenorhabditis elegans GN=rae-1 PE=1 SV=1 |
| *tufm-2* | Mitochondrial elongation factor Tu homologue OS=Caenorhabditis elegans GN=tufm-2 PE=1 SV=1 |
| *sup-35* | SUP-35 OS=Caenorhabditis elegans GN=sup-35 PE=1 SV=1 |
| *CELE_Y69H2* | Y69H2.7, isoform a OS=Caenorhabditis elegans GN=CELE_Y69H2.7 PE=1 SV=3 |
| *CELE_H43I07* | H43I07.3 OS=Caenorhabditis elegans GN=CELE_H43I07.3 PE=1 SV=1 |
| *C17G10* | C17G10.2 OS=Caenorhabditis elegans GN=C17G10.2 PE=1 SV=2 |
| *cht-3* | Chitinase-like protein C25A8.4 OS=Caenorhabditis elegans GN=cht-3 PE=1 SV=3 |
| *xpb-1* | XPB-1 OS=Caenorhabditis elegans GN=xpb-1 PE=1 SV=1 |
| *exc-4* | Chloride intracellular channel exc-4 OS=Caenorhabditis elegans GN=exc-4 PE=1 SV=2 |
| *pbrm-1* | PBRM-1, isoform a OS=Caenorhabditis elegans GN=pbrm-1 PE=1 SV=1 |
| *sun-1* | Sun domain-containing protein 1 OS=Caenorhabditis elegans GN=sun-1 PE=3 SV=1 |
| *CELE_M01B12* | M01B12.4, isoform b OS=Caenorhabditis elegans GN=CELE_M01B12.4 PE=1 SV=1 |
| *CELE_F17H10* | F17H10.1 OS=Caenorhabditis elegans GN=CELE_F17H10.1 PE=1 SV=1 |
| *tim-16* | Mitochondrial import inner membrane translocase subunit tim-16 OS=Caenorhabditis elegans GN=tim-16 PE=3 SV=1 |
| *CELE_ZK418* | ZK418.9, isoform a OS=Caenorhabditis elegans GN=CELE_ZK418.9 PE=1 SV=2 |
| *acl-9* | ACL-9 OS=Caenorhabditis elegans GN=acl-9 PE=1 SV=2 |
| *gly-9* | Probable N-acetylgalactosaminyltransferase 9 OS=Caenorhabditis elegans GN=gly-9 PE=2 SV=1 |
| *CELE_F30F8* | F30F8.9, isoform a OS=Caenorhabditis elegans GN=CELE_F30F8.9 PE=1 SV=1 |
| *C04H5* | Multifunctional methyltransferase subunit TRM112-like protein OS=Caenorhabditis elegans GN=C04H5.1 PE=3 SV=1 |
| *CELE_Y75B8A* | Y75B8A.24 OS=Caenorhabditis elegans GN=CELE_Y75B8A.24 PE=1 SV=2 |
| *did-2* | DID-2 OS=Caenorhabditis elegans GN=did-2 PE=1 SV=1 |
| *CELE_T08B1* | T08B1.1 OS=Caenorhabditis elegans GN=CELE_T08B1.1 PE=1 SV=2 |
| *CELE_Y87G2A* | Y87G2A.19 OS=Caenorhabditis elegans GN=CELE_Y87G2A.19 PE=1 SV=1 |

| | |
|---|---|
| *smg-6* | SMG-6, isoform a OS=Caenorhabditis elegans GN=smg-6 PE=1 SV=3 |
| *B0205* | B0205.6 OS=Caenorhabditis elegans GN=B0205.6 PE=1 SV=2 |
| *CELE_F33G12* | F33G12.3 OS=Caenorhabditis elegans GN=CELE_F33G12.3 PE=1 SV=1 |
| *C15C7* | C15C7.5 OS=Caenorhabditis elegans GN=C15C7.5 PE=1 SV=1 |
| *CELE_Y55F3AR* | Y55F3AR.2 OS=Caenorhabditis elegans GN=CELE_Y55F3AR.2 PE=1 SV=2 |
| *orc-5* | ORC-5, isoform a OS=Caenorhabditis elegans GN=orc-5 PE=4 SV=2 |
| *rpb-11* | Probable DNA-directed RNA polymerase II subunit RPB11 OS=Caenorhabditis elegans GN=rpb-11 PE=3 SV=1 |
| *CELE_VF13D12L* | VF13D12L.3 OS=Caenorhabditis elegans GN=CELE_VF13D12L.3 PE=1 SV=1 |
| *C34F6* | C34F6.9 OS=Caenorhabditis elegans GN=C34F6.9 PE=1 SV=1 |
| *riok-3* | Putative RIO-type serine/threonine-protein kinase 3 OS=Caenorhabditis elegans GN=riok-3 PE=3 SV=1 |
| *rpc-25* | RPC-25 OS=Caenorhabditis elegans GN=rpc-25 PE=1 SV=1 |
| *tag-10* | P52CeGA (Fragment) OS=Caenorhabditis elegans GN=tag-10 PE=2 SV=1 |
| *CELE_F53C3* | F53C3.13, isoform c OS=Caenorhabditis elegans GN=CELE_F53C3.13 PE=1 SV=1 |
| *mut-16* | MUT-16, isoform a OS=Caenorhabditis elegans GN=mut-16 PE=1 SV=4 |
| *mrps-11* | MRPS-11 OS=Caenorhabditis elegans GN=mrps-11 PE=1 SV=1 |
| *CELE_F59A3* | F59A3.12 OS=Caenorhabditis elegans GN=CELE_F59A3.12 PE=4 SV=2 |
| *nuaf-3* | NUAF-3, isoform a OS=Caenorhabditis elegans GN=nuaf-3 PE=1 SV=2 |
| *lin-61* | Protein lin-61 OS=Caenorhabditis elegans GN=lin-61 PE=1 SV=1 |
| *jmjc-1* | Bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66 OS=Caenorhabditis elegans GN=jmjc-1 PE=3 SV=2 |
| *pho-5* | Putative acid phosphatase 5 OS=Caenorhabditis elegans GN=pho-5 PE=1 SV=3 |
| *abts-2* | ABTS-2 OS=Caenorhabditis elegans GN=abts-2 PE=1 SV=1 |
| *D1044* | Uncharacterized kinase-like protein D1044.1 OS=Caenorhabditis elegans GN=D1044.1 PE=3 SV=2 |
| *fbxa-101* | FBXA-101 OS=Caenorhabditis elegans GN=fbxa-101 PE=1 SV=1 |
| *gst-41* | GST-41 OS=Caenorhabditis elegans GN=gst-41 PE=1 SV=1 |
| *F54D5* | Probable glutaryl-CoA dehydrogenase, mitochondrial OS=Caenorhabditis elegans GN=F54D5.7 PE=1 SV=1 |
| *aph-1* | Gamma-secretase subunit aph-1 OS=Caenorhabditis elegans GN=aph-1 PE=1 SV=1 |
| *CELE_R144* | R144.13 OS=Caenorhabditis elegans GN=CELE_R144.13 PE=1 SV=1 |
| *dox-1* | DOX-1 OS=Caenorhabditis elegans GN=dox-1 PE=1 SV=1 |
| *CELE_Y110A7A* | Y110A7A.15, isoform a OS=Caenorhabditis elegans GN=CELE_Y110A7A.15 PE=1 SV=1 |
| *Y67D2* | Probable ATP-dependent RNA helicase DHX35 homolog OS=Caenorhabditis elegans GN=Y67D2.6 PE=3 SV=1 |
| *cbs-1* | CBS-1, isoform a OS=Caenorhabditis elegans GN=cbs-1 PE=1 SV=1 |
| *C49A9* | C49A9.9, isoform a OS=Caenorhabditis elegans GN=C49A9.9 PE=1 SV=1 |
| *adbp-1* | 2L737 OS=Caenorhabditis elegans GN=adbp-1 PE=1 SV=1 |
| *CELE_W02G9* | W02G9.3 OS=Caenorhabditis elegans GN=CELE_W02G9.3 PE=1 SV=1 |
| *tag-131* | Neuferricin homolog OS=Caenorhabditis elegans GN=tag-131 PE=3 SV=1 |
| *exl-1* | Chloride intracellular channel exl-1 OS=Caenorhabditis elegans GN=exl-1 PE=2 SV=2 |
| *exos-9* | EXOS-9, isoform a OS=Caenorhabditis elegans GN=exos-9 PE=1 SV=2 |

| | |
|---|---|
| *aps-3* | APS-3 OS=Caenorhabditis elegans GN=aps-3 PE=1 SV=1 |
| *acbp-5* | ACBP-5 OS=Caenorhabditis elegans GN=acbp-5 PE=1 SV=2 |
| *C06H2* | C06H2.2 OS=Caenorhabditis elegans GN=C06H2.2 PE=1 SV=3 |
| *tnt-3* | TNT-3, isoform b OS=Caenorhabditis elegans GN=tnt-3 PE=1 SV=1 |
| *CELE_Y39A1A* | Y39A1A.22 OS=Caenorhabditis elegans GN=CELE_Y39A1A.22 PE=1 SV=1 |
| *CELE_F59B1* | F59B1.8 OS=Caenorhabditis elegans GN=CELE_F59B1.8 PE=1 SV=2 |
| *CELE_Y69A2AR* | Y69A2AR.18, isoform a OS=Caenorhabditis elegans GN=CELE_Y69A2AR.18 PE=1 SV=1 |
| *ZK1098* | Probable translation initiation factor eIF-2B subunit alpha OS=Caenorhabditis elegans GN=ZK1098.4 PE=1 SV=1 |
| *rfp-1* | E3 ubiquitin-protein ligase bre-1 OS=Caenorhabditis elegans GN=rfp-1 PE=1 SV=2 |
| *hint-1* | Histidine triad nucleotide-binding protein 1 OS=Caenorhabditis elegans GN=hint-1 PE=3 SV=1 |
| *CELE_F10C1* | F10C1.9, isoform b OS=Caenorhabditis elegans GN=CELE_F10C1.9 PE=1 SV=1 |
| *B0513* | Proline dehydrogenase 1, mitochondrial OS=Caenorhabditis elegans GN=B0513.5 PE=3 SV=2 |
| *trpp-10* | TRPP-10, isoform b OS=Caenorhabditis elegans GN=trpp-10 PE=1 SV=2 |
| *crls-1* | Probable cardiolipin synthase (CMP-forming) OS=Caenorhabditis elegans GN=crls-1 PE=3 SV=1 |
| *pir-1* | PIR-1, isoform a OS=Caenorhabditis elegans GN=pir-1 PE=1 SV=2 |
| *CELE_ZK1320* | ZK1320.9 OS=Caenorhabditis elegans GN=CELE_ZK1320.9 PE=1 SV=1 |
| *alh-12* | ALH-12, isoform a OS=Caenorhabditis elegans GN=alh-12 PE=1 SV=1 |
| *CELE_Y57A10A* | Y57A10A.23 OS=Caenorhabditis elegans GN=CELE_Y57A10A.23 PE=1 SV=1 |
| *zig-1* | Secreted 2-immunoglobulin-domain protein ZIG-1 OS=Caenorhabditis elegans GN=zig-1 PE=1 SV=1 |
| *ran-4* | Probable nuclear transport factor 2 OS=Caenorhabditis elegans GN=ran-4 PE=3 SV=1 |
| *ppk-2* | PPK-2 OS=Caenorhabditis elegans GN=ppk-2 PE=1 SV=1 |
| *tre-1* | Trehalase OS=Caenorhabditis elegans GN=tre-1 PE=1 SV=1 |
| *nekl-3* | Serine/threonine-protein kinase nekl-3 OS=Caenorhabditis elegans GN=nekl-3 PE=1 SV=1 |
| *CELE_F32D8* | F32D8.5, isoform b OS=Caenorhabditis elegans GN=CELE_F32D8.5 PE=1 SV=1 |
| *F25B4* | Hydroxymethylglutaryl-CoA synthase OS=Caenorhabditis elegans GN=F25B4.6 PE=2 SV=2 |
| *aat-3* | AAT-3 OS=Caenorhabditis elegans GN=aat-3 PE=1 SV=2 |
| *parn-2* | PARN-2 OS=Caenorhabditis elegans GN=parn-2 PE=1 SV=1 |
| *ZK652* | Uncharacterized protein ZK652.6 OS=Caenorhabditis elegans GN=ZK652.6 PE=3 SV=2 |
| *CELE_T08B6* | T08B6.4 OS=Caenorhabditis elegans GN=CELE_T08B6.4 PE=1 SV=2 |
| *CELE_F58B3* | F58B3.4 OS=Caenorhabditis elegans GN=CELE_F58B3.4 PE=1 SV=3 |
| *CELE_F25B5* | 5'-nucleotidase OS=Caenorhabditis elegans GN=CELE_F25B5.3 PE=1 SV=1 |
| *cacn-1* | Cactin OS=Caenorhabditis elegans GN=cacn-1 PE=1 SV=1 |
| *F52B11* | Probable phosphomannomutase OS=Caenorhabditis elegans GN=F52B11.2 PE=3 SV=2 |
| *CELE_Y106G6H* | Y106G6H.14 OS=Caenorhabditis elegans GN=CELE_Y106G6H.14 PE=1 SV=1 |

| | |
|---|---|
| *dhrs-4* | Dehydrogenase/reductase SDR family member 4 OS=Caenorhabditis elegans GN=dhrs-4 PE=1 SV=1 |
| *CELE_T22B7* | T22B7.7 OS=Caenorhabditis elegans GN=CELE_T22B7.7 PE=1 SV=1 |
| *aakb-1* | AAKB-1 OS=Caenorhabditis elegans GN=aakb-1 PE=1 SV=1 |
| *CELE_Y41C4A* | Y41C4A.9 OS=Caenorhabditis elegans GN=CELE_Y41C4A.9 PE=1 SV=1 |
| *fbxa-64* | FBXA-64 OS=Caenorhabditis elegans GN=fbxa-64 PE=1 SV=2 |
| *C49A9* | C49A9.9, isoform a OS=Caenorhabditis elegans GN=C49A9.9 PE=1 SV=1 |
| *ugt-9* | UGT-9, isoform b OS=Caenorhabditis elegans GN=ugt-9 PE=1 SV=1 |
| *dct-11* | DCT-11 OS=Caenorhabditis elegans GN=dct-11 PE=1 SV=2 |
| *cdk-7* | Cyclin-dependent kinase 7 OS=Caenorhabditis elegans GN=cdk-7 PE=1 SV=1 |
| *CELE_F59D12* | F59D12.5 OS=Caenorhabditis elegans GN=CELE_F59D12.5 PE=1 SV=1 |
| *CELE_F18H3* | F18H3.1 OS=Caenorhabditis elegans GN=CELE_F18H3.1 PE=1 SV=2 |
| *tftc-3* | TFTC-3 OS=Caenorhabditis elegans GN=tftc-3 PE=1 SV=1 |
| *hda-3* | Histone deacetylase OS=Caenorhabditis elegans GN=hda-3 PE=1 SV=1 |
| *lbp-5* | Fatty acid-binding protein homolog 5 OS=Caenorhabditis elegans GN=lbp-5 PE=3 SV=1 |
| *ptr-11* | PTR-11, isoform a OS=Caenorhabditis elegans GN=ptr-11 PE=1 SV=1 |
| *ooc-5* | Torsin-like protein OS=Caenorhabditis elegans GN=ooc-5 PE=1 SV=1 |
| *C39B5* | C39B5.5 OS=Caenorhabditis elegans GN=C39B5.5 PE=1 SV=1 |
| *CELE_F22B8* | F22B8.7 OS=Caenorhabditis elegans GN=CELE_F22B8.7 PE=1 SV=2 |
| *snf-3* | Sodium- and chloride-dependent betaine transporter OS=Caenorhabditis elegans GN=snf-3 PE=1 SV=1 |
| *CELE_T20F5* | T20F5.6 OS=Caenorhabditis elegans GN=CELE_T20F5.6 PE=1 SV=2 |
| *npp-22* | Nucleoporin ndc-1 OS=Caenorhabditis elegans GN=npp-22 PE=3 SV=1 |
| *adr-2* | Probable double-stranded RNA-specific adenosine deaminase OS=Caenorhabditis elegans GN=adr-2 PE=2 SV=2 |
| *C45G9* | Uncharacterized protein C45G9.5 OS=Caenorhabditis elegans GN=C45G9.5 PE=4 SV=1 |
| *wapl-1* | WAPL-1, isoform b OS=Caenorhabditis elegans GN=wapl-1 PE=1 SV=1 |
| *E02H1* | Probable dimethyladenosine transferase OS=Caenorhabditis elegans GN=E02H1.1 PE=3 SV=2 |
| *skr-4* | SKR-4 OS=Caenorhabditis elegans GN=skr-4 PE=1 SV=1 |
| *CELE_Y71H2B* | Y71H2B.5 OS=Caenorhabditis elegans GN=CELE_Y71H2B.5 PE=1 SV=4 |
| *plk-1* | Serine/threonine-protein kinase plk-1 OS=Caenorhabditis elegans GN=plk-1 PE=1 SV=3 |
| *ama-1* | DNA-directed RNA polymerase II subunit RPB1 OS=Caenorhabditis elegans GN=ama-1 PE=1 SV=3 |
| *CELE_D1081* | D1081.7, isoform b OS=Caenorhabditis elegans GN=CELE_D1081.7 PE=1 SV=1 |
| *bub-1* | Mitotic checkpoint serine/threonine-protein kinase BUB1 OS=Caenorhabditis elegans GN=bub-1 PE=1 SV=1 |
| *Y66H1A* | Probable H/ACA ribonucleoprotein complex subunit 1-like protein OS=Caenorhabditis elegans GN=Y66H1A.4 PE=3 SV=2 |
| *CELE_Y82E9BR* | Y82E9BR.18 OS=Caenorhabditis elegans GN=CELE_Y82E9BR.18 PE=1 SV=2 |
| *gld-2* | Poly(A) RNA polymerase gld-2 OS=Caenorhabditis elegans GN=gld-2 PE=1 SV=2 |
| *rab-5* | Rab-5 (Fragment) OS=Caenorhabditis elegans GN=rab-5 PE=2 SV=1 |
| *ger-1* | GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase OS=Caenorhabditis elegans GN=ger-1 PE=1 SV=1 |

| | |
|---|---|
| *pfd-2* | Prefoldin subunit 2 OS=Caenorhabditis elegans GN=pdf-2 PE=3 SV=1 |
| *pus-1* | tRNA pseudouridine synthase OS=Caenorhabditis elegans GN=pus-1 PE=1 SV=1 |
| *skr-3* | SKR-3 OS=Caenorhabditis elegans GN=skr-3 PE=1 SV=1 |
| *tac-1* | 2P40 OS=Caenorhabditis elegans GN=tac-1 PE=1 SV=1 |
| *CELE_Y79H2A* | Y79H2A.3, isoform f OS=Caenorhabditis elegans GN=CELE_Y79H2A.3 PE=1 SV=1 |
| *swsn-2* | SWSN-2.2 OS=Caenorhabditis elegans GN=swsn-2.2 PE=1 SV=2 |
| *F32A5* | Uncharacterized serine carboxypeptidase F32A5.3 OS=Caenorhabditis elegans GN=F32A5.3 PE=1 SV=1 |
| *CELE_T05B9* | T05B9.1 OS=Caenorhabditis elegans GN=CELE_T05B9.1 PE=1 SV=1 |
| *lact-9* | LACT-9, isoform a OS=Caenorhabditis elegans GN=lact-9 PE=1 SV=1 |
| *tbc-2* | TBC-2 OS=Caenorhabditis elegans GN=tbc-2 PE=1 SV=2 |
| *acs-2* | ACS-2 OS=Caenorhabditis elegans GN=acs-2 PE=1 SV=1 |
| *ech-8* | ECH-8 OS=Caenorhabditis elegans GN=ech-8 PE=1 SV=2 |
| *gst-20* | GST-20 OS=Caenorhabditis elegans GN=gst-20 PE=1 SV=1 |
| *alh-6* | ALH-6, isoform a OS=Caenorhabditis elegans GN=alh-6 PE=1 SV=1 |
| *C14A4* | Putative glycosyltransferase C14A4.3 OS=Caenorhabditis elegans GN=C14A4.3 PE=1 SV=2 |
| *wago-4* | Piwi-like protein OS=Caenorhabditis elegans GN=wago-4 PE=1 SV=1 |
| *F40A3* | Phosphatidylethanolamine-binding protein homolog F40A3.3 OS=Caenorhabditis elegans GN=F40A3.3 PE=3 SV=1 |
| *kin-20* | Casein kinase I isoform delta OS=Caenorhabditis elegans GN=kin-20 PE=2 SV=3 |
| *unc-94* | Tropomodulin OS=Caenorhabditis elegans GN=unc-94 PE=1 SV=2 |
| *rpoa-49* | RPOA-49 OS=Caenorhabditis elegans GN=rpoa-49 PE=1 SV=1 |
| *CELE_K12C11* | K12C11.1, isoform a OS=Caenorhabditis elegans GN=CELE_K12C11.1 PE=1 SV=2 |
| *CELE_H20J04* | H20J04.4, isoform b OS=Caenorhabditis elegans GN=CELE_H20J04.4 PE=1 SV=1 |
| *F08C6* | Putative choline-phosphate cytidylyltransferase OS=Caenorhabditis elegans GN=F08C6.2 PE=3 SV=2 |
| *pkn-1* | PKN-1, isoform b OS=Caenorhabditis elegans GN=pkn-1 PE=1 SV=1 |
| *chat-1* | CHAT-1, isoform a OS=Caenorhabditis elegans GN=chat-1 PE=1 SV=1 |
| *CELE_T08G11* | T08G11.1, isoform a OS=Caenorhabditis elegans GN=CELE_T08G11.1 PE=1 SV=1 |
| *cyn-12* | Peptidyl-prolyl cis-trans isomerase OS=Caenorhabditis elegans GN=cyn-12 PE=1 SV=1 |
| *dhs-12* | DHS-12 OS=Caenorhabditis elegans GN=dhs-12 PE=1 SV=1 |
| *C34B2* | C34B2.8 OS=Caenorhabditis elegans GN=C34B2.8 PE=1 SV=1 |
| *unc-43* | UNC-43, isoform r OS=Caenorhabditis elegans GN=unc-43 PE=1 SV=1 |
| *rpac-40* | RPAC-40 OS=Caenorhabditis elegans GN=rpac-40 PE=1 SV=2 |
| *CELE_F53H1* | F53H1.1, isoform a OS=Caenorhabditis elegans GN=CELE_F53H1.1 PE=1 SV=1 |
| *rpl-28* | 60S ribosomal protein L28 OS=Caenorhabditis elegans GN=rpl-28 PE=1 SV=3 |
| *msi-1* | MSI-1 OS=Caenorhabditis elegans GN=msi-1 PE=1 SV=1 |
| *T19C3* | Uncharacterized protein T19C3.4 OS=Caenorhabditis elegans GN=T19C3.4 PE=3 SV=1 |
| *CELE_M01G5* | M01G5.3 OS=Caenorhabditis elegans GN=CELE_M01G5.3 PE=1 SV=2 |

| | |
|---|---|
| *C35C5* | Putative selT-like protein C35C5.3 OS=Caenorhabditis elegans GN=C35C5.3 PE=3 SV=2 |
| *cmd-1* | Calmodulin OS=Caenorhabditis elegans GN=cmd-1 PE=1 SV=3 |
| *ubr-1* | E3 ubiquitin-protein ligase ubr-1 OS=Caenorhabditis elegans GN=ubr-1 PE=1 SV=2 |
| *CELE_W07E6* | W07E6.2 OS=Caenorhabditis elegans GN=CELE_W07E6.2 PE=1 SV=1 |
| *cyp-37a1* | CYP-37A1 OS=Caenorhabditis elegans GN=cyp-37a1 PE=1 SV=2 |
| *C02F5* | Uncharacterized GTP-binding protein C02F5.3 OS=Caenorhabditis elegans GN=C02F5.3 PE=3 SV=2 |
| *CELE_F21C10* | F21C10.9 OS=Caenorhabditis elegans GN=CELE_F21C10.9 PE=1 SV=3 |
| *CELE_F17A9* | F17A9.4 OS=Caenorhabditis elegans GN=CELE_F17A9.4 PE=1 SV=1 |
| *ehbp-1* | EHBP-1 OS=Caenorhabditis elegans GN=ehbp-1 PE=1 SV=1 |
| *mec-7* | Tubulin beta-1 chain OS=Caenorhabditis elegans GN=mec-7 PE=2 SV=1 |
| *CELE_F42A10* | F42A10.5 OS=Caenorhabditis elegans GN=CELE_F42A10.5 PE=1 SV=1 |
| *gex-3* | Membrane-associated protein gex-3 OS=Caenorhabditis elegans GN=gex-3 PE=1 SV=2 |
| *C18E9* | C18E9.4 OS=Caenorhabditis elegans GN=C18E9.4 PE=1 SV=1 |
| *acs-14* | ACS-14 OS=Caenorhabditis elegans GN=acs-14 PE=1 SV=1 |
| *ncl-1* | B-box type zinc finger protein ncl-1 OS=Caenorhabditis elegans GN=ncl-1 PE=2 SV=1 |
| *C39D10* | C39D10.8, isoform a OS=Caenorhabditis elegans GN=C39D10.8 PE=1 SV=1 |
| *klc-1* | KLC-1, isoform a OS=Caenorhabditis elegans GN=klc-1 PE=1 SV=2 |
| *smc-5* | SMC-5 OS=Caenorhabditis elegans GN=smc-5 PE=1 SV=2 |
| *CELE_Y48G10A* | Y48G10A.3 OS=Caenorhabditis elegans GN=CELE_Y48G10A.3 PE=1 SV=1 |
| *Y48B6A* | Ribosome biogenesis protein BOP1 homolog OS=Caenorhabditis elegans GN=Y48B6A.1 PE=3 SV=1 |
| *C27F2* | C27F2.8 OS=Caenorhabditis elegans GN=C27F2.8 PE=1 SV=4 |
| *dis-3* | Probable exosome complex exonuclease RRP44 OS=Caenorhabditis elegans GN=dis-3 PE=3 SV=2 |
| *ND1* | NADH-ubiquinone oxidoreductase chain 1 OS=Caenorhabditis elegans GN=ND1 PE=3 SV=1 |
| *str-101* | STR-101 OS=Caenorhabditis elegans GN=str-101 PE=4 SV=3 |
| *glt-1* | Excitatory amino acid transporter OS=Caenorhabditis elegans GN=glt-1 PE=1 SV=2 |
| *CELE_T28C6* | T28C6.7, isoform b OS=Caenorhabditis elegans GN=CELE_T28C6.7 PE=1 SV=1 |
| *cyk-3* | CYK-3, isoform a OS=Caenorhabditis elegans GN=cyk-3 PE=1 SV=1 |
| *soap-1* | SOAP-1 OS=Caenorhabditis elegans GN=soap-1 PE=1 SV=1 |
| *C27H6* | UPF0160 protein C27H6.8 OS=Caenorhabditis elegans GN=C27H6.8 PE=3 SV=2 |
| *atg-2* | ATG-2 OS=Caenorhabditis elegans GN=atg-2 PE=1 SV=3 |
| *lbp-6* | Fatty acid-binding protein homolog 6 OS=Caenorhabditis elegans GN=lbp-6 PE=1 SV=1 |
| *cyp-29a2* | CYP-29A2 OS=Caenorhabditis elegans GN=cyp-29a2 PE=1 SV=1 |
| *him-10* | Kinetochore protein Nuf2 homolog OS=Caenorhabditis elegans GN=him-10 PE=1 SV=1 |
| *sec-16* | Protein transport protein Sec16 OS=Caenorhabditis elegans GN=sec-16 PE=3 SV=1 |
| *CELE_F18F11* | F18F11.1 OS=Caenorhabditis elegans GN=CELE_F18F11.1 PE=1 SV=1 |
| *ani-1* | Anillin-like protein 1 OS=Caenorhabditis elegans GN=ani-1 PE=2 SV=2 |

| | |
|---|---|
| *eif-3* | Eukaryotic translation initiation factor 3 subunit C OS=Caenorhabditis elegans GN=eif-3.C PE=3 SV=2 |
| *csn-6* | CSN-6, isoform b OS=Caenorhabditis elegans GN=csn-6 PE=1 SV=1 |
| *mrps-5* | Putative 28S ribosomal protein S5, mitochondrial OS=Caenorhabditis elegans GN=mrps-5 PE=3 SV=3 |
| *CELE_Y17G9B* | Y17G9B.5 OS=Caenorhabditis elegans GN=CELE_Y17G9B.5 PE=1 SV=1 |
| *COX1* | Cytochrome c oxidase subunit 1 OS=Caenorhabditis elegans GN=COX1 PE=3 SV=1 |
| *farl-11* | FARL-11 OS=Caenorhabditis elegans GN=farl-11 PE=1 SV=2 |
| *chd-7* | CHD-7 OS=Caenorhabditis elegans GN=chd-7 PE=1 SV=3 |
| *mel-28* | MEL-28 OS=Caenorhabditis elegans GN=mel-28 PE=1 SV=2 |
| *frl-1* | FRL-1 OS=Caenorhabditis elegans GN=frl-1 PE=1 SV=3 |
| *hpo-3* | HPO-3, isoform b OS=Caenorhabditis elegans GN=hpo-3 PE=1 SV=2 |
| *umps-1* | UMPS-1 OS=Caenorhabditis elegans GN=umps-1 PE=1 SV=1 |
| *CELE_F35G12* | F35G12.12 OS=Caenorhabditis elegans GN=CELE_F35G12.12 PE=1 SV=3 |
| *CELE_Y23H5B* | Y23H5B.5 OS=Caenorhabditis elegans GN=CELE_Y23H5B.5 PE=1 SV=2 |
| *ugt-13* | UGT-13 OS=Caenorhabditis elegans GN=ugt-13 PE=1 SV=2 |
| *djr-1* | Glutathione-independent glyoxalase DJR-1.1 OS=Caenorhabditis elegans GN=djr-1.1 PE=1 SV=1 |
| *lin-33* | LIN-33, isoform a OS=Caenorhabditis elegans GN=lin-33 PE=1 SV=1 |
| *asf-1* | Probable histone chaperone asf-1 OS=Caenorhabditis elegans GN=asf-1 PE=3 SV=2 |
| *bet-1* | BET-1, isoform a OS=Caenorhabditis elegans GN=bet-1 PE=1 SV=3 |
| *C52A10* | Carboxylic ester hydrolase OS=Caenorhabditis elegans GN=C52A10.1 PE=1 SV=1 |
| *kcc-2* | KCC-2, isoform a OS=Caenorhabditis elegans GN=kcc-2 PE=1 SV=1 |
| *pbs-2* | Proteasome subunit beta type OS=Caenorhabditis elegans GN=pbs-2 PE=1 SV=1 |
| *C08F8* | C08F8.3, isoform a OS=Caenorhabditis elegans GN=C08F8.3 PE=1 SV=2 |
| *CELE_F31D4* | F31D4.2 OS=Caenorhabditis elegans GN=CELE_F31D4.2 PE=1 SV=1 |
| *rpl-24* | 60S ribosomal protein L24 OS=Caenorhabditis elegans GN=rpl-24.1 PE=3 SV=1 |
| *CELE_F49C12* | F49C12.7, isoform a OS=Caenorhabditis elegans GN=CELE_F49C12.7 PE=1 SV=1 |
| *CELE_F13E9* | F13E9.1 OS=Caenorhabditis elegans GN=CELE_F13E9.1 PE=1 SV=1 |
| *CELE_F17C11* | F17C11.12, isoform b OS=Caenorhabditis elegans GN=CELE_F17C11.12 PE=1 SV=1 |
| *3E324* | 3E324 OS=Caenorhabditis elegans GN=3E324 PE=1 SV=1 |
| *C05D11* | Uncharacterized protein C05D11.1 OS=Caenorhabditis elegans GN=C05D11.1 PE=4 SV=2 |
| *nbet-1* | BET1 OS=Caenorhabditis elegans GN=nbet-1 PE=1 SV=1 |
| *ugt-25* | UDP-glucuronosyltransferase OS=Caenorhabditis elegans GN=ugt-25 PE=1 SV=1 |
| *egg-6* | Leucine-rich repeat-containing protein egg-6 OS=Caenorhabditis elegans GN=egg-6 PE=2 SV=2 |
| *B0495* | Putative endoplasmic reticulum metallopeptidase 1-A OS=Caenorhabditis elegans GN=B0495.7 PE=1 SV=2 |
| *gld-4* | Poly(A) RNA polymerase gld-4 OS=Caenorhabditis elegans GN=gld-4 PE=1 SV=1 |
| *hyl-1* | Ceramide synthase hyl-1 OS=Caenorhabditis elegans GN=hyl-1 PE=1 SV=1 |

| | |
|---|---|
| *C09G9* | C09G9.1, isoform b OS=Caenorhabditis elegans GN=C09G9.1 PE=1 SV=1 |
| *CELE_Y113G7C* | Protein-tyrosine-phosphatase OS=Caenorhabditis elegans GN=CELE_Y113G7C.1 PE=1 SV=1 |
| *asp-2* | ASP-2, isoform a OS=Caenorhabditis elegans GN=asp-2 PE=1 SV=1 |
| *wrk-1* | WRK-1, isoform a OS=Caenorhabditis elegans GN=wrk-1 PE=1 SV=1 |
| *CELE_Y76B12C* | Y76B12C.6, isoform b OS=Caenorhabditis elegans GN=CELE_Y76B12C.6 PE=1 SV=1 |
| *tut-1* | Cytoplasmic tRNA 2-thiolation protein 1 OS=Caenorhabditis elegans GN=tut-1 PE=1 SV=1 |
| *ceh-44* | Protein CASP OS=Caenorhabditis elegans GN=ceh-44 PE=3 SV=1 |
| *CELE_F48E8* | F48E8.4 OS=Caenorhabditis elegans GN=CELE_F48E8.4 PE=1 SV=4 |
| *cyp-35a3* | CYP-35A3 OS=Caenorhabditis elegans GN=cyp-35a3 PE=1 SV=1 |
| *rep-1* | REP-1, isoform a OS=Caenorhabditis elegans GN=rep-1 PE=1 SV=1 |
| *C01G10* | C01G10.8 OS=Caenorhabditis elegans GN=C01G10.8 PE=1 SV=1 |
| *mau-2* | Maternal uncoordinated protein 2 OS=Caenorhabditis elegans GN=mau-2 PE=1 SV=1 |
| *far-1* | Fatty-acid and retinol-binding protein 1 OS=Caenorhabditis elegans GN=far-1 PE=3 SV=1 |
| *sup-1* | Protein SUP-1 OS=Caenorhabditis elegans GN=sup-1 PE=1 SV=1 |
| *iws-1* | IWS1-like protein OS=Caenorhabditis elegans GN=iws-1 PE=3 SV=1 |
| *CELE_F13B9* | F13B9.1, isoform a OS=Caenorhabditis elegans GN=CELE_F13B9.1 PE=1 SV=1 |
| *gsp-4* | Serine/threonine-protein phosphatase OS=Caenorhabditis elegans GN=gsp-4 PE=3 SV=1 |
| *sna-2* | SNA-2 OS=Caenorhabditis elegans GN=sna-2 PE=1 SV=1 |
| *CELE_Y53G8AR* | Y53G8AR.9 OS=Caenorhabditis elegans GN=CELE_Y53G8AR.9 PE=1 SV=1 |
| *snb-2* | SNB-2 OS=Caenorhabditis elegans GN=snb-2 PE=1 SV=3 |
| *ctg-1* | CTG-1, isoform a OS=Caenorhabditis elegans GN=ctg-1 PE=1 SV=2 |
| *lpd-6* | LPD-6 OS=Caenorhabditis elegans GN=lpd-6 PE=1 SV=2 |
| *mrps-23* | Probable 28S ribosomal protein S23, mitochondrial OS=Caenorhabditis elegans GN=mrps-23 PE=3 SV=1 |
| *rnp-2* | RNP-2 OS=Caenorhabditis elegans GN=rnp-2 PE=1 SV=2 |
| *ndx-3* | Nudix hydrolase 3 OS=Caenorhabditis elegans GN=ndx-3 PE=3 SV=2 |
| *plrg-1* | PLRG-1, isoform a OS=Caenorhabditis elegans GN=plrg-1 PE=1 SV=1 |
| *cni-1* | Protein cornichon homolog 1 OS=Caenorhabditis elegans GN=cni-1 PE=1 SV=2 |
| *mrps-25* | Probable 28S ribosomal protein S25, mitochondrial OS=Caenorhabditis elegans GN=mrps-25 PE=3 SV=1 |
| *alkb-8* | ALKB-8 OS=Caenorhabditis elegans GN=alkb-8 PE=1 SV=2 |
| *bcat-1* | Branched-chain-amino-acid aminotransferase, cytosolic OS=Caenorhabditis elegans GN=bcat-1 PE=2 SV=2 |
| *CELE_W05H9* | W05H9.1, isoform a OS=Caenorhabditis elegans GN=CELE_W05H9.1 PE=1 SV=1 |
| *cdc-73* | Cell division cycle protein 73 OS=Caenorhabditis elegans GN=cdc-73 PE=3 SV=4 |
| *sto-1* | Stomatin-1 OS=Caenorhabditis elegans GN=sto-1 PE=3 SV=2 |
| *F13C5* | Uncharacterized protein F13C5.5 OS=Caenorhabditis elegans GN=F13C5.5 PE=1 SV=1 |
| *tba-4* | TBA-4 OS=Caenorhabditis elegans GN=tba-4 PE=1 SV=1 |

| | |
|---|---|
| *CELE_F44E7* | F44E7.4, isoform c OS=Caenorhabditis elegans GN=CELE_F44E7.4 PE=1 SV=1 |
| *CELE_F45F2* | F45F2.9, isoform b OS=Caenorhabditis elegans GN=CELE_F45F2.9 PE=1 SV=1 |
| *CELE_F35H10* | F35H10.6 OS=Caenorhabditis elegans GN=CELE_F35H10.6 PE=1 SV=2 |
| *F59E10* | Probable coatomer subunit zeta OS=Caenorhabditis elegans GN=F59E10.3 PE=3 SV=1 |
| *CELE_F49E8* | F49E8.7, isoform a OS=Caenorhabditis elegans GN=CELE_F49E8.7 PE=1 SV=1 |
| *cku-80* | ATP-dependent DNA helicase II subunit 2 OS=Caenorhabditis elegans GN=cku-80 PE=1 SV=1 |
| *CELE_W06B4* | W06B4.1 OS=Caenorhabditis elegans GN=CELE_W06B4.1 PE=1 SV=1 |
| *chs-1* | CHS-1 OS=Caenorhabditis elegans GN=chs-1 PE=2 SV=1 |
| *mdt-9* | MDT-9, isoform b OS=Caenorhabditis elegans GN=mdt-9 PE=1 SV=1 |
| *hsp-43* | HSP-43, isoform b OS=Caenorhabditis elegans GN=hsp-43 PE=1 SV=2 |
| *mrps-26* | MRPS-26 OS=Caenorhabditis elegans GN=mrps-26 PE=1 SV=1 |
| *dhs-20* | DHS-20 OS=Caenorhabditis elegans GN=dhs-20 PE=1 SV=3 |
| *npp-8* | NPP-8, isoform b OS=Caenorhabditis elegans GN=npp-8 PE=1 SV=4 |
| *CELE_F32A5* | F32A5.8 OS=Caenorhabditis elegans GN=CELE_F32A5.8 PE=1 SV=2 |
| *CELE_T05E7* | T05E7.3 OS=Caenorhabditis elegans GN=CELE_T05E7.3 PE=1 SV=1 |
| *aph-2* | APH-2, isoform a OS=Caenorhabditis elegans GN=aph-2 PE=1 SV=1 |
| *capg-1* | CAPG-1 OS=Caenorhabditis elegans GN=capg-1 PE=1 SV=1 |
| *CELE_F56D3* | F56D3.1 OS=Caenorhabditis elegans GN=CELE_F56D3.1 PE=1 SV=2 |
| *wrm-1* | Armadillo repeat-containing protein wrm-1 OS=Caenorhabditis elegans GN=wrm-1 PE=1 SV=2 |
| *CELE_F15G9* | F15G9.1, isoform c OS=Caenorhabditis elegans GN=CELE_F15G9.1 PE=1 SV=1 |
| *mos-2* | Protein mos-2 homolog OS=Caenorhabditis elegans GN=mos-2 PE=1 SV=1 |
| *mec-17* | Alpha-tubulin N-acetyltransferase 1 OS=Caenorhabditis elegans GN=mec-17 PE=2 SV=1 |
| *C27A7* | C27A7.6, isoform c OS=Caenorhabditis elegans GN=C27A7.6 PE=4 SV=1 |
| *F12F6* | Probable DNA polymerase delta small subunit OS=Caenorhabditis elegans GN=F12F6.7 PE=3 SV=1 |
| *CELE_T10E9* | T10E9.4 OS=Caenorhabditis elegans GN=CELE_T10E9.4 PE=1 SV=2 |
| *cpg-2* | Chondroitin proteoglycan-2 OS=Caenorhabditis elegans GN=cpg-2 PE=1 SV=3 |
| *CELE_F56G4* | F56G4.6 OS=Caenorhabditis elegans GN=CELE_F56G4.6 PE=1 SV=3 |
| *C24D10* | C24D10.6 OS=Caenorhabditis elegans GN=C24D10.6 PE=1 SV=1 |
| *exos-1* | EXOS-1 OS=Caenorhabditis elegans GN=exos-1 PE=1 SV=1 |
| *Sep-01* | SEP-1 OS=Caenorhabditis elegans GN=sep-1 PE=1 SV=1 |
| *trcs-2* | TRCS-2 OS=Caenorhabditis elegans GN=trcs-2 PE=1 SV=3 |
| *trpp-1* | TRPP-1 OS=Caenorhabditis elegans GN=trpp-1 PE=1 SV=2 |
| *CELE_F58F9* | F58F9.3, isoform b OS=Caenorhabditis elegans GN=CELE_F58F9.3 PE=1 SV=2 |
| *ced-4* | Cell death protein 4 OS=Caenorhabditis elegans GN=ced-4 PE=1 SV=2 |
| *C35D10* | C35D10.5 OS=Caenorhabditis elegans GN=C35D10.5 PE=1 SV=2 |
| *CELE_Y49F6B* | Y49F6B.2 OS=Caenorhabditis elegans GN=CELE_Y49F6B.2 PE=1 SV=1 |
| *cpsf-1* | Probable cleavage and polyadenylation specificity factor subunit 1 OS=Caenorhabditis elegans GN=cpsf-1 PE=3 SV=2 |

| | |
|---|---|
| *CELE_Y110A2AR* | Y110A2AR.3, isoform a OS=Caenorhabditis elegans GN=CELE_Y110A2AR.3 PE=1 SV=1 |
| *CELE_F55A11* | F55A11.6, isoform c OS=Caenorhabditis elegans GN=CELE_F55A11.6 PE=1 SV=1 |
| *wrb-1* | WRB-1, isoform a OS=Caenorhabditis elegans GN=wrb-1 PE=1 SV=2 |
| *C48B4* | Probable peroxisomal acyl-coenzyme A oxidase OS=Caenorhabditis elegans GN=C48B4.1 PE=3 SV=1 |
| *F10E9* | Uncharacterized protein F10E9.4 OS=Caenorhabditis elegans GN=F10E9.4 PE=4 SV=1 |
| *T23D8* | Protein LTV1 homolog OS=Caenorhabditis elegans GN=T23D8.3 PE=3 SV=1 |
| *shc-1* | SHC-transforming protein homolog 1 OS=Caenorhabditis elegans GN=shc-1 PE=1 SV=2 |
| *CELE_Y71H2AM* | Y71H2AM.4 OS=Caenorhabditis elegans GN=CELE_Y71H2AM.4 PE=1 SV=1 |
| *B0495* | Putative endoplasmic reticulum metallopeptidase 1-A OS=Caenorhabditis elegans GN=B0495.7 PE=1 SV=2 |
| *gstk-2* | Glutathione s-transferase kappa 2 OS=Caenorhabditis elegans GN=gstk-2 PE=2 SV=1 |
| *Y48A5A* | Protein SHQ1 homolog OS=Caenorhabditis elegans GN=Y48A5A.1 PE=3 SV=4 |
| *CELE_D2085* | D2085.3 OS=Caenorhabditis elegans GN=CELE_D2085.3 PE=1 SV=1 |
| *dnj-21* | Mitochondrial import inner membrane translocase subunit TIM14 OS=Caenorhabditis elegans GN=dnj-21 PE=3 SV=1 |
| *CELE_R144* | R144.13 OS=Caenorhabditis elegans GN=CELE_R144.13 PE=1 SV=1 |
| *CELE_M60* | M60.4, isoform b OS=Caenorhabditis elegans GN=CELE_M60.4 PE=1 SV=1 |
| *CELE_Y43F8A* | Y43F8A.2 OS=Caenorhabditis elegans GN=CELE_Y43F8A.2 PE=1 SV=1 |
| *otub-3* | OTUB-3, isoform b OS=Caenorhabditis elegans GN=otub-3 PE=1 SV=1 |
| *CELE_F43G9* | F43G9.4 OS=Caenorhabditis elegans GN=CELE_F43G9.4 PE=1 SV=1 |
| *CELE_Y57G11B* | Y57G11B.5 OS=Caenorhabditis elegans GN=CELE_Y57G11B.5 PE=1 SV=1 |
| *zhit-3* | ZHIT-3 OS=Caenorhabditis elegans GN=zhit-3 PE=1 SV=2 |
| *lron-9* | LRON-9, isoform a OS=Caenorhabditis elegans GN=lron-9 PE=1 SV=1 |
| *hoe-1* | Ribonuclease Z OS=Caenorhabditis elegans GN=hoe-1 PE=2 SV=2 |
| *CELE_T26E3* | T26E3.4 OS=Caenorhabditis elegans GN=CELE_T26E3.4 PE=1 SV=3 |
| *ify-1* | IFY-1 OS=Caenorhabditis elegans GN=ify-1 PE=1 SV=1 |
| *cnb-1* | CNB-1 OS=Caenorhabditis elegans GN=cnb-1 PE=1 SV=1 |
| *CELE_F54E12* | F54E12.2 OS=Caenorhabditis elegans GN=CELE_F54E12.2 PE=1 SV=1 |
| *paf-2* | Platelet-activating factor acetylhydrolase homolog 2 OS=Caenorhabditis elegans GN=paf-2 PE=2 SV=2 |
| *ugt-5* | UDP-glucuronosyltransferase OS=Caenorhabditis elegans GN=ugt-5 PE=1 SV=2 |
| *CELE_R07E5* | R07E5.7 OS=Caenorhabditis elegans GN=CELE_R07E5.7 PE=1 SV=1 |
| *CELE_F28C1* | F28C1.3, isoform c OS=Caenorhabditis elegans GN=CELE_F28C1.3 PE=1 SV=1 |
| *CELE_F17E9* | F17E9.4 OS=Caenorhabditis elegans GN=CELE_F17E9.4 PE=1 SV=2 |
| *mrpl-11* | Probable 39S ribosomal protein L11, mitochondrial OS=Caenorhabditis elegans GN=mrpl-11 PE=3 SV=1 |
| *ent-2* | ENT-2 OS=Caenorhabditis elegans GN=ent-2 PE=1 SV=1 |
| *gfi-2* | GFI-2, isoform a OS=Caenorhabditis elegans GN=gfi-2 PE=1 SV=1 |
| *nsun-5* | Y53F4B.4, isoform b OS=Caenorhabditis elegans GN=nsun-5 PE=1 SV=2 |

| | |
|---|---|
| *ZK20* | Cytochrome c oxidase assembly factor 7 homolog OS=Caenorhabditis elegans GN=ZK20.4 PE=3 SV=1 |
| *CELE_T09B4* | T09B4.8 OS=Caenorhabditis elegans GN=CELE_T09B4.8 PE=1 SV=3 |
| *C34B2* | C34B2.8 OS=Caenorhabditis elegans GN=C34B2.8 PE=1 SV=1 |
| *CELE_F53E10* | F53E10.1 OS=Caenorhabditis elegans GN=CELE_F53E10.1 PE=1 SV=1 |
| *pssy-1* | PSSY-1 OS=Caenorhabditis elegans GN=pssy-1 PE=1 SV=1 |
| *ubc-7* | Probable ubiquitin-conjugating enzyme E2 7 OS=Caenorhabditis elegans GN=ubc-7 PE=1 SV=1 |
| *CELE_Y47G6A* | Y47G6A.18 OS=Caenorhabditis elegans GN=CELE_Y47G6A.18 PE=1 SV=2 |
| *CELE_F33C8* | F33C8.4 OS=Caenorhabditis elegans GN=CELE_F33C8.4 PE=1 SV=3 |
| *kbp-5* | KBP-5 OS=Caenorhabditis elegans GN=kbp-5 PE=1 SV=1 |
| *CELE_F32A7* | F32A7.5, isoform d OS=Caenorhabditis elegans GN=CELE_F32A7.5 PE=1 SV=1 |
| *CELE_R186* | R186.3 OS=Caenorhabditis elegans GN=CELE_R186.3 PE=1 SV=1 |
| *acs-6* | ACS-6 OS=Caenorhabditis elegans GN=acs-6 PE=1 SV=1 |
| *CELE_T07A9* | T07A9.10 OS=Caenorhabditis elegans GN=CELE_T07A9.10 PE=1 SV=1 |
| *cdc-14* | Probable tyrosine-protein phosphatase cdc-14 OS=Caenorhabditis elegans GN=cdc-14 PE=1 SV=2 |
| *Oct-02* | OCT-2, isoform a OS=Caenorhabditis elegans GN=oct-2 PE=1 SV=4 |
| *CELE_Y38C1AA* | Y38C1AA.7 OS=Caenorhabditis elegans GN=CELE_Y38C1AA.7 PE=1 SV=1 |
| *arf-6* | ARF-6 OS=Caenorhabditis elegans GN=arf-6 PE=1 SV=1 |
| *erv-46* | ERV-46 OS=Caenorhabditis elegans GN=erv-46 PE=1 SV=1 |
| *CELE_Y57G11A* | Y57G11A.2 OS=Caenorhabditis elegans GN=CELE_Y57G11A.2 PE=1 SV=1 |
| *C34C6* | C34C6.4 OS=Caenorhabditis elegans GN=C34C6.4 PE=1 SV=1 |
| *CELE_Y37D8A* | Y37D8A.19 OS=Caenorhabditis elegans GN=CELE_Y37D8A.19 PE=1 SV=1 |
| *CELE_F43C9* | F43C9.2 OS=Caenorhabditis elegans GN=CELE_F43C9.2 PE=1 SV=2 |
| *CELE_D1086* | D1086.8, isoform a OS=Caenorhabditis elegans GN=CELE_D1086.8 PE=1 SV=1 |
| *pcp-4* | PCP-4 OS=Caenorhabditis elegans GN=pcp-4 PE=1 SV=1 |
| *pes-4* | PES-4 OS=Caenorhabditis elegans GN=pes-4 PE=1 SV=3 |
| *fmo-5* | Dimethylaniline monooxygenase [N-oxide-forming] OS=Caenorhabditis elegans GN=fmo-5 PE=1 SV=1 |
| *Y56A3A* | Zinc finger protein 593 homolog OS=Caenorhabditis elegans GN=Y56A3A.18 PE=3 SV=1 |
| *nepr-1* | NEPR-1, isoform b OS=Caenorhabditis elegans GN=nepr-1 PE=1 SV=1 |
| *mrps-24* | 28S ribosomal protein S24, mitochondrial OS=Caenorhabditis elegans GN=mrps-24 PE=3 SV=1 |
| *tnc-2* | Troponin C, isoform 2 OS=Caenorhabditis elegans GN=tnc-2 PE=2 SV=1 |
| *CELE_K08E7* | K08E7.8, isoform a OS=Caenorhabditis elegans GN=CELE_K08E7.8 PE=4 SV=1 |
| *F09G8* | Uncharacterized protein F09G8.7 OS=Caenorhabditis elegans GN=F09G8.7 PE=4 SV=1 |
| *F44B9* | Probable replication factor C subunit 5 OS=Caenorhabditis elegans GN=F44B9.8 PE=3 SV=3 |
| *CELE_Y66A7A* | Ribonuclease P/MRP protein subunit POP5 OS=Caenorhabditis elegans GN=CELE_Y66A7A.2 PE=1 SV=1 |
| *tbh-1* | Tyramine beta-hydroxylase OS=Caenorhabditis elegans GN=tbh-1 PE=1 SV=3 |
| *teg-1* | TEG-1 OS=Caenorhabditis elegans GN=teg-1 PE=1 SV=2 |

| | |
|---|---|
| *zig-9* | ZIG-9 OS=Caenorhabditis elegans GN=zig-9 PE=1 SV=3 |
| *ugt-64* | UGT-64 OS=Caenorhabditis elegans GN=ugt-64 PE=1 SV=1 |
| *hrpf-2* | HRPF-2, isoform a OS=Caenorhabditis elegans GN=hrpf-2 PE=1 SV=1 |
| *CELE_F37B12* | F37B12.3 OS=Caenorhabditis elegans GN=CELE_F37B12.3 PE=1 SV=2 |
| *CELE_W08E3* | W08E3.2 OS=Caenorhabditis elegans GN=CELE_W08E3.2 PE=1 SV=1 |
| *CELE_F58E6* | F58E6.13, isoform d OS=Caenorhabditis elegans GN=CELE_F58E6.13 PE=1 SV=1 |
| *aly-1* | ALY-1 OS=Caenorhabditis elegans GN=aly-1 PE=1 SV=1 |
| *VPS34* | Phosphatidylinositol 3-kinase OS=Caenorhabditis elegans GN=VPS34 PE=2 SV=1 |
| *CELE_K07A1* | K07A1.9, isoform a OS=Caenorhabditis elegans GN=CELE_K07A1.9 PE=1 SV=2 |
| *dsh-1* | DSH-1, isoform c OS=Caenorhabditis elegans GN=dsh-1 PE=1 SV=1 |
| *taf-1* | TAF-1 OS=Caenorhabditis elegans GN=taf-1 PE=1 SV=1 |
| *cccp-1* | CCCP-1 OS=Caenorhabditis elegans GN=cccp-1 PE=2 SV=1 |
| *M01F1* | Lipoyl synthase, mitochondrial OS=Caenorhabditis elegans GN=M01F1.3 PE=3 SV=1 |
| *CELE_Y56A3A* | Acyl carrier protein OS=Caenorhabditis elegans GN=CELE_Y56A3A.19 PE=1 SV=1 |
| *C06A8* | C06A8.3 OS=Caenorhabditis elegans GN=C06A8.3 PE=1 SV=1 |
| *ttr-26* | TTR-26 OS=Caenorhabditis elegans GN=ttr-26 PE=1 SV=1 |
| *ttr-32* | TTR-32 OS=Caenorhabditis elegans GN=ttr-32 PE=1 SV=1 |
| *C49F8* | C49F8.3 OS=Caenorhabditis elegans GN=C49F8.3 PE=1 SV=1 |
| *CELE_F10C1* | F10C1.9, isoform b OS=Caenorhabditis elegans GN=CELE_F10C1.9 PE=1 SV=1 |
| *syp-3* | SYP-3 OS=Caenorhabditis elegans GN=syp-3 PE=1 SV=2 |
| *gst-35* | GST-35 OS=Caenorhabditis elegans GN=gst-35 PE=1 SV=2 |
| *CELE_R144* | R144.13 OS=Caenorhabditis elegans GN=CELE_R144.13 PE=1 SV=1 |
| *CELE_F29C6* | F29C6.1, isoform a OS=Caenorhabditis elegans GN=CELE_F29C6.1 PE=1 SV=1 |
| *CELE_Y49E10* | Y49E10.21, isoform a OS=Caenorhabditis elegans GN=CELE_Y49E10.21 PE=1 SV=1 |
| *san-1* | SAN-1 OS=Caenorhabditis elegans GN=san-1 PE=1 SV=2 |
| *F21D5* | Uncharacterized protein F21D5.5 OS=Caenorhabditis elegans GN=F21D5.5 PE=2 SV=2 |
| *C17C3* | C17C3.1, isoform e OS=Caenorhabditis elegans GN=C17C3.1 PE=1 SV=1 |
| *CELE_T01G1* | T01G1.2 OS=Caenorhabditis elegans GN=CELE_T01G1.2 PE=1 SV=1 |
| *CELE_ZK899* | ZK899.2 OS=Caenorhabditis elegans GN=CELE_ZK899.2 PE=1 SV=2 |
| *mvb-12* | MVB-12 OS=Caenorhabditis elegans GN=mvb-12 PE=1 SV=1 |
| *syx-16* | SYX-16 OS=Caenorhabditis elegans GN=syx-16 PE=1 SV=2 |
| *klo-1* | KLO-1 OS=Caenorhabditis elegans GN=klo-1 PE=1 SV=1 |
| *pfkb-1* | PFKB-1.1, isoform a OS=Caenorhabditis elegans GN=pfkb-1.1 PE=1 SV=2 |
| *sek-3* | SEK-3, isoform a OS=Caenorhabditis elegans GN=sek-3 PE=1 SV=3 |
| *CELE_Y37H9A* | Y37H9A.1, isoform b OS=Caenorhabditis elegans GN=CELE_Y37H9A.1 PE=4 SV=1 |
| *acs-11* | ACS-11 OS=Caenorhabditis elegans GN=acs-11 PE=1 SV=3 |
| *CELE_Y54E10BL* | Y54E10BL.3 OS=Caenorhabditis elegans GN=CELE_Y54E10BL.3 PE=4 SV=2 |

| | |
|---|---|
| *CELE_Y62H9A* | Y62H9A.6 OS=Caenorhabditis elegans GN=CELE_Y62H9A.6 PE=1 SV=1 |
| *mig-23* | Nucleoside-diphosphatase mig-23 OS=Caenorhabditis elegans GN=mig-23 PE=1 SV=2 |
| *B0513* | Proline dehydrogenase 1, mitochondrial OS=Caenorhabditis elegans GN=B0513.5 PE=3 SV=2 |
| *C46F11* | C46F11.4 OS=Caenorhabditis elegans GN=C46F11.4 PE=1 SV=2 |
| *CELE_Y37E11AL* | Y37E11AL.3, isoform a OS=Caenorhabditis elegans GN=CELE_Y37E11AL.3 PE=4 SV=1 |
| *CELE_F45G2* | F45G2.7 OS=Caenorhabditis elegans GN=CELE_F45G2.7 PE=4 SV=1 |
| *CELE_Y43C5B* | Y43C5B.3 OS=Caenorhabditis elegans GN=CELE_Y43C5B.3 PE=1 SV=1 |
| *tads-1* | TADS-1 OS=Caenorhabditis elegans GN=tads-1 PE=1 SV=1 |
| *adpr-1* | ADPR-1, isoform a OS=Caenorhabditis elegans GN=adpr-1 PE=1 SV=1 |
| *his-3* | Histone H2A OS=Caenorhabditis elegans GN=his-3 PE=1 SV=2 |
| *rga-1* | RGA-1, isoform a OS=Caenorhabditis elegans GN=rga-1 PE=1 SV=1 |
| *CELE_F54A3* | F54A3.2 OS=Caenorhabditis elegans GN=CELE_F54A3.2 PE=1 SV=5 |
| *algn-14* | HPO-16, isoform b OS=Caenorhabditis elegans GN=algn-14 PE=1 SV=1 |
| *B0001* | Probable uridine-cytidine kinase OS=Caenorhabditis elegans GN=B0001.4 PE=3 SV=2 |
| *CELE_F44G4* | F44G4.3 OS=Caenorhabditis elegans GN=CELE_F44G4.3 PE=4 SV=1 |
| *mak-1* | MAK-1, isoform a OS=Caenorhabditis elegans GN=mak-1 PE=1 SV=1 |
| *BE0003N10* | BE0003N10.1 OS=Caenorhabditis elegans GN=BE0003N10.1 PE=1 SV=3 |
| *gst-40* | GST-40 OS=Caenorhabditis elegans GN=gst-40 PE=1 SV=2 |
| *C41D11* | Uncharacterized protein C41D11.5 OS=Caenorhabditis elegans GN=C41D11.5 PE=4 SV=2 |
| *prp-4* | PRP-4 OS=Caenorhabditis elegans GN=prp-4 PE=1 SV=1 |
| *CELE_R01B10* | R01B10.6 OS=Caenorhabditis elegans GN=CELE_R01B10.6 PE=1 SV=1 |
| *dmd-8* | DMD-8 OS=Caenorhabditis elegans GN=dmd-8 PE=4 SV=2 |
| *CELE_K07C5* | K07C5.2 OS=Caenorhabditis elegans GN=CELE_K07C5.2 PE=1 SV=1 |
| *CELE_E01B7* | E01B7.1 OS=Caenorhabditis elegans GN=CELE_E01B7.1 PE=1 SV=1 |

**Supplemental Table 5** Summary of all proteins precipitated from N2 animals following biotinylated isoxazole treatment that show a greater than 50% change in abundance compared to DMSO treated control animals (OS = Organism Name; GN = Gene Name; PE = Evidence of Protein Existence; SV = Sequence Version)