# Different low-complexity regions of SFPQ play distinct roles in the formation of biomolecular condensates

Andrew C. Marshall[1], Jerry Cummins[1], Simon Kobelke[2], Tianyi Zhu[3], Jocelyn Widagdo[3], Victor Anggono[3], Anthony Hyman[4], Archa H. Fox[1,2]*, Charles S. Bond[1]* and Mihwa Lee[5]*.

[1] School of Molecular Sciences, The University of Western Australia, Crawley, WA, 6009, Australia.

[2] School of Human Sciences, The University of Western Australia, Crawley, WA, 6009, Australia.

[3] Clem Jones Centre for Ageing Dementia Research, Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, 4072, Australia.

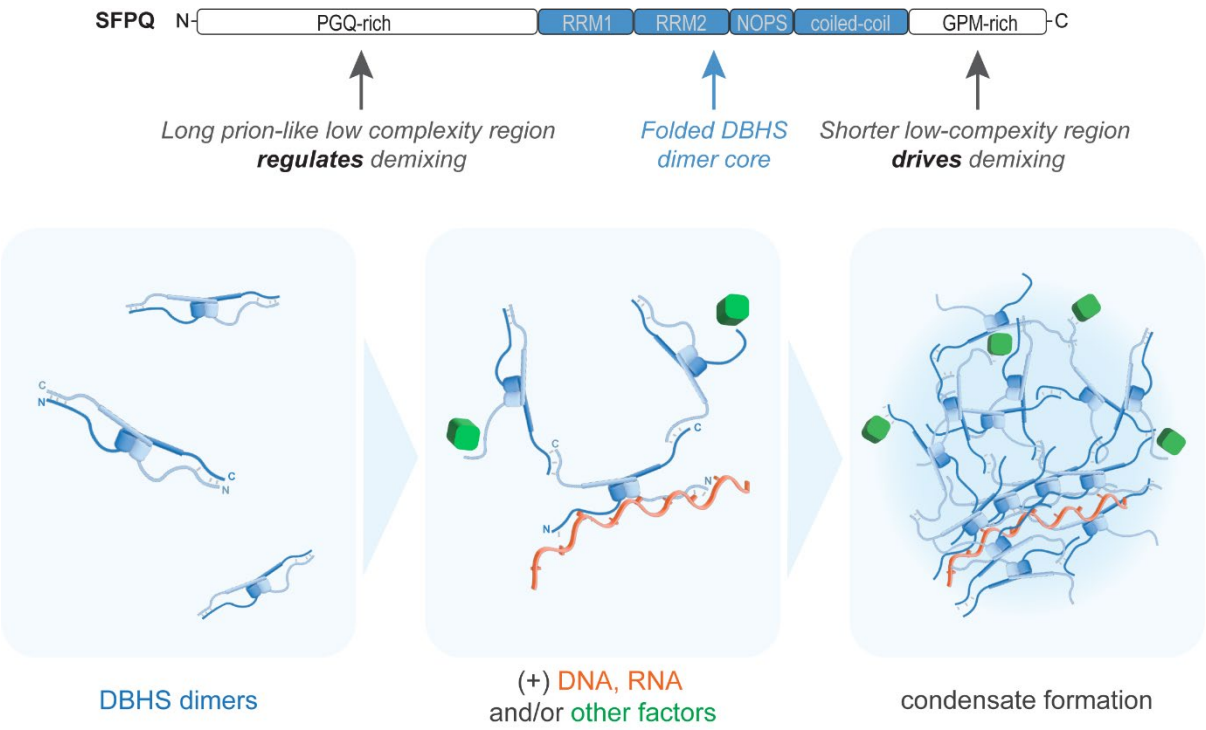
[4] Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.

[5] Department of Biochemistry and Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia.

*To whom correspondence should be addressed. Email: mihwa.lee@latrobe.edu.au, charles.bond@uwa.edu.au, archa.fox@uwa.edu.au.

## ABSTRACT

Demixing of proteins and nucleic acids into condensed liquid phases is rapidly emerging as a ubiquitous mechanism underlying the complex spatiotemporal organisation of molecules within the cell. Long disordered regions of low sequence complexity (LCRs) are a common feature of proteins that form liquid-like microscopic biomolecular condensates. In particular, RNA-binding proteins with prion-like composition have been highlighted as key drivers of liquid demixing to form condensates such as the nucleolus, paraspeckles and stress granules. Splicing factor proline- and glutamine-rich (SFPQ) is an RNA- and DNA-binding protein essential for DNA repair and paraspeckle formation. SFPQ contains two LCRs of different length and composition. Here, we show that the shorter C-terminal LCR of SFPQ is the main region responsible for the condensation of SFPQ *in vitro* and in the cell nucleus. In contrast, we find that, unexpectedly, the longer N-terminal prion-like LCR of SFPQ actually attenuates condensation of the full-length protein, suggesting a more regulatory role in preventing aberrant condensate formation in the cell. Our data add nuance to the emerging understanding of biomolecular condensate formation, by providing the first example of a common multifunctional nucleic acid-binding protein with an extensive prion-like region that serves to regulate rather than drive condensate formation.

## INTRODUCTION

Protein low-complexity regions (LCRs) are most simply defined as sequences enriched in only a few amino acids, and are commonly intrinsically disordered. Enormous interest in disordered LCRs has arisen in recent years due to the observation that they are commonly associated with the condensation of proteins and nucleic acids to form distinct membraneless cellular bodies with defined molecular compositions and physical properties. Biomolecular condensates such as the nucleolus, Cajal bodies, transcriptional hubs, nuclear speckles and paraspeckles are central to the spatial and temporal regulation of biochemical reactions in the mammalian cell nucleus (1-3). Mounting evidence suggests that the formation of these condensed bodies can be readily described using the soft matter physics concepts of phase separation (4), microphase separation (5) and surface wetting (6). The amino acid sequence and composition of LCRs have been shown to be critical for determining the condensation propensity of a protein and maintaining the reversible liquid-like state of the condensate (7-11). Although the establishment of rules relating sequence to material state is still in its infancy, various sequence features of LCRs have been shown to facilitate intermolecular interactions important for driving liquid demixing/condensation. These features are often overlapping and include prion-like domains (PLDs) (8,12), low-complexity aromatic-rich kinked segments (LARKS) (13), charge-patterning (14) and regions rich in basic and aromatic residues (11). These sequence features are over-represented in nucleic acid-binding proteins involved in subnuclear organisation and regulation of gene expression (15,16). In addition, both RNA and DNA have been suggested to provide multivalent interactions to drive condensation (3,6,17,18), providing a conceptual framework for understanding the formation of condensates containing proteins and nucleic acids.

Splicing factor proline and glutamine-rich (SFPQ; previously known as PSF) is an abundant nuclear DNA- and RNA-binding protein with essential roles in transcriptional elongation, mRNA processing and modification, and DNA repair (19,20). SFPQ is also an essential component of subnuclear bodies called paraspeckles, which are biomolecular condensates shown to be important for cellular stress response in the contexts of viral infection, cancer, early development and neurodegeneration (2,17). Paraspeckle abundance scales with the level of NEAT1_2 long noncoding RNA (21). SFPQ plays an essential role in determining NEAT1_2 levels by binding and protecting the RNA (22,23). Paraspeckles can be either singular spheroids of 360 nm diameter, or chains of spheroids with a constant width of 360 nm, and length up to 1.5 µm (23,24). In humans, SFPQ has two paralogues: NONO and PSPC1. They are all members of the '*Drosophila*-Behaviour, Human Splicing' (DBHS) protein family, and exist as obligate homo- or heterodimers in the nucleus (25). DBHS proteins are composed of a central globular region flanked on either side by long LCRs (**Figure 1a**). The central region, which is highly conserved between DBHS family members, contains tandem RNA-recognition motif (RRM) domains followed by a unique NonA/paraspeckle (NOPS) domain and an extended coiled-coil region (26). Multiple in-depth structural studies have shown that these domains pack together to form a compact, well-ordered DBHS core, which is essential for RNA-binding and protein dimerisation (**Figure 1c**) (27-30). In addition, the extended coiled-coil drives the polymerisation of SFPQ dimers, which is critical for the functions of SFPQ in transcriptional regulation and subnuclear body formation (31). In contrast, the flanking LCRs of DBHS proteins are highly divergent and remain largely unstudied. Phylogenetic analysis suggests that, of the three DBHS paralogues, SFPQ is the most divergent from their common ancestral DBHS protein, and that this is largely due to low selection pressure within its conspicuously longer N-terminal LCR (26).

Using *IUPred2A* (32) we find that both LCRs of SFPQ, which flank the globular conserved DBHS region, are predicted to be intrinsically disordered (**Figure 1b**). However, each is very different in length and displays a different amino acid enrichment profile. The shorter C-terminal LCR (C-LCR) of SFPQ is similar in length to the corresponding regions of NONO and PSPC1 (~100 amino acids), and is enriched in glycine, proline and methionine residues. The N-terminal LCR (N-LCR) is approximately 275 amino acids long, is enriched in glycine, proline and glutamine residues, and is characterised by multiple poly-Q and poly-P tracts. The length of the N-LCR of SFPQ is the most

obvious feature distinguishing SFPQ from NONO and PSPC1, which contain N-LCRs of only 50-60 amino acids (**Figure 1a**). Using the *PLAAC* algorithm (33) we find that a large proportion of the SFPQ N-LCR is predicted to have a "prion-like" amino acid composition (**Figure 1b**), suggesting a role in facilitating protein-protein interactions that may be important for driving condensation of the protein (8,15,34).

Here, we have used *in vitro* assays, fluorescence cell imaging and DNA damage recruitment assays to investigate the roles of the N- and C-LCRs of SFPQ in driving protein condensation. Unexpectedly, we find that the shorter C-LCR of SFPQ is the main driver of condensation, whereas the extended prion-like N-LCR attenuates the demixing of the full-length protein. Taken together these data show that the different LCRs of SFPQ have different and possibly opposing roles in determining the material state, subnuclear localisation and functions of SFPQ inside the cell.

## MATERIALS AND METHODS

### Plasmids for heterologous protein expression

For the production of proteins of interest in *Escherichia coli* for *in vitro* experiments, an open reading frame (ORF) for full-length SFPQ protein (residues 1-707; reference sequences: NCBI NM_005066.3 / Genbank X70944.1 / UniProt P23246-1) with an N-terminal 6×His-mEGFP-TEV tag was inserted using Gibson assembly between the ribosome binding site and T7 terminator of pET-28a(+) (Novagen). Expression vectors for H6-mEGFP-SFPQ ΔN-LCR (SFPQ residues 276-707), ΔC-LCR (residues 1-598) and ΔN&ΔC (residues 276-598) truncations were then constructed from this plasmid by site-directed deletion of the appropriate regions of the SFPQ ORF using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs). The mammalian expression construct encoding GFP-tagged full-length SFPQ was described previously (31,35). The truncated constructs [pEGFP-SFPQ ΔN-LCR (residues 276-707), pEGFP-SFPQ ΔC-LCR (residues 1-598 plus 700-707), and pEGFP-SFPQ ΔN- and ΔC-LCR (residues 276-598 plus 700-707)] were generated by site-directed deletion using the Q5 site-directed mutagenesis method (New England Biolabs). To ensure nuclear localisation of the truncated constructs, the endogenous nuclear localisation signal (NLS) of SFPQ (residues 700-707) was kept in ΔC-LCR constructs. All constructs were verified by DNA sequencing.

### Protein expression and purification for *in vitro* experiments

Plasmids for IPTG-inducible over-expression of H6-mEGFP-SFPQ proteins were transformed into *E. coli* Rosetta™ 2(DE3) cells (Novagen). Transformed cells were cultured overnight at 37°C in lysogeny broth (LB) with selection (50 µg/mL kanamycin, 34 µg/mL chloramphenicol). 30 mL of the overnight culture was then added to 3 L of fresh LB + 50 µg/mL kanamycin + 17 µg/mL chloramphenicol in six 2.5 L conical flasks and incubated at 37°C, shaking (180 rpm). Protein expression was induced at $OD_{600}$ ~ 0.5 by the addition of 0.5 mM IPTG, before continuing growth for 18-20 h at 20°C. All subsequent purification steps were performed at ambient temperature. Cells were harvested and resuspended in 100 mL of lysis buffer (1 M KCl, 5% glycerol, 10 mM imidazole, 50 mM Tris-HCl, 250 mM L-arginine, 1 mM PMSF, 0.05 mg DNase I) per L of cell culture. Cells were then lysed using either an EmulsiFlex-C5 high pressure homogenizer (Avestin) at ~ 15,000 psi, or by sonication on ice (cycles of 7 sec on, 3 sec off, for a total on-time of 6 min; 30% amplitude). Clarified lysate was loaded onto a 5 mL $Ni^{2+}$-charged HisTrap HP column (Cytiva), washed with 10 column volumes (CV) of binding buffer (1 M KCl, 5% glycerol, 10 mM imidazole, 50 mM Tris-HCl, 250 mM L-arginine, 1 mM PMSF, pH 7.4) followed by 10 CV wash buffer (binding buffer with 58 mM imidazole). His-tagged protein was then eluted in ~ 1 CV of elution buffer (binding buffer with 250 mM imidazole), filtered and passed through a 120 mL HiLoad Superdex 200 16/600 pg size exclusion column (Cytiva) pre-equilibrated in storage buffer (0.5 M KCl, 5% glycerol, 20 mM HEPES, 1 mM DTT, pH 7.4). Storage buffer was supplemented with 0.5 mM EDTA for SFPQ ΔN-LCR to prevent

degradation by proteolysis from the C-terminus (observed during a preliminary purification attempt). Peak fractions containing H6-mEGFP-tagged protein were pooled, concentrated using a 100 kDa MWCO centrifugal filter (Amicon), flash frozen in 20-50 µL aliquots in liquid nitrogen and stored at -80°C. To remove the H6-mEGFP tag, tagged protein was incubated with tobacco etch virus protease (TEV; produced in-house) at a 5:1 mass ratio of tagged protein:TEV for 2 hrs at ambient temperature in storage buffer. Untagged SFPQ was then isolated from TEV and H6-mEGFP tag by passing the sample through either a 120 mL HiLoad Superdex 200 16/600 pg or a 24 mL Superdex 200 Increase 10/300 GL size exclusion column (depending on the sample size) pre-equilibrated in storage buffer. Peak fractions containing untagged SFPQ were pooled and concentrated using a 30 kDa MWCO centrifugal filter (Amicon) to a final protein concentration of: 7.9 mg/mL for SFPQ full-length, 0.34 mg/mL for SFPQ ΔN-LCR, 20.5 mg/mL for SFPQ ΔC-LCR, 42.7 mg/mL for SFPQ ΔN&ΔC. Protein concentrations were calculated by absorbance at 280nm using the theoretical extinction coefficient for each protein as calculated from its amino acid sequence using *ProtParam* (36). Protein purity was assessed by SDS-PAGE (**Figure 2d**).

### *In vitro* phase separation assays

Storage buffer was added to protein stocks to produce a series of samples spanning a range of protein concentrations in PCR tubes. To induce phase separation, each sample was then diluted 3/10 by adding low salt buffer (20 mM HEPES, 1 mM DTT, pH 7.4) to give a final solution volume of 10 µL containing 150 mM KCl, 1.5% glycerol, 20 mM HEPES, 1 mM DTT, and a protein concentration of 30% of the pre-diluted sample. Low salt buffer was added rapidly but carefully to avoid bubbles and samples were flick mixed immediately after dilution. For negative controls, storage buffer was added to samples instead of low salt buffer.

To visualise condensed phase droplets in suspension, each sample was transferred to a well of a 384-well non-binding microplate (µClear®, Greiner Bio-One) immediately after dilution with low salt buffer and imaged by differential interference microscopy (DIC) using a Nikon Ti2 inverted microscope with a 60X oil immersion objective. All images were taken with the focal plane ~ 100 µm above the surface of the bottom of the well. Initially, samples containing mEGFP-SFPQ (full-length) droplets were imaged at 5 min intervals to assess the time-dependence of droplet formation and settling (**Figure S2**). Droplet size and number appeared to stabilise after 10-15 min; the number of droplets visibly decreased (due to settling) after ~ 35 min. Therefore, all samples were imaged ~ 20 min post-dilution. For SFPQ ΔN-LCR, initial experiments produced droplets that were very small (~ 1 µm or less in diameter) making it difficult to visualise droplet shape or fusion. Therefore, samples were rotated slowly for 1 h immediately after transfer to the microplate to allow time for droplets to coalesce in suspension, before imaging by DIC as described above.

To determine $C_{sat}$ from the dilute phase protein concentration, samples were incubated at room temperature for 10 min after inducing phase separation in PCR tubes as described above. Tubes were centrifuged at 2,000 × *g* for 10 min to accelerate settling of the condensed phase to the bottom of the tube. The protein concentration in the dilute phase was then determined by carefully transferring 2 µL of supernatant (i.e. the dilute phase) to a NanoDrop™ Lite spectrophotometer (Thermo Scientific) and measuring the absorbance at 280nm.

To determine $C_{sat}$ from the amount of condensed phase (**Figure S1**), samples containing mEGFP-tagged SFPQ were diluted to 150 mM KCl as described above, transferred immediately to a 384-well microplate, sealed and left overnight on the bench to allow all droplets to settle to the bottom of the plate. Fluorescent images of the surface of the plate were captured using a Nikon Ti2 inverted microscope with a 20X objective and 470nm excitation LED.

Analyses of all microscopy images were performed using ImageJ (v1.53q, NIH). To estimate the relative amount of condensed GFP-tagged SFPQ in samples shown in **Figure S1**, the threshold intensity above which pixels were counted as part of the condensed phase was set using a similar

approach to (11). Briefly, a Gaussian curve was fit to the background peak in the histogram for each image to estimate the mean and standard deviation of the background intensity. The threshold was then set as the mean background intensity plus three standard deviations. In this way, images could be thresholded in an unbiased way, even if no droplets were visible in the sample. Given that the condensed protein appeared to form puddles on the plate surface (rather than remain as spherical droplets of varying height), the proportion of plate surface area covered by condensed protein was used as a proxy for the relative condensed phase volume.

**DNA damage site recruitment assay**

HeLa cells were transfected with pEGFP-SFPQ constructs (either full-length, ΔN-LCR, ΔC-LCR or ΔN&ΔC) for 48 h. A day prior to the experiment, cells were treated with 10 µM BrdU in DMEM without phenol red (Gibco) containing 10% FBS and 1% Pen/Strep. The recruitment of GFP-SFPQ was imaged at 5 Hz on a Zeiss LSM710 confocal microscope. DNA damage was induced by irradiating 10 pixel-width of the nucleus with a 405 nm laser (line scan mode). The laser was set with a line attenuator transmission of 92%, 1 iteration and a pixel dwell time of 177.32 µs. Regions of interest were drawn to the DNA damage site to measure the GFP fluorescence over time using the ImageJ software (NIH). DNA damage-induced changes in fluorescence were calculated by normalizing the background subtracted values to the baseline fluorescence prior to laser microirradiation ($\Delta F/F_o$).

**Paraspeckle quantitation via fluorescence *in situ* hybridisation**

HeLa cells were maintained in DMEM containing high glucose and pyruvate (GIBCO) supplemented with 10% FBS (Sigma Aldrich) and Penicillin-Streptomycin (GIBCO). HeLa cells (7 x 104 cells/well) were seeded onto 1.5 thickness coverslips in a 12-well tray. The following day transfections were performed using 200 ng plasmid, 1 µl P3000 Reagent and 2.5 µl of Lipofectamine 3000 (Thermo Fisher) per well. Transfection media was replaced with fresh media after 6 h. 24 h post-transfection, coverslips were fixed in 4% formaldehyde for 10 min, washed three times with PBS and then permeabilised in 70% ethanol overnight at 4°C. RNA FISH was performed against NEAT1 using NEAT1 Middle Segment with Quasar® 570 Dye (Stellaris FISH probes). After removing 70% ethanol, coverslips were incubated in 1 mL of wash buffer A (Biosearch Technologies) for 5 min. Coverslips were then placed cell side down on 50 µl of hybridization buffer containing probe in a humidified chamber and incubated in the dark at 37°C for 16 h. After hybridisation, coverslips were transferred back to a 12-well tray, cell side up, and washed twice in 1 mL of wash buffer A incubated at 37°C in the dark for 30 min. Coverslips were incubated in buffer B (Biosearch Technologies) containing DAPI for 5 min at room temperature, washed once with buffer B, then mounted onto a glass slide using Vectashield antifade mounting medium (Vector Laboratories) and sealed using nail polish.

Fluorescent images were acquired using a Delta Vision Elite Deconvolution Microscope with a 60X objective. Images were acquired as 70 step z-stacks at intervals of 0.2 µm, with 3D deconvolution and maximum intensity projection performed to generate a final image for analysis. Brightness and contrast of all images were normalised using the ImageJ FIJI software package (NIH). Paraspeckle analysis was performed in CellProfiler using a custom pipeline. In brief, nuclear staining was mapped to determine regions of interest. Manual thresholding of NEAT1 FISH signal was then used to determine paraspeckle number and area per nuclei. Statistical analysis of paraspeckle number and area was performed using an ordinary one-way ANOVA with multiple comparisons in GraphPad Prism.

***In silico* sequence analysis**

The *IUPred2A* webserver (https://iupred2a.elte.hu/) (32) was used with default parameters (IUPred2 long disorder) to predict intrinsically disordered regions within the amino acid sequence of human SFPQ (UniProt accession P23246). The *PLAAC* webserver (http://plaac.wi.mit.edu/) (33) was used to find prion-like regions, with a minimal length ('core length', *c*) of 60 residues (default). The α value

was set to 0.5, such that the background amino acid frequencies used were the average of the frequencies across the entire *Homo sapiens* and *Saccharomyces cerevisiae* proteomes, as recommended by the program developers when scanning non-*S. cerevisiae* proteins (33).

## RESULTS

### SFPQ can form a condensed liquid-like phase *in vitro*

The ability to form homotypic liquid-like droplets *in vitro* has been described previously for many RNA-binding proteins with extended LCRs (11,37) (see also (7,38) for reviews). We therefore first sought to determine the phase behaviour of SFPQ *in vitro*. Recombinant full-length human SFPQ with an N-terminal GFP tag was purified from bacterial cell culture; the GFP tag was removed and untagged SFPQ was concentrated in a buffer containing 500 mM KCl (**Figure 2d**). Upon dilution to physiological salt concentration (150 mM KCl), micron-sized droplets formed in a concentration-dependent manner (**Figure 2a**). As droplets settled they were observed to fuse with the bulk condensed phase at the bottom of the sample container, confirming their liquid-like state (**Movie S1**). Furthermore, by measuring the absorbance at 280 nm of the supernatant after centrifuging to remove all droplets, we could determine the protein concentration in the dilute phase (i.e. the protein concentration outside of droplets). We observed that the dilute phase protein concentration was unchanged in samples with total protein concentrations greater than the concentration at which droplets began to form (**Figure 2b**). This is behaviour is consistent with the expectation for phase separation in a binary mixture. We identify the invariant concentration measured in the coexisting dilute phase with the saturation concentration ($C_{sat}$) (39). Under these conditions (150 mM KCl, pH 7.4, ambient temperature), we determined a $C_{sat}$ of 11.6 μM for full-length SFPQ. This $C_{sat}$ is in the same range as the estimated nuclear concentration of similar RNA-binding proteins (FUS, EWSR1, TAF15, TDP43 hnRNPA1 (37)). The total concentration of SFPQ in HeLa cells has been estimated using mass spectrometry to be 1.8 μM (40). However, given that SFPQ is predominantly localised to the nucleus, its concentration within the nucleus is likely to be much greater than this. Furthermore, the increase in local concentration due to recruitment to DNA and/or RNA sites, along with the highly crowded intracellular environment, makes condensation of SFPQ in the nucleus very plausible.

The method described above for $C_{sat}$ determination relies on measuring the protein concentration in the dilute phase. To further verify this result, we used an alternative method to instead quantify the amount of condensed phase. For this, we used GFP-tagged SFPQ to facilitate imaging by fluorescence. By imaging after all droplets had settled to the bottom of the sample we were able to quantify the relative amount of condensed phase in each sample over a range of concentrations (**Figure S1a** and **S1b**). For samples containing droplets, the relationship between amount of condensed phase and total protein concentration was approximately linear, as expected for a binary two-phase system (39). The minimum concentration required for phase separation to occur (i.e. the $C_{sat}$) was then calculated by extrapolation to zero. For comparison we also determined the $C_{sat}$ of GFP-SFPQ using the centrifuge-based assay as described above (**Figure S1c**). Both methods gave a $C_{sat}$ for GFP-SFPQ of 1.5-1.6 μM, providing confidence in the accuracy of these approaches. This also showed that the N-terminal GFP tag increases the propensity of SFPQ to condense *in vitro*. Therefore, for further *in vitro* experiments we used untagged proteins only and the centrifuge-based assay for $C_{sat}$ determination.

### The C-terminal LCR of SFPQ drives phase separation, which is attenuated by the N-terminal prion-like LCR *in vitro*

To determine the contribution of each LCR of SFPQ independently on the *in vitro* condensation behaviour of the protein, recombinant SFPQ truncations lacking the N-LCR, C-LCR or both N- and C-LCR were purified from bacterial cell culture (**Figure 2d**) and assayed for *in vitro* phase separation

in the same way as for full-length SFPQ. Unsurprisingly, deletion of both N- and C-LCRs (SFPQ ΔN&ΔC) produced a protein that would not phase separate to form droplets (**Figure 2a**), remaining as a single soluble phase even at a protein concentration almost 30-fold greater that the $C_{sat}$ of full-length SFPQ (**Figure 2b**). Even after further dilution to lower the KCl concentration to 75 mM, SFPQ ΔN&ΔC remained soluble up to a concentration of ~ 50 μM, at which point it precipitated as visible solid-looking aggregates (**Figure 2c**), indicating that without its LCRs, SFPQ cannot form a condensed liquid phase under physiological-like conditions (ionic strength, pH, [protein], temperature) *in vitro*.

Deletion of the C-LCR only (amino acids 599-707) also resulted in a protein which would not phase separate under the same conditions as for the full-length protein (**Figure 2a**). Even at a total protein concentration eight-fold greater than the $C_{sat}$ of full-length SFPQ, SFPQ ΔC-LCR remained as a single dilute phase, as determined using the centrifuge assay (**Figure 2b**). This clearly implies that the C-LCR is crucial for driving SFPQ phase separation. To assess whether SFPQ ΔC-LCR could phase separate at all, we decreased the salt concentration further by dilution. In contrast to SFPQ ΔN&ΔC, liquid droplets were clearly observed to form in a sample containing 75 mM KCl and 50 μM SFPQ ΔC-LCR (**Figure 2c** and **Movie S2**), demonstrating that the N-LCR of SFPQ can facilitate condensation of the protein as a liquid, albeit requiring more extreme conditions (i.e. less like physiological ionic strength and [protein]) than for full-length SFPQ.

Unexpectedly, deletion of the longer prion-like N-LCR (amino acids 1-275) resulted in a protein with an increased propensity to phase separate, suggesting that the N-LCR actually attenuates demixing (promotes mixing) of SFPQ in the context of the full-length protein. The $C_{sat}$ of SFPQ ΔN-LCR in 150 mM KCl buffer was determined to be 0.8 μM (**Figure 2b, bottom-right**); a ~15-fold decrease as compared to the full-length protein. Indeed, we could not access protein concentrations comparable to the full-length protein as, even in buffer containing 500 mM KCl, condensation (observed as turbidity) occurred while concentrating SFPQ ΔN-LCR during purification (data not shown). The SFPQ ΔN-LCR droplets that we were able to observe were smaller and more sparse than for the full-length protein (**Figure 2a, bottom**), as expected for a lower condensed volume in samples containing protein concentrations that we could readily achieve. Despite this, droplet fusion events could still be observed, indicating that this condensed phase formed by SFPQ ΔN-LCR was liquid-like (**Movie S3**).

Together, these *in vitro* data suggest that the LCRs of SFPQ have distinct roles in determining the phase behaviour of the protein. Independently, the N-LCR and the C-LCR can each facilitate homotypic condensation of the conserved globular DBHS region of SFPQ as a liquid; in the absence of both LCRs, the protein precipitates at high concentrations as a solid. However, the shorter C-LCR is clearly the main driver of condensation, whereas, in the absence of the C-LCR, much more extreme conditions are required to observe liquid condensation mediated by the longer prion-like N-LCR. Perhaps most significantly, we find that the N-LCR in fact attenuates SFPQ condensation in the context of the full-length protein at physiological salt concentration.

**The C-terminal LCR is critical for driving the recruitment of SFPQ to DNA damage sites**

High avidity liquid-like interactions between proteins with intrinsically disordered regions and nucleic acids have been proposed as a key mechanism driving spatiotemporal condensation of macromolecules within the cell nucleus (4,6,37), including the assembly of DNA repair complexes (41). Previous studies have shown that SFPQ is directly involved in DNA damage repair via the non-homologous end joining (NHEJ) pathway. *In vitro*, the SFPQ-NONO dimer forms a committed pre-ligation complex with another NHEJ factor (Ku) and the DNA substrate, which is essential for DNA ligase IV/XRCC4-mediated joining of double-stranded DNA ends (42). *In vivo*, depletion of SFPQ severely delays repair of double-strand breaks (43), and results in a marked increase in radiation-sensitivity in HeLa cells (44). In addition, SFPQ can bind to DNA in the absence of other DBHS proteins (31,45), and is the DBHS partner which drives DNA binding *in vitro* (46) and *in vivo* (44).

To probe for roles of the LCRs of SFPQ in driving functional condensation, we assessed the recruitment of SFPQ to DNA damage sites in HeLa cells overexpressing GFP-tagged SFPQ. Within minutes of causing localised damage to DNA in the cell nucleus using a laser microbeam, GFP signal accumulated at damage sites, indicating rapid SFPQ recruitment as shown previously (43,44) (**Figure 3a**). GFP signal peaked approximately 3 minutes after inducing DNA damage, and decreased gradually back to baseline after 10-15 minutes (**Figure 3b**). Compared to full-length SFPQ, no significant difference in recruitment to damage sites was observed for SFPQ ΔN-LCR (**Figure 3c**). Removal of the C-LCR, however, reduced the relative recruitment by approximately two-fold; the magnitude of this effect was similar for both SFPQ ΔC-LCR and SFPQ ΔN&ΔC. Interestingly, although DNA damage recruitment was unchanged for SFPQ ΔN-LCR, additional bright GFP puncta appeared to be present these in cells compared to those expressing full-length SFPQ (**Figure 3b, row 2**). Together, these results suggest that deletion of the C-LCR, but not the N-LCR impairs SFPQ condensate formation at DNA damage sites, and that removal of the N-LCR possibly increases the condensation propensity of SFPQ in the nucleus, consistent with our *in vitro* data.

**The C-terminal LCR of SFPQ is critical for paraspeckle formation *in vivo*, whereas the N-LCR increases SFPQ nuclear dispersal**

It has been observed that prion-like domains are enriched in human nuclear DNA- and RNA-binding proteins (16). Prion-like regions have been suggested to be important mediators of functional aggregation to form membraneless organelles, with proteins containing prion-like LCRs shown to be over-represented in paraspeckles (15) – membraneless organelles present in the nucleus of many cell types. SFPQ is one of seven proteins shown to be indispensable for the formation of paraspeckles (47). We have previously used FISH to specifically locate the long non-coding RNA, NEAT1_2, upon which paraspeckles are scaffolded (23,29,31). In addition, SFPQ subnuclear localisation has been observed in live cells previously by over-expression of GFP-tagged SFPQ (48). Therefore, to probe for roles of the LCRs of SFPQ in subnuclear organisation *in vivo*, we performed quantitative analysis on paraspeckles – identified as NEAT1 FISH puncta – in cultured HeLa cells overexpressing GFP-tagged SFPQ truncations (**Figure 4**).

SFPQ localises along with other essential paraspeckle proteins, NONO and FUS, within the core of paraspeckles in cells where they are present (49). Therefore, as expected, in cells expressing full-length SFPQ, GFP signal was predominantly observed as discrete puncta co-localised with NEAT1_2 (**Figure 4b**). Paraspeckles can be either singular spheroids of 360 nm diameter, or chains of spheroids with a constant width and length up to 1.5 μm. To determine the effect of SFPQ over-expression on paraspeckle number and size, we quantitated paraspeckle number, total paraspeckle area (i.e. the sum of pixels associated with paraspeckles), and the average paraspeckle area (total paraspeckle area/paraspeckle number) in cells where either GFP, or GFP-SFPQ were overexpressed. Compared to GFP alone, cells over-expressing GFP-SFPQ had a small but significant decrease in paraspeckle number, accompanied by an increase in paraspeckle area by approximately 1.5 to 2-fold (**Figure S3**). Thus, overexpressed GFP-SFPQ is targeted to paraspeckles and increases their average size, but does not increase singular paraspeckle numbers.

To determine the effect of removing the C-LCR from SFPQ on paraspeckle size and number, a truncation was constructed lacking residues 599-699; the C-terminal NLS (residues 700-707) was included to ensure nuclear localisation. Compared to cells expressing full-length GFP-SFPQ, a marked decrease in the number and size of paraspeckles was observed for cells expressing this SFPQ ΔC-LCR construct (**Figure 4a**). Given the presence of endogenous DBHS proteins (i.e. no knock-down was performed), this demonstrates that SFPQ ΔC-LCR acts as a dominant negative in regards to paraspeckle abundance. GFP signal was observed in the nucleus as expected, albeit with a tendency to be dispersed more evenly throughout the nucleus than for the full-length protein (**Figure 4b**). This again suggests that the C-LCR is a key driver of functional condensation of SFPQ in the cell nucleus.

In contrast to the C-LCR, which is clearly required for paraspeckle maintenance, loss of the N-LCR had no significant effect on paraspeckle size and number compared to full-length SFPQ (**Figure 4a**). However, upon inspection of the fluorescent images we noted that the paraspeckles (containing NEAT1 and GFP-SFPQ ΔN-LCR) appeared more rounded than for full-length GFP-SFPQ (**Figure 4b**). Furthermore, it appeared that SFPQ ΔN-LCR showed a greater tendency to non-specifically condense in cells, with many more GFP puncta than normal, many of which were not colocalised with NEAT1 and therefore not paraspeckles. These non-paraspeckle puncta were of various sizes (~ 200 - 2000 nm) and all approximately spherical. This is consistent with our *in vitro* data showing that SFPQ ΔN-LCR has a greater tendency to phase separate to form condensed droplets, and consistent with the additional GFP puncta observed for SFPQ ΔN-LCR cells in our DNA damage assay (**Figure 3a**). Similar bright non-paraspeckle puncta have been observed for SFPQ truncations missing the N-LCR previously (44,48).

Finally, over-expression of a double-deletion missing both LCRs (SFPQ ΔN&ΔC; includes NLS) resulted in a decrease in the number and size of paraspeckles, similar to SFPQ ΔC-LCR (**Figure 4a**). However, this decrease was significantly less than for SFPQ ΔC-LCR. Our interpretation is that the N-LCR has an inhibitory effect on paraspeckle formation, at least in the case where the C-LCR is absent. In contrast to the relatively dispersed nuclear localisation observed for SFPQ ΔC-LCR, the double deletion appeared as relatively few discrete, mostly non-paraspeckle puncta, more similar to those observed for SFPQ ΔN-LCR (**Figure 4b**). This is again consistent with the N-LCR attenuating SFPQ condensate formation, increasing its tendency to disperse throughout the nucleus rather than condensing to form discrete puncta. Given that the ability to condense is very likely to be important for driving paraspeckle formation, this suggests that SFPQ ΔC-LCR acts as a stronger dominant negative than SFPQ ΔN&ΔC due to the inclusion of the N-LCR.

## DISCUSSION

Demixing of proteins and nucleic acids is emerging as a key mechanism describing the dynamic and reversible organisation of molecules in the cell. The central involvement of nucleic acid-binding proteins with disordered, low-complexity regions – particularly those with prion-like composition – tends to be a common feature of many studies to date, with condensate formation driven by prion-like domains emerging as a key theme (15,17,23,41,50-52). SFPQ is an abundant, ubiquitous and multifunctional RNA- and DNA-binding protein with an extended N-terminal prion-like LCR. Unexpectedly, here we show that, rather than driving phase separation, this prion-like N-LCR attenuates SFPQ phase separation, which is in fact driven by the shorter, non-prion-like, C-terminal LCR. We have observed this using *in vitro* droplet-forming assays and using two different cell-based assays quantifying two different functional biomolecular condensates: paraspeckles and DNA damage repair foci.

Our *in vitro* data show that, under the range of conditions tested herein, the core globular DBHS region of SFPQ alone cannot condense as a liquid, and instead precipitates at high concentrations as solid-like aggregates. Under the same conditions, addition of the prion-like N-LCR allows the protein to instead condense as a liquid. However, our results urge caution regarding the interpretation of observed *in vitro* protein liquid condensation: in the absence of the C-LCR, much more extreme conditions (i.e. higher protein concentration and lower ionic strength) were required to observe droplet formation than for the full-length protein, indicating a large decrease in the propensity to undergo demixing compared to the full-length protein. That is, although the N-LCR can *allow* condensation of the protein as a liquid under conditions in which the protein would otherwise precipitate as a solid, it is not the *driver* of protein condensation under physiological conditions. Indeed, even though our SFPQ ΔC-LCR truncation contained all of the domains thus far identified as essential for DNA-binding and paraspeckle localisation (31), we observed significant impairment in

both DNA damage site localisation and paraspeckle formation, indicating that alone the N-LCR is insufficient to drive functional condensation of the protein in the cell. The C-LCR also allowed the protein to condense as a liquid *in vitro*. However, in contrast to the N-LCR, the C-LCR caused the core DBHS region to condense at much lower protein concentrations, under conditions much closer to physiological. Furthermore, to our surprise, we found that the removal of the prion-like N-LCR from the full-length protein resulted in a significant increase in the phase separation propensity *in vitro*, and the formation of additional condensates *in vivo*. Together, this shows that the C-LCR drives the formation of functional condensates, whereas the N-LCR attenuates condensation of the full-length protein. This is, to our best knowledge, the first description of an extensive prion-like LCR that serves to regulate rather than drive a protein's condensation propensity; most of the recent studies on LCRs of RNA-binding proteins have focused solely on their roles in promoting protein phase separation.

It is well established that DBHS proteins are obligate dimers – able to form homo- and heterodimers – with a preference for heterodimerisation (27-29,54). SFPQ, NONO and PSPC1 co-immunoprecipitate (55,56), and SFPQ is essential for the localisation of NONO both to DNA damage sites (44) and to paraspeckles (29). Previous in-depth structural studies have identified the minimal region necessary and sufficient for DBHS dimer formation as composed of RRM2, NOPS and coiled-coil domains (30,31). It is important to note that the entire core conserved DBHS region, which includes all of the domains necessary for dimer formation, was left intact in the truncations of SFPQ used herein. This provides a basis for interpreting the results of our *in vivo* experiments, in which SFPQ truncations were over-expressed in the presence of endogenous wild-type DBHS proteins and therefore would be very likely to exist as heterodimers with endogenous DBHS proteins in the cell.

The putative DNA binding domain (DBD) of SFPQ, which is part of the N-LCR and loosely defined as amino acids 214-275, has been shown to be essential for DNA binding *in vitro* (31) and localisation of SFPQ/NONO and other NHEJ factors to DNA damage sites *in vivo* (44). This latter study was performed using cells in which endogenous SFPQ was knocked-down (44). In contrast, our results show that SFPQ ΔN-LCR, for which most of the DBD is absent, localises to DNA damage sites to the same degree as full-length SFPQ. This suggests that interactions with endogenous SFPQ can compensate for the absent DBD of SFPQ ΔN-LCR, allowing for the wild-type-like DNA damage localisation we observe. In addition to direct SFPQ-DNA interactions, it is possible that other factors also contribute to SFPQ recruitment to DNA damage sites: Recruitment of the RNA-binding protein FUS to DNA damage sites precedes SFPQ, and its absence strongly reduces SFPQ recruitment (41). In light of this, the significant impairment of DNA-damage localisation we observe upon deletion of the C-LCR is striking, as it shows that even the presence of endogenous SFPQ cannot compensate for the lack of this domain, suggesting that DBHS dimers containing SFPQ ΔC-LCR have a significantly lower propensity to condense on DNA than those that contain full-length SFPQ. That is, SFPQ ΔC-LCR attenuates condensation, even in the presence of endogenous proteins. This is consistent with our paraspeckle quantitation data showing that over-expressed SFPQ ΔC-LCR impairs paraspeckle formation and displays a more dispersed granular nuclear distribution compared to full-length SFPQ. Furthermore, the fact that SFPQ ΔN&ΔC did not display a similar dispersed nuclear distribution, but rather localised to discrete rounded puncta, points again to the N-LCR as the region that attenuates condensation of DBHS proteins in the cell.

Altogether, this points to a regulatory role for the N-LCR, rather than a role as a driver of phase separation. Interestingly, data from two independent studies involving SFPQ truncations missing the N-LCR have hinted at this previously. In a study focusing on nuclear speckles, upon over-expression of SFPQ truncations lacking amino acids 1-278 or 1-337 in Hela cells, the authors observed discrete nuclear puncta, which appeared larger and more rounded than nuclear speckles (48). Similarly, authors of a study focussing on the localisation of SFPQ to DNA damage sites noted the accumulation of SFPQ truncations lacking 1-297 or 1-372 in discrete, bright non-paraspeckle puncta (44). These

observations are consistent with our results and suggest that, rather than driving condensation, the prion-like N-LCR is required to prevent aberrant protein condensation in the nucleus.

Despite being much smaller than the N-LCR, we show that the C-LCR is the main driver of homotypic liquid demixing *in vitro*. This same region is critical for both recruitment to DNA damage sites and paraspeckle formation *in vivo*, strongly suggesting that liquid phase condensation of SFPQ, via its C-LCR, is essential for achieving its normal functions in the cell nucleus. Further studies aiming to dissect the sequence features within the C-LCR important for driving demixing are currently underway. In addition, it is expected that interactions with nucleic acids via the DNA binding and RRM domains of SFPQ, and also polymerisation via the coiled-coil (31), will emerge as important contributors to the condensation of DBHS proteins with other macromolecules within the complex intracellular environment.

In summary, our data provide nuance to the emerging paradigm regarding phase separation driven by low-complexity prion-like domains by showing that the two LCRs of SFPQ have contrasting and possibly opposing effects on condensate formation. It is striking that the shorter C-LCR appears to be the main driver of condensation, whereas the much longer N-LCR of SFPQ, which makes up about one-third of the total polypeptide length and contains a long prion-like region, attenuates condensation of the full-length protein. It may be plausible to propose a model in which direct interaction between the different LCRs of SFPQ may be dynamically modulated to provide a mechanism for spatial and temporal regulation of condensation (**Figure 5**). That is, in the absence of competing factors, the long N-LCR attenuates SFPQ condensation by directly interfering with intermolecular interactions mediated by the C-LCR. Conversely, in the presence of competing factors such as nucleic acids or other protein interaction partners which occupy the N-LCR, the C-LCR is then free to form intermolecular interactions to drive demixing leading to condensate formation. This presents a mechanism in which phase separation can be regulated by competing intra- and intermolecular interactions between different LCRs within the same protein. A conceptually similar mechanism has been proposed recently for the self-regulation of phase separation of the stress-granule protein, G3BP1 (57). The highly variable N- and C-terminal LCRs of the three human DBHS homologs, coupled with their intrinsic combinatorial heterodimerisation nature, present additional avenues for complex self-regulation of condensate formation in the cell. This, along with their multifunctional nature, presents DBHS proteins as an attractive model system for further studies describing relationships between the composition of LCRs and the emergent material state of proteins that contain intrinsically disordered regions.

## DATA AVAILABILITY

The ImageJ macro used to threshold fluorescent images and calculate condensed area is available from https://github.com/acmarshall88/Drop-Counter/blob/main/Drop_Counter.ijm. Fitting of linear models for calculation of $C_{sat}$ from centrifuge assay data and fluorescent image data was performed using simple custom scripts written in R (v4.1.0), available from https://github.com/acmarshall88/Csat-CalculatoRs.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

This work was funded by the Australian Research Council (FT180100204 to A.H.F.: DP160102435 to C.S.B. and A.H.F.; DP220103667 to C.S.B. and A.H.F.; LE120100092 and LE140100096 to C.S.B.), the National Health and Medical Research Council of Australia (APP1147496 to C.S.B. and A.H.F.), Motor Neurone Disease Research Australia (the Judy Mitchell MND Research Grant to V.A., M.L., and J.W.), and Tracey Banivanua Mar Fellowship (La Trobe University, Melbourne, Australia to

M.L.). J.W. was supported by a University of Queensland (UQ) Amplify Fellowship. T.Z. is a recipient of a UQ Research Training Scholarship. A.C.M. was supported by the Clifford Bradley Robertson and Gwendoline Florence Anne Robertson Research Endowment Fund, established through Dr Glen Robertson's bequest to The University of Western Australia.

*Conflict of interest statement.* A.H. is a founder of Dewpoint Therapeutics and a member of the board as well as a shareholder in Caraway Therapeutics. All other authors have no competing interests.

## ACKNOWLEDGEMENTS

The authors thank both Dr Hiroyuki Uechi and Dr Patrick McCall (Max Planck Institute of Mol. Cell Biology and Genetics, Dresden) for helpful feedback while drafting the manuscript. The authors are grateful for free access to UCSF ChimeraX (v1.3) (58) (used for molecular graphics in **Figure 1b**), developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, NIAID.

*Author contributions:* M.L., A.H.F., C.S.B. and A.C.M. designed the project. A.C.M. and J.C. purified proteins and performed *in vitro* experiments with guidance from M.L. S.K. performed HeLa cell culture, imaging and analysis. T.Z., J.W. and V.A. performed DNA damage recruitment assays. A.C.M. drafted the manuscript. All authors contributed to data interpretation, manuscript review and editing.

## REFERENCES

1. Banani, S.F., Lee, H.O., Hyman, A.A. and Rosen, M.K. (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, **18**, 285-298.
2. McCluggage, F. and Fox, A.H. (2021) Paraspeckle nuclear condensates: Global sensors of cell stress? *BioEssays*, **43**, 2000245.
3. Wei, M.-T., Chang, Y.-C., Shimobayashi, S.F., Shin, Y., Strom, A.R. and Brangwynne, C.P. (2020) Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol.*, **22**, 1187-1196.
4. Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R.V. and Brangwynne, C.P. (2016) Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, **165**, 1686-1697.
5. Yamazaki, T., Yamamoto, T., Yoshino, H., Souquere, S., Nakagawa, S., Pierron, G. and Hirose, T. (2020) Paraspeckles are constructed as block copolymer micelles through microphase separation. *bioRxiv*, doi: https://doi.org/10.1101/2020.1111.1102.366021, 3 Nov 2020, pre-print: not peer-reviewed.
6. Morin, J.A., Wittmann, S., Choubey, S., Klosin, A., Golfier, S., Hyman, A.A., Jülicher, F. and Grill, S.W. (2022) Sequence-dependent surface condensation of a pioneer transcription factor on DNA. *Nat. Phys.*, **18**, 271–276.
7. Borcherds, W., Bremer, A., Borgia, M.B. and Mittag, T. (2021) How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation? *Curr. Opin. Struct. Biol.*, **67**, 41-50.
8. Bremer, A., Farag, M., Borcherds, W.M., Peran, I., Martin, E.W., Pappu, R.V. and Mittag, T. (2022) Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.*, **14**, 196-207.
9. Martin, E.W. and Mittag, T. (2018) Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry*, **57**, 2478-2487.
10. Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T. and Taylor, J.P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell*, **163**, 123-133.
11. Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D. *et al.* (2018) A Molecular Grammar Governing

the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*, **174**, 688-699.e16.
12. Guenther, E.L., Cao, Q., Trinh, H., Lu, J., Sawaya, M.R., Cascio, D., Boyer, D.R., Rodriguez, J.A., Hughes, M.P. and Eisenberg, D.S. (2018) Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat. Struct. Mol. Biol.*, **25**, 463-471.
13. Hughes, M.P., Sawaya, M.R., Boyer, D.R., Goldschmidt, L., Rodriguez, J.A., Cascio, D., Chong, L., Gonen, T. and Eisenberg, D.S. (2018) Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science*, **359**, 698-701.
14. Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D. *et al.* (2015) Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell*, **57**, 936-947.
15. Hennig, S., Kong, G., Mannen, T., Sadowska, A., Kobelke, S., Blythe, A., Knott, G.J., Iyer, K.S., Ho, D., Newcombe, E.A. *et al.* (2015) Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.*, **210**, 529-539.
16. March, Z.M., King, O.D. and Shorter, J. (2016) Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.*, **1647**, 9-18.
17. Fox, A.H., Nakagawa, S., Hirose, T. and Bond, C.S. (2018) Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends Biochem. Sci*, **43**, 124-135.
18. Zhang, H., Elbaum-Garfinkle, S., Langdon, E.M., Taylor, N., Occhipinti, P., Bridges, A.A., Brangwynne, C.P. and Gladfelter, A.S. (2015) RNA Controls PolyQ Protein Phase Transitions. *Mol. Cell*, **60**, 220-230.
19. Lim, Y.W., James, D., Huang, J. and Lee, M. (2020) The Emerging Role of the RNA-Binding Protein SFPQ in Neuronal Function and Neurodegeneration. *Int. J. Mol. Sci.*, **21**, 7151.
20. Yarosh, C.A., Iacona, J.R., Lutz, C.S. and Lynch, K.W. (2015) PSF: nuclear busy-body or nuclear facilitator? *Wiley Interdiscip. Rev. RNA*, **6**, 351-367.
21. Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Bénard, M., Fox, A.H. *et al.* (2014) NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell*, **25**, 169-183.
22. Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T. and Hirose, T. (2009) MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2525-2530.
23. West, J.A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R.E., Hirose, T., Bond, C. *et al.* (2016) Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.*, **214**, 817-830.
24. Souquere, S., Beauclair, G., Harper, F., Fox, A. and Pierron, G. (2010) Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. *Mol. Biol. Cell*, **21**, 4020-4027.
25. Knott, G.J., Bond, C.S. and Fox, A.H. (2016) The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res.*, **44**, 3989-4004.
26. Knott, G.J., Lee, M., Passon, D.M., Fox, A.H. and Bond, C.S. (2015) Caenorhabditis elegans NONO-1: Insights into DBHS protein structure, architecture, and function. *Protein Sci.*, **24**, 2033-2043.
27. Huang, J., Casas Garcia, G.P., Perugini, M.A., Fox, A.H., Bond, C.S. and Lee, M. (2018) Crystal structure of a SFPQ/PSPC1 heterodimer provides insights into preferential heterodimerization of human DBHS family proteins. *J. Biol. Chem.*, **293**, 6593-6602.
28. Knott, G.J., Chong, Y.S., Passon, D.M., Liang, X.-H., Deplazes, E., Maria, Andrew, Lee, M., Archa and Charles. (2022) Structural basis of dimerization and nucleic acid binding of human DBHS proteins NONO and PSPC1. *Nucleic Acids Res.*, **50**, 522-535.
29. Lee, P.W., Marshall, A.C., Knott, G.J., Kobelke, S., Martelotto, L., Cho, E., McMillan, P.J., Lee, M., Bond, C.S. and Fox, A.H. (2022) Paraspeckle subnuclear bodies depend on dynamic heterodimerisation of DBHS RNA-binding proteins via their structured domains. *J. Biol. Chem.*, (In Press, Available Online 7 Oct 2022).

30. Passon, D.M., Lee, M., Rackham, O., Stanley, W.A., Sadowska, A., Filipovska, A., Fox, A.H. and Bond, C.S. (2012) Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 4846-4850.
31. Lee, M., Sadowska, A., Bekere, I., Ho, D., Gully, Benjamin S., Lu, Y., Iyer, K.S., Trewhella, J., Fox, A.H. and Bond, C.S. (2015) The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Res.*, **43**, 3826-3840.
32. Erdős, G. and Dosztányi, Z. (2020) Analyzing Protein Disorder with IUPred2A. *Curr. Protoc. Bioinform.*, **70**.
33. Lancaster, A.K., Nutter-Upham, A., Lindquist, S. and King, O.D. (2014) PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics*, **30**, 2501-2502.
34. Peran, I. and Mittag, T. (2020) Molecular structure in biomolecular condensates. *Curr. Opin. Struct. Biol.*, **60**, 17-26.
35. Huang, J., Ringuet, M., Whitten, A.E., Caria, S., Lim, Y.W., Badhan, R., Anggono, V. and Lee, M. (2020) Structural basis of the zinc-induced cytoplasmic aggregation of the RNA-binding protein SFPQ. *Nucleic Acids Res.*, **48**, 3356-3365.
36. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) In Walker, J. M. (ed.), *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, pp. 571-607.
37. Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M. *et al.* (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*, **360**, 918–921.
38. Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L. *et al.* (2018) Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.*, **28**, 420-435.
39. Alberti, S., Gladfelter, A. and Mittag, T. (2019) Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*, **176**, 419-434.
40. Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F. *et al.* (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell*, **163**, 712-723.
41. Levone, B.R., Lenzken, S.C., Antonaci, M., Maiser, A., Rapp, A., Conte, F., Reber, S., Mechtersheimer, J., Ronchi, A.E., Mühlemann, O. *et al.* (2021) FUS-dependent liquid–liquid phase separation is important for DNA repair initiation. *J. Cell Biol.*, **220**.
42. Bladen, C.L., Udayakumar, D., Takeda, Y. and Dynan, W.S. (2005) Identification of the Polypyrimidine Tract Binding Protein-associated Splicing Factor·p54(nrb) Complex as a Candidate DNA Double-strand Break Rejoining Factor. *J. Biol. Chem.*, **280**, 5205-5210.
43. Salton, M., Lerenthal, Y., Wang, S.-Y., Chen, D.J. and Shiloh, Y. (2010) Involvement of Matrin 3 and SFPQ/NONO in the DNA damage response. *Cell cycle*, **9**, 1568-1576.
44. Ha, K., Takeda, Y. and Dynan, W.S. (2011) Sequences in PSF/SFPQ mediate radioresistance and recruitment of PSF/SFPQ-containing complexes to DNA damage sites in human cells. *DNA Repair*, **10**, 252-259.
45. Song, X., Sun, Y. and Garen, A. (2005) Roles of PSF protein and VL30 RNA in reversible gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12189-12193.
46. Zhang, W.W., Zhang, L.X., Busch, R.K., Farrés, J. and Busch, H. (1993) Purification and characterization of a DNA-binding heterodimer of 52 and 100 kDa from HeLa cells. *Biochemical J.*, **290**, 267-272.
47. Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N. and Hirose, T. (2012) Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.*, **31**, 4020-4034.
48. Dye, B.T. and Patton, J.G. (2001) An RNA recognition motif (RRM) is required for the localization of PTB-associated splicing factor (PSF) to subnuclear speckles. *Exp. Cell Res.*, **263**, 131-144.

49. Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G. and Hirose, T. (2018) Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol. Cell*, **70**, 1038-1053.e1037.
50. Dorone, Y., Boeynaems, S., Jin, B., Bossi, F., Flores, E., Lazarus, E., Michiels, E., De Decker, M., Baatsen, P., Holehouse, A.S. *et al.* (2020). *bioRxiv*, doi: https://doi.org/10.1101/2020.08.07.242172, 7 Aug 2020, pre-print: not peer-reviewed.
51. Franzmann, T.M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A.S., Nüske, E., Richter, D., Baumeister, W., Grill, S.W., Pappu, R.V. *et al.* (2018) Phase separation of a yeast prion protein promotes cellular fitness. *Science*, **359**.
52. Harrison, A.F. and Shorter, J. (2017) RNA-binding proteins with prion-like domains in health and disease. *Biochemical J.*, **474**, 1417-1438.
53. Tsuboyama, K., Osaki, T., Matsuura-Suzuki, E., Kozuka-Hata, H., Okada, Y., Oyama, M., Ikeuchi, Y., Iwasaki, S. and Tomari, Y. (2020) A widespread family of heat-resistant obscure (Hero) proteins protect against protein instability and aggregation. *PLoS Biol.*, **18**, e3000632.
54. Laurenzi, T., Palazzolo, L., Taiana, E., Saporiti, S., Ben Mariem, O., Guerrini, U., Neri, A. and Eberini, I. (2022) Molecular Modelling of NONO and SFPQ Dimerization Process and RNA Recognition Mechanism. *Int. J. Mol. Sci.*, **23**, 7626.
55. Bottini, S., Hamouda-Tekaya, N., Mategot, R., Zaragosi, L.-E., Audebert, S., Pisano, S., Grandjean, V., Mauduit, C., Benahmed, M., Barbry, P. *et al.* (2017) Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic Sfpq. *Nat. Commun.*, **8**, 1189.
56. De Silva, H.C., Lin, M.Z., Phillips, L., Martin, J.L. and Baxter, R.C. (2019) IGFBP-3 interacts with NONO and SFPQ in PARP-dependent DNA damage repair in triple-negative breast cancer. *Cell. Mol. Life Sci.*, **76**, 2015-2030.
57. Guillén-Boixet, J., Kopach, A., Holehouse, A.S., Wittmann, S., Jahnel, M., Schlüßler, R., Kim, K., Trussina, I.R.E.A., Wang, J., Mateju, D. *et al.* (2020) RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell*, **181**, 346-361.e317.
58. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H. and Ferrin, T.E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.*, **30**, 70-82.

## FIGURES

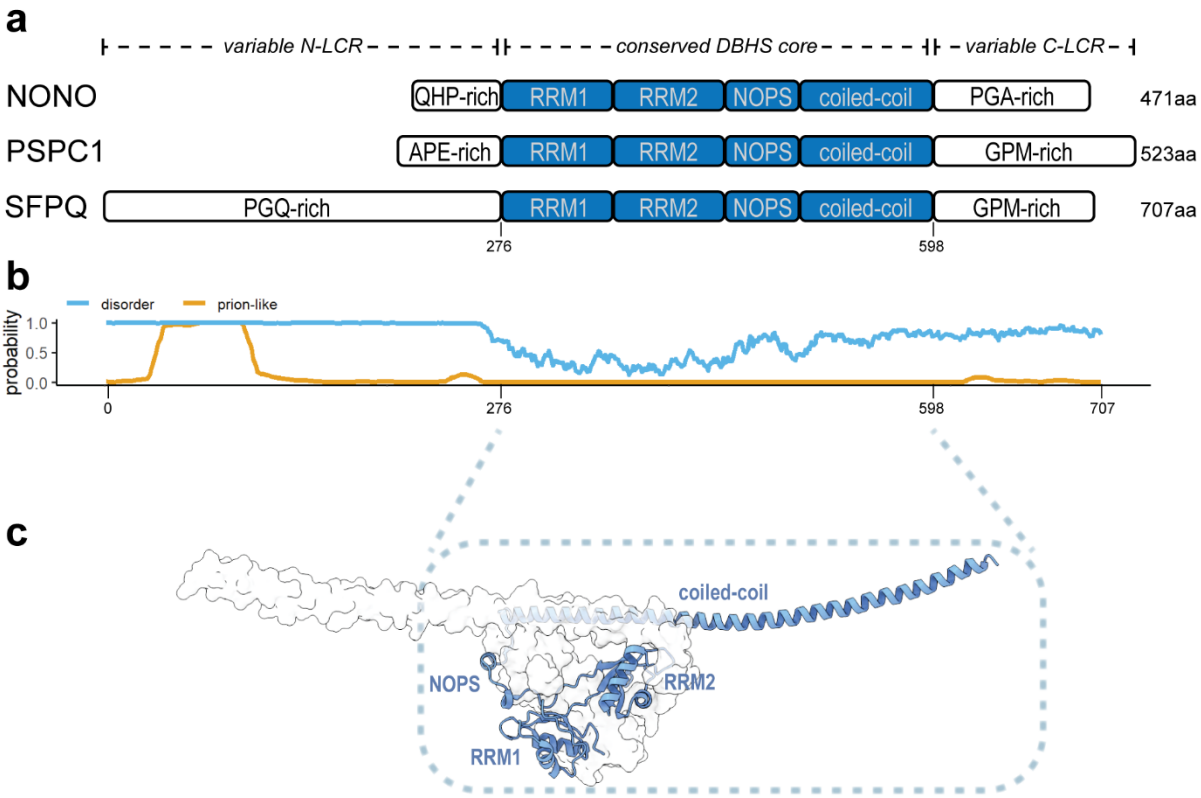



**Figure 1. SFPQ is a DBHS protein with an extended N-terminal prion-like low-complexity region. (a)** Schematic showing domain topologies of the three human DBHS proteins. The well-ordered folded domains of the DBHS core (blue) are flanked on either side by variable regions (white) enriched in specific amino acids, as indicated. Relative lengths of each region are shown to scale, highlighting the extended N-LCR of SFPQ. The total number of amino acids in each protein is indicated on the right. **(b)** *In silico* prediction of intrinsically disordered regions (IUPred2A; blue) and regions that are prion-like (PLAAC; orange) within SFPQ suggest that both LCRs are disordered, and that a large section of the N-LCR is prion-like. **(c)** Crystal structure of the homodimer formed by the core conserved DBHS region of SFPQ (PDB 4wij). Structural domains of one dimeric partner (shown in cartoon representation) are labelled. The other dimeric partner is shown in surface representation.

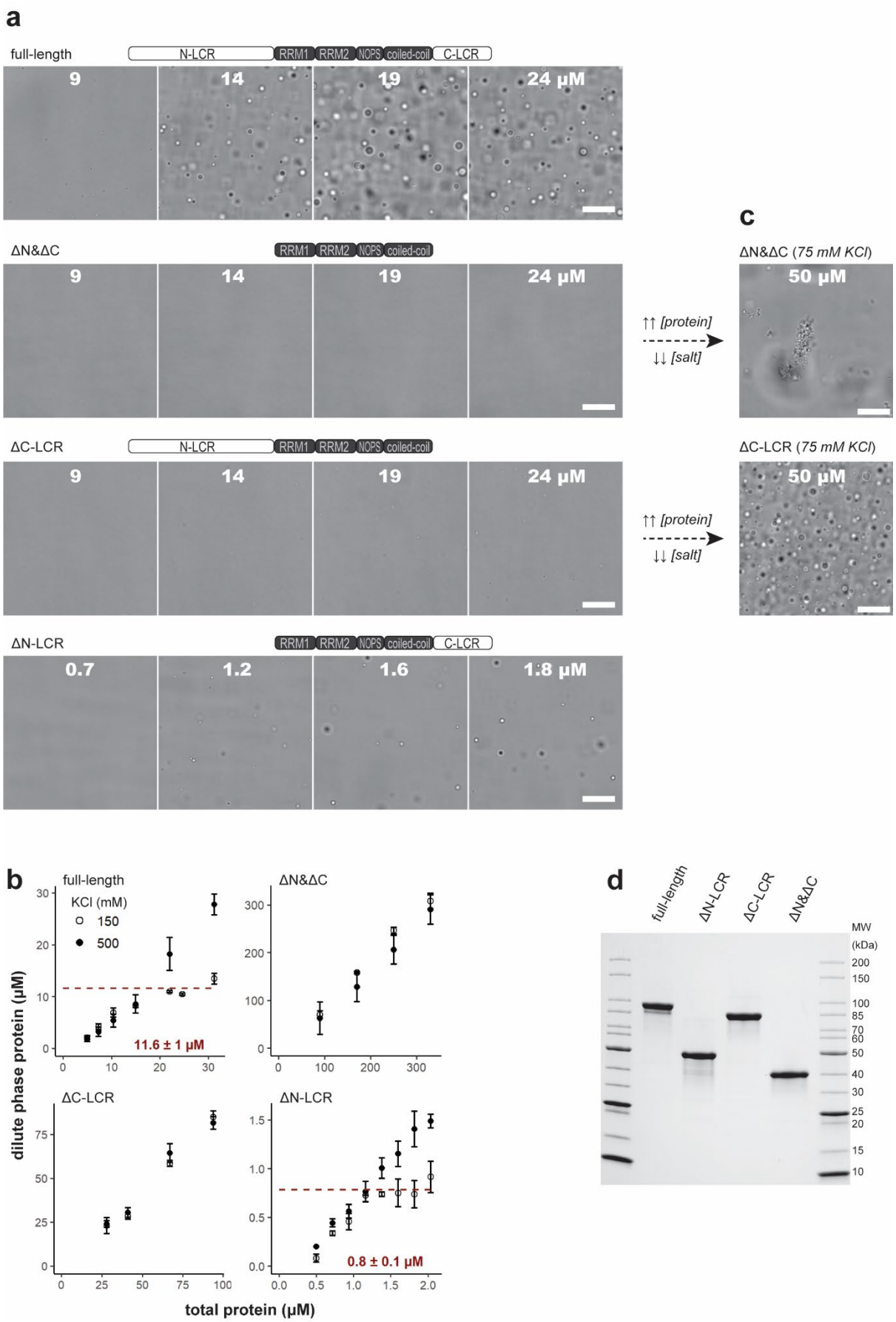

a
full-length
N-LCR
RRM1
RRM2
NOPS
coiled-coil
C-LCR
9
14
19
24 μM
ΔN&ΔC
ΔC-LCR
ΔN-LCR
0.7
1.2
1.6
1.8 μM
c
ΔN&ΔC (75 mM KCl)
50 μM
↑↑ [protein]
↓↓ [salt]
ΔC-LCR (75 mM KCl)
b
full-length
KCl (mM)
150
500
11.6 ± 1 μM
ΔN&ΔC
ΔC-LCR
ΔN-LCR
0.8 ± 0.1 μM
dilute phase protein (μM)
total protein (μM)
d
full-length
ΔN-LCR
ΔC-LCR
ΔN&ΔC
MW
(kDa)
200
150
100
85
70
60
50
40
30
25
20
15
10

**Figure 2. *In vitro* liquid demixing of SFPQ is facilitated by its LCRs. However, the shorter C-LCR is the main driver of condensation, whereas the longer prion-like N-LCR attenuates condensation of the full-length protein. (a)** DIC microscopy images show that, at physiological pH and salt concentration (150 mM KCl), both full-length SFPQ (top) and SFPQ ΔN-LCR (bottom) condense to form droplets in a concentration-dependent manner. SFPQ ΔN-LCR condenses at a much lower protein concentration than full-length SFPQ. Removal of both LCRs, or the C-LCR only, (middle panels) abolished droplet formation under these conditions. **(b)** By measuring protein concentration in the dilute phase after dilution to 150 mM KCl, the $C_{sat}$ was determined to be 11.6 μM for full-length SFPQ and 0.8 μM for SFPQ ΔN-LCR. For SFPQ ΔC-LCR and ΔN&ΔC, all proteins remained in the dilute phase, even at protein concentrations 8- and 28-fold greater than the $C_{sat}$ of full-length SFPQ, respectively. Dilutions in protein storage buffer (500 mM KCl) served as a negative control (solid circles). Data points were collected in triplicate; error bars show standard deviation; $C_{sat}$ uncertainty expressed as 95% confidence interval. **(c)** DIC images showing that at high protein concentrations in a solution containing 75 mM KCl, SFPQ ΔN&ΔC formed solid-like fibrous aggregates, whereas SFPQ ΔC-LCR condensed to form spherical liquid droplets. For all DIC images, the total protein concentration (μM) in each sample is indicated on the image. All scale-bars are 20 μm. **(d)** SDS-PAGE of recombinant full-length SFPQ and truncations after purification from bacterial cell culture (Coomassie-stained).

---

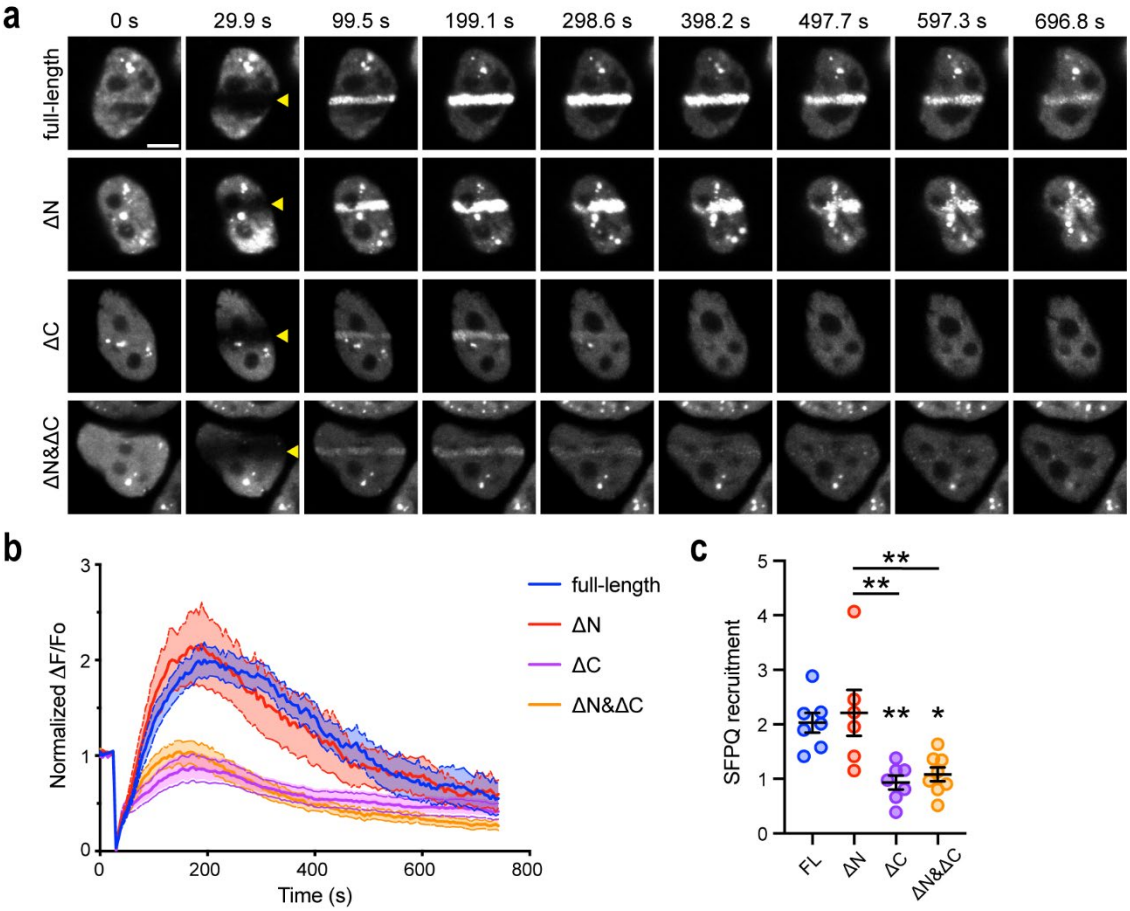



**Figure 3. Efficient recruitment of SFPQ to DNA damage sites *in vivo* requires its C-LCR, but not its N-LCR. (a)** HeLa cells overexpressing GFP-SFPQ, either full-length (FL), ΔN-LCR, ΔC-LCR or ΔN&ΔC-LCR, were subjected to laser microirradiation. Scale bar = 5 µm. **(b)** Traces showing relative changes in GFP fluorescent intensity at sites of DNA damage over time. Solid and dashed lines indicate means and SEM, respectively. FL, n = 7 cells; ΔN-LCR, n = 6 cells; ΔC-LCR, n = 7 cells; and ΔN&ΔC-LCR, n = 8 cells. **(c)** Average peak recruitment of GFP-SFPQ to DNA damage sites. Data are presented as means ± SEM. *p<0.05, **p<0.01 (one-way ANOVA with Tukey's multiple comparison test).

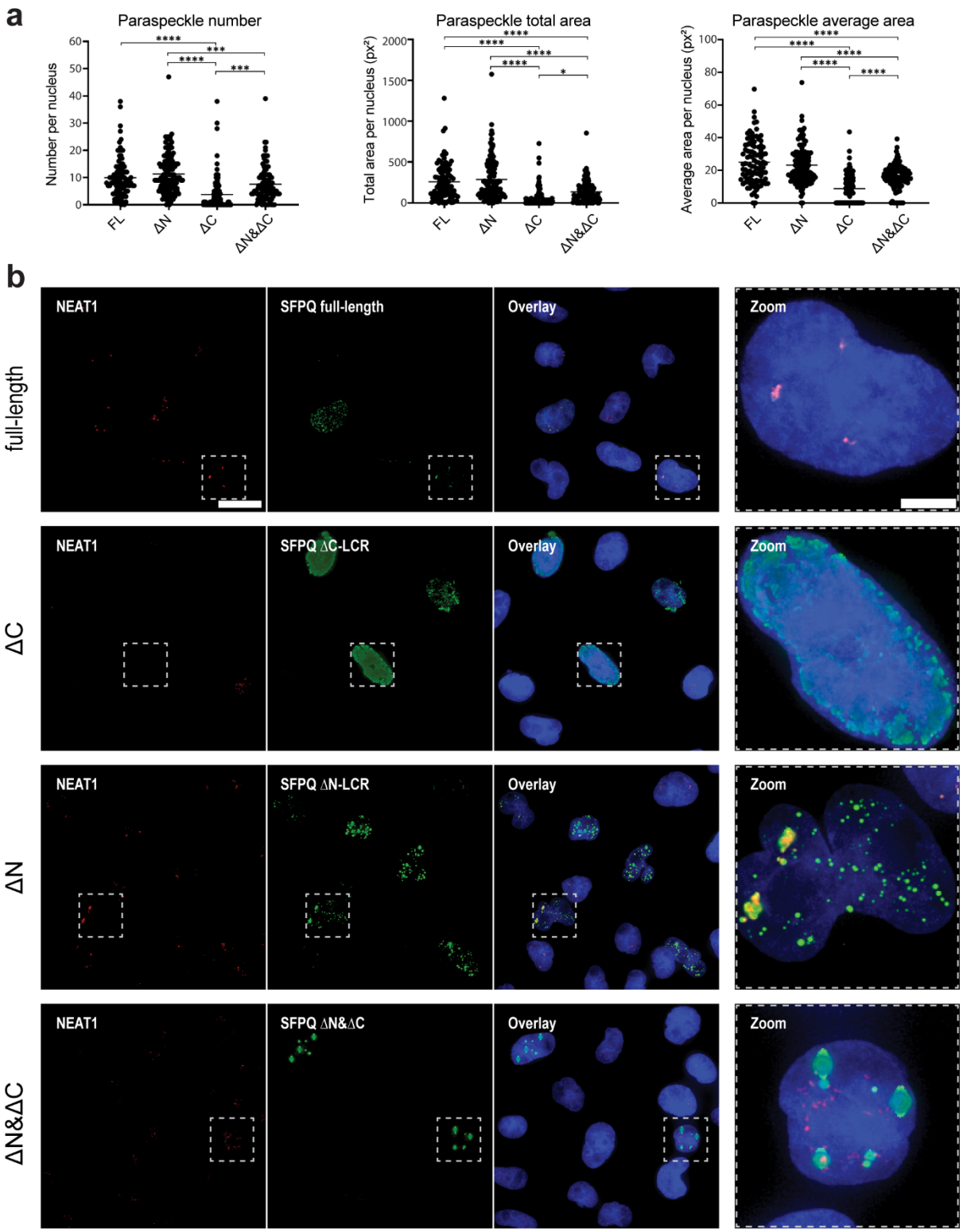

a
Paraspeckle number
Number per nucleus
Paraspeckle total area
Total area per nucleus (px²)
Paraspeckle average area
Average area per nucleus (px²)
FL
ΔN
ΔC
ΔN&ΔC
b
full-length
ΔC
ΔN
ΔN&ΔC
NEAT1
SFPQ full-length
SFPQ ΔC-LCR
SFPQ ΔN-LCR
SFPQ ΔN&ΔC
Overlay
Zoom

**Figure 4. Contrasting effects of over-expressing SFPQ LCR truncations on nuclear paraspeckle size and abundance. (a)** Quantification of paraspeckle size and number in cells expressing GFP-tagged SFPQ truncations shows that removing the C-LCR (Δ599-699) results in fewer and smaller paraspeckles, whereas ΔN-LCR (Δ1-275) results in no significant differences compared to full-length SFPQ. Expression of SFPQ ΔN&ΔC also reduces paraspeckle size and abundance, but not to the same extent as ΔC-LCR (**** $p<0.0001$, *** $p<0.001$ and * $p<0.05$ by ANOVA). **(b)** Representative fluorescence images of Hela cells overexpressing GFP-tagged full-length SFPQ, ΔC-LCR, ΔN-LCR and ΔN&ΔC. Removal of the N-LCR results in a large number of droplet-like non-paraspeckle puncta. Removal of the C-LCR results in a more diffuse nuclear localisation. Scale-bar = 20 µm for main panels; scale-bar = 5 µm for zoomed panels (right).

---

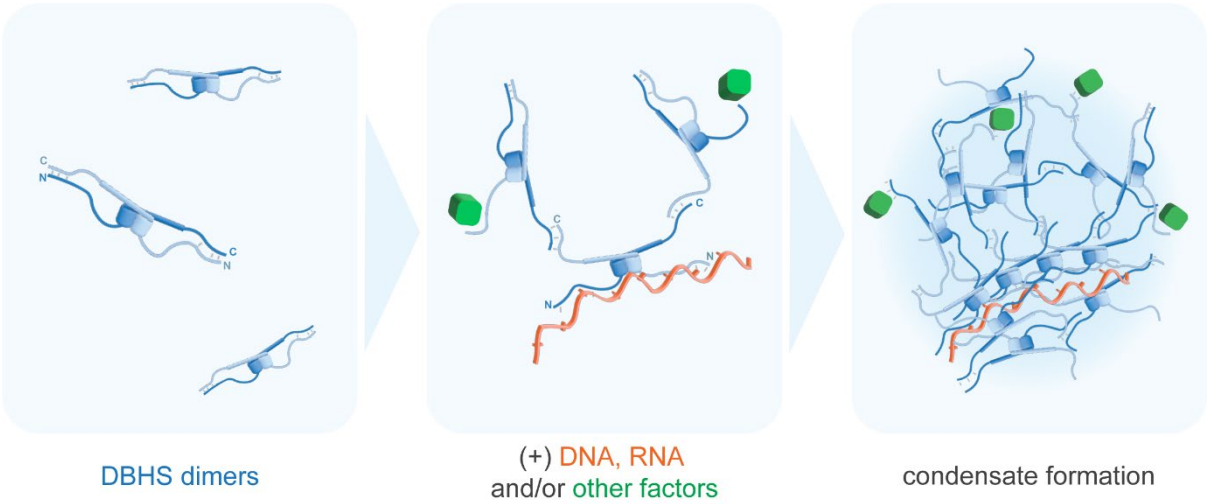



**Figure 5. Self-regulation of DBHS protein phase separation via competing intra- and intermolecular interactions between N- and C-terminal LCRs.** In the absence of other factors, direct intra-dimer interactions between the N- and C-LCRs preclude inter-dimer interactions, preventing demixing (left). Interactions with DNA, RNA and/or other factors sequester the N-LCR (middle), allowing the C-LCR to form productive inter-dimer interactions, nucleating phase separation to facilitate condensate formation (right).