

Phosphorylation regulates arginine-rich RNA-binding protein solubility and oligomerization

Sean R. Kundinger¹, Eric B. Dammer¹, Luming Yin¹, Cheyenne Hurst¹, Lingyan Ping¹, Duc M. Duong¹ and Nicholas T. Seyfried^{1,2*}

¹Department of Biochemistry, Emory University, Atlanta, GA, USA, ²Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA

*Address correspondence to: Nicholas T. Seyfried, Department of Biochemistry, Emory University School of Medicine, 505J Whitehead Biomedical Research Building, 615 Michael St., Atlanta, GA 30322, USA; e-mail: nseyfri@emory.edu; Phone: 1-404.712.8703

*Running title: Phosphorylation regulates SRSF2 aggregation

Keywords: RNA-binding proteins, phosphorylation, post-translational modifications (PTMs), mass spectrometry, protein interactions, SRSF2, SC35

ABSTRACT

Post-translational modifications (PTMs) within splicing factor RNA-binding proteins (RBPs), such as phosphorylation, regulate several critical steps in RNA metabolism including spliceosome assembly, alternative splicing and mRNA export. Notably, the arginine-/serine-rich (RS) domains in SR proteins are densely modified by phosphorylation compared with the remainder of the proteome. Previously, we showed that dephosphorylation of SRSF2 regulated increased interactions with similar arginine-rich RBPs U1-70K and LUC7L3. In this work, we dephosphorylated nuclear extracts using phosphatase *in vitro* and analyzed equal amounts of detergent-soluble and -insoluble fractions by mass spectrometry-based proteomics. Correlation network analysis resolved 27 distinct modules of differentially soluble nucleoplasm proteins. We found classes of arginine-rich RBPs that decrease in solubility following dephosphorylation and enrich to the insoluble pelleted fraction, including the SR protein family and the SR-like LUC7L RBP family. Importantly, increased insolubility was not observed across broad classes of RBPs. Phosphorylation regulated SRSF2 structure,

as dephosphorylated SRSF2 formed high molecular weight oligomeric species *in vitro*. Reciprocally, phosphorylation of SRSF2 by serine-/arginine protein kinase 2 (SRPK2) *in vitro* prevented high molecular weight SRSF2 species formation. Furthermore, we pharmacologically inhibited SRPKs in mammalian cells and observed increased cytoplasmic granules as well as the formation of cytoplasmic SRSF2 tubular structures that associate with microtubules by immunocytochemical staining. Collectively, these findings demonstrate that phosphorylation may be a critical modification that prevents arginine-rich RBP insolubility and oligomerization.

INTRODUCTION

RNA-binding proteins (RBPs) cooperatively engage both RNA and protein (1). RBPs frequently contain an RNA-binding domain, typically K-homology (KH) or RNA recognition motif (RRM) domains, that allow the RBP to achieve sequence-specific binding to target RNA molecules (2). Unbiased RNA interactome studies have identified many RBPs containing low complexity (LC) domains that participate in both RNA and protein interactions (2-5). LC domains are typically composed of a select few residues out of the entire amino acid

Phosphorylation regulates SRSF2 aggregation

code, giving rise to protein domains that are intrinsically disordered (6). However, LC RBPs exist in a dynamic continuum of native states that range from soluble monomers to liquid-liquid phase separated (LLPS) granules to insoluble fibrils (7) *in vitro* and *in vivo* (8). These assembly states are believed to be influenced in large part by RNA molecules (9,10) and post-translational modifications (PTMs) (11,12). Although we are beginning to decipher a “molecular grammar” regulating LLPS (13), the conditions that give rise to irreversible aggregation are incompletely known.

Recently it has been discovered that the progression of several neurodegenerative diseases is promoted by the aggregation of RBPs (14-21). Interestingly, LC domains are necessary for RBP LLPS and fibrillization (16,19,22,23), processes found to be regulated by PTM. LC RBPs are commonly modified by reversible PTM in the physiological milieu (13,24), yet in neurodegenerative disease phosphorylation PTMs increasingly occupy RBPs such as TDP-43 (25-27). It remains unclear whether phosphorylation is a trigger, or rather a consequence, of pathogenic RBP aggregation.

A major gap in our understanding of arginine-rich proteins is our inability to accurately measure absolute, site-specific phosphorylation levels. Recently our group used middle-down proteomic approaches to demonstrate arginine-rich RBPs have high steady-state levels of PTMs, particularly phosphorylation (28). One such group of arginine-rich RBPs with high levels of phosphorylation is the serine-/arginine-rich (SR) splicing factor family of RBPs (29). This twelve member RBP family is known to contain at least one RRM RNA-binding domain (30) at the N-terminus and a C-terminal arginine-/serine-rich (RS) domain distinguished by an expanded tract of RS dipeptide motifs, a phosphomotif conserved from yeast to man (3). The most extensively studied regulator of SR protein function is phosphorylation, primarily catalyzed by nuclear cdc2-like kinases (CLKs) (31) and cytoplasmic SR protein kinases (SRPKs) (32-34). Phosphorylation regulates nearly every facet of SR protein function, including splicing (35), coupling to sites of active transcription (36,37), subcellular localization (38-40), nuclear speckle

compartmentalization (31,32,41) and binding partner selection and affinity (30,39,42,43). Importantly, it is not fully understood whether excessive, or rather, insufficient phosphorylation alters the stability of SR proteins.

Our group (28) and others (40,44-46) suggest SR proteins may increasingly bind together and aggregate when insufficiently phosphorylated. Importantly, SR proteins and proteins that harbor homologous domains can aggregate under native conditions (47). Collectively, these data support a hypothesis that dephosphorylation would result in SR proteins becoming insoluble, as well as those RBPs with SR-like LC domains.

Here, we sought to understand the role of phosphorylation in regulating RBP solubility. We enriched for RBPs by biochemical fractionation from mammalian cell lines and incubated with calf intestinal alkaline phosphatase (CIP), which catalyzes the removal of phosphate PTMs from proteins (48). We conducted liquid-chromatography coupled with tandem mass spectrometry (LC-MS/MS) on detergent-soluble and -insoluble pellet fractions of dephosphorylated and mock-treated nucleoplasm extracts and used a network based approach to identify groups of RBPs that exhibited similar solubility changes that were regulated by phosphorylation. Importantly, we found that SRSF2 and related SR proteins co-aggregated to the insoluble fraction, while other nuclear RBPs such as TDP-43 did not. Moreover, we found that phosphorylation regulates SRSF2 assembly states *in vitro*. Finally, we show that SRPK inhibition in cells results in an increase in the number of cells harboring cytoplasmic SRSF2 granules as well as filamentous-like structures that co-localize with microtubules. Collectively, this work reinforces phosphorylation as an important regulator of SR protein solubility and structure and suggests that phosphorylation may be a preventative cellular mechanism against arginine-rich RBP aggregation.

RESULTS

Phosphorylation prevents SRSF2 aggregation

Phosphorylation regulates SRSF2 aggregation

Phosphorylation is a critical regulator of RNA-binding protein (RBP) protein function (45,49), condensation (12,50,51) and overall structure (52,53). The serine-/arginine-rich (SR) splicing factor family of RBPs is observed to be substantially phosphorylated, although a wide range of phosphorylation states are predicted for SR proteins *in vivo* (28). Recently, we reported that dephosphorylation of serine-/arginine-rich splicing factor 2 (SRSF2), a prototypical arginine-rich nuclear speckle phosphoprotein (28,33,54,55), enhanced interactions with fellow arginine-rich RBPs (28), leading us to hypothesize that phosphorylation regulates the solubility of structurally similar arginine-rich RBPs. Here, we use SRSF2 as a paradigm to study the regulation of arginine-rich RBP solubility, structure and morphology by phosphorylation. In SRSF2, the RS domain is highly phosphorylated (28), a region with high probability of intrinsic disorder (**Supplemental Fig.S1a**) (56,57). We incubated lysates containing recombinant SRSF2-myc, a known phosphoprotein, with calf intestinal alkaline phosphatase (CIP) which corresponded to a faster migrating SRSF2 band (**Supplemental Fig.S1b**) by SDS-PAGE, suggesting substantial dephosphorylation of SRSF2 (58,59). To further validate dephosphorylation of SRSF2, we immunoblotted with an antibody raised against the C-terminus of SRSF2 that preferentially labels hypophosphorylated SRSF2 (*hypo*SRSF2) (60-62) (**Supplemental Fig.1c**). We once again observed increased migration of SRSF2. Furthermore, we saw an increase in *hypo*SRSF2 labeling, demonstrating that SRSF2 is indeed dephosphorylated.

We then asked whether phosphorylation regulates the solubility of SRSF2. Detergent-soluble (S) and -insoluble pelleted (P) fractions were isolated following mock (-CIP) or phosphatase (+CIP) treatment (**Fig.1a**). We resolved equal amounts of total (T), soluble (S) and pelleted (P) fractions by SDS-PAGE and immunoblotted for SRSF2-myc (**Fig.1b**). While phosphorylated SRSF2-myc was primarily soluble (68% of total) in the mock condition, dephosphorylated SRSF2-myc significantly decreased in solubility, enriching to the

detergent-insoluble pellet fraction (89% of total, **Fig.1c**). This suggests similar arginine-rich RBPs or groups of RBPs may experience altered solubility following dephosphorylation as well. We next sought to globally identify and quantify RBPs that aggregate following dephosphorylation.

Proteomics reveals RBPs that aggregate following dephosphorylation

The soluble and pellet samples were analyzed by label-free quantitative proteomics, using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) (**Fig.1a, Supplemental Table S1**). Notably, dephosphorylation did not induce global aggregation of the nuclear proteome, as insoluble pellet fraction protein concentrations were unchanged after phosphatase incubation (**Supplemental Table S2**). Following database search and removal of proteins with >50% missing values, we identified 4,366 unique proteins (**Supplemental Table S3**).

To discover proteins with the largest changes in solubility following dephosphorylation, we calculated the log₂ fold differences of fraction insoluble values between phosphatase and mock treatments and visualized this as a volcano plot (**Fig.1d, Supplemental Tables S4-5**). Proteins were highlighted as enriched/depleted from a fraction if a greater than or equal to a two-fold change in fraction insoluble values was observed with a p value less than 0.05 (two-tailed paired t-test). Relatively few proteins (*n*=10) enriched to the soluble fraction following dephosphorylation. In contrast, many more proteins (*n*=734) experienced increased aggregation following dephosphorylation. Among these were SRSF2 and fellow SR proteins (SRSF3/4/5/6/7), as well as the SR-like SRRM1 and LUC7L RBP family. These observations suggest phosphorylation is an important PTM that regulates the solubility of SRSF2, as well as the solubilities of similar arginine-/serine-rich RBPs.

Phosphorylation regulates SRSF2 aggregation

Arginine-/lysine-rich RNA-binding proteins with positive net charge preferentially aggregate following dephosphorylation

Recently, the Jedd group generated a synthetic, recombinant arginine-rich low complexity protein equally balanced in positively-charged and negatively-charged residues in a dipeptide repeat pattern, and found that increasingly incorporating positively-charged residues resulted in increased aggregation of the recombinant protein (44). This suggests that increased net positive charge by virtue of increased arginine and lysine content may increase the condensation of low complexity RBPs. Importantly, phosphorylation significantly alters the net charge of a protein, adding a -2 charge with each phosphorylated residue at physiological pH (63). We hypothesized that proteins with highly positive net charge (high densities of arginine/lysine) may be predisposed to aggregate when not sufficiently phosphorylated.

We generated separate histograms of net charge per residue (NCPR) values for the nuclear proteome sequenced ($n=4,366$ proteins) and of those, the top 10% most insoluble proteins ($n=435$ proteins) following dephosphorylation (**Fig.1e**). The average NCPR of the top 10% most insoluble proteins (0.028) was significantly increased from the proteome studied (0.009) (p value = $8.2e-15$, Welch Two Sample t-test). The top 10% most insoluble proteins exhibited a bimodal distribution of NCPR, with a shoulder of proteins with a highly positive mean NCPR value (0.076). These data suggest proteins with a high NCPR (> 0.076) are susceptible to increased aggregation following dephosphorylation. We then plotted the distribution of NCPR of the nucleoplasm proteome sequenced and highlighted those proteins with $NCPR > 0.076$ (**Fig.1f**). Interestingly, 11/12 members of the SR protein family surpass this threshold, with SRSF2 being the most positively charged overall. The SR RBP family has an abnormally high positive average net charge, relative to other RBP families (**Supplemental Table S6**). SRSF2, as well as SRSF3/4/5/6/7 all experienced significantly aggregation following dephosphorylation. This suggests that phosphorylation may be an

especially important mechanism to regulate the solubility of proteins with high concentrations of arginine and lysine, among which are many nuclear RBPs involved in splicing.

Systems analysis identifies modules of proteins with solubility impacted by phosphorylation

We hypothesized that if we applied systems biology approaches to the protein abundance data we collected, we could discover groups of structurally-similar proteins that may co-aggregate when not sufficiently phosphorylated. To test this, we performed Weighted Gene Correlation Network Analysis (WGCNA) (64) to group proteins with highly correlated soluble and insoluble fraction abundance patterns. To define functionally divergent protein groups, we plotted a dendrogram that was segregated by hierarchical clustering into modules of related proteins (**Supplemental Fig.S2**). The network reduced our proteome into 27 modules [rank ordered by size, M1 (largest) – M27 (smallest)] each assigned a representative color (**Supplemental Table S7**). Each module was classified by the strength of associations to GO terms linked to discrete and generalizable cellular functions.

To understand which module of proteins experienced the greatest alterations in solubility following dephosphorylation, we performed a one-dimensional hypergeometric Fisher's exact test (FET) for enrichment within each module of those proteins with the most increased insolubility following CIP-treatment ($n=734$) (**Fig.2a**). Several key modules (M5, M7, M16) were enriched with proteins that aggregated following dephosphorylation. To identify protein drivers behind module solubility changes, eigenprotein values were plotted according to protein fraction and treatment condition (**Fig.2b-c**). Eigenproteins are defined as the first principal component of a module and serve as a representative, weighted module expression profile. As expected, some modules did not experience appreciable solubility changes following dephosphorylation. Among these were the soluble module M1 (protein folding) and

Phosphorylation regulates SRSF2 aggregation

insoluble module M9 (mRNA splicing), both unchanged in solubility profile after dephosphorylation (**Fig.2b**). Other modules, however, including M5 (dephosphorylation), M7 (cell cycle phase) and M16 (DNA repair) represented a prominent group of modules comprised of proteins that are soluble when phosphorylated, yet aggregate upon dephosphorylation (**Fig.2c**). The three members of the SR-like LUC7L family were each hub proteins of the M7 module, which experienced a large decrease in solubility. SRSF2 was a member of the M16 module, which interestingly also contained several cytoskeletal components (**Supplemental Table S7**). Given the robust insolubility of individual modules following dephosphorylation, we sought to validate individual proteins that significantly changed in solubility as well.

Confirmation of hub protein solubility changes following dephosphorylation

To examine the effect of dephosphorylation on the solubility of individual RBPs, we plotted the mass spectrometry protein abundance measurements (**Fig.3a**) and compared with immunoblot signals of select endogenous proteins from total, soluble and pelleted fractions (**Fig.3b**). The classical SR proteins SRSF1 and SRSF2 were depleted from the soluble fraction which was verified by western blot. LUC7L and LUC7L3, both hubs of the M7 ‘cell cycle phase’ module, were significantly depleted from soluble fractions and enriched to insoluble pellet fractions. Interestingly, the U1 small nuclear ribonucleoprotein C (SNRNPC), a member of the U1 snRNP subunit of the spliceosome, experienced increased solubility following dephosphorylation. Other nuclear RBPs in HEK293 nucleoplasm lysates including TDP-43 (M5), hnRNPAB (M21) and ZC3H18 (M9), however, showed no significant solubility changes following dephosphorylation. Mass spectrometry examination of soluble and pellet fractions revealed global RBP solubility changes in response to phosphatase co-incubation, revealing RBPs susceptible to destabilization following dephosphorylation.

Phosphorylation decreases NCPR and regulates the oligomerization of arginine-rich SRSF2

With no consideration of post-translational modifications (PTMs), SRSF2 is among the most positively charged proteins in the entire proteome (**Supplemental Table S8**). In a theoretical exercise of how phosphorylation changes the net charge of the SRSF2, we calculated the average SRSF2 NCPR (**Fig.4a**) and local charge density in a 21 residue sliding window range (**Fig.4b**) assuming three separate states: no phosphorylation (*hypo*SRSF2), MS-observed phosphorylation (pSRSF2+) (28), and full phosphorylation (*hyper*pSRSF2) (**Supplemental Table S9**). Although substantially positively-charged within the RS domain, SRSF2 adopts increasing negative charge with increasing phosphorylation, such that the RS domain becomes net negatively-charged at full phosphorylation occupancy. As we observed increased SRSF2 aggregation upon dephosphorylation, we asked whether phosphorylation could similarly regulate the oligomerization of recombinant SRSF2.

Given the aggregation we observed for SRSF2, we hypothesize that a critical structural change occurs as a result of dephosphorylation. To test this, we analyzed phosphorylated and dephosphorylated lysates containing SRSF2-myc by both denaturing SDS-PAGE and non-denaturing Blue native PAGE followed by western blotting for the myc-tag of recombinant SRSF2 (**Fig.4c-d**). In contrast to SDS-PAGE, which separates proteins under denaturing conditions, native PAGE resolves native protein masses in high molecular weight oligomeric states as well as protein complexes formed by physiological protein-protein interactions (65). The mock-treated SRSF2-myc sample had a recognizable monomer species band at ~ 37 kDa, whereas dephosphorylated SRSF2 exhibited dimer and tetramer species formation, as well as high molecular weight oligomeric species (**Fig.4d**). These data suggest that dephosphorylation leads to higher molecular weight oligomeric SRSF2 structures.

Phosphorylation regulates SRSF2 aggregation

To reciprocally demonstrate how SRSF2 structure changes with phosphorylation, we performed an *in vitro* kinase reaction using the SR protein kinase 2 (SRPK2) (**Fig.4e-f**) (66). Both SRPK2 and SRSF2 were expressed and purified in *E. coli*, an organism with a low abundance of phosphoproteins and only three known Ser/Thr-directed kinases (67,68). Despite not observing an overt change in molecular mass via SDS-PAGE following phosphorylation (**Supplemental Fig.S3a**), the increased migration of SRSF2 by isoelectric focusing (IEF) (**Supplemental Fig.S3b**) is consistent with reduced net charge and increased phosphorylation. We then separated these reaction products by non-denaturing native PAGE and found that unphosphorylated SRSF2 formed high molecular weight oligomers, with a prominent high molecular weight band at ~720 kDa (**Fig.4f**). In contrast, SRPK2-phosphorylated SRSF2 existed primarily as monomeric species. These data demonstrate that phosphorylation regulates SRSF2 homotypic interactions and prevents high molecular weight oligomeric species formation.

SRPK inhibitor SRPIN340 decreases SR protein phosphorylation and increases SRSF2 granule and tubule formation

Having established that SRSF2 aggregates and forms high molecular weight oligomers upon dephosphorylation *in vitro*, we attempted to inhibit SR protein phosphorylation in cells. We incubated HEK293 cells in media containing the compound SRPIN340, a cell-permeable, ATP-competitive selective SRPK inhibitor (69). To validate SRPIN340 we analyzed cell lysates by western blot using mAb104 (70), an antibody that labels phosphoepitopes on numerous SR proteins (71). We also co-labeled with an antibody that recognizes phosphorylated SRSF2 (pSRSF2) epitopes called SC35 (54), a synonym for the SRSF2 gene product protein (**Fig.5a**). Quantification of mAb104 immunoblot signal intensities normalized to Histone H3 demonstrate that several SR proteins, particularly SRSF2, exhibit decreased phosphorylation with increasing concentrations of SRPIN340 (**Fig.5b**). Independently, we compared SC35 labeling of

pSRSF2 with an antibody anti-phosphoTDP-43 (pTDP-43). We found that while pTDP-43 did not quantitatively change with SRPIN340 treatment, pSRSF2 signal was significantly decreased at 50 μ M SRPIN340 (**Fig.5c**). These results demonstrate that inhibition of SRPKs using SRPIN340 successfully reduces pSRSF2 levels, as well as overall SR protein phosphorylation.

As we observed high molecular weight species of dephosphorylated SRSF2 *in vitro*, we hypothesized that inhibiting SR protein kinases within cells would induce an increase in cytoplasmic granules and/or the formation of fibril-like species, hallmarks of various RBP proteinopathies (72). Immunocytochemical (ICC) staining using the nuclear speckle antibody SC35 demonstrated canonical nuclear speckle morphology of pSRSF2 in vehicle-treated cells (**Fig.5d**). In cells incubated with SRPIN340, we observed unusual cytoplasmic granules (**Fig.5e**). We performed quantified these and observed a significant increase in the average number of cytoplasmic granules per cell in the SRPIN340 condition (**Fig.5f**). We next labeled cells using the *hypo*SRSF2 antibody, observing dephosphorylated SRSF2 dispersed largely throughout the cytoplasm and infrequently observed as small tubules extending from the cell body outwards (**Fig.5g**). In cells incubated with SRPIN340, however, we observed a significant increase in cells harboring cytoplasmic SRSF2 tubule structures, which were substantially larger in size (**Fig.5h-i**). As expected, granule condensation and formation of tubule-like filaments was not observed for TDP-43 which showed a consistent pattern of nuclear distribution in both conditions (**Fig.5j-k**).

SRSF2 interacts with microtubule subunit proteins α - and β -tubulin

Given the unusual tubular morphology we observed of hypophosphorylated SRSF2 in cells and enhanced aggregation of SRSF2 following dephosphorylation *in vitro*, we further examined constituents of M16, of which SRSF2 is a member (**Fig.6a**). Interestingly, several proteins within this module are components of

Phosphorylation regulates SRSF2 aggregation

the cytoskeleton, including microtubule subunit protein β -tubulin 8 (TUBB8). To answer whether SRSF2 interacts with microtubules, we immunopurified recombinant SRSF2 and immunoblotted for TUBB8 and TUBA1A. We indeed found that SRSF2 interacts with TUBB8 and TUBA1A (**Fig.6b**). Furthermore, ICC staining of SRSF2 with either TUBA1A or TUBB8 demonstrates that dephosphorylated SRSF2 preferentially associates with microtubule structures (**Fig.6c-d**). Thus, network-based proteomics revealed an association between arginine-rich RBPs like SRSF2 and microtubules, confirmed by both biochemical and cell imaging approaches.

DISCUSSION

Here we used a mass spectrometry-based proteomics approach to investigate how phosphorylation affects protein solubility and oligomerization. Using systems biology analyses we identified modules of proteins with shared biology and sequence homology that co-aggregate following dephosphorylation. Arginine-rich proteins, including SRSF2, were among the proteins with the most decreased solubility following dephosphorylation. SRSF2 was used as a paradigm to investigate the relationship between phosphorylation and protein structure and solubility. We discovered that dephosphorylation regulates higher order multimer formation of SRSF2, whereas phosphorylated SRSF2 exists largely as a monomer species *in vitro*. Inhibition of SR protein kinases within mammalian cells decreased SR protein phosphorylation, resulting in increased cytoplasmic SRSF2 granule formation, as well as cytoplasmic tubule SRSF2 structures which co-localize with cytoskeletal proteins.

SRSF2 is a well-studied classical SR protein with integral roles in constitutive and alternative splicing (30). This study highlights numerous behaviors of SRSF2 that are associated with RBP dysregulation in neurodegenerative disease, including decreased solubility, high molecular weight oligomer formation and cytoplasmic granule and tubule-like morphologies. Although mutations in SRSF2 are frequently observed in individuals with

myelodysplastic syndromes (MDS) or chronic myelomonocytic leukemia (CMML) (73-77), SRSF2 has not been commonly associated with human neurodegenerative disease. Recently, however, the McKnight group demonstrated that SRSF2 does indeed form condensates *in vitro*, a hallmark feature of RBPs that aggregate in neurodegenerative disease, which importantly was reversible by phosphorylation (50,78).

Collectively, these data suggest that phosphorylation tunes SRSF2 net charge, solubility and structure by virtue of multimer disassembly *in vitro* and *in vivo* (**Fig.7**). While it is generally considered that phosphorylation promotes protein aggregation, a notable finding of this study is that arginine-rich RBPs, such as the SR and LUC7L protein families, exhibit remarkably similar aggregation patterns when dephosphorylated. Indeed, our group and others have observed that arginine-rich splicing proteins mislocalize and aggregate in AD brain (14,15,79,80). As kinases are dysregulated in AD (25), SRSF2 is a protein susceptible to solubility dysregulation in disease. Remarkably, SRSF2 was identified as a novel protein that associates with phosphorylated tau in AD brain by proteomic examination of microdissected neurofibrillary tangles (81). Our group has not identified SRSF2 as a protein with significantly increased insolubility in AD (82). However, our studies have been limited to sarkosyl-insoluble fractions only. Future studies that include internal comparisons of insoluble to soluble fractions may implicate SRSF2 and similar RBPs that are depleted from AD brain homogenate soluble fractions, and may reveal novel proteins that undergo altered solubility in AD.

We highlight phosphorylation as a critical feature of RBPs that strongly influences protein solubility. An important consideration is the influence that similarly negatively charged RNA molecules may hold over the solubility of nuclear proteins. Future investigations that explore the role of RNA in regulating the solubility of RBPs may yield a parallel interpretation of RBP stability. Further still, studies that investigate the role of phosphorylation in the regulation of arginine-rich RBP aggregation may hold promise to reveal the mechanism underlying RBP aggregation in neurodegenerative diseases.

EXPERIMENTAL PROCEDURES

Materials

The primary antibodies used in this study include: rabbit polyclonal anti-myc (Cell Signaling Technology 2272S), rabbit polyclonal anti-SRSF2 (Abcam ab229473), mouse monoclonal anti-SRSF2 (Clone 1Sc-4F11, Millipore Sigma 04-1550), mouse monoclonal anti-pSRSF2 (Abcam ab11826), rabbit polyclonal anti-LUC7L (ThermoFisher 17085-1-AP), rabbit polyclonal anti-LUC7L3 (ThermoFisher PA5-53816), rat monoclonal anti-SNRPC (Abcam ab122901), rabbit polyclonal anti-SRSF1 (Abcam ab38017), rabbit polyclonal anti-TDP-43 (ProteinTech Group 10782-2-AP), rabbit anti-pS409-410 TDP-43 (CosmoBio, CAC-TIP-PTD-P02), rabbit polyclonal anti-hnRNPAB (Santa Cruz sc-98723), rabbit polyclonal anti-ZC3H18 (ThermoFisher PA5-59322), rabbit polyclonal anti-SRRM1 (ab221061), rabbit polyclonal anti-Histone H3 (Abcam ab1791), rat monoclonal anti-alpha-tubulin clone YL1/2 (MAB1864), rabbit polyclonal anti-TUBB8 (ab97880) and mouse IgM anti-pSR proteins (Clone mAb104, ATCC CRL-2067, see below isolation method). Each antibody save for mAb104 was used at a 1:1,000 dilution in blocking buffer for western blotting. Each antibody used for ICC was diluted 1:500 in normal horse serum (NHS). The pcDNA3.1-SC35-cMyc SRSF2 plasmid was a gift from Kathleen Scotto (83) (Addgene plasmid #44721).

mAb104 Isolation

Immortalized mouse hybridomas were purchased from ATCC (CRL-2067). Cells were thawed and then passaged twice before collection. Forty-eight hours post-passaging, the cell suspension media was aliquoted into a 15 ml conical tube then spun at 250xg for 5 min. The supernatant was aspirated and cells were resuspended in 2%FBS in DMEM. Cells were cultured for 3 more days then the cell suspension was aliquoted into a 15 mL conical tube then spun at 1,500xg for 15 min. The supernatant was transferred to a fresh 15 mL conical tube, then aliquoted into microcentrifuge tubes. The supernatant aliquot

volumes were then reduced under SpeedVac (Labconco 731022) to 50% of original volume and restored to the original volume with 100% glycerol. Tubes were then frozen at -20C. The antibody was diluted 1:5 in blocking buffer for western blotting.

Local Charge Density and NCPR

Residue charge at physiological pH (7.4) was calculated using a simple algorithm modeled from EMBOSS. The residues D/E were counted with charge = -1, residues K/R with charge = +1, and H with charge = +0.5. Residues S/T/Y were counted as either charge = 0 or -2, to mimic the charge assumed after phosphorylation post-translational modification at physiological pH (63). To generate the local charge density, the average charge over a window of 21 residues starting at the C-terminus was calculated, sliding towards the N-terminus of the protein. To calculate the Net Charge per Residue (NCPR) of full proteins, the residue charges, as calculated above, were summed and this number was divided by the number of residues in the protein. The NCPR of the average protein of the known human proteome is ~ +0.013, spanning -0.40 to 0.32 (minimum protein size = 100 residues).

Cell Culture and Transfection

HEK293 cells were cultured in Dulbecco's Modified Eagle Medium [DMEM, high glucose (Gibco)] supplemented with 10% (v/v) fetal bovine serum (Gibco) and 1% penicillin-streptomycin (Gibco), and maintained at 37°C under a humidified atmosphere of 5% (v/v) CO₂ in air. Cells were grown to 70-80% confluency in 10 cm² culture dishes and transfected with 10 µg SRSF2-myc plasmid and 30 µg linear polyethylenimine (PEI). Cells were harvested and fractionated to enrich for nucleoplasm as described below. For western blotting SRPIN340-inhibitor treated cells, HEK293 cells were incubated for 12 hours (84) with equivolume amounts of either DMSO (vehicle) or SRPIN340 (final concentration = 50 µM) starting 24 hours post-passage. Afterwards, cells were harvested in IP lysis buffer (50 mM HEPES pH7.4, 150 mM NaCl, 5% Glycerol, 1 mM EDTA, 0.5 (v/v) NP-40, 0.5% (v/v) CHAPS) and sonicated at 25% amplitude for 3 × 10 sec on/off

Phosphorylation regulates SRSF2 aggregation

cycles. Lysates were then cleared after a 15,600xg centrifugation step at 4°C. Supernatants were transferred to new tubes and run by SDS-PAGE followed by western blotting.

Nucleoplasm Enrichment

This cellular extraction procedure was adapted from the Gozani group (85) and modified to include NP-40 detergent. In short, forty eight hours post-transfection, HEK293 cells were rinsed with cold PBS then scraped in PBS+1xHALT protease/phosphatase inhibitor (ThermoFisher). The cell slurry was centrifuged for 5 min at 1,000xg at 4°C to pellet cells. The supernatant was aspirated and the cells were washed with 1 mL PBS+1xHALT (100 µL taken as total fraction) and centrifuged again to pellet cells. The supernatant was aspirated and the cells were swelled in 150 µL Hypotonic Lysis Buffer+1xHALT (10 mM HEPES pH 7.9, 20 mM KCl, 0.1 mM EDTA, mM dithiothreitol (DTT), 5% Glycerol, 0.5 mM PMSF, 10 µg/mL Aprotinin, 10 µg/mL Leupeptin, 0.1% NP-40) and incubated on ice for 5 min. The sample was then centrifuged for 10 min at 15,600xg at 4°C. The supernatant was collected as the cytoplasmic fraction, and the resulting pellet was incubated in 100 µL High Salt Buffer+1xHALT (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.5 mM PMSF, 10 µg/mL Aprotinin, 10 µg/mL Leupeptin) for 30 min on ice to extract nuclei. The samples were then sonicated for 5 sec at 25% amplitude and centrifuged at 18,213xg for 10 min at 4°C. The supernatant was collected as the nucleoplasm fraction, and the pellet (chromatin fraction) was resuspended and sonicated in Nuclei lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS). All fractions were frozen at -70°C, and only the cytoplasm and nucleoplasm fraction were used for following applications.

Calf Intestinal Phosphatase (CIP) Treatment

For dephosphorylation assays, nucleoplasm fractions (100 µg) were incubated with either mock (distilled water) or 50 Units of calf intestinal alkaline phosphatase (QuickCIP, NEB M0525L) for 1 hour at 37°C. Following this, Laemmli sample buffer (8% glycerol, 2% SDS, 50mM Tris pH 6.8, 3.25% beta-mercaptoethanol)

was added to each sample to 1X and boiled at 95°C for 10 minutes and run by SDS PAGE. For sedimentation assays, nucleoplasm fractions (100 µg) were incubated with either distilled water (mock) or 50 Units of calf intestinal alkaline phosphatase (QuickCIP, NEB M0525L) and brought up to 100 µL with PBS. The samples were heated at 37°C for 1 hour.

Sedimentation Assay

Following QuickCIP dephosphorylation, 50 µL of the 100 µL sample was added to polycarbonate ultracentrifuge tubes. Samples were spun at 100,000xg for 1 hour at 4°C. The supernatant (vol=50 µL) was transferred to a LoBind tube (Eppendorf 0030108442) and the insoluble pellet was resuspended in 50 µL 8M Urea and transferred to a separate LoBind tube. The pellet sample was sonicated for 1 sec on/off cycles at 25% amplitude until the pellet disappeared. For western blotting, 7.5 µL of total (pre-spin) fractions were added and 15 µL of soluble and pellet fractions were added. For mass spectrometry analysis 30 µL of soluble and pellet fractions were used.

Western Blotting

Western Blotting was performed according to standard protocol as previously described (28). In short, samples were boiled in Laemmli sample buffer (8% glycerol, 2% SDS, 50mM Tris pH 6.8, 3.25% beta-mercaptoethanol) for 10 minutes, then resolved on a Bolt® 4-12% Bis-tris gel (Invitrogen NW04120BOX) by SDS-PAGE and semi-dry transferred to a nitrocellulose membrane with the iBlot2 system (ThermoFisher IB21001). Membranes were blocked with TBS Starting Block Blocking Buffer (ThermoFisher 37542) and probed with primary antibodies (1:1,000 dilutions) overnight at 4°C. Membranes were then incubated with secondary antibodies conjugated to either Alexa Fluor 680 or 800 (Invitrogen) fluorophores for one hour at RT. Membranes were imaged using an Odyssey Infrared Imaging System (Li-Cor Biosciences) and band intensities were calculated using Odyssey imaging software.

Sample preparation for mass spectrometry analyses

Phosphorylation regulates SRSF2 aggregation

Thirty microliters of soluble and pellet fractions were normalized to 50 μ L with 8M urea buffer. Next, 10 mM Dithiothreitol (DTT) in 50 mM ammonium bicarbonate (ABC) was added to a final concentration of 1 mM and incubated for 30 min at room temperature (RT), then 50 mM iodoacetamide (IAA) in 50 mM ABC was added to a final concentration of 5 mM and incubated in the dark for 30 min at RT. Samples were then digested overnight with 1 μ g of Lys-C (Wako 121-05063). Following Lys-C digestion, samples were diluted to 1M urea and digested overnight with 1 μ g of Trypsin (ThermoFisher 90057). The next day samples were incubated with acidifying buffer [10% Formic Acid (FA), 1% trifluoroacetic acid (TFA)] and centrifuged for 2 min. Sample pH was verified as less than 3 using pH strips. Samples were desalted on an Oasis® PRIME HLB 10 mg plate (Oasis 186008053) and washes were flowed through columns by a 96-well Positive Pressure processor (Waters 186006961). Samples were washed first with methanol, then Buffer A (0.1% TFA). The digested samples were then loaded onto Oasis® PRIME HLB 10 mg plates (Oasis 186008053), washed with Buffer A twice, and then peptides were eluted with Buffer C [50% acetonitrile (ACN), 0.1% FA]. The elutant was lyophilized using a SpeedVac (Labconco 731022).

Mass spectrometry analysis

Lyophilized peptides were resuspended in loading buffer (0.1% FA, 0.03% TFA, 1% ACN) and separated on a self-packed C18 (1.9 μ m Dr. Maisch, Germany) fused silica column (20 cm \times 75 μ m internal diameter; New Objective, Woburn, MA) by a NanoAcquity UHPLC (Waters). Linear gradient elution was performed using Buffer A (0.1% formic acid, 0% acetonitrile) and Buffer B (0.1% formic acid, 80% acetonitrile) starting from 3% Buffer B to 40% over 100 min at a flow rate of 300 nl/min. Mass spectrometry was performed on an Orbitrap Fusion Lumos Mass Spectrometer in top speed mode. One full MS1 scan was collected followed by as many data-dependent MS/MS scans that could fit within a three second cycle. MS1 scans (400-1600 m/z range, 400,000 AGC, 50 ms maximum ion time) were collected in the Orbitrap at a resolution of 60,000 in profile mode

with FAIMS CV set at -45. The MS/MS spectra (1.6 m/z isolation width, 35% collision energy, 10,000 AGC) were acquired in the ion trap. Dynamic exclusion was set to exclude previous sequenced precursor ions for 30 seconds with a mass tolerance of 10 ppm.

Database Searching

Data files for the 16 samples were analyzed using MaxQuant v1.6.17.0 with Thermo Foundation 2.0 for RAW file reading capability. The search engine Andromeda (86) was used to build and search a concatenated target-decoy UniProt Knowledgebase (UniProtKB) containing both Swiss-Prot and TrEMBL human protein sequences (86,395 sequences, downloaded August 9, 2020) with 245 contaminant proteins as a parameter for the search (87). Methionine oxidation (+ 15.9949 Da), protein N-terminal acetylation (+ 42.0106 Da) and STY-phosphorylation (+ 79.966 Da) were included as variable modifications (up to 5 allowed per peptide); cysteine was assigned a fixed carbamidomethyl modification (+ 57.0215 Da). Fully tryptic peptides were considered with up to 2 miscleavages allowed in the search. A precursor mass tolerance of \pm 20 ppm was applied prior to mass accuracy calibration and \pm 4.5 ppm after internal MaxQuant calibration. The false discovery rate (FDR) for peptide spectral matches, proteins and site decoy fraction were all set to 1%. Phosphorylated peptides were not quantified. Protein group intensity was used for protein quantitation. The quantitation method did not consider reverse, contaminant, and by site only protein identifications, leaving 5,017 proteins for downstream analysis. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (88) with the dataset identifier PXD026894 on 6/23/2021.

Weighted Gene Correlation Network Analysis of Nuclear Fractionation Proteome

Prior to network analysis, the nucleoplasm proteome (5,017 proteins) was culled to select for proteins with less than or equal to 50% missing values (intensity values in 8 out of 16 samples; 4,366 proteins). The intensity values were then \log_2 -transformed. Missing protein intensity

Phosphorylation regulates SRSF2 aggregation

values were imputed using random numbers drawn from a normal distribution of observed protein intensity values of each sample (width = 0.3, down shift = 1.8, calculated separately for each column) in Perseus v1.6.15.0. The R package Weighted Gene Correlation Network Analysis (WGCNA) v1.68 was used to cluster proteins by abundance into groups of proteins with similar solubility patterns using a dissimilarity metric for clustering distance based on 1 minus the topology overlap matrix (1-TOM), a calculation based on an adjacency matrix of correlations of all pairs of proteins in the abundance matrix supplied to WGCNA (64). The weighted protein co-expression network was built using the \log_2 -transformed intensity values using the blockwiseModules function with the following parameters: soft threshold power beta = 20, deepSplit = 4, minimum module size = 15, merge cut height = 0.07, signed network partitioning about medioids respecting the dendrogram, and a reassignment threshold of $p = 0.05$. GO Elite v1.2.5 python package was used as previously published to categorize summary biological functions of individual modules (89,90). Z scores were determined by one-tailed Fisher's exact test (Benjamini-Hochberg FDR corrected) to demonstrate overrepresentation of ontologies in the nucleoplasm proteome of each module. The filters included a cutoff for Z scores as 1.96, P value cut off of 0.01 and a minimum of 5 genes per ontology.

Differential Solubility Analysis

Proteins either differentially soluble or differentially insoluble following dephosphorylation were calculated using a two-tailed paired t-test on the fraction insolubility values according to a previous study by our group (91):

$$\log_2[\text{insoluble}/(\text{insoluble} + \text{soluble})]_{+CIP} - \log_2[\text{insoluble}/(\text{insoluble} + \text{soluble})]_{\text{mock}}$$

Proteins were highlighted as enriched/depleted from a fraction if greater than or equal to a two-fold change in fraction insoluble values ($\pm \log_2(2)=1$) was observed with a p-Value less than 0.05 (two-tailed paired t-test).

SRPK2 in vitro kinase reaction

Human recombinant SRPK2 (0.4 ug/reaction; EMD Millipore 14-666; Glu46-end) and SRSF2 (0.8 ug/reaction; MyBioSource, Inc. MBS2029592; Thr14-end) expressed in *E. coli* were purchased and added to Kinase Buffer I (25 mM MOPS pH7.2, 12.5 mM β -glycerol phosphate, 25 mM $MgCl_2$, 5 mM EGTA, 2 mM EDTA, 0.25 mM DTT; Abcam ab189135). Either water (mock) or ATP was added to a final concentration of 100 μ M to a final volume of 25 μ L and incubated for 30 min at 30°C. Reactions were then separated by denaturing or non-denaturing Blue native Gel Electrophoresis or by Isoelectric focusing (IEF).

Blue Native Gel Electrophoresis

Blue native PAGE was carried out as described previously (16). Purified recombinant SRPK2 and SRSF2 proteins (0.4 and 0.8 μ g, respectively) were incubated with blue native gel loading buffer [5% glycerol, 50mM TCEP (Sigma-Aldrich 646547), 0.02% (w/v) Coomassie G-250 (Invitrogen BN2004), 1X NativePAGE Sample Buffer (Invitrogen BN2003)] in LoBind tubes (Eppendorf 0030108442) for 30 min at room temperature. Mock and CIP-treated nucleoplasm lysate (50 μ g) were incubated with the same blue native gel loading buffer and incubated on ice for 30 min. Samples and Blue Native Protein Ladder (Thermo LC0725) were loaded onto NativePAGE Bis-Tris Gels (3-12%, Invitrogen BN1001BOX) and gel electrophoresis was performed using anode NativePAGE Running Buffer (Invitrogen BN2001) and cathode buffer (Invitrogen BN2002) with additive (Invitrogen BN2004) and run for 15 min at 150V. The Dark Blue Cathode Buffer was then interchanged with Light Blue Cathode Buffer, proceeding for another 90 min at 150V. Following this, the gel was gently rocked in 50 mM Tris-HCl, pH7.5 and 1% SDS for 30 min and transferred using the semidry iBolt transfer system (Invitrogen) with PVDF membrane (Invitrogen IB24002) for 7 min at 20V. Following transfer, the PVDF membrane was rocked in 8% acetic acid solution for 5 min, rinsed with distilled water, air dried, then rinsed with methanol. The membrane was then blocked, western blotted and imaged on the Odyssey Infrared Imaging System (Li-Cor Biosciences).

Phosphorylation regulates SRSF2 aggregation

Isoelectric Focusing (IEF) Gel Electrophoresis

In vitro kinase reactions (10 μ L; 400 ng SRSF2, 200 ng SRPK2) were mixed with IEF Sample Buffer pH 3-10 (Fisher Scientific LC5311) and dH₂O. Samples were loaded onto Novex IEF pH 3-10 gels (ThermoFisher EC6655BOX) and gel electrophoresis was run with 1X IEF Anode Buffer (ThermoFisher LC5300) and 1X Cathode Buffer pH 3-100 (ThermoFisher LC5310). Protein marker ladder was included to benchmark protein pI (ThermoFisher 3921201). The gel was run by 100V constant for 1 hr, 200V constant for 1 hr, then 500V for 30 min. Following electrophoresis, the gel was fixed in 12% TCA containing 3.5% sulfosalicylic acid. After fixation, the gel was washed 3 x dH₂O. The gel was then stained overnight with Coomassie G-250.

Immunocytochemistry

Immunocytochemical staining was carried out as previously described (92). HEK293 cells were grown on Nunc Lab-Tek II Chamber Slide systems (ThermoFisher 154534PK). Twenty four hours post-passage, HEK293 cells were incubated with equivolume amounts of either DMSO (vehicle) or SRPIN340 (final concentration = 50 μ M). Cells were incubated at 37°C under a humidified atmosphere of 5% (v/v) CO₂ in air for 4 hours. Cell media was aspirated then washed 3 \times warm sterile PBS for 5 min each. Cells were then fixed with 4% paraformaldehyde (Electron Microscopy Sciences 15713S) diluted in PBS for 45 min at RT and afterwards washed 3 \times warm sterile PBS for 5 min each. Cells were then incubated with 0.05% Triton X-100 diluted in PBS for 20 min at RT. Afterwards, cells were blocked with 10% normal horse serum (NHS) in PBS for 45 min at RT. The blocking solution was aspirated and excess liquid was removed with a kimwipe. The cells were then incubated in primary antibody diluted in 2% NHS/1xPBS overnight at 4°C. The next day, cells were washed

3 \times warm sterile PBS for 5 min each. Cells were then incubated in secondary antibody diluted in 2% NHS/1xPBS for 1 hour at RT then washed 3 \times warm sterile PBS for 5 min each. PBS was aspirated and a droplet of DAPI-containing mounting media (Abcam ab104139) was placed on cells which were coverslipped and sealed with clear nail polish. Images were captured on a Keyence BZ-X810 widefield laser scanning microscope (Keyence). At least 14 images were taken for each condition per biological replicate at random areas of the slide.

Imaging Quantification and Statistical Analysis

Images were analyzed with FIJI. Graphs were developed with GraphPad Prism. Over 100 cells each were counted (DAPI+ nuclei) and analyzed per three biological replicates. Then, the number of SC35+ (pSRSF2) granules were counted in each image and divided by total cells observed. The average number of granules per cell per image ($n=14$) was collected for each biological replicate. At least 120 cells were counted in each condition per replicate. The number of cytoplasmic granules per cell was averaged for each replicate and conditions were statistically compared using a two-tailed paired t-test. To measure the percent of cells with cytoplasmic SRSF2 tubules, the total number of DAPI+ nuclei were counted as the number of cells per image. A total of 15 images per biological replicate were recorded, with four biological replicates analyzed. Then, the number of cells that contained an SRSF2 tubule morphology were counted. The number of SRSF2 tubule-containing cells were then divided by the number of DAPI+ nuclei and expressed as a decimal. The number of SRSF2 tubules per cell was averaged in each replicate and the vehicle and drug conditions were statistically compared using a two-tailed paired t-test.

ACKNOWLEDGMENTS

We thank Anita Corbett and Zachary T. McEachin for their helpful comments in discussion of this work.

Conflict of interest statement. None declared.

Author contributions: SK designed and carried out the experiments. DD, LP, LY and NTS advised experimental procedures. SK, ED and CH performed the data analyses. SK and ED did the computational analyses. SK drafted the manuscript and figures. SK and NS wrote and edited the manuscript. SK, ED, CH, LY, LP, DD and NTS reviewed and edited the manuscript. NTS carried out funding acquisition. All authors contributed to the article and approved the submitted version.

References

1. Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M. W. (2021) RNA-binding proteins in human genetic disease. *Nature Reviews Genetics* **22**, 185-198
2. Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**, 327-341
3. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsvelde, J., and Hentze, M. W. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406
4. Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., and Landthaler, M. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**, 674-690
5. Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., Curk, T., Krijgsvelde, J., and Hentze, M. W. (2016) Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**, 696-710
6. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**, 739-756
7. Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., and McKnight, S. L. (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767
8. St George-Hyslop, P., Lin, J. Q., Miyashita, A., Phillips, E. C., Qamar, S., Randle, S. J., and Wang, G. (2018) The physiological and pathological biophysics of phase separation and gelation of RNA binding proteins in amyotrophic lateral sclerosis and fronto-temporal lobar degeneration. *Brain Research* **1693**, 11-23
9. 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
10. Schwartz, Jacob C., Wang, X., Podell, Elaine R., and Cech, Thomas R. (2013) RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Reports* **5**, 918-925
11. Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., and Zweckstetter, M. (2017) Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature communications* **8**
12. Monahan, Z., Ryan, V. H., Janke, A. M., Burke, K. A., Rhoads, S. N., Zerze, G. H., O'Meally, R., Dignon, G. L., Conicella, A. E., Zheng, W., Best, R. B., Cole, R. N., Mittal, J., Shewmaker, F., and Fawzi, N. L. (2017) Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *The EMBO journal* **36**, 2951-2967
13. Wang, J., Choi, J. M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., and Hyman, A. A. (2018) A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 e616
14. Bai, B., Hales, C. M., Chen, P. C., Gozal, Y., Dammer, E. B., Fritz, J. J., Wang, X., Xia, Q., Duong, D. M., Street, C., Cantero, G., Cheng, D., Jones, D. R., Wu, Z., Li, Y., Diner, I., Heilman, C. J., Rees, H. D., Wu, H., Lin, L., Szulwach, K. E., Gearing, M., Mufson, E. J., Bennett, D. A., Montine, T. J., Seyfried, N. T., Wingo, T. S., Sun, Y. E., Jin, P., Hanfelt, J., Willcock, D. M., Levey, A., Lah, J. J., and Peng, J. (2013) U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proc Natl Acad Sci U S A* **110**, 16562-16567
15. Hales, C. M., Dammer, E. B., Deng, Q., Duong, D. M., Gearing, M., Troncoso, J. C., Thambisetty, M., Lah, J. J., Shulman, J. M., Levey, A. I., and Seyfried, N. T. (2016) Changes in the detergent-

Phosphorylation regulates SRSF2 aggregation

- insoluble brain proteome linked to amyloid and tau in Alzheimer's Disease progression. *Proteomics* **16**, 3042-3053
16. Bishof, I., Dammer, E. B., Duong, D. M., Kundinger, S. R., Gearing, M., Lah, J. J., Levey, A. I., and Seyfried, N. T. (2018) RNA-binding proteins with basic-acidic dipeptide (BAD) domains self-assemble and aggregate in Alzheimer's disease. *J Biol Chem* **293**, 11047-11066
 17. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (New York, N.Y.)* **314**, 130-133
 18. Neumann, M., Rademakers, R., Roeber, S., Baker, M., Kretzschmar, H. A., and Mackenzie, I. R. (2009) A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain : a journal of neurology* **132**, 2922-2931
 19. Liu-Yesucevitz, L., Lin, A. Y., Ebata, A., Boon, J. Y., Reid, W., Xu, Y. F., Kobrin, K., Murphy, G. J., Petrucelli, L., and Wolozin, B. (2014) ALS-linked mutations enlarge TDP-43-enriched neuronal RNA granules in the dendritic arbor. *J Neurosci* **34**, 4167-4174
 20. Vanderweyde, T., Yu, H., Varnum, M., Liu-Yesucevitz, L., Citro, A., Ikezu, T., Duff, K., and Wolozin, B. (2012) Contrasting pathology of the stress granule proteins TIA-1 and G3BP in tauopathies. *J Neurosci* **32**, 8270-8283
 21. Johnson, E. C. B., Carter, E. K., Dammer, E. B., Duong, D. M., Gerasimov, E. S., Liu, Y., Liu, J., Betarbet, R., Ping, L., Yin, L., Serrano, G. E., Beach, T. G., Peng, J., De Jager, P. L., Gaiteri, C., Bennett, D. A., Gearing, M., Wingo, T. S., Wingo, A. P., Lah, J. J., Levey, A. I., and Seyfried, N. T. (2021) Large-Scale Deep Multi-Layer Analysis of Alzheimer's Disease Brain Reveals Strong Proteomic Disease-Related Changes Not Observed at the RNA Level. *bioRxiv*, 2021.2004.2005.438450
 22. Furukawa, Y., Kaneko, K., Matsumoto, G., Kurosawa, M., and Nukina, N. (2009) Cross-seeding fibrillation of Q/N-rich proteins offers new pathomechanism of polyglutamine diseases. *J Neurosci* **29**, 5153-5162
 23. Waelter, S., Boeddrich, A., Lurz, R., Scherzinger, E., Lueder, G., Lehrach, H., and Wanker, E. E. (2001) Accumulation of mutant huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation. *Mol Biol Cell* **12**, 1393-1407
 24. Liu, Q., and Dreyfuss, G. (1995) In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol* **15**, 2800-2808
 25. Ping, L., Kundinger, S. R., Duong, D. M., Yin, L., Gearing, M., Lah, J. J., Levey, A. I., and Seyfried, N. T. (2020) Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci Data* **7**, 315
 26. Hasegawa, M., Arai, T., Nonaka, T., Kametani, F., Yoshida, M., Hashizume, Y., Beach, T. G., Buratti, E., Baralle, F., Morita, M., Nakano, I., Oda, T., Tsuchiya, K., and Akiyama, H. (2008) Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann Neurol* **64**, 60-70
 27. Neumann, M., Kwong, L. K., Lee, E. B., Kremmer, E., Flatley, A., Xu, Y., Forman, M. S., Troost, D., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2009) Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. *Acta neuropathologica* **117**, 137-149
 28. Kundinger, S. R., Bishof, I., Dammer, E. B., Duong, D. M., and Seyfried, N. T. (2020) Middle-Down Proteomics Reveals Dense Sites of Methylation and Phosphorylation in Arginine-Rich RNA-Binding Proteins. *J Proteome Res*
 29. Fu, X. D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA (New York, N.Y.)* **1**, 663-680
 30. Tacke, R., and Manley, J. L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO journal* **14**, 3540-3551

Phosphorylation regulates SRSF2 aggregation

31. Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., and Duncan, P. I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *The EMBO journal* **15**, 265-275
32. Gui, J. F., Lane, W. S., and Fu, X. D. (1994) A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* **369**, 678-682
33. Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C., and Fu, X. D. (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *The Journal of cell biology* **140**, 737-750
34. Kannan, N., and Neuwald, A. F. (2004) Evolutionary constraints associated with functional specificity of the CMGC protein kinases MAPK, CDK, GSK, SRPK, DYRK, and CK2alpha. *Protein Sci* **13**, 2059-2077
35. Mermoud, J. E., Cohen, P., and Lamond, A. I. (1992) Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res* **20**, 5263-5269
36. Misteli, T., Cáceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *The Journal of cell biology* **143**, 297-307
37. Misteli, T., Cáceres, J. F., and Spector, D. L. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523-527
38. Botti, V., McNicoll, F., Steiner, M. C., Richter, F. M., Solovyeva, A., Wegener, M., Schwich, O. D., Poser, I., Zarnack, K., Wittig, I., Neugebauer, K. M., and Muller-McNicoll, M. (2017) Cellular differentiation state modulates the mRNA export activity of SR proteins. *J Cell Biol* **216**, 1993-2009
39. Yeakley, J. M., Tronchère, H., Olesen, J., Dyck, J. A., Wang, H. Y., and Fu, X. D. (1999) Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *The Journal of cell biology* **145**, 447-455
40. Aubol, B. E., Serrano, P., Fattet, L., Wuthrich, K., and Adams, J. A. (2018) Molecular interactions connecting the function of the serine-arginine-rich protein SRSF1 to protein phosphatase 1. *J Biol Chem* **293**, 16751-16760
41. Aubol, B. E., Keshwani, M. M., Fattet, L., and Adams, J. A. (2018) Mobilization of a splicing factor through a nuclear kinase-kinase complex. *The Biochemical journal* **475**, 677-690
42. Xiao, S. H., and Manley, J. L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & development* **11**, 334-344
43. Aubol, B. E., Hailey, K. L., Fattet, L., Jennings, P. A., and Adams, J. A. (2017) Redirecting SR Protein Nuclear Trafficking through an Allosteric Platform. *Journal of molecular biology* **429**, 2178-2191
44. Greig, J. A., Nguyen, T. A., Lee, M., Holehouse, A. S., Posey, A. E., Pappu, R. V., and Jedd, G. (2020) Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol Cell*
45. Cho, S., Hoang, A., Sinha, R., Zhong, X. Y., Fu, X. D., Krainer, A. R., and Ghosh, G. (2011) Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S A* **108**, 8233-8238
46. Peng, T. Y., Lee, K. R., and Tarn, W. Y. (2008) Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *Febs j* **275**, 4152-4163
47. Nikolakaki, E., Drosou, V., Sanidas, I., Peidis, P., Papamarcaki, T., Iakoucheva, L. M., and Giannakouros, T. (2008) RNA association or phosphorylation of the RS domain prevents aggregation of RS domain-containing proteins. *Biochimica et biophysica acta* **1780**, 214-225
48. Robison, R. (1923) The Possible Significance of Hexosephosphoric Esters in Ossification. *The Biochemical journal* **17**, 286-293

Phosphorylation regulates SRSF2 aggregation

49. Fung, P. A., Labrecque, R., and Pederson, T. (1997) RNA-dependent phosphorylation of a nuclear RNA binding protein. *Proc Natl Acad Sci U S A* **94**, 1064-1068
50. Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S. L. (2014) Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139-1145
51. Wang, A., Conicella, A. E., Schmidt, H. B., Martin, E. W., Rhoads, S. N., Reeb, A. N., Nourse, A., Ramirez Montero, D., Ryan, V. H., Rohatgi, R., Shewmaker, F., Naik, M. T., Mittag, T., Ayala, Y. M., and Fawzi, N. L. (2018) A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *The EMBO journal* **37**
52. Xiang, S., Gapsys, V., Kim, H. Y., Bessonov, S., Hsiao, H. H., Mohlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., Luhrmann, R., de Groot, B., and Zweckstetter, M. (2013) Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure (London, England : 1993)* **21**, 2162-2174
53. Díaz-Moreno, I., Hollingworth, D., Frenkiel, T. A., Kelly, G., Martin, S., Howell, S., García-Mayoral, M., Gherzi, R., Briata, P., and Ramos, A. (2009) Phosphorylation-mediated unfolding of a KH domain regulates KSRP localization via 14-3-3 binding. *Nature Structural & Molecular Biology* **16**, 238-246
54. Fu, X. D., and Maniatis, T. (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437-441
55. Saitoh, N., Spahr, C. S., Patterson, S. D., Bubulya, P., Neuwald, A. F., and Spector, D. L. (2004) Proteomic Analysis of Interchromatin Granule Clusters. *Molecular Biology of the Cell* **15**, 3876-3890
56. Jones, D. T., and Cozzetto, D. (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857-863
57. Haynes, C., and Iakoucheva, L. M. (2006) Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res* **34**, 305-312
58. Keshwani, M. M., Aubol, B. E., Fattet, L., Ma, C. T., Qiu, J., Jennings, P. A., Fu, X. D., and Adams, J. A. (2015) Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *The Biochemical journal* **466**, 311-322
59. Aubol, B. E., Wu, G., Keshwani, M. M., Movassat, M., Fattet, L., Hertel, K. J., Fu, X. D., and Adams, J. A. (2016) Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol Cell* **63**, 218-228
60. Saitoh, N., Sakamoto, C., Hagiwara, M., Agredano-Moreno, L. T., Jiménez-García, L. F., and Nakao, M. (2012) The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. *Molecular Biology of the Cell* **23**, 1115-1128
61. Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stévenin, J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA (New York, N.Y.)* **5**, 468-483
62. Kadri, F., Pacifici, M., Wilk, A., Parker-Struckhoff, A., Del Valle, L., Hauser, K. F., Knapp, P. E., Parsons, C., Jeansonne, D., Lassak, A., and Peruzzi, F. (2015) HIV-1-Tat Protein Inhibits SC35-mediated Tau Exon 10 Inclusion through Up-regulation of DYRK1A Kinase. *J Biol Chem* **290**, 30931-30946
63. Cooper, J. A., Sefton, B. M., and Hunter, T. (1983) Detection and quantification of phosphotyrosine in proteins. *Methods Enzymol* **99**, 387-402
64. Langfelder, P., and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559
65. Wittig, I., Braun, H. P., and Schägger, H. (2006) Blue native PAGE. *Nat Protoc* **1**, 418-428
66. Edmond, V., Moysan, E., Khochbin, S., Matthias, P., Brambilla, C., Brambilla, E., Gazzeri, S., and Eymin, B. (2011) Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. *The EMBO journal* **30**, 510-523

67. Enami, M., and Ishihama, A. (1984) Protein phosphorylation in *Escherichia coli* and purification of a protein kinase. *J Biol Chem* **259**, 526-533
68. Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., and Mann, M. (2008) Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol Cell Proteomics* **7**, 299-307
69. Fukuhara, T., Hosoya, T., Shimizu, S., Sumi, K., Oshiro, T., Yoshinaka, Y., Suzuki, M., Yamamoto, N., Herzenberg, L. A., Herzenberg, L. A., and Hagiwara, M. (2006) Utilization of host SR protein kinases and RNA-splicing machinery during viral replication. *Proc Natl Acad Sci U S A* **103**, 11329-11333
70. Roth, M. B., Murphy, C., and Gall, J. G. (1990) A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *The Journal of cell biology* **111**, 2217-2223
71. Hatcher, J. M., Wu, G., Zeng, C., Zhu, J., Meng, F., Patel, S., Wang, W., Ficarro, S. B., Leggett, A. L., Powell, C. E., Marto, J. A., Zhang, K., Ki Ngo, J. C., Fu, X. D., Zhang, T., and Gray, N. S. (2018) SRPKIN-1: A Covalent SRPK1/2 Inhibitor that Potently Converts VEGF from Pro-angiogenic to Anti-angiogenic Isoform. *Cell Chem Biol* **25**, 460-470.e466
72. Conlon, E. G., and Manley, J. L. (2017) RNA-binding proteins in neurodegeneration: mechanisms in aggregate. *Genes & development* **31**, 1509-1528
73. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W. K., Miyawaki, S., Sugano, S., Haferlach, C., Koeffler, H. P., Shih, L. Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S., and Ogawa, S. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-69
74. Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., Yoshida, K., Roller, A., Nadarajah, N., Shiraishi, Y., Shiozawa, Y., Chiba, K., Tanaka, H., Koeffler, H. P., Klein, H. U., Dugas, M., Aburatani, H., Kohlmann, A., Miyano, S., Haferlach, C., Kern, W., and Ogawa, S. (2014) Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**, 241-247
75. Thol, F., Kade, S., Schlarmann, C., Löffeld, P., Morgan, M., Krauter, J., Wlodarski, M. W., Kölkling, B., Wichmann, M., Görlich, K., Göhring, G., Bug, G., Ottmann, O., Niemeyer, C. M., Hofmann, W. K., Schlegelberger, B., Ganser, A., and Heuser, M. (2012) Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* **119**, 3578-3584
76. Wu, S. J., Kuo, Y. Y., Hou, H. A., Li, L. Y., Tseng, M. H., Huang, C. F., Lee, F. Y., Liu, M. C., Liu, C. W., Lin, C. T., Chen, C. Y., Chou, W. C., Yao, M., Huang, S. Y., Ko, B. S., Tang, J. L., Tsay, W., and Tien, H. F. (2012) The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood* **120**, 3106-3111
77. Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C. J., Ellis, P., Wedge, D. C., Pellagatti, A., Shlien, A., Groves, M. J., Forbes, S. A., Raine, K., Hinton, J., Mudie, L. J., McLaren, S., Hardy, C., Latimer, C., Della Porta, M. G., O'Meara, S., Ambaglio, I., Galli, A., Butler, A. P., Walldin, G., Teague, J. W., Quek, L., Sternberg, A., Gambacorti-Passerini, C., Cross, N. C., Green, A. R., Boulton, J., Vyas, P., Hellstrom-Lindberg, E., Bowen, D., Cazzola, M., Stratton, M. R., and Campbell, P. J. (2013) Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616-3627; quiz 3699
78. Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., Spille, J. H., Afeyan, L. K., Zamudio, A. V., Shrinivas, K., Abraham, B. J., Boija, A., Decker, T. M., Rimel, J. K., Fant, C. B., Lee, T. I., Cisse, II, Sharp, P. A., Taatjes, D. J., and Young, R. A. (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543-548

Phosphorylation regulates SRSF2 aggregation

79. Hales, C. M., Dammer, E. B., Diner, I., Yi, H., Seyfried, N. T., Gearing, M., Glass, J. D., Montine, T. J., Levey, A. I., and Lah, J. J. (2014) Aggregates of small nuclear ribonucleic acids (snRNAs) in Alzheimer's disease. *Brain Pathol* **24**, 344-351
80. Lester, E., Ooi, F. K., Bakkar, N., Ayers, J., Woerman, A. L., Wheeler, J., Bowser, R., Carlson, G. A., Prusiner, S. B., and Parker, R. (2021) Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. *Neuron*
81. Drummond, E., Pires, G., MacMurray, C., Askenazi, M., Nayak, S., Bourdon, M., Safar, J., Ueberheide, B., and Wisniewski, T. (2020) Phosphorylated tau interactome in the human Alzheimer's disease brain. *Brain : a journal of neurology* **143**, 2803-2817
82. Guo, Q., Dammer, E. B., Zhou, M., Kundinger, S. R., Gearing, M., Lah, J. J., Levey, A. I., Shulman, J. M., and Seyfried, N. T. (2021) Targeted Quantification of Detergent-Insoluble RNA-Binding Proteins in Human Brain Reveals Stage and Disease Specific Co-aggregation in Alzheimer's Disease. *Front Mol Neurosci* **14**, 623659
83. Shi, J., Hu, Z., Pabon, K., and Scotto, K. W. (2008) Caffeine regulates alternative splicing in a subset of cancer-associated genes: a role for SC35. *Mol Cell Biol* **28**, 883-895
84. Heaton, B. E., Trimarco, J. D., Hamele, C. E., Harding, A. T., Tata, A., Zhu, X., Tata, P. R., Smith, C. M., and Heaton, N. S. (2020) SRSF protein kinases 1 and 2 are essential host factors for human coronaviruses including SARS-CoV-2. *bioRxiv*
85. O., X. G. (2005) Small Scale Nuclear Protein Extraction Without the Use of a Detergent.
86. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372
87. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805
88. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., Ternent, T., Brazma, A., and Vizcaíno, J. A. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-d450
89. Zambon, A. C., Gaj, S., Ho, I., Hanspers, K., Vranizan, K., Evelo, C. T., Conklin, B. R., Pico, A. R., and Salomonis, N. (2012) GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* **28**, 2209-2210
90. Seyfried, N. T., Dammer, E. B., Swarup, V., Nandakumar, D., Duong, D. M., Yin, L., Deng, Q., Nguyen, T., Hales, C. M., Wingo, T., Glass, J., Gearing, M., Thambisetty, M., Troncoso, J. C., Geschwind, D. H., Lah, J. J., and Levey, A. I. (2017) A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Syst* **4**, 60-72 e64
91. Seyfried, N. T., Gozal, Y. M., Dammer, E. B., Xia, Q., Duong, D. M., Cheng, D., Lah, J. J., Levey, A. I., and Peng, J. (2010) Multiplex SILAC analysis of a cellular TDP-43 proteinopathy model reveals protein inclusions associated with SUMOylation and diverse polyubiquitin chains. *Mol Cell Proteomics* **9**, 705-718
92. Dammer, E. B., Fallini, C., Gozal, Y. M., Duong, D. M., Rossoll, W., Xu, P., Lah, J. J., Levey, A. I., Peng, J., Bassell, G. J., and Seyfried, N. T. (2012) Coaggregation of RNA-binding proteins in a model of TDP-43 proteinopathy with selective RGG motif methylation and a role for RRM1 ubiquitination. *PLoS One* **7**, e38658

Phosphorylation regulates SRSF2 aggregation

FOOTNOTES

Support for this research was provided by funding from the National Institute on Aging (R01AG053960, R01AG061800, and RF1AG062181), the Accelerating Medicine Partnership for AD (U01AG046161 and U01AG061357). S.K. was supported by the Training Program in Biochemistry, Cell and Developmental Biology (T32GM008367) and Emory Neurology Department Training Fellowship (5T32NS007480-19).

FIGURE LEGENDS

Figure 1. Phosphorylation regulates arginine-rich RNA-binding protein solubility. (a) Sample preparation and proteomic workflow. Nucleoplasm extracts of HEK293 cells expressing recombinant SRSF2-myc protein were incubated with either calf intestinal phosphatase (+CIP) or distilled water (-CIP) at 37°C for 1 hour. Following this, samples were ultracentrifuged at 100,000xg for 1 hour. The soluble and insoluble pellet fractions were desalted and run by either western blot or liquid chromatography coupled with tandem mass spectrometry. (b) Pre-spin input (total, *T*), supernatant (soluble, *S*) and insoluble pellets (*P*) were run by SDS-PAGE and western blotted for myc. The average percent soluble ($sol./[sol. + insol.]$) and insoluble ($insol./[sol. + insol.]$) values were calculated for five biological replicates and displayed below the representative western blot. (c) Band densitometry of soluble and pellet fraction \log_2 -transformed SRSF2-myc band intensities normalized to the total signal in -CIP and +CIP conditions (5 biological replicates; Soluble p value = 0.0123; Pellet p value = 0.0372; two-tailed paired t-test). (d) Differential abundance of proteins in the soluble fractions. Fold-change, displayed on the x-axis, was the \log_2 value for fraction of signal that was insoluble [$insoluble/(insoluble+soluble)$] for the pairwise comparison +CIP/-CIP. The t-statistic ($-\log_{10}(p\text{-Value})$) was calculated for all proteins and displayed on the y-axis. Insoluble-enriched proteins were highlighted in *red* ($\log_2(\text{fold change}) \geq 1$, p value < 0.05) and proteins depleted from the insoluble fractions upon dephosphorylation were highlighted in *blue* ($\log_2(\text{fold change}) \leq -1$, p-Value < 0.05) squares, respectively. (e) Histogram plot of the net charge per residue (NCPR) of all measured nuclear proteins ($n=4366$; *green*) overlaid with the top 10% most insoluble proteins ($n=435$; *magenta*). The left y-axis corresponds to protein bin counts of all measured nuclear proteins (*green*) whereas the right y-axis corresponds to the protein bin counts of the top 10% insoluble proteins (*magenta*). Gaussian curves were fit to the histogram plots, with two curves fit to the top 10% insoluble proteins. The NCPR range right of the right-most gaussian curve mid-point (NCPR=0.076) was shaded *purple*. (f) An S-graph ranking each protein by the NCPR value. SR proteins are highlighted, ranking among the highest NCPR value proteins in the proteome. The region of proteins with NCPR values greater than 0.076 was shaded *purple*, while proteins with NCPR<0.076 were colored *green*.

Figure 2. A correlation network approach groups proteins into modules with discrete gene ontologies and solubility patterns in response to dephosphorylation. (a) Protein modules were clustered to assess module relatedness based on correlation with abundances in detergent-soluble and -insoluble pellet fractions in -CIP and +CIP conditions. \log_2 values for fraction of signal that was insoluble [$insoluble/(insoluble+soluble)$] for the pairwise comparison +CIP/-CIP displayed as a heatmap for each module and compared with the average net charge per residue (NCPR) of all members of that module, also displayed as a heat map. Significance of enrichment to top 734 insoluble proteins displayed as asterisks, determined by one-dimensional hypergeometric Fisher's exact test (FET, BH corrected; *p < 0.05, **p < 0.01, ***p < 0.001). Modules with positive average NCPR are colored *purple* while those with negative average NCPR are colored *yellow*. (b-c) Eigenprotein abundance box and whisker plots of selected modules. The farthest data points, up to 1.5 times the interquartile range away from box edges, define the extent of whiskers (error bars). (b) Protein modules that are unchanged in solubility phosphorylated or dephosphorylated. (c) Selected modules with increased insolubility abundances following dephosphorylation.

Figure 3. RNA-binding proteins have variable abundance patterns in soluble and pellet fractions following dephosphorylation. (a) Box and whisker plots of mass spectrometry abundance measurements ($n=4$) of RNA-binding proteins in soluble and pellet fractions in -CIP and +CIP conditions. (two-tailed paired t-test; *p < 0.05, **p < 0.01, ***p < 0.001). Module number and color are indicated next to each gene symbol. Whiskers range from min to max values. (b) Western blot validation of solubility changes of well-described RNA-binding proteins. Modules, colored accordingly, are paired with the protein name.

Phosphorylation regulates SRSF2 aggregation

Figure 4. SRSF2 net charge and insolubility, respectively, increases substantially with dephosphorylation. (a) SRSF2 net charge per residue (NCPR) calculated according to phosphorylation state, i.e., no phosphorylation (*hypo*SRSF2), phosphorylation sites previously observed by middle-down mass spectrometry (pSRSF2⁺, Kunding & Bishof *et al.*, 2020) or full phosphorylation (*hyper*pSRSF2). Phosphorylation sites represented by turquoise ball and stick. (b) Line plots of average charge density (window = 21 residues) from C-terminus to N-terminus of SRSF2 that is either non-phosphorylated (*black*), observed phosphorylation by MS (*grey*, Kunding & Bishof *et al.*, 2020) or fully phosphorylated (*turquoise*) in the RS domain. The SRSF2 protein map is included below the line plot, with the RNA-recognition motif (RRM) domain (*green box*) and RS domain (*black box*) annotated. (c) Nucleoplasm extracts of HEK293 cells expressing recombinant SRSF2-myc protein were incubated with either calf intestinal alkaline phosphatase (+CIP) or distilled water (-CIP) at 37°C for 1 hour. Following this, samples were split and run by both denaturing and non-denaturing native PAGE and western blotted for myc (*n*=3). (d) By denaturing SDS-PAGE (*left*), CIP-treated SRSF2-myc has increased electrophoretic mobility. Equal loading is demonstrated by Histone H3 labeling. Immunoblotting for SRSF2-myc after non-denaturing Blue native PAGE (*right*) identifies various dephosphorylated SRSF2 species not observed in mock-treated samples, including monomer (~37 kDa, *light grey triangle*), dimer (~74 kDa, *dark grey triangle*), tetramer (~148 kDa, *black triangle*) and high molecular weight (HMW) species. (e) Diagram of *in vitro* kinase reaction using kinase SRPK2 (*black*) and substrate SRSF2 (*green*). Both SRPK2 and SRSF2 were expressed and purified from *E. coli* and were mixed in the presence (+) or lack thereof (-) ATP and incubated at 30°C for 30 min. Phosphorylated SRSF2 was represented by a turquoise ring. (f) The *in vitro* reaction was separated by non-denaturing native PAGE which was transferred and western blotted for SRSF2. Monomeric (*grey triangle*), tetrameric (*black triangle*) and high molecular weight (HMW) oligomeric (*line*) species were unequally observed in the - and + ATP conditions by non-denaturing native PAGE.

Figure 5. Inhibiting SRPKs decreases SR protein phosphorylation and increases cells harboring cytoplasmic SRSF2 granule and tubule structures in HEK293 cells. (a) HEK293 cells were incubated with either DMSO (vehicle, VEH) or increasing concentrations of SRPK inhibitor SRPIN340 for a length of 12 hours. Immediately following treatment, cells were harvested, run by SDS-PAGE and Western blotted for phosphoSR (pSR) proteins SRSF4 (*purple*), SRSF6 (*blue*), SRSF10 (*red*), SRSF2/SRSF7 (*gold*), SRSF1/9 (*turquoise*) and SRSF3 (*green*) (mAb104 antibody, *red signal*). These membranes were also co-labeled using the SC35 antibody (*green signal*) that independently recognizes pSRSF2 (*gold*). The putative pSRSF2 band is indeed labeled by both mAb104 (*red signal*) and SC35 (*green signal*), such that the 35 kDa band appears a *yellow* color, bolstering confidence that pSRSF2 is indeed that protein. A p-TDP-43 antibody and a Histone H3 antibody were used to confirm specificity of SRPIN340 and equal protein loading, respectively. (b) mAb104 pSR protein band intensities for SRSF4 (*purple*), SRSF6 (*blue*), SRSF10 (*red*), SRSF2/SRSF7 (*gold*), SRSF1/9 (*turquoise*) and SRSF3 (*green*) of SRPIN340 treatments (10 μ M, 50 μ M) were quantified and normalized to the DMSO condition (artificially set to value=1). Error bars indicate maximum and minimum ranges of band values. (c) Band intensities of pTDP-43 (*black*) and pSRSF2 (*gold*) were quantified and normalized to Histone H3 labeling. The DMSO condition was artificially normalized to equal 1.00 and compared with normalized pTDP-43 and pSRSF2 values in SRPIN340 conditions. The pSRSF2 signal was significantly decreased compared with pTDP-43 at 50 μ M SRPIN340 concentration (multiple t tests, *p value=0.0099). (d-e) Immunocytochemical (ICC) staining of HEK293 cells for phosphoSRSF2 (pSRSF2)-positive nuclear speckles (*green*) and DAPI+ nuclei (*blue*) in vehicle-treated cells (d) and 50 μ M SRPIN340-treated cells (e). (f) The number of cytoplasmic granules observed were divided by the number of cells counted and averaged for 14 independent images in three independent replicates (*p value = 0.0191, two-tailed paired t-test). A minimum of 120 cells were counted in each condition per replicate. The error bars represent the range of standard deviation. (g-h) ICC staining of HEK293 cells for hypophosphorylated SRSF2 (*hypo*SRSF2) (*green*) and DAPI+ nuclei (*blue*) in vehicle-treated cells (g) and 50 μ M SRPIN340-treated cells (h). (i) The fraction of cells harboring cytoplasmic SRSF2 tubule structures was quantified in four biological replicates and compared (*p value = 0.0178, two-

Phosphorylation regulates SRSF2 aggregation

tailed paired t-test). (j) Vehicle- and (k) 50 μ M SRPIN340-treated HEK293 cells were stained by ICC for TDP-43 (*green*) and DAPI-stained.

Figure 6. Association of SRSF2 with microtubule proteins. (a) I-graph of module 16 (M16) representing hub proteins and corresponding gene symbols as nodes. Node size and edges (gray) are reflective of the degree of intramodular connectivity in WGCNA (kME). SRSF2 (*green*) and cytoskeletal-associated proteins (*red*) are highlighted. (b) Representative western blot of an immunoprecipitation (IP) of recombinant SRSF2-myc in cytoplasm and nucleoplasm extracts isolated from HEK293 cells ($n=3$ replicates per fraction). Co-IP complexes were blotted for α -tubulin (TUBA1A) and β -tubulin (TUBB8). (c-d) ICC staining of HEK293 cells for hypophosphorylated SRSF2 (hypoSRSF2) (*green*) and both TUBA1A (*red*; c) and TUBB8 (*red*; d) in vehicle- cells or 50 μ M SRPIN340-treated cells.

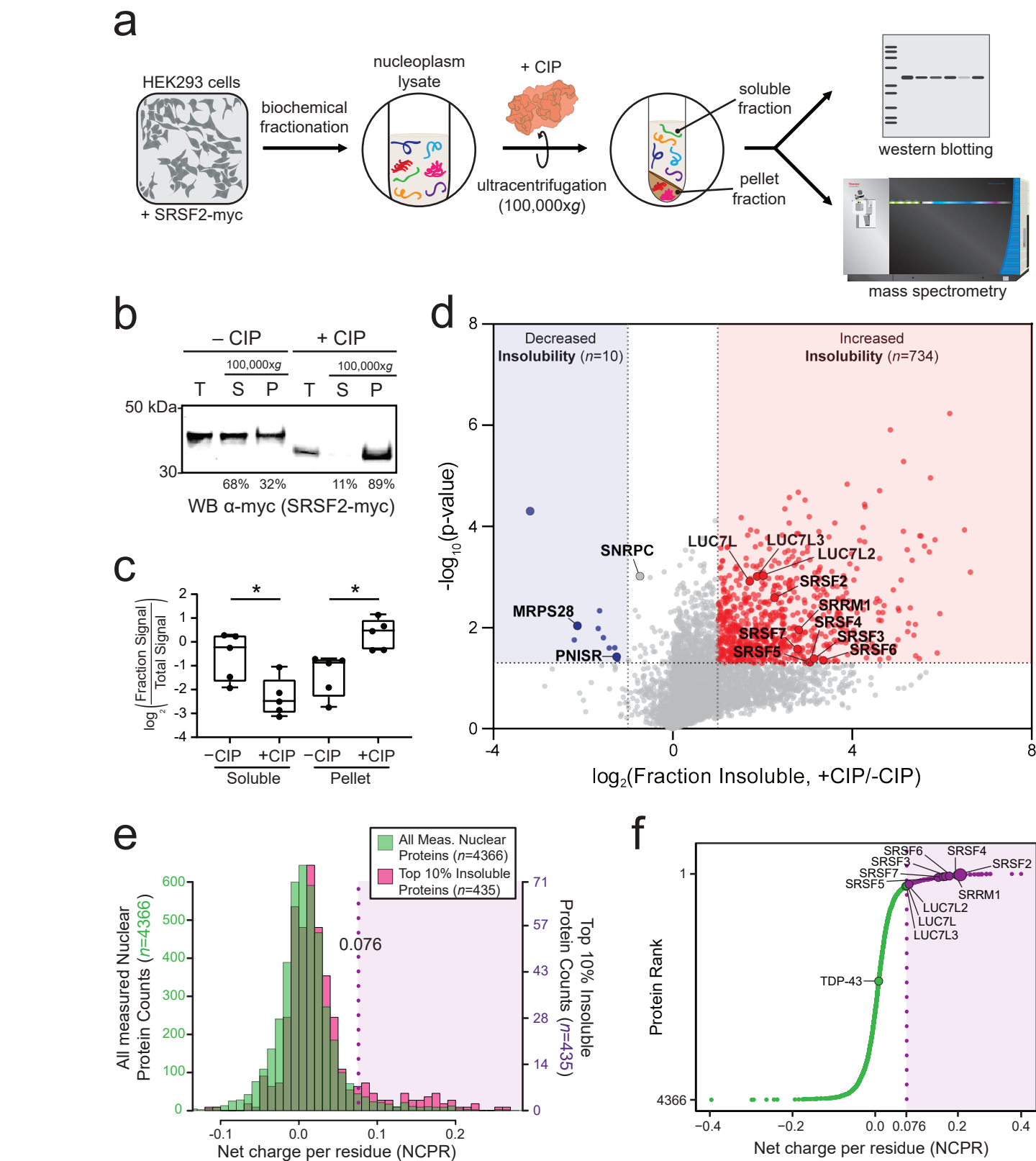
Figure 7. Proposed model of SRSF2 solubility and oligomerization changes regulated by phosphorylation. Equilibrium of structural states (monomer, multimer, oligomer) of SRSF2 (*green*) in varying states of phosphorylation (*dark turquoise border*). The magnitude of positive (*purple triangle*) SRSF2 net charge is illustrated, inverse to the degree of phosphorylation (*turquoise triangle*). Phosphorylation regulates higher order SRSF2 solubility, oligomerization and structure formation.

Supplemental Figures

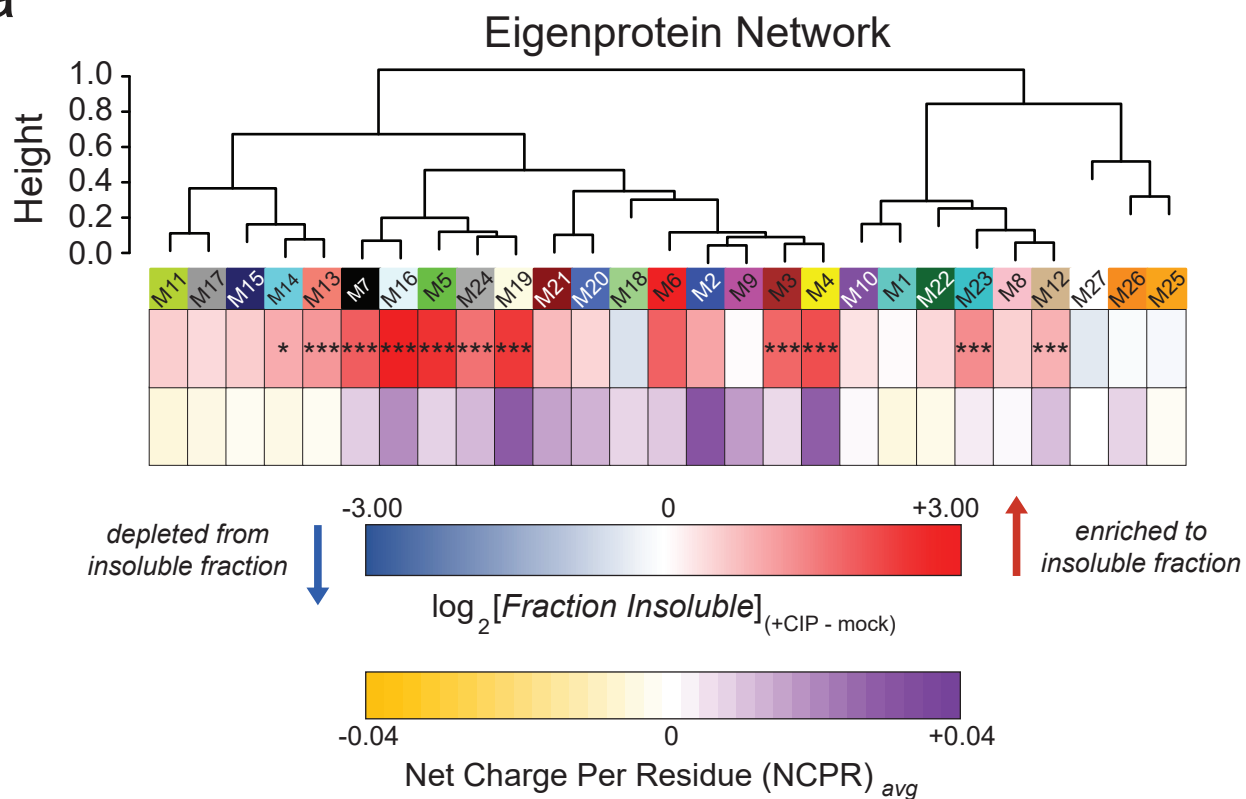
Supplemental Figure S1. The arginine-/serine-rich (RS) domain of SRSF2 is predicted to be highly disordered. The DISOPRED 3.0 algorithm predicts intrinsically disordered regions of proteins based on primary sequence alone. The protein architecture of SRSF2 below is sized to match the x-axis of the disorder plot. SRSF2 harbors an N-terminal RNA-recognition motif (RRM) domain (*green box*) and a C-terminal arginine-/serine-rich (RS) domain (*black box*). (b) Nucleoplasm extracts of HEK293 cells expressing recombinant SRSF2-myc protein were incubated with either distilled calf intestinal alkaline phosphatase (+CIP) for increasing time lengths (10 min, 30 min, 1 hour) or water (-CIP) or at 37°C and separated by denaturing SDS-PAGE and immunoblotted for the myc tag. (c) Nucleoplasm fractions of HEK293 cells transiently-expressing SRSF2-myc were treated with either dH₂O (-) or CIP (+), separated by SDS-PAGE and immunoblotted for myc tag (*red*) and hypophosphorylated SRSF2 (*green*, hypoSRSF2) as well as Histone H3 for loading control. Increased electrophoretic mobility of SRSF2-myc was observed in +CIP treatment, along with increased labeling by hypoSRSF2-specific antibody (1Sc-4F11).

Supplemental Figure S2. The nuclear proteome was separated into discrete groups by weighted gene correlation network analyses. Weighted Gene Correlation Network Analysis (WGCNA) cluster dendrogram groups all proteins ($n=4,366$) measured by hierarchical clustering into 27 different protein modules (M1-M27). The top generalizable biological process gene ontology term was given as a title to each module.

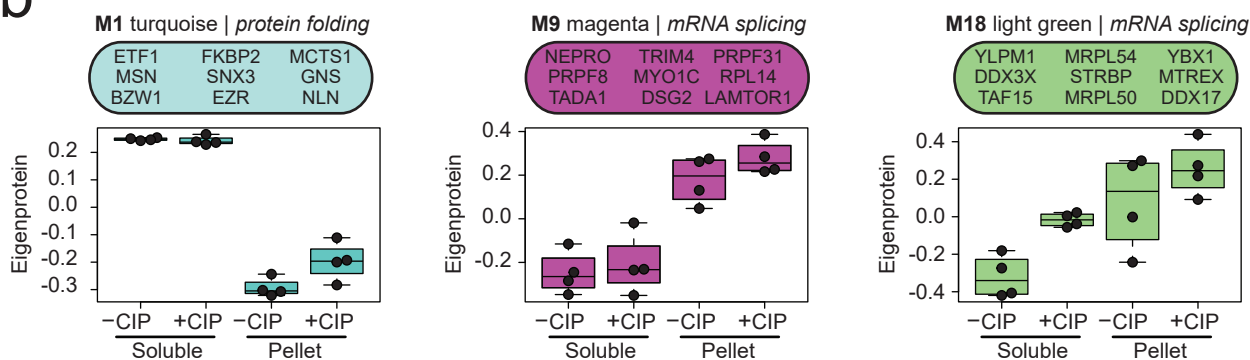
Supplemental Figure S3. Validation of *in vitro* SRPK2-SRSF2 kinase reactions. (a-b) Serine-/arginine protein kinase 2 (SRPK2) and SRSF2, expressed and purified from *E. coli* were mixed in the presence or lack thereof (-/+, respectively) ATP and incubated at 30°C for 30 min and run by denaturing SDS-PAGE or isoelectric focusing (IEF) gel electrophoresis. (a) Equal loading of SRPK2 was validated by Coomassie staining (~110 kDa). (b) The pI of non-phosphorylated SRSF2 (-ATP) is 12.4, while SRSF2 co-incubated with SRPK2 and ATP had an apparent pI range of 10-12.4.



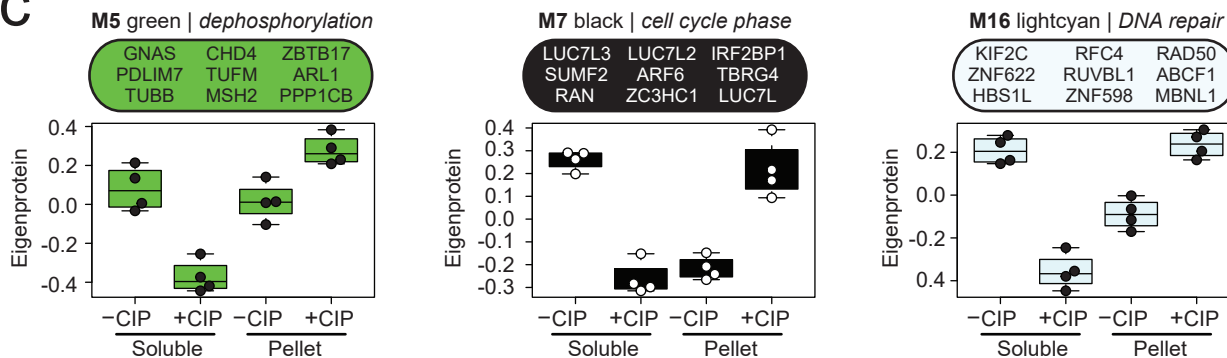
a

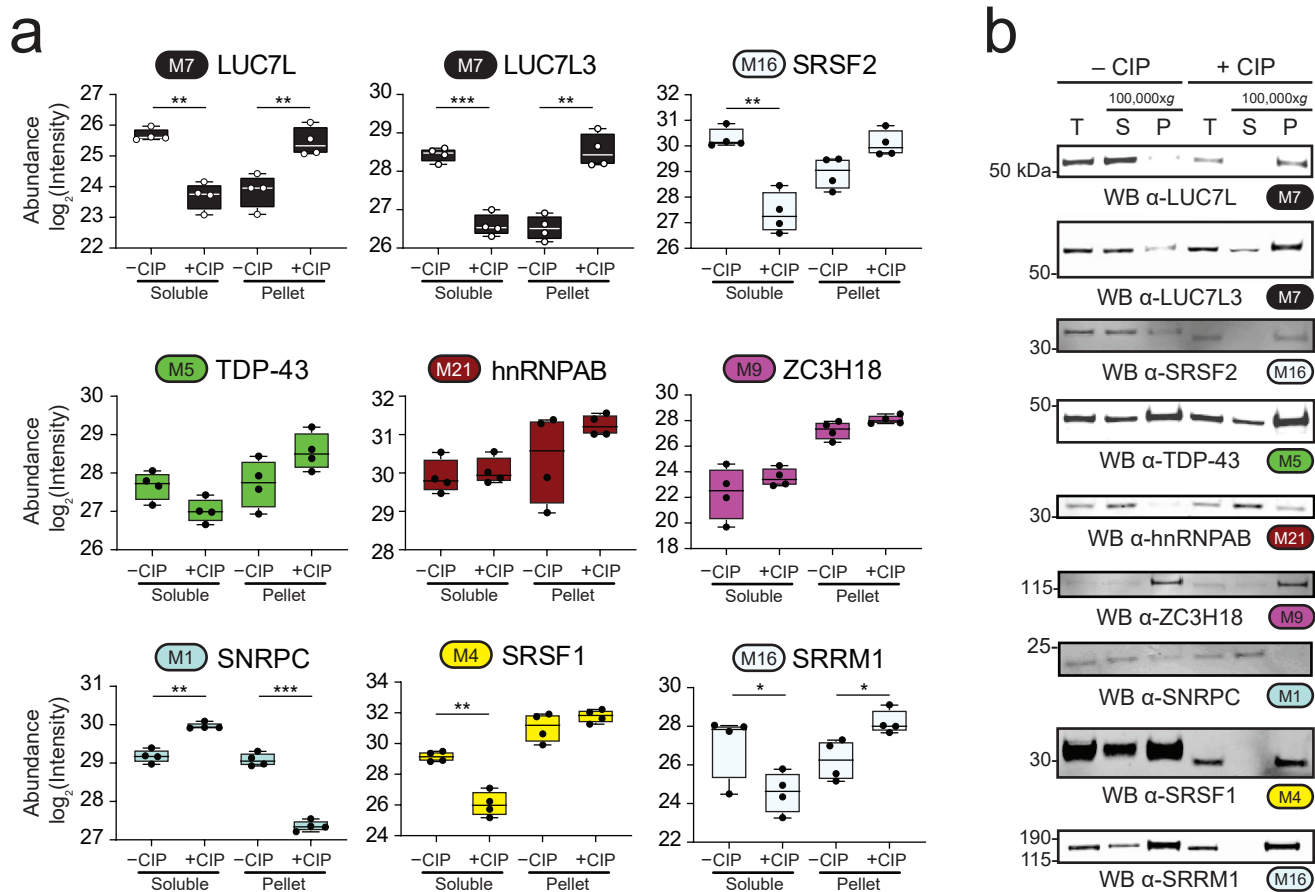


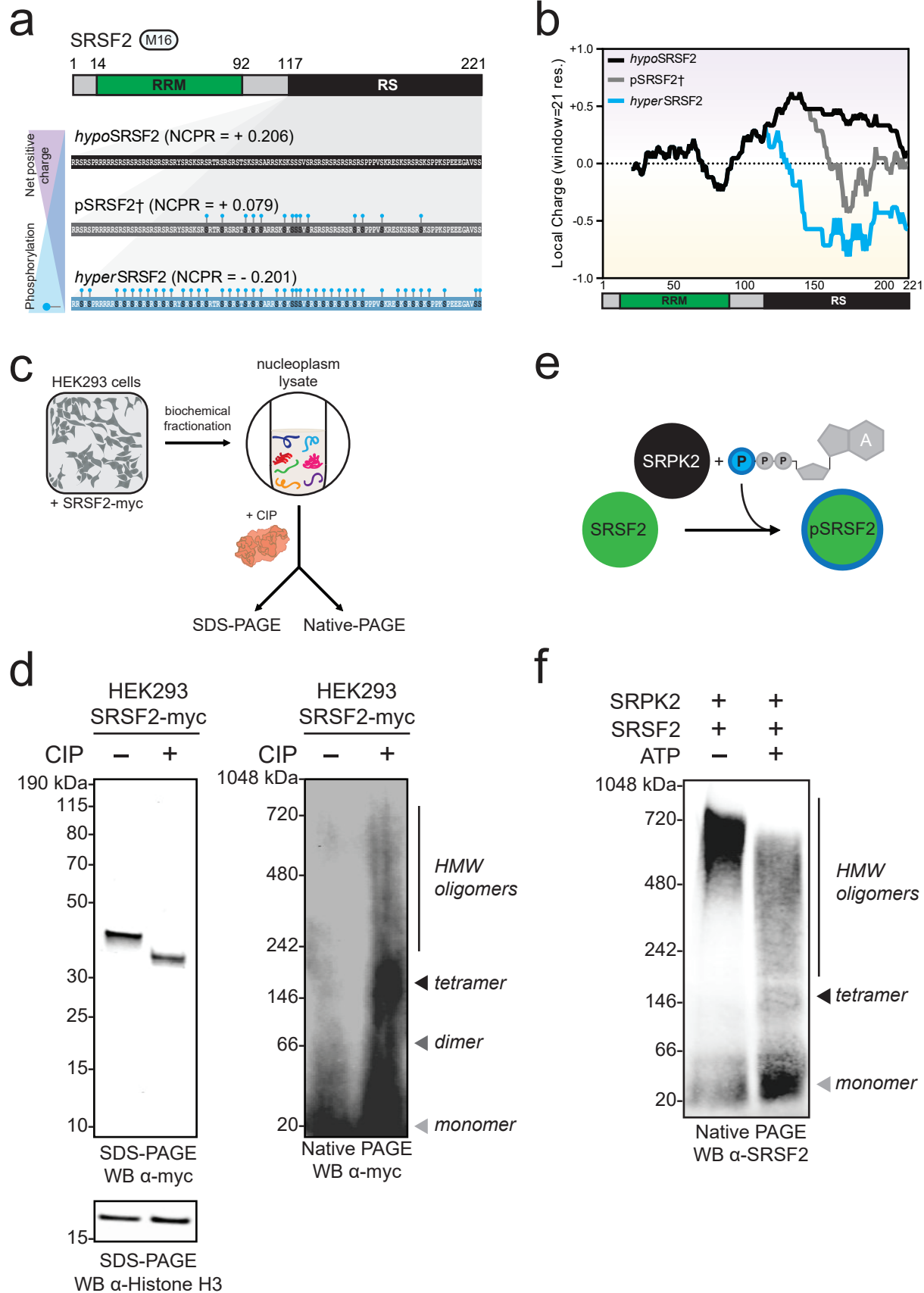
b

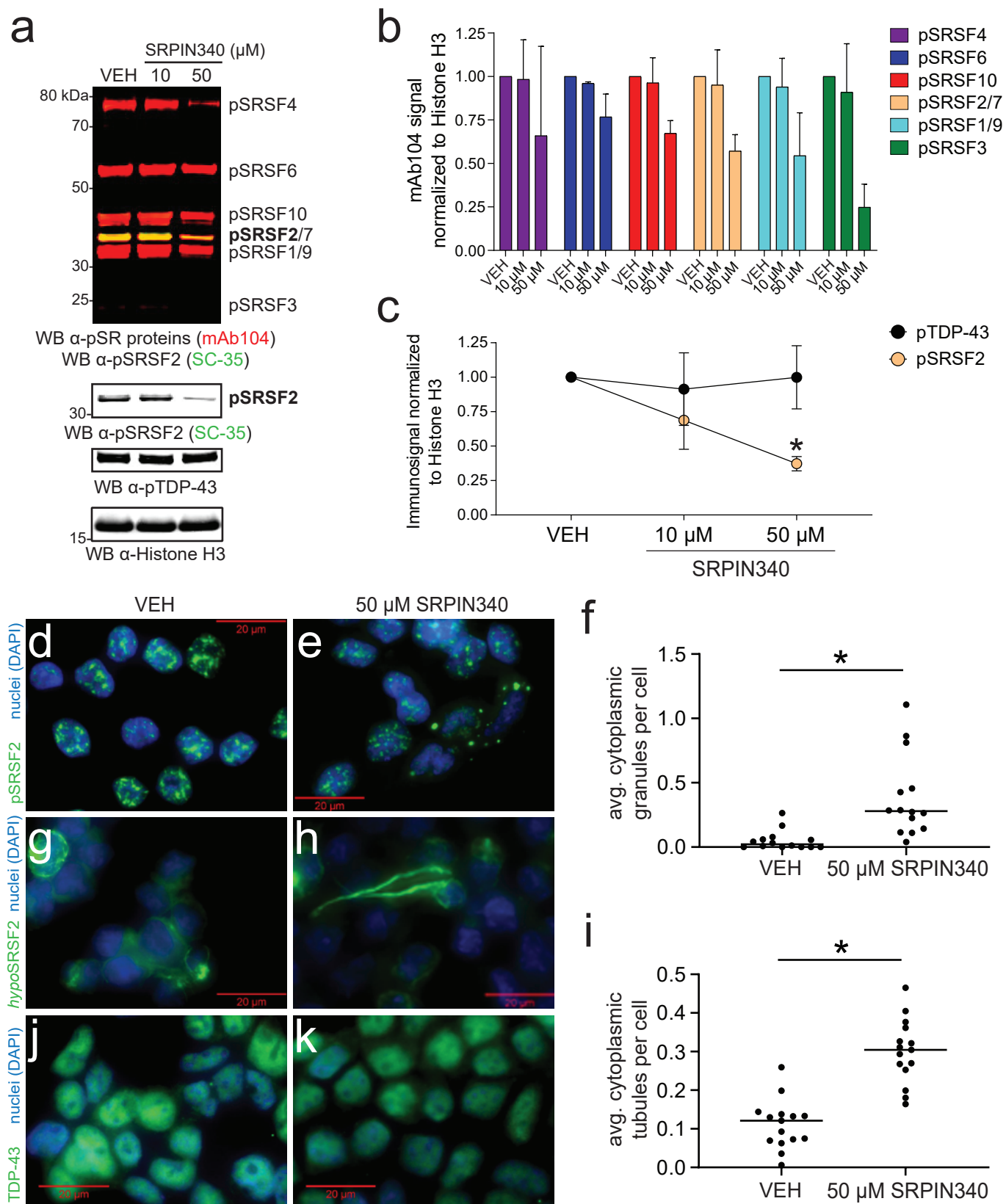


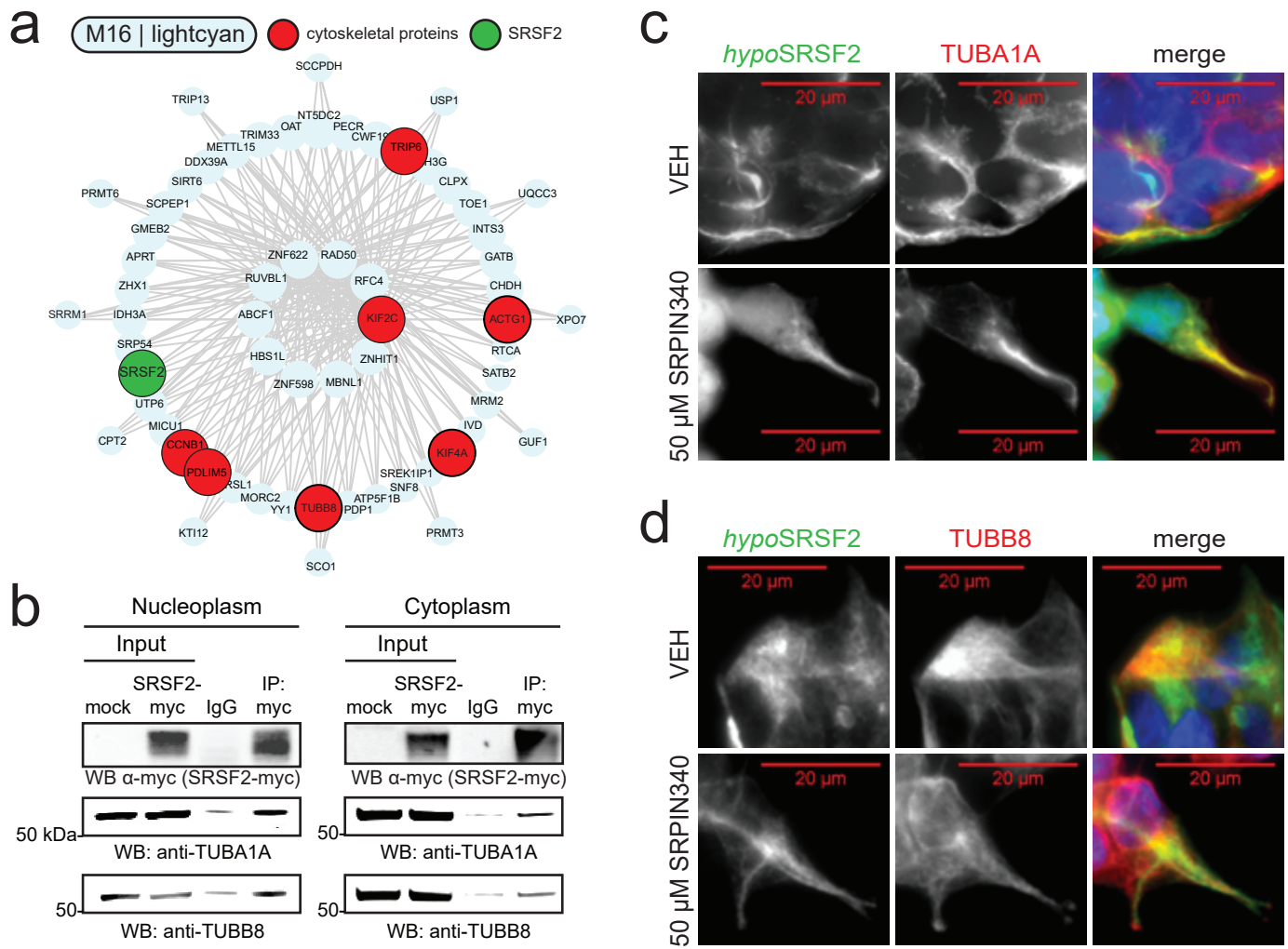
c

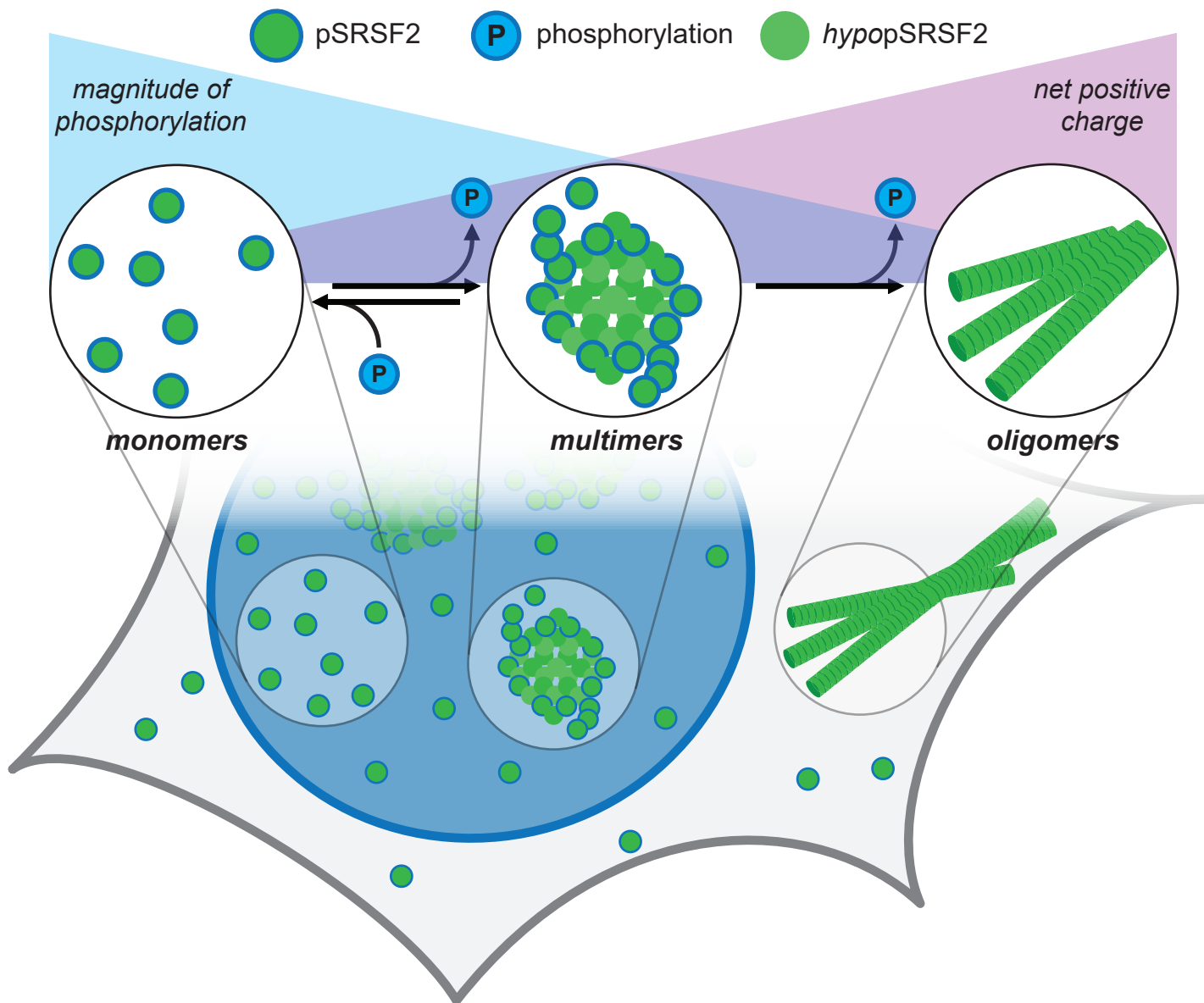


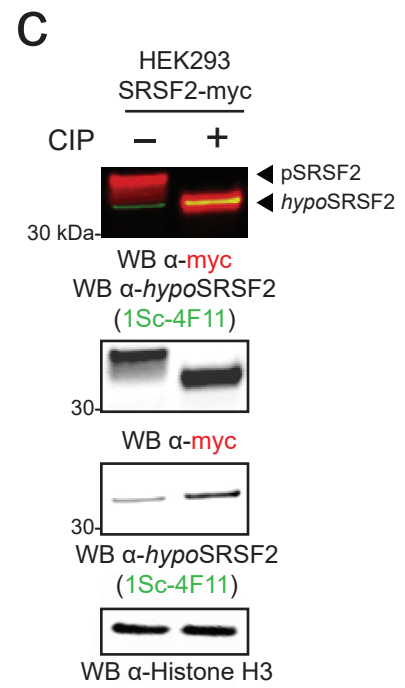
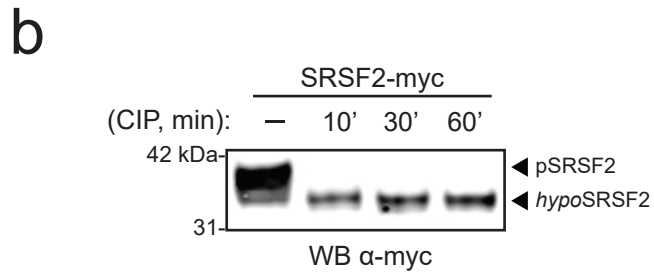
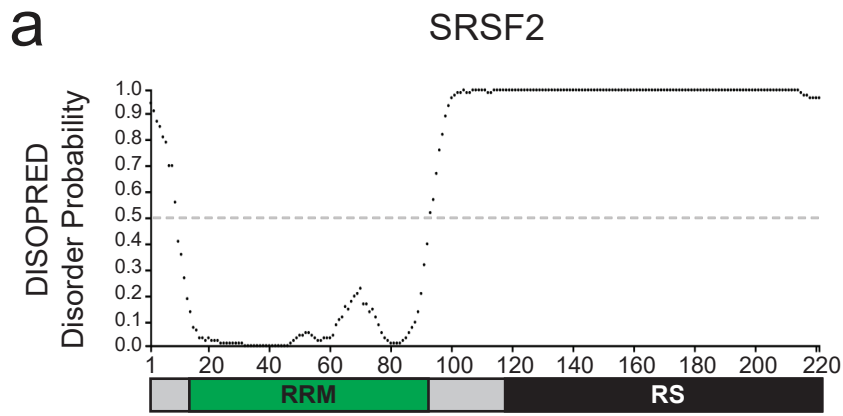




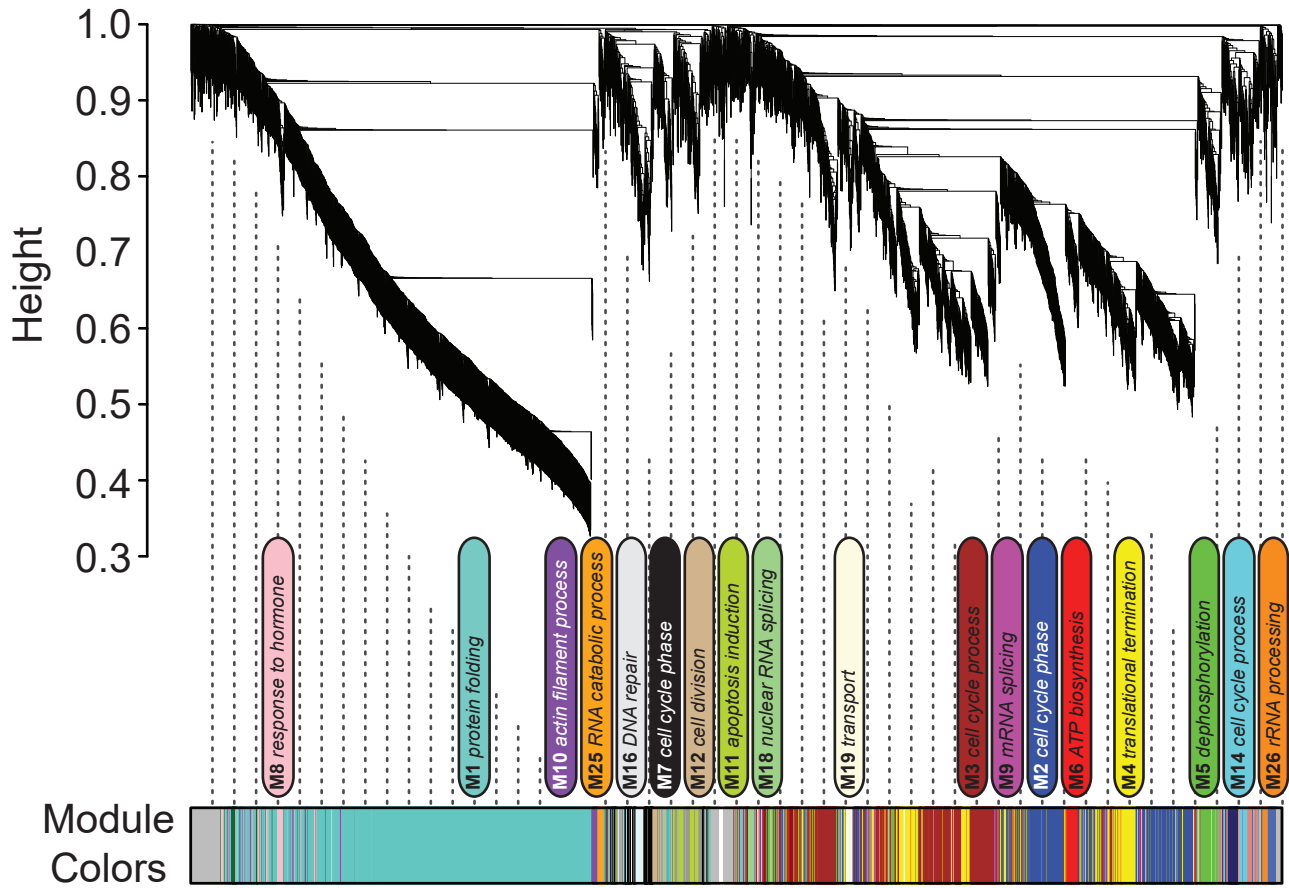




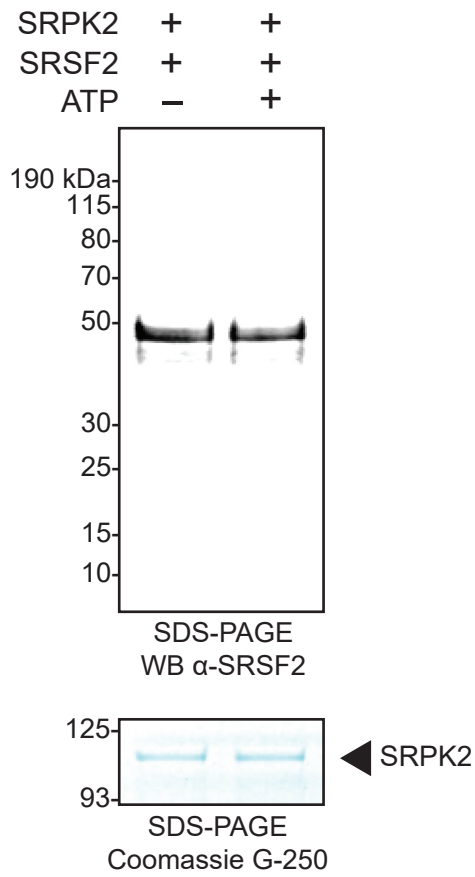




Cluster Dendrogram



a



b

