

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Sean Robert Kunding

Date

THE REGULATION OF THE SOLUBILITY, OLIGOMERIZATION AND STRUCTURE OF
ARGININE-RICH RNA-BINDING PROTEINS BY POST-TRANSLATIONAL
MODIFICATION IN ALZHEIMER'S DISEASE

By

Sean Robert Kunding
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Biochemistry, Cell, and Developmental Biology

Nicholas Seyfried, Ph.D.
Advisor

Gary Bassell, Ph.D.
Committee Member

Anita H. Corbett, Ph.D.
Committee Member

Homa Ghalei, Ph.D.
Committee Member

Daniel Reines, Ph.D.
Committee Member

Accepted:

Kimberly J. Arriola, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

The regulation of the solubility, oligomerization and structure of arginine-rich RNA-binding proteins by post-translational modification in Alzheimer's disease

by

Sean Robert Kunding

B.S., University of Wisconsin-Madison

Advisor: Nicholas T. Seyfried, Ph.D.

An abstract of
a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science (GDBBS)
Biochemistry, Cell and Developmental Biology (BCDB)
2021

ABSTRACT

The regulation of the solubility, oligomerization and structure of arginine-rich RNA-binding proteins by post-translational modification in Alzheimer's disease

By Sean Robert Kunding

Post-translational modifications (PTMs) within splicing factor RNA-binding proteins, such as phosphorylation, regulate several critical steps in RNA metabolism including spliceosome assembly, alternative splicing and mRNA export. Previously, the application of conventional mass spectrometry methods have been insufficient to sequence arginine-rich domains of RNA-binding proteins. Here we report a middle-down proteomic approach coupled with electron transfer dissociation (ETD) mass spectrometry to map previously unknown sites of phosphorylation and methylation within the arginine-rich domains of U1-70K, SRSF2 and structurally similar RNA-binding proteins from nuclear extracts of HEK-293 cells. Notably, the arginine-rich LC domains in RNA-binding proteins are densely modified by methylation and phosphorylation compared with the remainder of the proteome, with methylation and phosphorylation favoring RSRS motifs. Analysis of combinatorial PTMs within RSRS motifs indicate that phosphorylation and methylation do not often co-occur, suggesting they may functionally oppose one another. Furthermore, we show that phosphorylation may modify interactions between Arg-rich proteins, as SRSF2 has stronger association with U1-70K and LUC7L3 upon dephosphorylation. We dephosphorylated nuclear extracts *in vitro* and analyzed equal amounts of detergent-soluble and -insoluble fractions by mass spectrometry-based proteomics. Correlation network analysis resolves 27 distinct modules of differentially soluble nucleoplasm proteins following dephosphorylation. We found classes of RNA-binding proteins with increased aggregation following dephosphorylation, including the SR protein family and the SR-like RBPs although increased aggregation was not observed across broad classes of RBPs. Phosphorylation regulated SRSF2 structure, as native 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 species formation of SRSF2. Furthermore, SRPK inhibition by SRPIN340 decreased SR protein phosphorylation *in vivo* and regulated the cytoplasmic mislocalization of SRSF2 and formation of tubular structures that colocalize with microtubules by immunocytochemical staining. Collectively, these findings demonstrate that phosphorylation is a critical determinant of SR and SR-like protein solubility, oligomerization, and structure. Furthermore, these findings suggest that the level of phosphorylation within arginine-rich domains of RNA-binding proteins may be among the highest in the proteome, and a possible critical suppressor of arginine-rich RNA-binding protein aggregation and mislocalization.

The regulation of the solubility, oligomerization and structure of arginine-rich RNA-binding proteins by post-translational modification in Alzheimer's disease

by

Sean Robert Kunding

B.S., University of Wisconsin-Madison

Advisor: Nicholas T. Seyfried, Ph.D.

a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science (GDBBS)
Biochemistry, Cell and Developmental Biology (BCDB)
2021

Acknowledgments

Thank you to P.R. for teaching me my first molecular biology lessons. Thank you to J.B. for taking a chance on me. Thank you to N.S. for your trust and your advocacy. Thank you to D.D. for taking an interest in me. Thank you to G.B., A.C., H.G. and D.R. for supporting my thoughts and efforts. Thank you to R.F., J.E., J.G., J.M., J.M., B.W., A.M., D.S., K.T., B.F., A.O. and P.H. for supporting me from Wisconsin. Thank you to B.N., D.N., J.N., J.N., B.N., A.N., D.N., M.N., N.N., B.N., M.N., K.N., B.W. and A.W. for making Atlanta feel like home. Thank you to K.S. for making every day feel like Christmas. Thank you to E.S. for showing me how to be stubborn about your passions. Thank you to A.S. for being the sweetest person I have ever known. Thank you to J.K., N.K. and Z.K. for teaching me teamwork and humility. Thank you to S.K. for teaching me everything you knew. Thank you to I.K., V.K. and E.K. for being sources of pure joy. Thank you to J.S., J.S. and D.S. for welcoming me with open arms. Thank you to I.S. for teaching me the meaning of love. Thank you to K.K. for unconditional love and support. Thank you to R.K. for teaching me to measure twice and cut once, that you get what you put into it and what it means to be a man.

Table of Contents	Page
1.0 RNA-binding protein aggregation in Alzheimer's disease	1
1.1 An Introduction to Alzheimer's disease	1
1.2 A History of Alzheimer's disease	1
1.2.1 Pathology of Alzheimer's disease	3
1.2.2 Genetic heterogeneity of Alzheimer's disease	5
1.3 Amyloid Cascade Hypothesis	7
1.3.1 Limitations of the Amyloid Cascade Hypothesis	8
1.4 Links between RNA-binding proteins and neurodegenerative disease	9
1.4.1 Granules to Fibrils: an equilibrium of Liquid-liquid phase separation (LLPS)	10
1.4.2 RNA-binding protein pathology in Alzheimer's disease	10
1.5 Using mass spectrometry-based proteomics to investigate the pathophysiology of Alzheimer's disease	12
1.5.1 Mass spectrometry-based proteomics reveals novel protein signatures in Alzheimer's disease pathogenesis	13
1.5.2 Identification of U1 small nuclear ribonucleoprotein spliceosome and arginine-rich RNA-binding proteins enriched in Alzheimer's disease insoluble fractions	13
1.6 The Spliceosome and associated arginine-rich splicing factor RNA-binding proteins	15
1.6.1 Discovery of arginine-rich splicing factor RNA-binding proteins	16
1.7 Regulation of arginine-rich RNA-binding protein function by post-translational modification (PTM)	16
1.7.1 PTM regulation of spliceosome assembly and canonical/alternative splicing	17
1.7.2 Post-translational modification regulation of RNA-binding protein LLPS	18
1.7.3 Limitations of using mass spectrometry-based proteomics to identify PTM sites in arginine-rich RNA-binding proteins	19
1.8 Research Focus and Innovation	20
1.9 Figures and Tables	21
2.0 Middle-down proteomics reveals dense sites of methylation	35

and phosphorylation in arginine-rich RNA-binding proteins	
2.1 Abstract	36
2.2 Introduction	37
2.3 Materials and Methods	42
2.4 Results	51
2.4.1 Preparation of Nucleoplasm Fractions Enriched with Arg-rich RBPs	53
2.4.2 Global Analysis of RNA-binding proteins from Nucleoplasm Extracts by Middle-down ETD MS	55
2.4.3 Enhanced Sequence Coverage of Arg-rich Proteins by Middle-Down ETD MS	56
2.4.4 Motif Algorithms Resolve RNA-Binding Protein Subgroups with Distinct Biological Properties	57
2.4.5 PTM Site Validation	59
2.4.6 Arg-Rich Domains in RNA-Binding Proteins Contain Combinatorial PTMs	60
2.4.7 Arg-rich Domains in RNA-Binding Proteins Are Densely Modified	61
2.4.8 Phosphorylation and Methylation Favor RSRS Motifs	63
2.4.9 Phosphorylation Regulates Protein-Protein Interactions between Structurally Similar Arg-Rich RNA-Binding Proteins	64
2.5 Discussion	65
2.6 Acknowledgments	70
2.7 Figures	71
3.0 Phosphorylation regulates arginine-rich RNA-binding protein solubility and oligomerization	95
3.1 Abstract	96
3.2 Introduction	97
3.3 Materials and Methods	99
3.4 Results	111
3.4.1 Phosphorylation prevents SRSF2 aggregation	111
3.4.2 Proteomics reveals RBPs that aggregate following dephosphorylation	112
3.4.3 Arginine-/lysine-rich RNA-binding proteins with positive net charge preferentially aggregate following dephosphorylation	113

3.4.4 Systems analysis identifies modules of proteins with solubility impacted by phosphorylation	115
3.4.5 Confirmation of hub protein solubility changes following dephosphorylation	116
3.4.6 Phosphorylation decreases NCPR and regulates the oligomerization of arginine-rich SRSF2	117
3.4.7 SRPK inhibitor SRPIN340 decreases SR protein phosphorylation and increases SRSF2 granule and tubule formation	118
3.4.8 SRSF2 interacts with microtubule subunit proteins α - and β -tubulin	120
3.5 Discussion	121
3.6 Acknowledgments	123
3.7 Figures	125
4.0 Discussion	147
4.1 Summary	147
4.2 Future directions of mass spectrometry to improve sequencing of arginine-rich RNA-binding proteins	147
4.3 Use of middle-down proteomic approaches to sequence post-translational modification sites in arginine-rich RNA-binding proteins in Alzheimer's disease	149
4.4 Recent discoveries of RNA-binding protein functions regulated by post-translational modifications	149
4.5 Is phosphorylation a trigger or a consequence of protein aggregation?	150
4.6 Does phosphorylation regulate the core splicing function of arginine-rich RNA-binding proteins?	152
4.7 The <i>in vivo</i> roles of phosphatases	153
4.8 Implications of phosphorylation dysregulation in AD	154
4.9 Future Directions	155
4.10 Figures	158
5.0 References	160

Figure List	Page
Figure 1.1. Jack curve model of AD.	21
Figure 1.2. Sternburg model of arginine-rich RNA-binding protein phase separation continuum.	23
Figure 1.3. Arginine-rich RNA-binding proteins aggregate in Alzheimer's disease.	25
Figure 1.4. A network of interactions between arginine-rich RNA-binding proteins facilitate spliceosome assembly and pre-mRNA splicing.	27
Figure 1.5. Comparison of prototypical BAD and SR arginine-rich RNA-binding proteins U1-70K and SRSF2.	29
Figure 1.6. Post-translational modifications linked to human neurodegenerative disease.	31
Table 1.1. A superfamily of arginine-rich pre-mRNA splicing factor RNA-binding proteins.	33
Fig.2.1. Electron Transfer Dissociation (ETD) more effectively fragments Arg-rich peptides compared with HCD.	71
Fig.2.2. Middle-down ETD MS strategy provides enhanced sequence coverage of U1-70K arginine-rich low complexity domains.	73
Fig.2.3. Isolation of nucleoplasm by cellular fractionation enriches for RNA-binding proteins.	75
Fig.2.4. Middle-down ETD analysis of nucleoplasm contributes to greater coverage of Arg-rich RNA-binding proteins.	77
Fig.2.5. HCD and ETD fragmentation of nucleoplasmic sample enhances identification of unique peptides.	79
Fig.2.6. Motif algorithms resolve RNA-binding protein subgroups with distinct biological properties.	81
Fig.2.7. A motif algorithm stringently selects high-confidence peptides.	83
Fig.2.8. PTM validation by phosphosite.org database comparison and manual inspection of spectra.	85
Fig.2.9. Arginine-rich RNA-binding proteins are enriched for PTMs.	87
Fig.2.10. Arginine methylation and serine phosphorylation are enriched in RSRS motifs.	89
Fig.2.11. Dephosphorylation increases protein interactions between SRSF2 and the BAD RNA-binding proteins U1-70K and LUC7L3.	91
Fig.2.12. Motif analysis of nucleoplasmic PTM-modified peptide sequences.	93
Fig.3.1. The arginine-/serine-rich (RS) domain of SRSF2 is predicted	125

to be highly disordered.	
Fig.3.2. Phosphorylation regulates arginine-rich RNA-binding protein solubility.	127
Fig.3.3. SR and SR-like proteins are enriched in detergent insoluble pellet when dephosphorylated.	129
Fig.3.4. The nuclear proteome was separated into discrete groups by weighted gene correlation network analyses.	131
Fig.3.5. A correlation network approach groups proteins into modules with discrete gene ontologies and solubility patterns in response to dephosphorylation.	133
Fig.3.6. RNA-binding proteins have variable abundance patterns in soluble and pellet fractions following dephosphorylation.	135
Fig.3.7. SRSF2 net charge and insolubility, respectively, increases substantially with dephosphorylation.	137
Fig.3.8. Validation of <i>in vitro</i> SRPK2-SRSF2 kinase reactions.	139
Fig.3.9. Inhibiting SRPKs decreases SR protein phosphorylation and increases cells harboring cytoplasmic SRSF2 granule and tubule structures in HEK293 cells.	141
Fig.3.10. Hypophosphorylated SRSF2 exhibits nuclear mislocalization phenotypes.	143
Fig.3.11. Association of SRSF2 with microtubule proteins.	145
Fig.4.1. Proposed model of SRSF2 solubility and oligomerization changes regulated by phosphorylation.	158

ABBREVIATIONS

AD – Alzheimer’s disease
A β – Amyloid beta
NFT – Neurofibrillary tangle
CERAD – Consortium to Establish a Registry for Alzheimer’s disease
IHC – Immunohistochemistry
fAD – familial Alzheimer’s disease
sAD – sporadic Alzheimer’s disease
GWAS – Genome Wide Association Studies
SNP – Single Nucleotide Polymorphism
FTLD – frontotemporal lobar dementia
AsymAD – asymptomatic Alzheimer’s disease
FDA – Food and Drug Administration
RBP – RNA-binding protein
FTD – frontotemporal dementia
ALS – amyotrophic lateral sclerosis
LC – low complexity
LLPS – liquid-liquid phase separation
FRAP – fluorescence recovery after photobleaching
PTM – post-translational modification
PD – Parkinson’s disease
MS – mass spectrometry
snRNP – U1 small nuclear ribonucleoprotein subunit
LLPS – liquid-liquid phase separation
IF - immunofluorescence
IP – immunoprecipitation
pre-mRNA – precursor messenger RNA
snRNA – small nuclear RNA
Sm – Smith antigen core
SR – serine-/arginine-rich
ETD - electron transfer dissociation
BAD – Basic Acidic Dipeptide
Arg-rich – arginine-rich
PRMT – protein arginine methyltransferases
RRM – RNA recognition motif
CID – collision induced dissociation
HCD – higher energy collision dissociation
ECD – electron capture dissociation
FDR – false discovery rate
PSM - peptide spectral match
GO – gene ontology

ESE – exon splicing enhancer
no-M – no-motif
EThCD – Electron transfer higher energy collisional dissociation
CIP – calf-intestinal phosphatase
NCPR – Net charge per residue
WGCNA – Weighted gene correlational network analysis
FET – Fisher’s exact test
ICC – Immunocytochemistry
MDS – myelodysplastic syndromes
CMML – chronic myelomonocytic leukemia
APP – Amyloid precursor protein
MAPT – microtubule associated protein Tau
APOE – Apolipoprotein E
SRSF2 – serine-/arginine-rich splicing factor 2
U1-70K – U1 small nuclear ribonucleoprotein complex
TDP-43 - TAR DNA-binding protein 43 kDa
FUS – Fused in Sarcoma
LUC7L3 – lethal unless Cap-binding-complex produced 7 like 3

CHAPTER 1.0 : RNA-binding protein aggregation in Alzheimer's disease

1.1 An Introduction to Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurodegenerative disease and the most common form of dementia, affecting over 6 million Americans^{1,2}. By 2050, this number is expected to double, with potential treatment and caregiving costs to exceed \$1 trillion. No medications have been discovered that cure AD let alone categorically prevent cognitive decline, leaving therapeutics that delay symptoms for a short period of time. Despite the recent approval of aducanumab by the FDA, clinicians are hesitant to prescribe the controversial drug, and moreover, some insurers such as Blue Cross Blue Shield are denying reimbursement claims³. Therefore, research is needed to advance our comprehensive understanding of AD with the goal of developing novel therapeutics that aid in the prevention and treatment of AD.

1.2 A History of Alzheimer's disease

Alois Alzheimer, the man who first described AD, was born in 1864 in the Bavarian town Markbreit⁴. Alois went on to study medicine at the German Universities of Berlin and Tubinga, and became skilled in the practice of histology at the University of Würzburg. In 1888, Dr. Alzheimer was hired as a resident at The Municipal Asylum for the Insane and Epileptic in Frankfurt am Main. Months later, Franz Nissl, who had recently discovered a novel technique to stain nerve cells, joined the Asylum staff. Together, Drs. Alzheimer and Nissl began a transformation of the clinic to focus on the use of histological examinations at autopsy to promote research into the understanding of mental disorders.

Records discovered at the City Historical Institute and the Psychiatric Unit at the University Clinic in Frankfurt in 1995 by Drs. Maurer, Volk and Gerbaldo provide evidence of the admission of a 51-year-old woman by the name of Auguste Deter on November 25th, 1901⁴. Information

provided by Mrs. Deter's rail worker husband indicate Auguste had never been seriously ill prior to her admission. While her mother had suffered convulsive attacks and her father throat anthrax, there was no family history of mental disorders. Mrs. Deter's husband described Auguste as "polite, hard-working, shy and slightly anxious". Eight months prior, Mrs. Deter had believed her husband was having an affair with a neighbor woman. Following this, Mrs. Deter began exhibiting memory problems, notably making errors while cooking and wandering throughout the house. Mrs. Deter's symptoms continued to deteriorate and she began to suffer from frightening delusions and episodes of intense paranoia.

Dr. Alzheimer recorded several unusual behaviors of Mrs. Deter, including confusing meat and vegetables while eating, repeating sentences, misspelling her name and having difficulty reading. Mrs. Deter's behavior became hostile and she began to hit physicians and fellow patients. By each month Ms. Deter's symptoms worsened until she became immobile, experiencing rapid weight loss and bed sores. She succumbed to pneumonia and died on April 8th, 1906, five years after being admitted.

Anatomical autopsy records indicate hydrocephalus, cerebral atrophy, arteriosclerosis of small cortical vessels, pneumonia of lower lobes and nephritis. Upon being notified of Mrs. Deter's death, Dr. Alzheimer requested autopsy samples and completed a medical examination history of Mrs. Deter on April 28th, 1906⁵. Using Bielchowsky's silver impregnation staining, Dr. Alzheimer described microscopic brain lesions in the cerebral cortex, now understood to be neurofibrillary tangles, consisting of thick bundles of parallel fibrils that were differentially stained relative to neurofibrils, suggesting these fibrils had undergone some chemical change. Many neurons in the upper layers of the cortex appeared to have disappeared, while the presence of foci were also remarked in this region. In November of 1906, Dr. Alzheimer presented his observations, entitled,

“*On a Peculiar, Severe Disease Process of the Cerebral Cortex*” at the 37th Assembly of the Southwest German Psychiatrists in Tübingen. Dr. Alzheimer concluded his presentation, “taken as a whole, I believe I have just presented a clearly defined and hitherto unrecognized disorder.” He received no questions from the audience⁴.

The tissue slides taken from Mrs. Auguste Deter were re-discovered in 1997 at the Institute of Neuropathology of the University of Munich⁶. Upon re-examination, the slides confirmed the abundant neurofibrillary tangles and neuritic plaques that Dr. Alzheimer described some 90 years earlier.

1.2.1 Pathology of Alzheimer’s disease

Dr. Alzheimer published descriptions of neurofibrillary tangles inside neurons and neuritic plaques surrounding neurons in his seminal paper “*Über einen eigenartigc Erkrankung der Hirnrinde (a peculiar disorder of the cerebral cortex)*” using innovative staining techniques and state-of-the-art Zeiss microscopes in 1907. It wasn’t until 1968 that these structures were inextricably linked to symptoms, when Drs. Blessed, Tomlinson and Roth established the relationship between plaque and neurofibrillary tangle lesions and quantitative declines in cognition and memory⁷.

Using a vibratome and the electron microscope in 1963, researcher Michael Kidd discovered that AD neurofibrillary tangles harbored a paired helical filament structure. In 1964, Robert Terry used the same methods to discover that the core of neuritic plaques were composed of amyloid protein⁸. It wasn’t until two decades later that the core molecules of these structures were identified. In 1984, George Glenner, together with Caine Wong, used a Beckman analyzer to sequence amyloid beta (A β), a 40-42 residue peptide⁹. The next year Colin Masters isolated A β from cerebral plaques in the brain of individuals with Down’s Syndrome, or those with three copies

of chromosome 21 which carries amyloid precursor protein (APP)¹⁰. The year after, Khalid Iqbal, Inge Grundke-Iqbal and Robert Terry isolated the protein constituent of the paired helical filaments of neurofibrillary tangles as Tau, which was originally identified in 1975 by Murray Weingarten and Marc Kirschner¹¹. The identification of the components of the two aggregate structures found in AD brain allowed researchers to focus investigations into the relationship of A β and Tau in the pathophysiology of AD.

Two criteria are currently applied to definitively diagnose brains as AD versus some other neurological disease in post-mortem examination. The first of these, is the Consortium to Establish a Registry for Alzheimer's disease (CERAD) standardized test for A β ¹². The CERAD criteria is a semi-quantitative test done by a neuropathologist to assess likelihood of AD diagnosis based on a count of the sheer number of A β plaques. Secondly, performing immunohistochemistry (IHC) of brain tissue sections using the antibody AT8¹³, which recognizes Tau that is phosphorylated at residues S202/T205, Heiko Braak established a semi-quantitative, chronological staging system of AD pathology¹⁴. Braak demonstrated that Tau pathology begins in the transentorhinal cortex (Stages I, II), before moving to the hippocampus (Stage III), temporal lobe (Stage IV), occipital lobe (Stage V), and finally, to primary areas of the neocortex (Stage VI). This assay has become a routine part of post-mortem examination of a demented individual, and the gross degree of Tau pathology is named the "Braak score" after its namesake. The enumeration of discrete stages of A β and Tau pathology has provided a means to stratify brain samples in the experimental design process and understand fundamental physiological changes as a result of A β and Tau aggregation.

Reduced clearance of A β oligomers in extracellular space results in the formation of A β plaques¹⁵. This event is thought to trigger pathogenesis through a network of signal transduction events, leading to the accumulation of neurofibrillary Tau tangles¹⁶. Despite the known role of A β

in AD pathogenesis the downstream events that contribute to neurotoxicity and cognitive decline still remain fundamental and poorly understood¹⁷. Notably, Tau correlates more strongly with cognitive decline in AD than A β plaques⁷. Defining the mechanisms that link A β plaque formation to downstream Tau aggregation may identify potential therapeutic targets and biomarkers that aid in the prevention and treatment of AD.

1.2.2 Genetic heterogeneity of Alzheimer's disease

Alzheimer's can be broadly characterized into those with familial, early onset AD (fAD) and late onset, sporadic AD (sAD)¹⁸. Autosomal dominant fAD accounts for ~2% of all cases, with over one thousand families worldwide harboring confirmed mutations¹⁹. Individuals with fAD develop symptoms between the ages of 30 and 60 years old¹⁸, experiencing rapid behavioral decline and dying within several years of symptom onset. In 1987 Jie Kang and others identified the gene located on chromosome 21 that encodes A β and named it APP²⁰. Using cell lines derived from members of multiple families with early onset fAD, researchers attempted to find autosomal dominant mutations that cause AD. In 1991, Alison Goate and John Hardy described a family with early onset fAD with a mutation in *APP*, called the "London Mutation", causing a V717I substitution in APP²¹. In 1995, Robin Sherrington and Peter St. George-Hyslop discovered a AD familial mutation in *PSEN1*²², a gene located on chromosome 14 which encodes for presenilin 1 protein. Investigation by Rudolph Tanzi and his team into a group of German families known to have settled on the Volga River identified a mutation in the *PSEN2* gene²³, located on chromosome 1, that encodes for presenilin 2. Future research established the relationship between these three genes, as PSEN1 and PSEN2 make up the catalytic components of γ -secretase, which assists in the cleavage of APP into A β ². Animal transgene models of the mutations described developed amyloid lesions that mirrored those observed in AD and exhibited learning disabilities^{24,25}.

Late onset, sporadic AD (sAD) is the most common form of AD, accounting for the remaining 98% of cases²⁶. Multiple gene loci are implicated in late onset sAD and it is thus understood as a polygenic disorder²⁶. The most common genetic susceptibility for sAD is the inherited variant of apolipoprotein E (APOE)²⁷. APOE is a protein expressed in astrocytes and oligodendrocytes that contributes to lipid transport and has a central role in neuronal injury repair^{28,29}. Humans inherit three different variants of APOE that differ at residues 112 and 158 of the protein: APOE2 (C112/C158), APOE3 (C112/R158) and APOE4 (R112/R158)^{30,31}. Importantly, studies in animal models demonstrate that APOE increases clearance of A β ³². Inheriting one copy of the APOE4 allele triples an individual's risk for developing sAD, while two copies increase risk 12-fold³³. In contrast, inheritance of the APOE3 allele is believed to have no effect on the risk of developing AD, and the APOE2 allele may be protective³⁴. Interestingly, in 2019 the investigation into a Columbian family with fAD pedigree revealed a PSEN1 E280A mutation carrier was discovered to also carry two copies of the "Christchurch" mutation in APOE that caused R136S substitution³⁵. This individual resisted cognitive decline until her mid-seventies, more than three decades after typical symptom onset.

Although genetic studies support that APOE exerts the largest role in modifying an individual's risk of developing sAD³⁶, approximately half of individuals with sAD do not harbor a APOE4 allele^{27,34} indicating that many other loci influence sAD development. Innovations in genomic sequencing technology have allowed the identification of genetic polymorphisms linked to AD using Genome Wide Association meta-analysis Studies (GWAS)³⁷⁻³⁹. Importantly, a caveat of GWAS is that genetic loci are discovered with proximity to single nucleotide polymorphisms (SNPs) which requires in-depth examination of the loci to determine which gene is truly associated with a disease or disorder⁴⁰. In light of this however, many new genes have been implicated with

AD in this manner, pointing to dysregulation of microglia (CD33, SHARPIN, TREM2), α -secretase (ADAM10, ADAM17, TSPAN14), endocytosis (BIN1, PICALM, WDR81), the lysosome (CTSB, CTSH, IDUA, TMEM106B) and sorting receptors (SORL1, SORT1, SNX1)³⁷⁻³⁹. More genes are expected to be associated with AD with increased GWAS sample sizes, although the genes that have been identified are implicated in processes that are already intense areas of study.

With the advent of next generation sequencing methods researchers have revisited samples taken from Auguste Deter and attempted to identify a hypothesized autosomal dominant mutation that brought on early onset AD that was described by Dr. Alzheimer. Genotyping of DNA extracted from histological sections from Auguste Deter confirm her genotype as APOE3/3, demonstrating that she had no increased risk of developing AD based on her APOE status⁶. Instead, German researchers discovered Mrs. Deter carried a novel T→C mutation in the *PSEN1* gene, causing an F176L substitution in the third transmembrane domain of the PSEN1 protein⁴¹. While a novel mutation, the F176 residue neighbors mutation-induced substitution events in other individuals who developed early onset fAD⁴². Despite many polymorphic genes associated with AD, there remain only three genes (*APP*, *PSEN1*, *PSEN2*) involved in APP processing which, when mutated, cause autosomal dominant forms of AD.

1.3 Amyloid Cascade Hypothesis

The Amyloid Cascade Hypothesis asserts that A β accumulation is the first event that triggers AD pathogenesis, leading to downstream Tau neurofibrillary tangle formation, neurodegeneration and dementia that follow thereafter as illustrated by the Jack model (**Fig.1.1**)¹⁶. Developed by John Hardy & Gerald Higgins, this hypothesis has become a key framework for understanding AD. This model separates individuals into three discrete stages of disease, including pathologically-

/clinically-normal controls followed by asymptomatic AD (AsymAD), a prodromal phase with robust A β deposition but normal cognition, and finally, the AD stage wherein Tau NFT load is substantial and cognition sharply declines^{12,14}. The molecular mechanisms that directly link early A β deposition to Tau NFT formation and cognitive decline are unknown, however. In spite of this, the hypothesis is buttressed foremost by the fact that only three genes (*APP*, *PSEN1*, *PSEN2*), when mutated, cause autosomal dominant inheritance of fAD through the unified biological mechanism of generating A β polypeptides³⁶. Any and all mutations in other genes confer increased risks for developing late onset sAD yet do not cause early-onset fAD. Secondly, mutations in microtubule associated protein Tau (*MAPT*) gene do not cause AD but instead causes a subset of a related but distinct tauopathy called frontotemporal lobar dementia (FTLD)⁴³ marked by the development of tangles, cell death and dementia with no A β pathology observed. Therefore, this demonstrates that Tau pathology is downstream from initial A β plaque deposition. Finally, crossing mice with human *APP* mutant transgenes onto a Tau transgenic background triggers Tau tangle formation⁴⁴, while reduction of *APP* dosage alleviates this process⁴⁵.

1.3.1 Limitations of the Amyloid Cascade Hypothesis

While many researchers attempt to understand AD through the Amyloid Cascade Hypothesis, there remains questions surrounding the mechanism of AD pathogenesis. First and foremost, we do not completely understand why many adults accumulate A β in their brain with no apparent cognitive deficit^{7,46,47}. This finding suggests that A β plaque deposition is necessary but not sufficient alone to fully trigger Tau aggregation AD. In support of this, mouse models that express human *APP*, *PSEN1* or *PSEN2* mutant transgenes have incompletely recapitulated the cellular and behavioral phenotypes of AD⁴⁸. In addition, none of these models in isolation develop Tau NFTs, the other core arm of AD pathology⁴⁵. Therefore, one may deduce that we do not currently

understand the events which directly link A β deposition to later Tau neurofibrillary tangle formation, and that these mechanisms may be unique to the human brain⁴⁷. This thesis in large part attempts to identify alternative proteins and mechanisms that link A β and Tau aggregation.

A significant piece of evidence that casts doubt on the Amyloid Cascade Hypothesis is the lack of success in A β -targeting therapies in human trials. As of 2021, 66 drugs that target A β have undergone clinical trials, and only one, aducanumab or Aduhelm, has been passed by the Food and Drug Administration (FDA) for use⁴⁹. As mentioned previously, aduhelm does lower cerebral A β levels as measured by (18)F-florbetapir positron emission tomography (PET) imaging during clinical trials⁵⁰. However, there are significant concerns relating to the validity of the effects on patient symptom recovery with the use of aducanumab despite FDA approval⁵¹. The failure of many of the drugs that target A β demonstrate that a better understanding of the mechanisms that lead to Tau neurofibrillary tangle formation is needed.

1.4 Links between RNA-binding proteins and neurodegenerative disease

Recently, many groups have discovered several RNA-binding proteins (RBPs) that appear to be involved in the pathology of neurodegenerative diseases^{52,53}, including AD⁵⁴. RBPs comprise a growing list of over 2,000 proteins, which are essential for RNA metabolism⁵⁵. RBPs surveil and coordinate the RNA life-cycle, regulating the transcription, splicing, transport, translation and stability of RNA molecules⁵⁶. RBPs are therefore critical to sustaining basic cellular health. Several neurodegenerative diseases, including Alzheimer's disease, are promoted by the cytoplasmic accumulation of RBP aggregates via disordered low complexity domains⁵⁷. Other examples include amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), wherein TAR DNA-binding protein of 43 kDa (TDP-43)^{52,58} and Fused in sarcoma (FUS)⁵⁹ accumulate as insoluble cytoplasmic inclusions in degenerating regions of the brain. Given RBP

mutations found associated with disease⁵⁹⁻⁶³ and the conspicuous presence of RBP aggregates in deteriorating regions of the brain and lack thereof in unaffected areas in different neurodegenerative diseases, it is hypothesized that RBPs play a key role in these diseases. It has previously been unclear, however, what shared property contributes to RBP aggregation in neurodegenerative disease.

1.4.1 Granules to Fibrils: an equilibrium of Liquid-liquid phase separation (LLPS)

RBPs commonly contain regions of intrinsically disordered low complexity (LC) domains^{55,64-66}. LC domains participate in a plurality of interactions between self, RNA molecules and other RBPs via weak, multivalent interactions^{67,68}. The multivalent interactions can thus facilitate liquid-liquid phase separation (LLPS), a process of molecules de-mixing into soluble molecules and other molecules that compartmentalize into insoluble condensates *in vitro*⁶⁷. This process also occurs *in vivo*, in which multivalent interactions between LC domains and proteins and RNA molecules drive the formation of membrane-less organelles, otherwise called granules⁶⁸. Granule structures in the cell include nuclear speckles, nucleoli, stress granules and P bodies⁶⁹. Cellular granule structures are characteristically round and highly dynamic *in vivo*, as fluorescence recovery after photobleaching (FRAP) experiments demonstrate rapid reorganization of granule structures⁷⁰. Although we have discovered the various dynamic states RBPs can form *in vivo*, it is not fully understood how RBPs are prevented from forming stable aggregate structures.

1.4.2 RNA-binding protein pathology in Alzheimer's disease

It is not completely known how proteins transition from oligomeric granule structures into stable, irreversible aggregates. Numerous chaperone proteins are thought to regulate the process of granule reorganization, through unfolding and re-folding granule protein constituents⁷¹. It is hypothesized, however, that deficiency of chaperone reorganization of granule structures and

proteins within them cause proteins with LC domains to irreversibly form inclusions and fibrils alike⁷². FRAP experiments on inclusion and fibril structures highlight the stability of these structures, as little to no reorganization (fluorescence recovery) occurs as compared with granule structures⁷³. Although LC domains are typically intrinsically unstructured under homeostatic conditions *in vivo*, in neurodegenerative disease these regions polymerize into detergent-insoluble fibril-like aggregates⁷⁴.

Across numerous neurodegenerative diseases, researchers have observed several shared, staple RBP pathologies which include: 1) increased evidence of phase transitions or aggregate formation; 2) altered cellular interactions; 3) cellular mislocalization; and, 4) RNA metabolism defects (**Fig.1.2**)⁷⁵. Several groups have identified numerous RNA-binding proteins that exhibit these pathological phenotypes in diseased regions of the brain. For example, TDP-43 pathology is observed in a large percentage of AD patients and is believed to contribute to poor clinical outcomes relative to individuals without TDP-43 immunoreactivity post-mortem⁷⁶. Also, the Wolozin group has identified several stress granule RNA-binding proteins that functionally and/or physically interact with Tau, including TIA1, EWSR1, HNRNPA2B1, and G3BP1⁷⁷⁻⁸⁰. What's more, modulation of the expression of stress granule RBPs TIA1 and G3BP1 prevents the accumulation of Tau oligomers⁸⁰. Recently, HNRNPA2B1 was found to associate with oligomeric Tau in neurons, animal models and in human AD brain, serving to link oligomeric Tau to methylated RNA transcripts⁷⁸. The Parker group identified the nuclear scaffold protein SRRM2 as an RBP that mislocalizes and associates with Tau in models of pathological Tau⁸¹. Immunoprecipitation of Tau from AD brain extracts suggests that other RBPs may in fact exhibit similar changes in solubility, cellular interaction and localization profiles⁸². In light of these

findings, the investigation of mechanisms that contribute to the process of RBP aggregation is a high priority in the field of neurodegenerative disease research.

1.5 Using mass spectrometry-based proteomics to investigate the pathophysiology of Alzheimer's disease

A relatively recent innovation in biochemistry was the introduction of mass spectrometry (MS) which revolutionized the field of proteomics⁸³. Proteomics is the study of the proteins that constitute a biological system or organism. The collection of proteins in a sample or a subject or an organism is defined as the “proteome”. Conventionally, a user collects a protein sample, then reduces and alkylates the proteins and digests the proteins into peptides with a protease which are run by liquid chromatography and sprayed into a mass spectrometer while being ionized and converted to gas phase⁸⁴. The peptide ions are then measured by their mass:charge ratio, selected and fragmented further to read fragment mass:charge ratios in order to “sequence” the residue-string of the peptide and “fingerprint” the peptide and the protein from which it originated in the sample. Mass spectrometry can provide not only information regarding protein identity and abundance but also cellular location, interactions, activity, modification state and turnover rate⁸⁵. Mass spectrometry remains the most sensitive and powerful proteomic tool, capable of measuring analytes in the range of attomolar (10^{-18} M) concentrations⁸⁶. Moreover, MS can measure thousands of different proteins in a single sample⁸⁷. The sensitivity of the technology also buttresses mass spectrometry as relatively unbiased, meaning that generally no one protein is preternaturally identified and measured over another protein. This allows one to perform proteome-wide weighted correlational network analysis (WGCNA)⁸⁸, as applied to AD recently, to identify novel proteins that may be involved in the disease⁸⁹. As neurodegenerative diseases are

chiefly proteinopathies, mass spectrometry-based proteomics is an ideal tool to discover disease mechanisms and pathological progression of AD.

1.5.1 Mass spectrometry-based proteomics reveals novel protein signatures in Alzheimer's disease pathogenesis

Although researchers could visualize protein aggregates in AD brain using histological methods for decades, there were many technical challenges that prevented the identification of the components of those aggregates. Adapting a procedure originally used to isolate tangles⁹⁰, research groups began to analyze detergent-insoluble extracts of tissue by MS in neurodegenerative diseases⁵⁴. Analysis of other neurodegenerative diseases were conducted, including Parkinson's and dementia with Lewy bodies^{91,92}, progressive supranuclear palsy⁹³, corticobasal degeneration⁹⁴, FTLN and ALS⁵² and chronic traumatic encephalopathy⁹⁵. Importantly, these studies identified the enrichment of RBPs to the insoluble proteome in neurodegenerative diseases. These studies demonstrate that specific groups of proteins, particularly RBPs, become insoluble in each neurodegenerative disease, suggesting unique mechanisms of disease in each.

1.5.2 Identification of U1 small nuclear ribonucleoprotein spliceosome and arginine-rich RNA-binding proteins enriched in Alzheimer's disease insoluble fractions

Using the simplified detergent fractionation protocol described earlier, we isolated aggregated proteins from AD brain extract as compared with tauopathy controls. Our group then leveraged MS-based proteomic investigations of these detergent-insoluble extracts^{54,95-97}. Among the most enriched proteins in AD extracts include APP and MAPT protein. A group of nuclear arginine-rich RBPs that constitute the U1 small nuclear ribonucleoprotein (snRNP) subunit of the spliceosome and proteins associated with the spliceosome, namely U1 snRNP 70kDa protein (U1-70K), U1 snRNP A protein (SNRPA) and lethal unless Cap-binding complex is produced like

protein 3 (LUC7L3), were similarly enriched in the detergent-insoluble fractions of AD brain. Western blot of the brain extracts confirmed the substantial enrichment of U1-70K to the detergent-insoluble pellet (**Fig.1.3a**)⁹⁸. Examination of brain samples from an early onset fAD individual with a *PSEN1* mutation and an individual with Down's Syndrome (Trisomy 21) demonstrated a similar enrichment of the same group of RBPs⁹⁹, demonstrating that arginine-rich RBP aggregation is a core pathological mechanism that is inherent to both familial and sporadic versions of AD. Intriguingly, these proteins were also enriched in AsymAD brains, with robust A β pathology yet little to no Tau pathology⁹⁶. Indeed, U1-70K protein is the most highly correlated with A β insolubility⁹⁶. This collection of evidence suggests that this group of RBPs may be mediators that connect an initial A β trigger of AD to the later formation of Tau neurofibrillary tangles.

Working from that hypothesis, our group set out to investigate whether these proteins associated with either A β or Tau and by what mechanisms. Both IHC and immunofluorescence (IF) staining confirmed that the nuclear RBP U1-70K mislocalizes to the cytoplasm and associates with Tau NFTs in AD brain sections (**Fig.1.3b-e**)^{54,98}. Immunopurifying myc-tagged U1-70K spiked into AD brain extracts shows that the LC1/BAD domain of U1-70K interacts with AD Tau (**Fig.1.3e**)¹⁰⁰. Similarly, cross-linking of the LC1/BAD domain of U1-70K with Tau was achieved by photo-activation of an NHS-diazirine linker⁹⁸, further demonstrating that U1-70K interacts with Tau via its LC1/BAD domain. In certain neurons, however, U1-70K can be found to aggregate with no evidence of Tau NFT formation⁹⁶, suggesting U1-70K can independently aggregate and may in fact precede or even seed Tau aggregate formation. Moreover, in collaboration with Joshua Shulman (Baylor College of Medicine), our lab has shown that in a *Drosophila* model where human Tau is ectopically over-expressed, loss of U1-70K exacerbates Tau-mediated toxicity and

neurodegeneration⁸². Unspliced and cryptic RNA splicing events are also found significantly elevated in *Drosophila* models of tauopathies and in AD brain and correlate with Tau pathology, further supporting the role of RBP loss-of-function in AD^{82,101,102}. Given the critical role of arginine-rich RBPs in regulating Tau toxicity and RNA splicing, and the large-scale RNA splicing deficits in AD brain, a better understanding of arginine-rich RBP aggregation may reveal new mechanisms that link amyloid plaques to Tau aggregation in AD pathogenesis¹⁰³.

1.6 The spliceosome and associated arginine-rich splicing factor RNA-binding proteins

A large proportion of the protein constituents of the spliceosome, the macromolecular machine that performs RNA splicing, are RBPs (**Fig.1.4**)¹⁰⁴. A core tenet regarding the genomes of different organisms is that the “complexity” of an organism is not proportional to the number of genes encoded. Instead, the genome utilizes several innovative post-transcriptional regulatory methods to expand the number and variety of transcribed and translated gene products including the process of pre-mRNA splicing, which removes non-coding intronic sequences from transcribed precursor messenger RNA (pre-mRNA) molecules¹⁰⁵. Although canonical forms of mRNA, otherwise called isoforms, exist for many genes, alternative splicing of pre-mRNA creates novel variations of mRNA and proteins. In fact, as many as 90% of all human genes undergo alternative splicing^{106,107}, and this process is hypothesized to contribute to organ and species differentiation.

Although over 2,000 proteins are believed to associate with mRNA, it is incompletely known which proteins directly bind mRNA and operate the cellular splicing machinery. The spliceosome is a macromolecular cellular machine composed of both nuclear RBPs and small nuclear RNAs (snRNA)¹⁰⁸. The spliceosome consists of five small snRNP subunits and a common Smith antigen (Sm) ring core¹⁰⁴. Many spliceosome-constituent and -associated RBPs contain domains with an elevated frequency of arginine residues that are able to activate splicing¹⁰⁹. These proteins have

been shown to alter splicing as both direct members of the spliceosome and as *trans*-acting factors, such as serine-/arginine-rich (SR) proteins, that bind *cis*-acting splicing regulatory elements¹¹⁰. Importantly, many arginine-rich RBPs including U1-70K and other SR proteins are increasingly insoluble in AD brain¹⁰⁰, suggesting a possible shared biological role in disease.

1.6.1 Discovery of arginine-rich splicing factor RNA-binding proteins

Proteins containing arrays of arginine-/serine-rich (RS) dipeptide domains were originally discovered in *Drosophila* as splicing regulators, including suppressor-of-white-apricot (su(w[a])) and the sex-determining regulators transformer (tra) and transformer-2 (tra2)¹¹¹. Afterwards, the proteins U1-70K^{112,113}, ASF/SF2 (SRSF1)^{114,115} then SC35 (SRSF2)¹¹⁶ were discovered to contain RS domains (**Fig.1.5, Table 1.1**). Historically, SR proteins have been formerly grouped according to numerous qualifications, which include but were not limited to: 1) containing a phosphoepitope labeled by mAb104 antibody which labels lateral loops on amphibian lampbrush chromosomes¹¹⁷; 2) purifiable by a salt (MgCl₂) precipitation procedure; 3) can complement splicing-incompetent S100 extracts; 4) contain at least one RRM domain and an RS domain; 5) their molecular sizes on SDS-PAGE are conserved from *Drosophila* to mammals^{118,119}. Due to the ever changing standards of what consisted an SR protein there is confusion about how, not only to categorize these proteins, but how to understand shared and non-shared biological roles. The current 12 member family of serine-/arginine-rich splicing factor proteins (SRSF1-SRSF12) were ultimately defined by protein sequence criteria put forth by Manley and Krainer which include: 1) containing one or two N-terminal RRM domain(s), and 2) containing a C-terminal RS domain greater than 50 residues comprised of more than 40% arginine and serine¹²⁰.

1.7 Regulation of arginine-rich RNA-binding protein function by post-translational modification (PTM)

Although we have known for decades that splicing factors are regulated broadly by post-translational modifications (PTMs), especially phosphorylation, we still do not understand residue-specific PTM regulation of the behavior of certain essential proteins. Arginine-rich proteins are known to be phosphorylated, methylated and acetylated¹²¹⁻¹²⁴. The most complete characterizations of PTM regulation of arginine-rich proteins have focused on phosphorylation. The earliest efforts to develop antibodies raised against SR proteins were discovered to recognize phosphoepitopes^{117,118}. The initial discovery of serine-/arginine-rich protein kinase 1 (SRPK1) revealed the phosphorylation status of SR proteins increased across discrete stages of the cycle, cratering during G₀ while peaking during mitosis¹²⁴. Interestingly, different cell types harbor different ranges of SR protein phosphorylation, with HeLa cells being relatively phosphatase resistant *in vitro* whereas other cell types (pluripotent P19 cells) harbored more varied levels of SR protein phosphorylation¹²⁵. Interestingly, the differentiation of P19 cells into neuronal cells increased phosphorylation of SRSF2 and SRSF5¹²⁵. These observations suggest phosphorylation is altered in different cellular contexts and in different tissues for some undiscovered, variable functions.

1.7.1 PTM regulation of spliceosome assembly and canonical/alternative splicing

Among the functions governed by phosphorylation is splicing^{126,127}, the core function shared between arginine-rich RBPs. Both phosphorylation and dephosphorylation regulate splicing and splice site selection¹²⁷. It was discovered early on that an important consideration of splicing is the importance of *reversible* phosphorylation, especially in the context of spliceosome assembly and activity. Splicing extracts, when dephosphorylated by PP1 and PP2A, could not complete splicing reactions of precursor mRNA (pre-mRNA)^{126,127}. Furthermore, the phosphatase PP1 was found to block the formation of the pre-spliceosome complex as well as the mature

spliceosome^{127,128}. Interestingly, while SRPKs are found to associate with multiple individual subunits of the spliceosome¹²⁹, phosphatases are not. Nevertheless, it has been concluded that a series of phosphorylation and dephosphorylation events are required for the formation of the spliceosome and splicing activity to complete¹³⁰, meaning that pools of differentially phosphorylated SR proteins exist in the cell.

1.7.2 Post-translational modification (PTM) regulation of RNA-binding protein LLPS

Importantly, PTMs not only regulate canonical RNA metabolism functions of RBPs, but they critically alter the potential of proteins to transition to solid-like fibrils in different neurodegenerative diseases (**Fig.1.6**)^{75,131}. In AD, for example, APP is post-translationally cleaved by beta- and gamma-secretase, generating A β 1-38, 1-40 and 1-42 oligomeric species¹³². These species are demonstrated to be aggregation-prone, with the 1-42 species being the most amyloidogenic¹³³. Furthermore, Tau has been found to be extensively modified by PTMs in AD, including phosphorylation¹³⁴, ubiquitination¹³⁵, acetylation¹³⁶ and methylation¹³⁷. Similarly, other proteins that aggregate in different neurodegenerative diseases, including alpha-synuclein in Parkinson's disease (PD) and FUS and TDP-43 in ALS and FTL, are robustly modified by PTMs (**Fig.1.6**)¹³⁸. What is not completely understood is whether PTMs accelerate protein aggregation, reverse aggregate conformational species, and/or serve to mark toxic species for degradation.

We hypothesize that alteration in physiological PTM status is a fundamental trigger of protein aggregation. In order to investigate this question, researchers have modeled LLPS *in vitro* and asked whether phase transition events are modified by PTMs¹³⁹. Perhaps more importantly, the question of whether addition or subtraction of PTMs fuel this process remains undetermined as there is a collection of conflicting evidence supporting either alternatives. Whereas multiple groups independently provide evidence that phosphorylation promotes Tau LLPS^{140,141}, many

groups have also found that multiple different types of PTMs actually prevent phase transitions in similar, yet independent proteins. For example, recently it was demonstrated that phosphorylation prevents Tau protein aggregation and LLPS *in vivo* and *in vitro*, respectively^{142,143}, contrary to prevailing evidence. Other groups demonstrated that hyperphosphorylated RBPs including FUS and TDP-43 show diminished capacity to phase separate and aggregate as compared to wild type control protein¹⁴⁴⁻¹⁴⁶. Arginine methylation generally tends to suppress LLPS, as it has been independently shown to prevent phase separation and promote chaperone binding in multiple RBPs including FUS, DDX4 and HNRNPA2¹⁴⁷⁻¹⁵⁰. Having established that PTMs indeed alter the kinetics of LLPS formation in many RBPs despite the conflicting reports of positive or negative regulation of this hypothesis, we hypothesized that PTMs could similarly alter arginine-rich RBP aggregation in AD. However, a limitation to understanding the role of protein phosphorylation within arginine-rich RBPs and related aggregation-prone RBPs in neurodegenerative disease, has been the inability to effectively map sites of phosphorylation using traditional MS approaches, due to high arginine density within these LC domains.

1.7.3 Limitations of using mass spectrometry-based proteomics to identify PTM sites in arginine-rich RNA-binding proteins

Although several lines of evidence highlight the critical importance of PTMs, especially that of phosphorylation, to modulate the behavior of arginine-rich RBPs, attempts at sequencing exact sites of PTMs within arginine-rich domains by proteomic methods have proved lacking. Modern, high-throughput mass spectrometry methods conventionally rely on the bottom-up proteomic approaches, which consists of complete digestion of protein samples with the serine protease trypsin, which cleaves C-terminal to arginine and lysine residues. Typically, this method digests proteins into peptides which are of appropriate length and charge to be faithfully sequenced by

mass spectrometry. Although this strategy is highly effective at sequencing proteins with a normal distribution of arginine and lysine, it proves inadequate at sequencing proteins especially enriched with arginine and lysine, namely arginine-rich RBPs. Additionally, the ability to identify PTMs that may regulate critical pathogenic processes in disease are lost with the inability to sequence arginine-rich LC domains of RBPs. Therefore, a large amount of information critical to understanding arginine-rich RBPs cannot be observed through current proteomic sequencing.

1.8 Research Focus and Innovation

In this thesis, I report a middle-down (partial proteomic digestion) proteomic approach that, when combined with electron transfer dissociation (ETD) fragmentation, yields significantly improved coverage of arginine-rich LC domains of RBPs in mammalian HEK293 cell extracts. The human embryonic kidney cell line was chosen over more disease-related neuronal cell lines for several reasons, including that RBPs are ubiquitous to all cell types, the ease of transfection, particularly high expression of RBPs and the high degree of phosphorylation of SR proteins in HEK293 cells compared to other cell lines (e.g., HeLa cells). We also find that in exemplary mammalian cell extracts, arginine-rich LC domains are particularly enriched with PTMs, including phosphorylation and methylation. In light of the observed enrichment of phosphorylation within arginine-rich LC domains, we find that dephosphorylation of arginine-rich RBPs contributes to enhanced collaborative interactions. Building on this, we use detergent fractionation, oligomerization assays, mass spectrometry, correlational network analysis, immunocytochemistry and immunoprecipitation to demonstrate that phosphorylation regulates arginine-rich RBP solubility and oligomerization, particularly that of SRSF2 protein. This work has led us to explore the magnitude of phosphorylation in arginine-rich RBPs and the extent to which phosphorylation regulates phase transitions, cellular localization and interactions of these critical proteins.

1.9 FIGURES AND TABLES

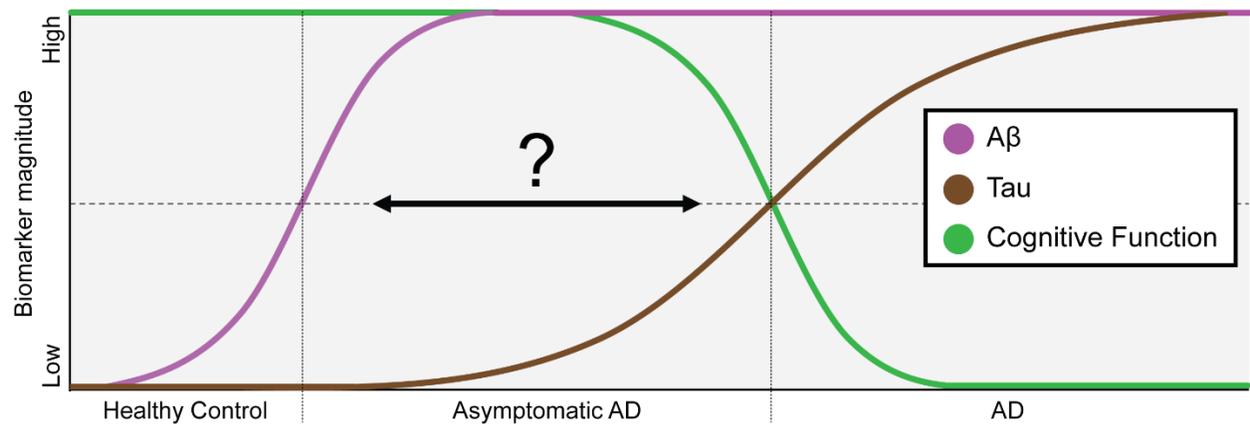


Figure 1.1. Jack curve model of AD. Theoretical model of AD biomarker magnitude over an individual's lifetime as measured by positron emission tomography (PET) imaging, ELISA and mass spectrometry ¹⁵¹. Elevated levels of amyloid beta ($A\beta$; *magenta*) deposited in the cortex begin to be observed in the third or fourth decade of an individual's lifetime. During a prodromal phase of AD, called asymptomatic AD (AsymAD), cognitive function (*green*) remains steady in the face of growing Tau neurofibrillary tangle (NFT) pathology (*brown*). Eventually, the increasing Tau pathology leads to increased neuronal death and neurodegeneration, at which point a clinical diagnosis of AD can be made.

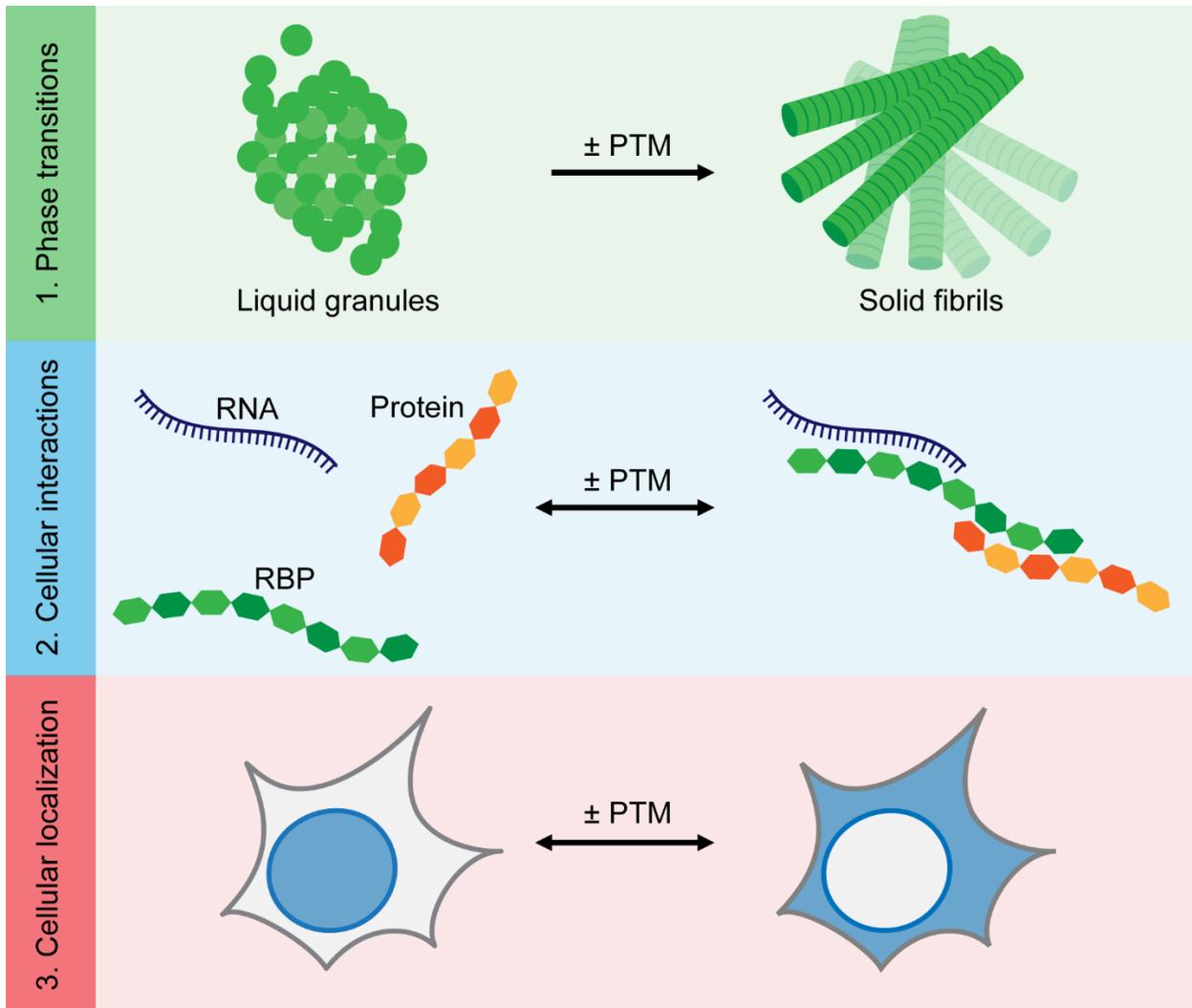


Figure 1.2. Sternburg model of arginine-rich RNA-binding protein phase separation continuum. Core pathological phenotypes of RNA-binding proteins (RBPs) observed in neurological disease ⁷⁵. RBPs use liquid-liquid phase separation to compartmentalize into non-membrane organelles or liquid granules, and can phase transition into solid fibrils if not constantly disassembled and reassembled. RBPs can also become mislocalized and form novel protein and RNA interactions under stress and disease conditions. All these processes are demonstrated to be altered by the presence or absence of post-translational modification (PTM).

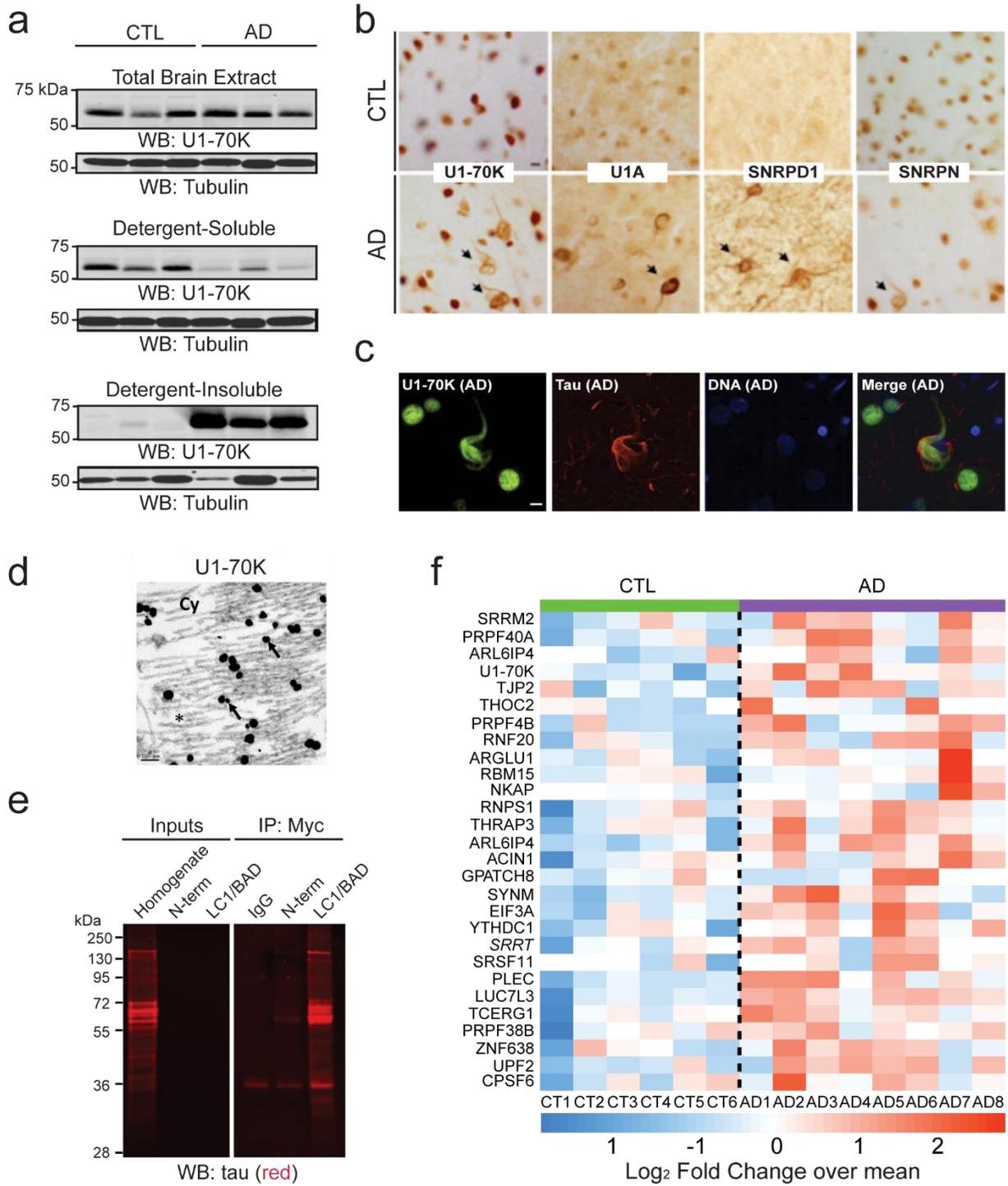


Figure 1.3. Arginine-rich RNA-binding proteins aggregate in Alzheimer's disease. (a) Immunoblotting of sarkosyl detergent fractions of extracts from control (CTL) and Alzheimer's disease (AD) brain tissue for U1 small nuclear ribonucleoprotein 70 kDa (U1-70K) and the fractionation procedure control, tubulin⁹⁸. (b) Immunohistochemical (IHC) staining of U1-70K, U1A, SNRPD1 and SNRPN in CTL and AD frontal cortex paraffin sections⁵⁴. (c) Immunofluorescence (IF) staining for U1-70K (*green*) and paired helical filament (PHF) Tau (*red*). Nuclei (*blue*) stained by Hoechst dye⁵⁴. (d) Electron microscopy imaging of immunogold labeled U1-70K in AD frontal cortex ultrathin sections⁹⁸. (e) Western blotting for Tau (*red*) of input and myc immunoprecipitation samples of recombinant truncation U1-70K proteins spiked into AD brain homogenate¹⁰⁰. (f) Heat map illustrating log₂ fold changes of U1-70K and homologous arginine-rich RNA-binding protein (RBP) abundances in the detergent insoluble pellets of CTL and AD brain tissue extracts. A list of 255 homologous arginine-rich RBPs were determined by sharing > 20% similarity to the LC1/BAD domain of U1-70K (E-value < 0.005) by the Uniprot protein BLAST¹⁰⁰.

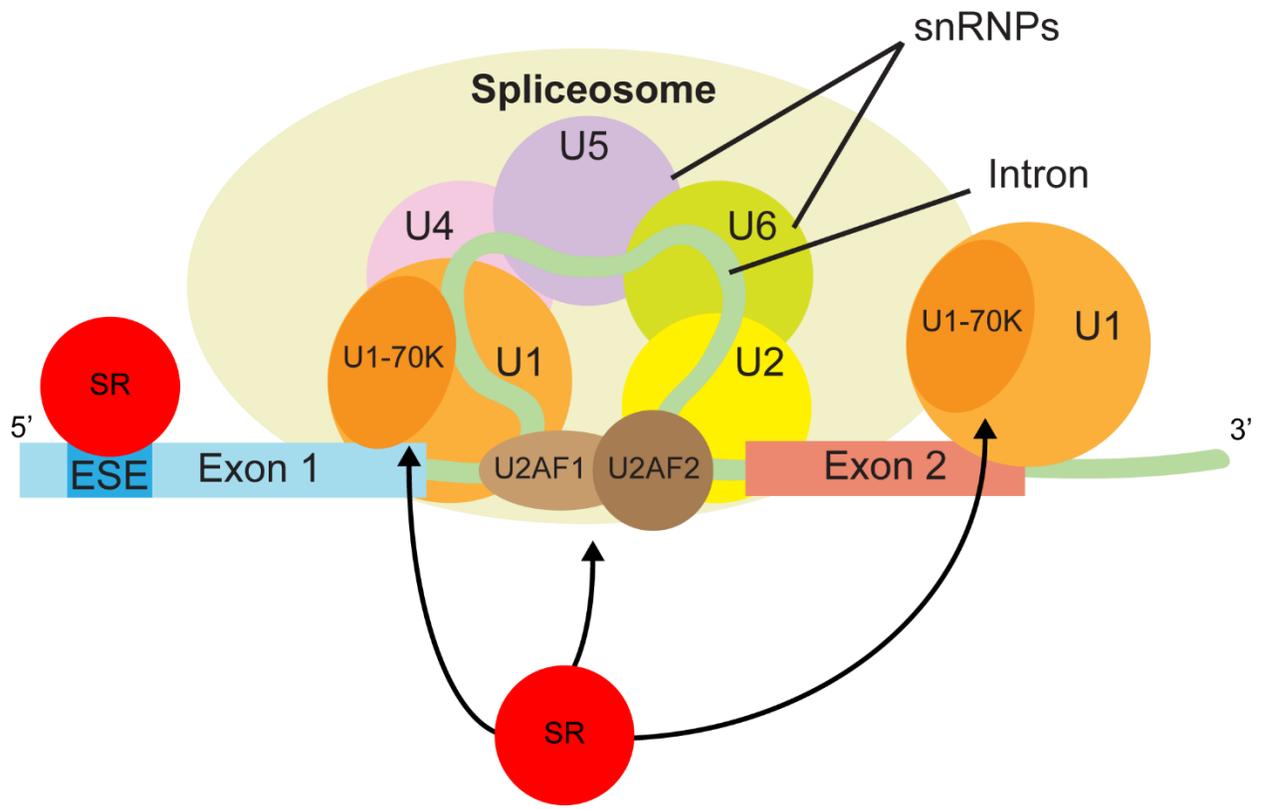


Figure 1.4. A network of interactions between arginine-rich RNA-binding proteins facilitate spliceosome assembly and pre-mRNA splicing. The spliceosome is composed of five different small nuclear ribonucleoprotein (snRNP) subunits (U1, U2, U4, U5, U6) and small nuclear RNAs that help to assemble the machine on precursor messenger RNA (pre-mRNA) molecules. The U1 snRNP binds to the 5' exon-intron splice site boundary through complementary base pairing of the U1 snRNA to the pre-mRNA. The U2AF heterodimer (U2AF1, U2AF2) bridge the 5' and 3' splice sites to maintain interactions between exons. The U1-70K (*orange*) and the U1 snRNP is recruited to exons by coordinated interactions with SR proteins (*red*) via arginine-rich low complexity domains.

U1-70K



SRSF2



Figure 1.5. Comparison of prototypical BAD and SR arginine-rich RNA-binding proteins U1-70K and SRSF2. Protein sequences of the Low Complexity 1 Basic-Acidic Di-peptide (LC1/BAD) domain of U1-70K and the aRginine-/Serine-rich domain of SRSF2.

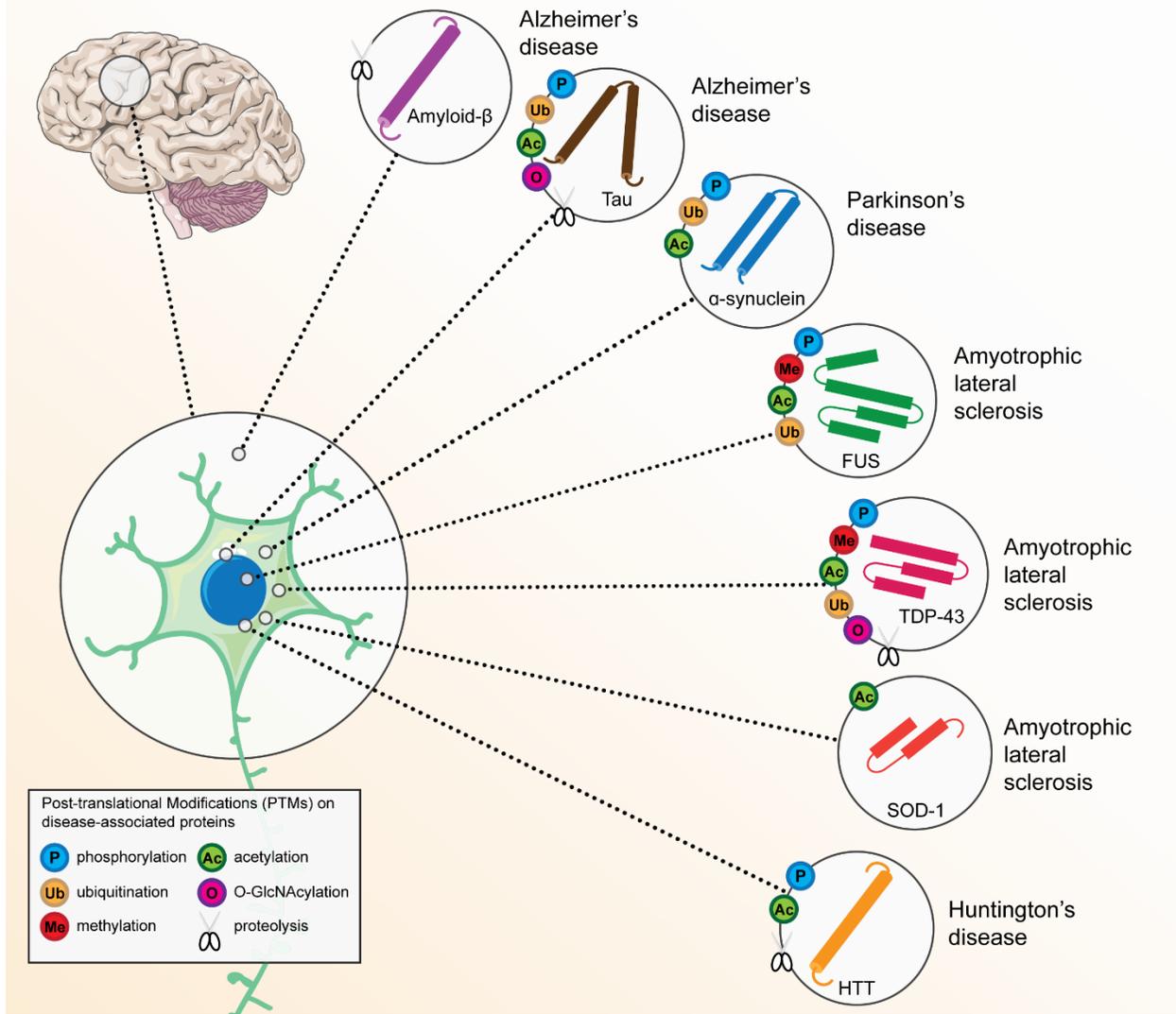


Figure 1.6. Post-translational modifications linked to human neurodegenerative disease. Many proteins causatively linked to neurodegenerative diseases harbor post-translational modifications that contribute to the pathology of the protein¹³⁸. Namely, amyloid precursor protein (APP) is proteolytically cleaved by beta- and gamma-secretase to produce A β polypeptides, which form oligomeric species and eventually A β plaques. Researchers have discovered that Tau in AD brain is hyperphosphorylated, ubiquitinated, acetylated and glycosylated. Proteins involved in amyotrophic lateral sclerosis (FUS, TDP-43, SOD-1) and Huntington's (HTT) also contain PTMs that modify LLPS and cellular localization and interactions.

TABLE 1. A superfamily of arginine-rich pre-mRNA splicing factor RNA-binding proteins.

Class I: SR proteins	References	Class II: Arg-rich proteins	References
SRSF1	Ge <i>et al.</i> (1991) Krainer <i>et al.</i> (1991)	U1-70K	Theissen <i>et al.</i> (1986) Spritz <i>et al.</i> (1987)
SRSF2	Fu & Maniatis (1992)	U2AF1	Zhang <i>et al.</i> (1992)
SRSF3	Zahler <i>et al.</i> (1992)	U2AF2	Zamore <i>et al.</i> (1992)
SRSF4	Zahler <i>et al.</i> (1993)	LUC7L	Tufarelli <i>et al.</i> (2001)
SRSF5	Screaton <i>et al.</i> (1995)	LUC7L2	Howell <i>et al.</i> (2007)
SRSF6	Screaton <i>et al.</i> (1995)	LUC7L3	Nishii <i>et al.</i> (2000)
SRSF7	Cavaloc <i>et al.</i> (1994)	SRRM1	Blencowe <i>et al.</i> (1998)
SRSF8	Soret <i>et al.</i> (1998)	SRRM2	Blencowe <i>et al.</i> (2000)
SRSF9	Screaton <i>et al.</i> (1995)		Sawada <i>et al.</i> (2000)
SRSF10	Yang <i>et al.</i> (1998)	RBM25	Fortes <i>et al.</i> (2007)
SRSF11	Chaudhary <i>et al.</i> (1991)	TRA2A	Dauwalder <i>et al.</i> (1996)
SRSF12	Cowper <i>et al.</i> (2001)	TRA2B	Beil <i>et al.</i> (1997)

Table 1.1. A superfamily of arginine-rich pre-mRNA splicing factor RNA-binding proteins.
A list of serine-/arginine-rich (SR) splicing factor family member proteins (SRSF1-SRSF12) as well as proteins with arginine-rich domains (Arg-rich proteins) ¹¹⁹.

CHAPTER 2.0 : Middle-down proteomics reveals dense sites of methylation and phosphorylation in arginine-rich RNA-binding proteins

Briefs:

Middle-down proteomics reveals arginine-rich RNA-binding proteins contain many sites of methylation and phosphorylation.

*These findings were published in the *Journal of Proteome Research* on April 3, 2020 [19(4);1574-1591]. Supplemental tables can be found online at: <https://pubmed.ncbi.nlm.nih.gov/31994892/>

2.1 Abstract

Post-translational modifications (PTMs) within arginine (Arg)-rich RNA-binding proteins, such as phosphorylation and methylation, regulate multiple steps in RNA metabolism. However, the identification of PTMs within Arg-rich domains with complete trypsin digestion is extremely challenging due to the high density of Arg residues within these proteins. Here, we report a middle-down proteomic approach coupled with electron transfer dissociation (ETD) mass spectrometry to map previously unknown sites of phosphorylation and methylation within the Arg-rich domains of U1-70K and structurally similar RNA-binding proteins from nuclear extracts of HEK-293T cells. Notably, the Arg-rich domains in RNA-binding proteins are densely modified by methylation and phosphorylation compared with the remainder of the proteome, with methylation and phosphorylation favoring RSRS motifs. Although they favor a common motif, analysis of combinatorial PTMs within RSRS motifs indicate that phosphorylation and methylation do not often co-occur, suggesting they may functionally oppose one another. Furthermore, we show that phosphorylation may modify interactions between Arg-rich proteins, as SRSF2 has stronger association with U1-70K and LUC7L3 upon dephosphorylation. Collectively, these findings suggest that the level of PTMs within Arg-rich domains may be among the highest in the proteome, and a possible unexplored regulator of RNA-binding protein interactions.

Keywords: Proteomics, middle-down, mass spectrometry, electron transfer dissociation, post-translational modifications, methylation, phosphorylation, RNA-binding proteins

2.2 Introduction

Key RNA-binding proteins (RBPs) that carry out specialized biological processes such as RNA splicing, polyadenylation and transport, contain domains disproportionately enriched with arginine^{152,153}. RBPs that harbor Arginine (Arg)-rich domains may be broadly classified into subsets based on residue composition. One class of these RBPs contains highly repetitive complementary repeats of basic (K/R) and acidic (D/E) residues, that we have previously referred to as Basic Acidic Dipeptide (BAD) domains¹⁵⁴. Importantly, BAD domains facilitate protein condensation and eventual aggregation^{73,155}, and in the context of Alzheimer's disease, regulate interactions with pathological Tau protein¹⁵⁴. A second subset, related to the BAD proteins, are the Arginine/serine-rich (RS) domains that are ubiquitous in the Serine/Arginine (SR) family of proteins¹⁵³. Upon serine phosphorylation, RS domains mimic BAD domains with a similarly alternating basic-acidic dipeptide sequence pattern. The RS domains are commonly found in splicing factors and are essential for alternative splicing, protein-protein interactions, and cellular localization^{123,156-160}. Most BAD and SR proteins share a similar primary domain structure, containing RNA recognition motif (RRM) domains that function in recognizing and binding RNA, and Arg-rich low complexity (LC) domains. While Arg-rich BAD and RS domains do bind pre-mRNA at the 5' splice site and branchpoint^{161,162}, Arg-rich domains primarily engage in a variety of protein-protein interactions^{154,163-165}.

Although BAD and RS domains are recognized for their central role in RBP protein-protein interactions within the spliceosome, the similar RGG/RG box is a repetitive motif also enriched within RNA-binding proteins. RG boxes contribute to binding primary and secondary RNA structures^{166,167} and facilitate protein binding, as methylation of RG motifs within Sm proteins of small nuclear ribonucleoproteins (snRNPs) regulate Sm core assembly, an essential step to snRNP

biogenesis^{168,169}. The intrinsically disordered YG/YGG box motif was also identified as an RNA-binding motif in an RNA interactome study, putatively involved in RNA binding through hydrophobic base stacking^{55,66}. Recent studies suggest that post-translational modifications (PTMs) within low complexity RNA binding motifs may regulate RBP liquid-liquid phase separation^{68,73,146}. Therefore, a more comprehensive examination of PTMs within intrinsically disordered RNA binding motifs is warranted.

The functional and structural diversity of the proteome is markedly increased through PTMs, which are covalent modifications of protein primary structure, with over 300 types identified to date¹⁷⁰. The RS domains are hypothesized to be key hubs of RBP regulation, finely regulated by reversible PTM^{123,158,171}. For example, serine-arginine protein kinases (SRPKs) and cyclin-like kinase (CLK/STY1) proteins phosphorylate RS domains to influence alternative splicing and localization of RS domain-containing proteins, indicating that the modification status of RS domains may regulate a vast range of essential RNA metabolism processes^{160,170,171}. The phosphorylation and subsequent de-phosphorylation of Arg-rich domains controls RBP subcellular localization, spliceosome assembly and splice site selection^{172,173}. Furthermore, a recent study identified arginine methylation sites throughout RBPs, including U1-70K and SRSF2, that regulate RBP speckle association and binding to RNA¹⁷⁴. However, Arg-rich RS and BAD domains, and PTMs within them, remain poorly resolved by mass spectrometry approaches. Indeed, RNA interactome studies identify RNA-binding domain sequence motifs such RG and YG boxes as ‘hot spots’ for PTMs⁶⁶. However, without complete coverage of the proteome, especially sequences expected to be extensively post-translationally modified such as the LC1/BAD domain of U1-70K¹⁷⁵, such analyses remain incomplete. Identifying and understanding the functional consequences of PTMs in Arg-rich domains is therefore critical to provide mechanistic insight into

the role of low complexity Arg-rich domains in specific molecular interactions under physiological and pathological states.

Although the functional role of Arg-rich domain PTMs have been broadly characterized, attempts to globally identify individual PTM sites by mass spectrometry (MS) have been stymied by technical challenges. This is largely due to the inability to sequence Arg-rich domains using bottom-up proteomic approaches. Standard bottom-up proteomic methods rely on a complete digestion of proteins using the serine protease trypsin, which cleaves at arginine and lysine residues to produce peptides generally of appropriate lengths and charge suitable for MS identification^{176,177}. While this strategy is appropriate for proteins with a normal distribution of arginine and lysine, it is ineffective when examining proteins with Arg-rich domains. Complete digestion of Arg-rich domains using trypsin produces peptides that are too small to map to a unique protein. Additionally, peptides produced from Arg-rich domains are highly protonated and thus difficult to fragment by standard approaches. Peptide fragmentation by tandem mass spectrometry with collision induced dissociation (CID) or higher-energy collision dissociation (HCD) disfavors multiply-protonated Arg-rich peptides¹⁷⁸⁻¹⁸⁰. Arginine, the most basic amino acid, contains three basic nitrogens that collectively reduce proton mobility along the peptide backbone, eventually sequestering protons and preventing dissociation^{181,182}.

Electron transfer dissociation (ETD) is a non-ergodic fragmentation technique, related to electron capture dissociation (ECD), that cleaves the peptide backbone N-C_α bond following the transfer of an electron from an anion (e.g., fluoranthene radical ion) of low electron affinity to a multiply-charged peptide^{179,180,183,184}. This dissociation event results in c- and z-type fragment ions instead of the typical b- and y-type ions observed in CID or HCD^{185,186}. Notably, ETD peptide fragmentation is not influenced by amino acid side-chain chemistry and thereby preserves PTMs

that are otherwise labile by CID, such as phosphorylation, glycosylation, nitrosylation and sulfonation, providing more complete PTM identification^{178,183,184,187-194}. ETD has been extensively utilized to map PTMs within highly basic N-terminal histone tails and similar high charge state peptides by both bottom-up and middle-down approaches¹⁹²⁻¹⁹⁵. Non-canonical ‘middle-down’ approaches may be leveraged to characterize longer peptides by employing shorter proteolysis time points or non-canonical proteolytic enzymes. This is in contrast to standard ‘bottom-up’ strategies that utilize overnight trypsin digestion, typically rendering Arg-rich domains extremely difficult to detect by mass spectrometry. Importantly, the middle-down approaches enhanced the ability to both sequence protein sequences rich in lysine and/or arginine and capture co-occurring or combinatorial PTMs^{192,193}. Although the combination of these approaches have been used extensively to map the histone code, few studies have focused on Arg-rich domains contained in many essential splicing RBPs.

Here we report a middle-down proteomic approach utilizing limited trypsin digestion and ETD on an Orbitrap Fusion mass spectrometer to map PTMs within Arg-rich domains¹⁹⁶. By performing cellular fractionation and enriching for the nucleoplasm we achieved near complete coverage and identification of multiple phosphorylation and methylation sites within the Arg-rich domains from the core snRNP protein U1 small nuclear ribonucleoprotein 70 kDa (U1-70K), among other key splicing RBPs within the nucleoplasm of HEK-293T cells. We found BAD, RS and RG motifs to be highly decorated with methylation of arginine and lysine residues and phosphorylation of serine and threonine, while YG motifs were infrequently modified, especially phosphorylation at tyrosine residues. RS motifs in particular were elevated in PTMs and were a top motif of both arginine dimethylation and serine phosphorylation. Peptides mapping within these domains contained a combinatorial pattern of PTM, enriched above background proteomic

levels. Furthermore, following phosphatase treatment, serine-arginine splicing factor 2 (SRSF2) had greater affinity for the Arg-rich RBPs U1-70K and LUC7L3, suggesting that phosphorylation could play a key role in modifying protein-protein interactions in these structurally similar RBPs. Overall our approach can be generally applied for mapping PTMs in Arg-rich domains found throughout the proteome by various RBP purification procedures.

2.3 Materials and Methods

Materials

The following are primary antibodies included in this study: an in-house rabbit polyclonal antibody raised against a synthetic keyhole limpet hemocyanin-conjugated peptide corresponding to an epitope C-terminal to the low complexity domains of U1-70K⁹⁸; anti-SRSF1 antibody (catalog no. ab38017, Abcam); anti-myc antibody (catalog no. 9B11, Cell Signaling); anti-LUC7L3 (catalog no. PA5-53816, ThermoFisher); and an IgG mouse control (catalog no. 550339, BD Pharmingen). Secondary antibodies were conjugated to either Alexa Fluor 680 (Invitrogen) or IRDye800 (Rockland) fluorophores.

Plasmids and Cloning

The U1-70K LC1/BAD plasmid used herein was generated in a previous study⁹⁸. The SRSF2 gene was cloned into the same plasmid backbone. Cloning was performed by the Emory Custom Cloning Core Facility (Oskar Laur), and plasmids were confirmed by DNA sequencing.

Cell Culture and Transfection

Human embryonic kidney (HEK) 293T cells (ATCC CRL-3216) 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. For transient transfection, the cells were grown to 80-90% confluency in 10 cm² culture dishes and transfected with 10 µg expression plasmid and 30 µg linear polyethylenimine (PEI).

Protein expression and purification

Recombinant U1-70K protein was expressed and purified as described previously¹⁹⁷. In short, HEK-293T cells were grown and transfected with a plasmid encoding the GST-LC1/BAD (AA231-310) domain of U1-70K using PEI, and harvested 72 hours post-transfection. The cells were resuspended in ice-cold lysis buffer (50 mM HEPES pH 7.4, 200 mM NaCl, 5% glycerol, 1 mM EDTA, 1% (w/v) sarkosyl and 1X HALT protease/phosphatase inhibitor cocktail). Lysates were sonicated using a microtip probe to shear nucleic acids. To make the lysates compatible with GST affinity purification, Triton X-100 (TX-100) was added to a concentration of 1.5% (v/v). The lysates were cleared by centrifugation at 14,000 x g for 10 min at 4°C. The resulting supernatants were incubated overnight at 4°C with 0.5-1 ml of swelled glutathione-agarose resin (Sigma G4510), after which the slurries were loaded onto a column. The resin was washed with 10 column volumes of wash buffer (50 mM HEPES pH 7.4, 200 mM NaCl and 1% TX-100), and eluted with 4 x 1 ml elution buffer (50 mM Tris-HCl pH 8.5, 500 mM NaCl, 20 mM reduced L-glutathione (Sigma G4251) and 0.1% TX-100). The eluted fractions were concentrated to ~200 µl using Amicon Ultra-0.5 ml 10K MWCO Centrifugal Filter Units (EMD Millipore), and dialyzed overnight against 50 mM HEPES pH 7.4, 200 mM NaCl and 0.1 mM PMSF using 10K MWCO Slide-A-Lyzer MINI Dialysis Units (Thermo). Protein concentration was determined by running each elution fraction on an SDS-PAGE gel with bovine serum albumin (BSA) standards ranging from 0.2-1 µg per lane and staining with Coomassie G-250¹⁹⁷. Densitometry of the BSA standards was used to calculate the concentration of GST affinity purified protein.

Western Blotting

Western Blotting was performed according to standard protocol as previously described in Bishop *et al.*¹⁵⁴. In short, samples were boiled in Laemmli sample buffer (8% glycerol, 2% SDS, 50mM Tris pH 6.8, 3.25% beta-mercaptoethanol) for 5 minutes, then resolved on a Bolt® 4-12% Bis-tris

gel (catalog no. NW04120BOX, Invitrogen) by SDS-PAGE and semi-dry transferred to a PVDF membrane with the iBlot2 system (ThermoFisher). Membranes were blocked with TBS Starting Block Blocking Buffer (catalog no. 37542, ThermoFisher) 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 (Invitrogen) or IRDye800 (Rockland) 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.

In-gel Limited Trypsin digestion

The GST-LC1/BAD (residues 231-310) purified protein (14 µg) was run onto a 10% acrylamide gel. The gel was stained with Coomassie G-250 and the GST-LC1/BAD band was compared to BSA standards to estimate protein concentration. The GST-LC1/BAD band was cut out and diced into small pieces. The gel pieces were divided among five tubes each receiving ~3 µg of GST-LC1/BAD. Gel pieces were de-stained until clear using 70% 50mM ammonium bicarbonate (ABC) and 30% acetonitrile. While on ice, each tube received 30 µL digest buffer [12.5 ng/µl trypsin (Pierce MS grade) in 50mM ABC buffer. After the addition of trypsin, samples were incubated on ice for 3 minutes then brought up to room temperature to start digestion. At 6 different time points (15, 30, 60, 120, 240 minutes, and overnight) excess trypsin solution was removed and the digestion reaction was stopped with 30 µl extraction buffer (50% acetonitrile, 5% acetic acid). After the addition of extraction buffer, samples were allowed to equilibrate for five minutes, then stored at -20 °C until peptide extraction. Peptides were shaken for 40 minutes at room temperature then spun at 20,000 x g for one minute. After a minute period following the first spin, peptides were spun again. This spin and relax cycle was repeated twice. The supernatant containing the extracted peptides was then collected into a new tube, in which 30 µl of extraction buffer was then

added. This extraction process was repeated two more times. Peptides were lyophilized using a SpeedVac (catalog no. 731022, Labconco) and resuspended in MS sample loading buffer (1% acetonitrile, 0.1% formic acid, and 0.03% trifluoroacetic acid).

Nucleoplasm Enrichment

This cellular extraction procedure was adapted from the Gozani group¹⁹⁸. In short, cells from eight 15-cm plates were combined and rinsed with cold PBS, and then scraped in 10 ml PBS + 1X Protein Inhibitor Cocktail buffer (catalog no. COUL-RO, Roche) and centrifuged for 5 min at 1,000 x g at 4°C. Cells were washed once in 1 ml 1X (cold) PBS and recovered by centrifugation for 5 min at 1,000 x g at 4°C, swelled in 75 µl of hypotonic lysis buffer (10 mM HEPES pH 7.9, 20 mM KCl, 0.1 mM EDTA, 1mM DTT, 5% Glycerol, 0.5 mM PMSF, 10 µg/ml Aprotinin, 10 µg/mL Leupeptin) and incubated on ice for 10 min. Samples were lysed by 0.1% NP-40, vortexed, and incubated on ice for 5 min. Nuclei were recovered by centrifugation for 10 min at 15,600 x g at 4°C. Nuclei were extracted for 30 min on ice in 40 µl high salt buffer (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). Samples were sonicated for 5 sec and extracts were collected by centrifugation for 10 min at 15,600 x g at 4°C. The supernatant obtained following the centrifugation step consisted of isolated nucleoplasm and the resulting pellet was the chromatin fraction.

In-solution digest of nucleoplasm fractions

For each time point, 150 µg of nucleoplasmic fraction was digested. Samples were brought to a final concentration of 1M urea, then dithiothreitol was added to a final concentration of 1 mM, and incubated for 30 minutes. Iodoacetamide was added to a final concentration of 1 mM, and incubated for 20 minutes in the absence of light. Samples were then diluted in digestion buffer, and digested with a 1:50 ratio of trypsin (Pierce MS grade) to total protein. The digestion reactions

were performed at room temperature and quenched at increasing time lengths (5, 10, 20, 40, 80, 160 minutes, overnight) with 0.1% formic acid, and 0.01% trifluoroacetic acid solution. Resulting peptides were cleaned up using an HLB column (Waters). Samples were washed first with methanol, then 0.1% trifluoroacetic acid. The digested samples were then loaded into the column, washed with 0.1% trifluoroacetic acid twice, and then eluted with Buffer C. The resulting elutant was lyophilized using a SpeedVac (catalog no. 731022, Labconco).

Mass spectrometry analysis

Lyophilized peptides were resuspended in loading buffer (0.1% formic acid, 0.03% TFA, 1% acetonitrile) and separated on a self-packed C18 (1.9 μm Dr. Maisch, Germany) fused silica column (20 cm \times 75 μm 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 Tribrid Mass Spectrometer. Data-dependent MS/MS analyses included a high resolving power MS1 step (120,000 at m/z 400) with an m/z range of 100-1000. MS1 scans were conducted in the Orbitrap, and the top 10 ions with highest charge, followed by the precursor ion with the greatest abundance, were given priority for fragmentation. A data-dependent decision tree was used¹⁹⁹ and MS/MS spectra from both HCD and ETD were collected in the ion-trap. Peptides with charge state of +2 were chosen for fragmentation by HCD only, while all charge states +3 and above were fragmented by both ETD and HCD. At +3 charge state, precursor ions under 650 m/z were fragmented by both ETD and HCD. Those equal to or greater than 650 m/z were fragmented by HCD only. At +4 charge state, precursor ions under 900 m/z were jointly fragmented by both HCD and ETD, and only by HCD at m/z values equal to greater than 900. At +5 charge state, precursor

ions under 950 m/z were fragmented by both HCD and ETD, and only by HCD at m/z values equal to or greater than 950. At + 6 charge state, all precursor ions were fragmented by ETD. Dynamic exclusion was set to exclude previous sequenced precursor ions for 30 seconds.

Database Searching

Data files for the time points were analyzed using MaxQuant v1.5.2.8 and v1.6.1.0 with Thermo Foundation 2.0 for RAW file reading capability. The search engine Andromeda was used to build and search a concatenated target-decoy UniProt Knowledgebase (UniProtKB) containing both Swiss-Prot and TrEMBL human reference protein sequences (90,411 target sequences downloaded April 21, 2015), plus 245 contaminant proteins included as a parameter for Andromeda search within MaxQuant²⁰⁰. Methionine oxidation (+15.995 Da), and protein N-terminal acetylation (+42.011 Da) were included as variable modifications (up to 5 allowed per peptide); cysteine was assigned a fixed carbamidomethyl modification (+57.022 Da) for the nucleoplasm. A two stage search was performed as described previously²⁰¹. This two-step method allows for a larger search space while limiting false-discovery rate (FDR)²⁰². For the first search only fully tryptic peptides were considered with up to 2 missed cleavages in the database search. A precursor mass tolerance of ± 20 ppm was applied prior to mass accuracy calibration and ± 4.5 ppm after internal MaxQuant calibration. Other search settings included a maximum peptide mass of 6,000 Da, a minimum peptide length of 6 residues and 0.6 Da tolerance for ion-trap ETD and HCD MS/MS scans. The false discovery rate (FDR) for peptide spectral matches, proteins, and site decoy fraction were all set to 1 percent. The proteins identified in the first database search were then used to make a targeted FASTA formatted database to re-search the raw files with an expanded missed cleavage window and variable modifications. The targeted database corresponding to purified U1-70K LC1/BAD protein contained a FASTA file of 366 proteins and

allowed for 6 mis-cleavages. The targeted database corresponding to the nucleoplasm contained a FASTA file of 4,307 unique protein groups (20,392 protein isoforms), allowing 6 mis-cleavages. Both targeted databases were used to search the spectra for variable PTMs including phosphorylation (S/T/Y) (+79.966 Da) and mono- and di-methylation (K/R) (+14.016 Da, +28.031 Da), in addition to N-terminal acylation and methionine oxidation. The MS/MS spectra were collected by a low-resolving power ion trap (product ion tolerance = 0.6 Da). As such, trimethylation of lysine (42.047 ± 0.002 Da) and acetylation of lysine (42.011 ± 0.004 Da) were not included in our search, given that the masses of these modifications are less than 20 ppm apart. The .raw and .txt files obtained from MaxQuant searches were uploaded to ProteomeXchange on 9/4/2019 (Accession ID: PXD015208). To prevent aberrant PTM localization assignments by MaxQuant, including isobaric peptides with multiple assignable residues, peptides with a PTM localization site scores greater than or equal to 0.6 were retained after using a self-written PTM-site filtering script.

Parameters for selecting Basic Acidic Dipeptide (BAD), Arginine-Serine (RS), Arginine-Glycine (RG) and Tyrosine-Glycine (YG) peptides

We applied a BAD score algorithm counting peptides that added 1.0 points for an alternating basic and acidic charge (+/- or -/+) and added 0.1057/0.0764/0.0475 points for S/T/Y residues, respectively, neighboring a basic residue. This attempted to reflect the average relative phosphorylation frequency across all nucleoplasmic peptides of each residue, which transforms the dipeptide into a BAD sequence. This sum was calculated and divided by the peptide length to give a “BAD score” for each peptide. A cutoff score of 0.3 or greater was chosen, identifying 574 BAD peptides in total. These selection criteria enriched for Lys/Arg/Ser/Thr/Tyr residues, and actual residue frequency was factored into normalized PTM fold changes. RS, RG and YG

peptides were identified through a similar algorithm, counting peptides with 1.0 points for alternating residues, normalized to peptide length. The sum was calculated and divided by the peptide length to give a score for each peptide. For RS peptides, a cutoff of greater or equal to 0.18 was chosen. There were 583 RS peptides identified by this selection algorithm. A cutoff of greater than or equal to 0.129 for RG peptides was chosen, selecting 539 peptides. A cutoff of greater than or equal to 0.103 for YG peptides was chosen, selecting 525 peptides. These selection criteria resulted in an enrichment for Lys/Arg/Ser/Thr/Tyr residues, which was factored into calculating normalized PTM fold changes.

MEME analysis

Phosphosite Motif Logo analysis was conducted (<https://www.phosphosite.org/sequenceLogoAction.action>)²⁰³ to identify sequence motifs of serine phosphorylation, arginine monomethylation and arginine dimethylation. Input peptide sequence windows with unique sites of PTM were assembled and entered for analysis. A list of 46 monomethylated arginine peptides, 252 dimethylated arginine peptides and 458 phosphorylated serine peptides generated by our MS data served as input sequences. Additionally, Motif analysis was performed for mono-/di-methylated lysines, and threonine/tyrosine phosphorylation (Fig.2.12).

Measurement of PTM Coverage and RSRS Motif PTM States

Modified sequences containing a 31-residue width window were matched to sequences downloaded (5/24/2019) from phosphosite.org²⁰³. Excel was used to match methylated or phosphorylated sequences previously observed by mass spectrometry methods only. A character length (len) function in Excel substitution function was used to search for post-translationally

modified RSRS/SRSR sequences. Occurrences were normalized to the number of RSRS/SRSR sequences within a peptide and calculated as percent of total occurrences.

Co-Immunoprecipitation

Co-immunoprecipitations were performed essentially as described¹⁵⁴. Briefly, HEK-293T cells were transiently transfected with either myc-tagged SRSF2 plasmid or a mock control pUC19 vector. Harvested cells were fractionated to isolate the nucleoplasm as described above. Nucleoplasm fractions were then homogenized in ice-cold. Samples were sonicated for 10 seconds on 10 seconds off at 30% amplitude for a total of 1 minute. Protein concentrations were determined using a bicinchoninic acid (BCA) assay (catalog no. 23225, ThermoFisher). Nucleoplasm fractions were then treated with either mock (H₂O) or calf-intestinal phosphatase (Quick CIP, catalog no. M0525L, NEB) for 1 hr at 37°C. The phosphatase reaction was quenched in a heat bath at 80°C. Then, 20 µL of Protein A Sepharose 4B beads (catalog no. 101042, Invitrogen) were washed twice in immunoprecipitation (IP) buffer [50 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 0.5% (v/v) NP-40, 0.5% (v/v) CHAPS, HALT phosphatase inhibitor cocktail (1:100, catalog no. 87786, ThermoFisher)] then blocked in 0.1 mg/ml bovine serum albumin (catalog no. 23209, Thermo), then washed an additional three times in IP buffer. Anti-myc (4 µg) mouse monoclonal antibody (catalog no. 2276, Cell Signaling) or IgG control (4 µg, catalog no. 550339, BD Pharmingen) was incubated with the bead slurry in IP buffer and rotated for 90 minutes to conjugate antibody to beads. The beads were then washed three times in IP buffer, then nucleoplasm lysates were added to beads (0.5 mg per IP) and incubated rotating overnight at 4°C. The beads were then washed three times in IP wash buffer (IP buffer lacking glycerol and CHAPS) then resuspended in IP wash buffer. Preceding the final wash, the bead slurry was transferred to a new microcentrifuge tube to limit contamination. The bead slurry was then centrifuged at 500 x g

for 5 min at 4°C and the pelleted beads were resuspended with 1X Laemmli sample buffer (8% glycerol, 2% SDS, 50 mM Tris pH 6.8, 3.25% β-mercaptoethanol) and run by SDS-PAGE. Nucleoplasm fractions (10 μg) were included as input loading controls. Three independent biological replicates were performed.

2.4 Results

The spliceosomal protein, U1-70K, contains two Arg-rich low complexity (LC) domains, LC1/BAD (residues 231-310) and LC2 (residues 317-407) (**Fig.2.1a**). The ‘BAD’ acronym of LC1/BAD stands for “Basic Acidic Dipeptide”, containing dipeptide repeats of a basic (K/R) residue adjacent to an acidic residue (D/E). The U1-70K LC1/BAD domain has both BAD and RS motifs, and can form condensates via liquid-liquid phase separation (LLPS) and eventually aggregates *in vitro* and *in vivo*^{73,155}. The LC1/BAD domain of U1-70K is of particular biological interest due to its central role in U1-70K nuclear localization, granule formation, and co-aggregation with tau in Alzheimer’s disease^{98,154,204}. On PeptideAtlas.org there are currently over 50,000 peptide spectral matches (PSMs) to U1-70K using standard bottom-up approaches with complete trypsin digestion, mapping approximately 70% of the protein²⁰⁵. Despite this number of PSMs, the LC1/BAD domain has only 7.5% coverage (6 out of 80 residues)^{206,207}. In comparison, the PTM repository Phosphosite.org shows several phosphorylated peptides within the LC1/BAD domain, albeit with few repeated observations²⁰³. Thus, we sought to develop a method to achieve more comprehensive sequence coverage of the U1-70K LC1/BAD domain and characterize PTMs within that would be generally applicable to other Arg-rich RNA-binding proteins in the proteome utilizing a middle-down proteomics approach.

Recombinant LC1/BAD domain was transiently expressed and GST-purified from human embryonic kidney (HEK-293T) cells. The recombinant LC1/BAD protein (~38 kDa) was

incubated with trypsin for varying time periods to achieve a partial digest. The digestion reaction was quenched with acetic acid at 15, 30, 60, 120, and 240 minutes. An additional sample was also allowed to proceed overnight for complete digestion (**Fig.2.1b**). The resulting peptides were then extracted and analyzed by LC-MS/MS on an Orbitrap Fusion mass spectrometer. A higher-energy collisional dissociation (HCD) and electron transfer dissociation (ETD) decision tree was used (**Fig.2.1c**)¹⁹⁰.

In order to sequence low complexity Arg-rich domains, we needed to account for a large number of missed cleavages and PTMs, which would increase the search space compared to traditional proteome database searches. A two-step database search method was utilized wherein proteins matched from a primary search were used to create a smaller, focused database. This strategy limits the false discovery rate (FDR) and the number of false negatives²⁰², resulting in higher confidence peptide spectral matches (PSMs) compared to traditional one-step database search methods²⁰². Our focused database was then used to perform a second search that included increased missed-cleavages, and PTMs (methylation and phosphorylation), using a <1% FDR cutoff.

A strength of ETD is the ability to fragment highly protonated peptides and retain PTMs^{178,187,189,190,192-194}. For example, when comparing the MS/MS spectra of the same precursor peptide fragmented by either ETD or HCD, the ETD method results in increased fragmentation (**Fig.2.1d-e**). These fragment ions (c- and z-ions) produced by ETD provide unique diagnostic ions, unafforded by canonical HCD/CID fragmentation methods, that enhance identification of the peptide (**Fig.2.1e**)^{208,209}. When examining the HCD spectra, only a neutral loss of 98 Da is observed. This mass shift is a signature of a phosphorylation PTM, indicating the loss of phosphate and water²¹⁰⁻²¹³ (**Fig.2.1d**). However, the HCD spectra contain few other additional fragment ions

(b- and y-ions), making it difficult to assign the linear sequence of amino acids of the precursor peptide. When observing unique peptides accumulated across all time points, as a result of increased fragmentation, ETD yielded 114 unique LC1/BAD peptides, as compared to HCD, which identified only 18 unique LC1/BAD peptides (**Fig.2.1f**). When examining LC1/BAD peptides containing PTMs, the difference between peptides identified by ETD and HCD fragmentation is similarly apparent. ETD identifies 55 PTM-containing LC1/BAD peptides alone compared with HCD, which identified just 13 such peptides (**Fig.2.1g**).

Collectively we identified 35 sites of PTMs across all proteolysis time lengths (**Fig. 2.2A**). Eight out of 10 possible serine residues within LC1/BAD were found to be phosphorylated. Additionally, 23 total sites of arginine methylation were discovered (12 mono-methylated and 13 di-methylated). Furthermore, four sites of lysine mono-methylation were identified. To visualize these findings, the LC1/BAD unique peptides identified by either HCD or ETD fragmentation were mapped onto the LC1/BAD domain (**Fig.2.1h**). ETD led to the identification of more LC1/BAD unique peptides than HCD, providing coverage of 79/80 residues (99%) within the LC1/BAD domain (**Online Supplemental Table S1**). If trypsin digestion is allowed to proceed overnight, the number of PSMs decreases and significant coverage is no longer obtained with either ETD or HCD. Thus, a combination of the middle-down strategy with ETD fragmentation achieved near complete coverage of the LC1/BAD domain of U1-70K *in vitro* and serves as proof-of-concept for enhancing coverage of Arg-rich proteins in complex mixtures.

2.4.1 Preparation of nucleoplasm fractions enriched with Arg-rich RBPs

We next sought to employ this approach globally to achieve widespread coverage of Arg-rich BAD and RS proteins, mapping PTMs therein, from complex cell extracts. Arginine is not evenly distributed throughout the proteome, but rather is densely concentrated in Arg-rich domains^{152,153}.

The snRNPs, associated heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine/arginine-rich splicing factors (SRSFs) form macromolecular spliceosome structures in the nucleus, many of which contain Arg-rich domains^{214,215}. Many recent RNA interactome approaches have utilized UV crosslinking or organic phase separation to enrich for RBPs that interact with adenylated RNAs that have built a substantial catalog of known and previously unknown RBPs^{64,216-218}. However, to avoid the co-enrichment of histone proteins and associated factors in such methods we chose to perform biochemical fractionation to enrich for RBPs isolated within the nucleoplasm. The nucleoplasm is rich in splicing RBPs, many of which undergo LLPS and aggregate in neurodegenerative disease^{54,154,197,219}. For downstream proteomic analysis, the over-representation of highly-expressed histone proteins in nuclear fractions would suppress the sensitivity of the mass spectrometer to identify comparatively less abundant RBPs that we sought to capture in our analyses. Therefore, simultaneous nucleoplasm isolation and histone depletion was performed to increase the probability of sequencing Arg-rich RBPs, especially low stoichiometry peptides within BAD and RS domains, many of which are expected to contain PTMs^{174,175}.

Differential centrifugation was performed to enrich cytoplasmic, nuclei, chromatin and nucleoplasmic fractions from HEK-293T cell extract (**Fig.2.3a**). Following nucleoplasm isolation, core splicing proteins U1-70K and SRSF1 were analyzed by SDS-PAGE and western blotting, showing a 2.75- and 7.02-fold enrichment in the nucleoplasmic fraction, respectively, relative to the original whole cell fraction (**Fig.2.3b-c**). Importantly, coomassie staining of an identical gel showed histone proteins were enriched by 5.26-fold in the chromatin fraction (**Fig.2.3b-c**), and depleted from the nucleoplasm by 0.45-fold, relative to the whole cell fraction. The enrichment of splicing factors and other RBPs make this complex nucleoplasm fraction an excellent source for the examination of the Arg-rich proteome.

2.4.2 Global analysis of RNA-binding proteins from nucleoplasm extracts by middle-down ETD MS

Nucleoplasm samples were incubated with trypsin, and reactions were quenched with acetic acid at 5, 10, 20, 40, 80, and 160 minutes, with the standard overnight digestion serving as a control (**Fig.2.3a**). Partially-digested peptides were extracted and analyzed by LC-MS/MS on an Orbitrap Fusion Tribrid mass spectrometer operating on an HCD-ETD decision tree as described for the recombinant U1-70K LC1/BAD domain expressed in HEK-293T cells¹⁹⁰. An initial conventional search identified 49,307 unique peptides, matching to 20,392 proteins, collapsed into 4,307 unique protein groups by parsimony. A second search using a smaller targeted database containing the 20,392 proteins matched in the first search was performed with mono-/ di-methylation of Arginine/Lysine and phosphorylation of Serine/Threonine/Tyrosine PTMs and a maximum of 6 missed cleavages. A total of 61,283 unique peptides were identified following a target-decoy search, using an FDR cutoff at less than 1%. Detected peptides matched to 4,140 unique protein groups. Although fewer input protein database entries were included in the second search, it yielded more matched unique peptides (~12,000) compared to the conventional database search, consistent with previous two-step search strategies²⁰¹. Peptides with three or more missed-cleavages (n=8,208) and modified peptides (n=6,949) comprised a significant proportion of the novel peptide matches in the two-step search strategy. Furthermore, our second database search matched 96% of protein groups identified in the first search (4,140/4,307). GO-term analysis of these proteins sequenced demonstrate that the nucleoplasm fraction was enriched with factors involved in 'RNA binding', consistent with our western blot results (**Fig.2.3d**). Moreover, the nucleoplasm fraction contained 687 RNA-binding proteins previously identified within a catalog of RNA-binding proteins captured using RNA interactome methods within HEK-293T cells

(**Fig.2.3e**)⁶⁴. Therefore, 687/796 (86%) of representative RNA-binding proteins within HEK-293T cells were identified in this study (**Fig.2.3e**). Thus, these proteomic and bioinformatic results confirm that the cellular fractionation method successfully enriched for native RNA-binding proteins, generating a complex sample amenable to middle-down ETD analysis.

2.4.3 Enhanced sequence coverage of Arg-rich proteins by middle-down ETD MS

We attempted to compare the influence of ETD or HCD fragmentation strategies on sequencing events across the complex nucleoplasm sample. Notably, ETD and HCD sequenced a similar number of unique peptides with relatively little overlap. A total of 22,134 and 27,746 unique peptides were sequenced by ETD and HCD, respectively, while 7,991 peptides were identified by both fragmentation methods (**Fig.2.4a**). Peptides above +2 charge are preferentially sequenced by ETD as directed by the ETD/HCD decision tree algorithm (**Fig.2.5d**)¹⁹⁰, contributing to this peptide sequencing disparity. Importantly, peptides sequenced following ETD fragmentation were generally more confidently scored and assigned a peptide sequence, averaging a higher Andromeda Score (147.57), a measure of the confidence of the peptide sequence assignment, as compared with HCD-fragmented peptides (103.24) (**Fig.2.4b**). This indicates that for complex sample mixtures, a decision tree approach that utilizes both ETD and HCD fragmentation platforms may provide complementary sequence information than currently achieved with standard sample preparation methods^{178,185-188}. For example, at shorter trypsin proteolysis times, ETD fragmented-peptides consisted of the majority of sequenced peptides (**Fig.2.4c**). Between 40-80 minutes of digestion, however, HCD fragmentation becomes the preferred method of fragmentation, contributing the majority of matched peptides (**Fig.2.4c**).

When analyzing peptides modified by methylation or phosphorylation, the reliance on shorter trypsin incubation lengths and ETD fragmentation strategies becomes more apparent. The

majority of peptides with at least one PTM are more frequently identified after ETD fragmentation compared with HCD fragmentation across all limited trypsin proteolysis time lengths tested (**Fig.2.5a**). These modified peptides fragmented by ETD do not suffer from significantly reduced confidence scores, as ETD-fragmented peptides average higher Andromeda Scores overall (93.07) compared with HCD-fragmented peptides (71.53), likely due to the increased length of the peptides (**Fig. 2.5c,f**). As methylation of arginine or lysine does not affect the charge of the residue, Arg-rich peptides are preferentially identified at earlier digestion time points by ETD (**Fig.2.6d, Fig.2.5i-k**)¹⁹⁰. Indeed, 1,158 out of 1,696 Arg-rich peptides (68%) were fragmented by ETD (**Fig.2.7c**). Thus, as expected, it is apparent that ETD fragmentation and middle-down strategies enhance of the identification of Arg-rich RNA binding proteins in complex mixtures¹⁸⁸.

2.4.4 Motif algorithms resolve RNA-binding protein subgroups with distinct biological properties

RNA-binding proteins with domains containing BAD, RS and RG motifs have multiple binding partners that can change due to cellular condition, hypothesized to be regulated by PTMs^{154,165,220-222}. In addition, tyrosine phosphorylation in YG motifs may promote electrostatic interactions with arginine in RG motifs to catalyze LLPS in RBPs such as FUS⁶⁸. To study the characteristics of similar RBP binding motif subgroups, we implemented a scoring algorithm to select for peptides with alternating basic-acidic “BAD”, arginine-serine “RS”, arginine-glycine “RG” or tyrosine-glycine “YG” dipeptides normalized to peptide length, and focused the following analyses for peptides that scored above a stringent cutoff score (**Fig.2.7a**). The BAD algorithm selected 574 peptides, approximately 0.99% of all unique peptides sequenced (**Fig.2.7a**). These peptides matched to 260 proteins (**Fig.2.6a, Online Supplemental Table S3**). The RS algorithm selected 583 peptides (1.00% of all peptides identified), identifying 129 RS proteins (**Fig.2.6a, Fig.2.7a,**

Online Supplemental Table S4). The RG algorithm selected 539 peptides, approximately 0.93% of all unique peptides sequenced (**Fig.2.7a**). These peptides matched to 323 proteins (**Fig.2.6a, Online Supplemental Table S5**). The YG algorithm selected 525 peptides, approximately 0.91% of all unique peptides sequenced (**Fig.2.7a**). These peptides matched to 326 proteins (**Fig.2.6a, Online Supplemental Table S6**). Although BAD, RS and RG peptides require several missed cleavages to be sequenced, each subgroup exhibits similar Andromeda Scores to no-motif (no-M) peptides (**Fig.2.7b, Online Supplemental Table S2**). A Venn diagram comparing proteins containing BAD, RS, RG or YG motifs show the overlap between groups (**Fig.2.6a**). The U1 snRNP member U1-70K (*red*) harbors BAD, RS and RG motifs. Six proteins (DHX38, PRPF4B, RBM39, SRSF4, SRSF5 and U2AF1) contain all four motifs (BAD, RS, RG and YG) and are either core members of the spliceosome or spliceosome-associated splicing factors^{214,215,223}.

To examine the biological function of proteins containing BAD/RS/RG/YG motifs, GO-elite analysis was performed, as compared with the nucleoplasmic proteome identified by our analysis (**Fig.2.6b-e**). Using the 3,900 total nucleoplasmic gene symbols identified in this study as background, BAD, RS, RG and YG proteins are all enriched in ‘RNA splicing’ and ‘mRNA processing’ functions. Both the BAD and RS proteins, in particular, are found enriched in spliceosomal complexes and nuclear speckles (**Fig.2.6b-c**). In contrast, RG- and YG-containing proteins identified in this study are exclusive to the hnRNP complex (**Fig.2.6d-e**). Although these proteins participate in splicing²²⁴, they operate in a distinct complex from BAD and RS proteins. In addition to occupying different cellular compartments, SR proteins and hnRNPs mutually oppose one another to regulate splice site selection by the spliceosome, with SR proteins binding exon splicing enhancers (ESEs) whereas hnRNPs such as hnRNPA1 antagonize this activity by binding to exonic splicing silencers (ESSs) to trigger alternative splicing²²⁴⁻²²⁷. Indeed, the second

most enriched Cellular Component term for RG- and YG-motif containing proteins is the ‘catalytic step 2 spliceosome’ (**Fig.2.6d-e**), which has undergone drastic rearrangements such that BAD-motif proteins such as U1-70K and RS-motif proteins such as SRSF2 have already been dissociated from the pre-mRNA complex²²⁸.

Additionally, functions tangential to canonical splicing regulatory roles, but essential to mRNA maturation, were parsed out by GO analysis. For example, RS proteins were identified as regulating mRNA export (**Fig.2.6c**)²²⁹⁻²³¹. In the case of SRSF3 and SRSF7, mRNA export is mediated by arginine-rich peptide sequences that do not bind RNA but instead the Tip-associated protein (TIP) export factor²³². Thus, by utilizing bioinformatics approaches we were able to infer the biological function of similar RBPs based on unique structural motifs.

2.4.5 PTM site validation

To avoid identifying sites of PTMs at the wrong sites in isobaric peptides, we retained PTM sites with PTM Localization scores scoring greater than or equal to 0.6 (60% likelihood) for both ETD- and HCD-fragmented peptides. The use of these threshold estimates a <5% false localization rate (FLR) of all PTMs identified, ensuring high-confidence PTM site-calling²³³. With increased proteomic coverage of Arg-rich sequences, we hypothesized we would identify a significant portion of PTMs corroborated by previous studies, as well as some that were unreported on proteomic databases. We thus compared PTM sites identified in our study to that of all PTMs previously identified by mass spectrometry and uploaded on phosphosite.org²⁰³, which allows us to define PTM sites identified as either ‘confirmed’ or ‘unreported’ by previous high-throughput mass spectrometry studies. We identified a total of 648 unique phosphorylated Ser/Thr/Tyr sites, 337 of which (52%) were validated in previous studies and published on phosphosite.org

(**Fig.2.8a**). This was further illustrated and validated when we performed peptide mapping of several highly-rated BAD and RS proteins, including U1-70K, SRSF2, SRSF4 and LUC7L2. The Arg-rich BAD/RS domains in the RBPs were uniquely enriched in methylation and phosphorylation (**Fig.2.9a, Fig.2.2b, Online Supplemental Table S7**). When comparing the Arg-rich peptides we sequenced within U1-70K, SRSF2, SRSF4 and LUC7L2 to phosphosite.org, we confirmed 35/94 (37%) sites we sequenced, and identified 35 unreported sites (**Online Supplemental Table S8**). In addition, we manually inspected individual spectra of several Arg-rich peptides within U1-70K, SRSF4 and LUC7L3 (**Fig.2.8b-d**). Thus, the high degree of overlap between PTMs identified in our dataset and previous studies provides further confidence in the PTMs identified using the middle-down proteomic approach.

2.4.6 Arg-rich domains in RNA-binding proteins contain combinatorial PTMs

BAD, RS, RG and YG motifs have a high density of modifiable residues (Lys/Arg/Ser/Thr/Tyr), and we therefore hypothesized these peptides may have an increased frequency of multiple modifications as compared with no-Motif peptides (no-M). Indeed, the majority of BAD and RS peptides contained two or more PTMs, while 36% of RG peptides and less than 10% of no-M and YG peptides were multiply-modified (**Fig.2.9b**). While the majority of BAD and RS peptides were modified, no-M and YG peptides were overwhelmingly unmodified (89%, 88% respectively) (**Fig.2.9c**). RS peptides, in particular, more frequently contained five PTMs on a single peptide, than one or no PTMs at all (**Fig.2.9c**).

We next sought to determine whether BAD and RS peptides contained a combination of several PTM subtypes at once. Peptides belonging to no-M, BAD, RS, RG and YG subgroups were classified according to the number of PTMs contained (0-6). The percentage of peptides

belonging to uniform PTM states (mono-methylation alone, di-methylation alone, phosphorylation alone), double-PTM states (mono- and di-methylation, monomethylation and phosphorylation, dimethylation and phosphorylation) or a triple-PTM state (mono- and di-methylation and phosphorylation) were then calculated, and percentages were displayed as a heat map (**Fig.2.9d**). The majority of no-M peptides were unmodified, although the most frequent PTM state was a single phosphorylation. YG peptides were also almost exclusively unmodified, while containing 32 peptides that were exclusively methylated (**Fig.2.9d**). While the majority of RG peptides were unmodified (64%), a large portion was exclusively methylated (28%) (**Fig.2.9d**). BAD peptides in contrast were enriched in combinatorial PTMs, with approximately 37% containing a combination of PTM subtypes on a single peptide (**Fig.2.9d**). RS domains were more frequently modified, as over 3 out of every 5 peptides were modified by a combination of PTM subtypes (**Fig.2.9d**). Namely, nearly a quarter (24%) of all RS peptides contained all three PTM subtypes searched for in a single peptide (**Fig.2.9d**). Compared to peptides mapping to surrounding domains, BAD and RS domains have a complex combinatorial signature of PTM. As PTMs are an essential regulator of the structure and function of RBPs²³⁴⁻²³⁸, we next sought to further characterize the frequency of these PTMs in the nucleoplasm proteome.

2.4.7 Arg-rich domains in RNA-binding proteins are densely modified

We next performed PTM enrichment on BAD, RS, RG and YG peptides, as compared to the background nucleoplasm proteome. For instance, we calculated the average frequency of a PTM within a peptide, divided by the number of potential modifiable (Lys/Arg/Ser/Thr/Tyr) residues within a peptide. On average, approximately 28% of Lys/Arg/Ser/Thr/Tyr residues within the BAD peptides are modified (**Fig.2.10a**). The RG peptides are similarly modified, with almost a quarter (24%) of Lys/Arg/Ser/Thr/Tyr residues being modified (**Fig.2.10a**). RS peptides are

particularly increased in methylation and phosphorylation, as nearly a third (31%) of Lys/Arg/Ser/Thr/Tyr residues within RS peptides are modified (**Fig.2.10a**). This is in stark contrast to just ~6.5% of residues modified in the remainder of the nucleoplasm proteome sequenced (**Fig.2.10a**). YG peptides, were the least modified peptide group, with only 4% of Lys/Arg/Ser/Thr/Tyr residues being modified (**Fig.2.10a**). In sum, the BAD, RS and RG peptides are all more than four-fold likely to contain lysine methylation, arginine methylation or serine phosphorylation, compared with no-M sequences identified in the nucleoplasm (**Fig.2.10a**).

Next, we sought to determine if the increased PTM density in the BAD, RS and RG regions is due purely to an increased number of modifiable residues or rather a true increase in PTM frequency. To assess this, we divided the PTM frequencies observed in BAD/RS/RG/YG peptides by the frequencies observed in the background nucleoplasm proteome (no-M) to estimate the relative fold increase in PTM frequency. We found markedly increased methylation and phosphorylation fold changes, indicating that the hyper-modification observed within BAD, RS and RG sequences is not simply caused by an increase in modifiable residue density within these regions, but rather a naturally increased likelihood for residues in these domains to be modified (**Fig.2.10b**). For example, serine in particular is more than six times more likely to be phosphorylated in BAD or RS domains, compared with the rest of the proteome, even when normalizing for the presence of serine by our selection criteria (**Fig.2.10b**). Interestingly, RG peptides were the most likely to contain dimethylated arginine, although they were less likely to contain monomethylated arginine as compared with BAD or RS peptides (**Fig.2.10b**). Surprisingly, YG peptides were less likely than no-M peptides to contain phosphorylation, though slightly more likely to contain arginine methylation (**Fig.2.10b**). Tyrosine phosphorylation was particularly infrequent within YG motifs, even less so than no-M sequences (**Fig.2.10b**). Overall,

YG peptides were much less likely to be modified when compared with BAD, RS or RG peptides. Thus, by increasing proteomic coverage with middle-down ETD, we discovered that Arg-rich domains, particularly RS motifs, contain markedly increased PTM densities.

2.4.8 Phosphorylation and methylation favor RSRS motifs

To determine if increased proteomic coverage influences phosphorylation and methylation PTM motif sites, a motif analysis was performed to determine if consensus sequences were observed at sites of arginine methylation and serine phosphorylation PTMs. Our nucleoplasm middle-down ETD MS PTM-containing sequences were searched in 31-residue width windows using the Motif Logo tool²⁰³ offered by phosphosite.org to detect increased frequencies of residues adjacent to a central modified residue for monomethylated arginine (mmR), dimethylated arginine (dmR) and phosphorylated serine (pS) (**Fig.2.10c**). Importantly, the PTM motif searches were conducted in isolation, without consideration of other PTM subtypes. As expected, the canonical RG/RGG motif was found as a favored consensus motif for both arginine monomethylation and arginine dimethylation^{166,239,240}. Unexpectedly, an “RSRS” motif featured prominently for not only serine phosphorylation²⁴¹, but also arginine dimethylation (**Fig.2.10c**). Thus, the difference of motifs between mono-methylation and di-methylation modification states of arginine highlights the utility of the middle-down ETD strategy and alternate MS approaches in uncovering unreported features of the proteome.

To assess whether serine phosphorylation and arginine methylation co-occur within RSRS motifs, a search was performed to select peptide sequences containing only RSRS motifs. Interestingly, although the combination of both arginine methylation and serine phosphorylation occur most frequently within BAD and RS peptides in nucleoplasmic RNA-binding proteins, these PTMs do not tend to co-occur within an RSRS motif. Namely, the most frequent modification

states of the RSRS motif are either serine phosphorylation in isolation (43.6%) or arginine methylation in isolation (23.5%) (**Fig.2.10d**). This disparity of modification states suggests a complex layer of co-regulatory cross-talk between arginine and serine residues within Arg-rich domains.

2.4.9 Phosphorylation regulates protein-protein interactions between structurally similar Arg-rich RNA binding proteins

In an effort to establish the biological role of the PTMs we identified in this study, we sought to test how phosphorylation within Arg-rich RBPs regulated protein-protein interactions. Phosphorylation of SR proteins disassembles nuclear speckles, freeing these proteins to the nucleoplasm to participate in RNA splicing^{124,242-245}. The phosphorylation state of core members of the spliceosome furthermore regulates assembly behavior and splicing activity of the overall spliceosome machinery^{54,96,165,214,215,219,220}. We identified 16 sites of phosphorylation on native SRSF2 protein in the nucleoplasm sample by our middle-down approach (**Fig.2.9d**). Previously, we have shown the SR-like LC1/BAD domain of U1-70K is essential for self-association and interactions with proteins involved in mRNA processing and snRNP assembly, including SRSF2²¹⁰. As phosphorylation is integral to the regulation of protein-protein interactions, we hypothesized that dephosphorylation of SRSF2 could influence these interactions.

To assess the impact phosphorylation has on SRSF2 protein interactions we transiently transfected HEK-293T cells with plasmids expressing recombinant SRSF2 followed by treatment with calf intestinal alkaline phosphatase (CIP), an alkaline phosphatase that catalyzes dephosphorylation of phospho-serine/-threonine/-tyrosine residues. Notably, SRSF2 in nucleoplasmic fractions migrated at a lower molecular weight following CIP treatment by immunoblot analysis (**Fig.2.11a**). This difference in molecular weight is likely due to the extensive

dephosphorylation of serine at 16 sites (**Fig.2.9d**). To assess the impact phosphorylation has on PPIs, Mock- or CIP-treated nucleoplasm fractions were immunoprecipitated for recombinant SRSF2 followed by immunoblotting for U1-70K and LUC7L3, which are two RNA-binding proteins with BAD motifs (**Fig.2.11b**). Notably, both U1-70K and LUC7L3 showed significantly increased interaction with recombinant SRSF2 following CIP treatment, compared with mock-treated (-CIP) nucleoplasm IPs or the IgG IP control. Thus, dephosphorylation of recombinant SRSF2 and structurally similar arginine-rich RBPs, U1-70K and LUC7L3, increases their respective interactions suggesting that phosphorylation (or lack thereof) regulates these interactions between these RBPs *in vivo*.

2.5 DISCUSSION

Mass spectrometry has achieved great advances towards the goal of globally characterizing PTMs within proteins²⁴⁶, yet significant gaps in understanding remain. Here we describe a method to examine the PTM diversity of Arg-rich proteins, re-purposing a technique first conceptualized by others to explore the biology of histone tails^{178,186,194,247}. The power of middle-down ETD is to identify unique combinatorial PTMs within Arg-rich proteins, many of which being RBPs that aggregate in neurodegenerative disease^{154,197,219}. This technique can be further used to determine the PTM profiles of Arg-rich domains across the proteome, which may illuminate key regulatory steps in RNA processing and metabolism.

Here, we achieved full proteomic sequence coverage of the recombinant arginine-rich U1-70K LC1/BAD domain by a combination of limited proteolysis and ETD approaches. We then expanded the approach for the proteomic analysis of nucleoplasm proteins isolated from mammalian cells, many of which are Arg-rich RBPs. Arg-rich sequences containing BAD, RS and RG motifs were highly modified in both methylation and phosphorylation, while peptides

containing YG motifs were infrequently modified, especially by phosphorylation. As phosphorylation may aid in the phase transition of YG-motif containing proteins⁶⁸, cells may contain mechanisms to negatively regulate tyrosine phosphorylation at these sites.

A vital discovery achieved by our analysis is the combinatorial nature of PTMs within Arg-rich domains, highlighted by a shared RSRS motif between arginine dimethylation and serine phosphorylation in nucleoplasm proteins^{239,248}. Serine-arginine protein kinases (SRPKs) and protein arginine methyltransferase (PRMT) enzymes modify arginine-rich domains in RNA-binding proteins^{172,234,249,250}. Due to the proximal relationship of methylation and phosphorylation PTMs, cross-talk between adjacent arginine and serine residues within RSRS motifs is likely. Phosphorylation and methylation have been previously shown to functionally oppose one another. For example, arginine methylation at RxRxxS/T motifs, similar to RSRS motifs, have emerged as an important negative regulator of serine phosphorylation. The mouse homolog of forkhead box O1 (FOXO1) is methylated *in vitro* and *in vivo* at its Akt recognition motif site at R248 and R250, blocking phosphorylation of nearby S253 and preventing FOXO1 nuclear export²⁵¹. Conversely, phosphorylation also blocks methylation in certain contexts. Namely, phosphorylation of RNA polymerase II (RNAPII) at its carboxy terminal domain prevents symmetric arginine dimethylation at R1810, integral to SMN protein interaction^{252,253}. It is evident that there is functional *cis*-acting crosstalk between site-specific methylation and phosphorylation PTMs that is context-dependent. Therefore, the overall ratio between methylation and phosphorylation in arginine-rich domains may critically define the protein interactions and function. The global identification of novel sites and further interrogation of the crosstalk between methylation and phosphorylation of RS motifs is warranted.

The appearance of a range of arginine methylation states reflect a dynamic PTM process, in which protein function may be tuned by varying degrees of methylation of contiguous arginines. Interestingly, PRMTs appear to have a broad yet essential role in regulating alternative splicing. Arg-rich RBPs are preferential substrates of the Type II methyltransferase PRMT5^{174,250}. PRMT5-depleted cells critically trigger changes in gene expression, cell-cycle de-regulation and alternative splicing^{250,254}. Another type II enzyme, PRMT9, regulates alternative splicing by methylation of spliceosome-associated protein 145 (SAP145) at R508, priming SAP145 for interaction with the protein SMN²⁵⁵. Thus, arginine methylation is likely an essential regulator of alternative splicing, and the identification of residue-specific arginine methylation sites in splicing RBPs can reveal novel and essential insights into the structural and functional inquiry of these proteins. MS analyses have been performed on PRMT inhibitor-treated cells to reveal site-specific targets to individual arginine residues¹⁷⁴, and middle-down ETD MS approaches may serve to expand these sites to those within domains containing BAD, RS and RG motifs. This may offer insight to the therapeutic capacity of individual PRMT enzyme inhibition on a disease-specific basis. Numerous lines of evidence suggest that further characterization of RBP modifications will increase our understanding of multiple steps in RNA processing^{158,160,172,256}.

Another fundamental observation made in this study is that Arg-rich domains are densely modified by PTMs. These modifications likely regulate numerous aspects of physiological RBP function and localization. Interestingly, these modifications were observed almost exclusively in BAD and RS motif-containing peptide sequences, many of which participate in protein-protein functions^{154,163-165}. For example, U1-70K was found to contain 25 PTMs, 23 of which were located in Arg-rich BAD and/or RS peptide sequences. U1-70K also has a role in nucleic acid binding, binding to Stem Loop I of the U1 snRNA through sequences within the N-terminal and RRM

domains (AA88-103, 122-131, 202-209)^{108,217,257}. While phosphosite.org lists three PTMs within these sequences (ggK103, pY126, ggK130)²⁰³, we found many more PTMs located in the Arg-rich LC1/BAD domain of U1-70K. Although there exists structural data for these RNA binding sequences within U1-70K^{108,257}, the structures of the low complexity LC1/BAD and LC2 domains of U1-70K are not solved, and similarly, few structures of other low complexity Arg-rich domains such as those within the SR family of proteins exist^{232,235,258,259}. Numerous studies point to the role of Arg-rich BAD and RS domains in protein-protein interactions^{154,163-165}, and we hypothesize that the high density of PTMs we observed within them regulate protein-protein interactions. Therefore, targeted amino acid substitution experiments, biochemical characterization and improved structural studies of Arg-rich low complexity domains may resolve the role of methylation and phosphorylation PTMs in U1-70K and other splicing RBPs.

In addition to regulating localization and molecular interactions, PTMs within Arg-rich domains may also trigger toxic gain-of-function pathological properties in neurodegenerative disease. For example, net positive charge within the disordered LC1/BAD domain of U1-70K mediates the condensation, insolubility, and eventually, aggregation of U1-70K^{73,155}. Importantly, phosphorylation PTMs within LC1/BAD may be a mechanism to regulate the condensation behavior of U1-70K by altering the overall charge within the LC1/BAD domain. Similarly, phosphorylation within the LC domain of FUS acts to prevent LLPS, aggregation and toxicity *in vitro*¹⁴⁶. We indeed show phosphatase-treated SRSF2 shows a stronger interaction with fellow Arg-rich proteins U1-70K and LUC7L3. Therefore, the localization, condensation and molecular interactions of these proteins is likely regulated by phosphorylation.

Future examinations with advanced search engines, novel fragmentation technologies and high resolving power Orbitrap MS/MS strategies will complement this study to increase the

identification of combinatorial PTM within Arg-rich RBPs. We performed our analyses on a Orbitrap mass spectrometer platforms, with a high-resolving power precursor (MS1) scan and a low-resolving power ion-trap MS/MS scan. Lysine acetylation (42.011 ± 0.004 Da) and trimethylation (42.047 ± 0.002 Da) variable modifications were therefore excluded in this current study, as the change in mass between the divergent PTM fragment ions cannot be differentiated using a low-resolving power ion-trap MS/MS scan (product ion mass tolerance = 0.6 Da). Thus, future studies resolving these biologically-significant PTMs is warranted, and may be achieved using a high-resolving power MS/MS scan. Furthermore, immobilized metal affinity chromatography (IMAC) and other PTM immunoaffinity approaches may further address the depth of PTM in the nucleoplasm proteome. New, powerful search engines and algorithms are currently being developed to handle the expanded search space of multiple PTMs, such as MSFragger and TagGraph^{260,261}. Importantly, the unique pattern of ETD fragmentation itself aids in the sequencing of modified peptides¹⁹⁵. The generation of many c- and z-type fragment ions contributes to the generation of unique fragment ions across a wide range of m/z values, increasing the likelihood of accurate peak calling by search engine algorithms, compared with HCD/CID fragmentation platforms. Complementary ‘electron-transfer/higher-energy collisional dissociation’ (EThCD) fragmentation technology, which generates both ETD fragment ions (c/z) and HCD ions (b/y), may further enhance the identification of low complexity Arg-rich peptides and PTM site localization within them^{262,263}.

Finally, by illuminating previously “dark” regions of the Arg-rich proteome we have discovered densely-modified regions within specific classes of RBPs. This dataset is an immense resource for the RBP community that investigates the roles of PTMs in RNA processing. The site-specific PTMs identified herein may be leveraged in future mechanistic studies in model systems

of these biologically-significant proteins. Increasing the catalog of PTMs could be key in understanding global regulation of the spliceosome and non-histone epigenetic influences on gene expression, highlighting the ability of middle-down proteomic approaches to discover previously unappreciated aspects of cell biology.

2.6 Acknowledgments

Support was provided by 5R01AG053960 and by the Accelerating Medicines Partnership for AD (U01AG046161). S.K. was supported by the Training Program in Biochemistry, Cell and Developmental Biology (T32GM008367) and Emory Neurology Department Training Fellowship (5T32NS007480-19). I.B. was supported by a pre-doctoral NINDS training grant (T32NS007480) and an individual NRSA grant (F31NS09385902). We thank Anita Corbett (Emory Department of Biology) for her helpful comments and suggestions.

2.7 FIGURES

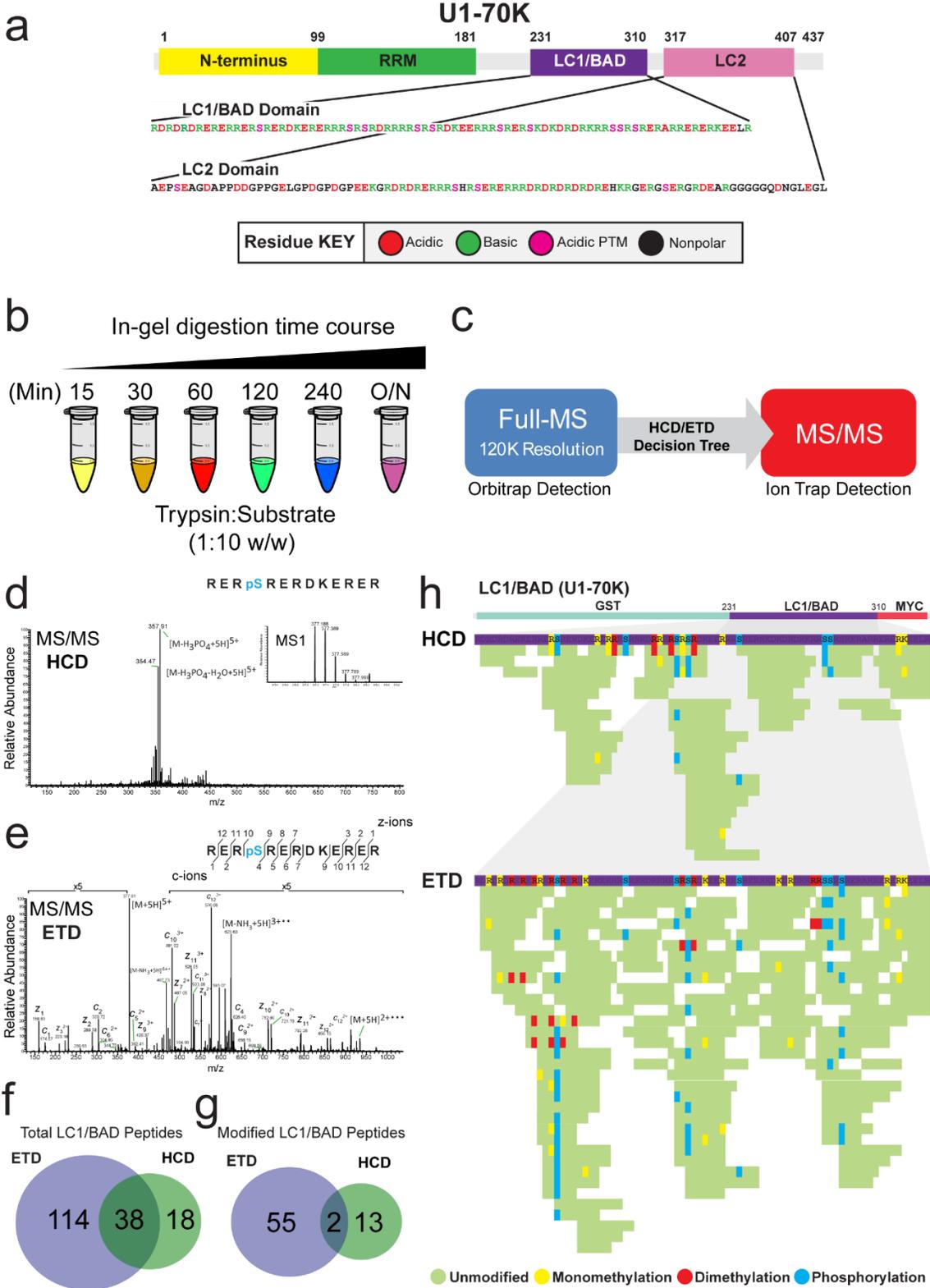


Figure 2.1. Electron Transfer Dissociation (ETD) more effectively fragments Arg-rich peptides compared with HCD. (a) Schematic of U1-70K protein domains. (b) At six different time points [15, 30, 60, 120, 240 minutes, and overnight (O/N)] the digestion reaction was quenched with acetic acid. (c) Extracted peptides were analyzed by LC-MS/MS on an Orbitrap Fusion using a data-dependent decision tree acquisition method¹⁹⁰ to alternatively select between HCD and ETD peptide fragmentation based on the charge state and m/z of the precursor ion. A two-step database search method using the Andromeda search engine was employed wherein proteins that were identified from a primary target-decoy search against human UniProtKB database were used to create a second smaller focused database. This focused database was then used to search for phosphorylation (serine, threonine or tyrosine) and mono-, di- and tri-methylation (arginine or lysine) with consideration of up to 6 missed trypsin cleavage events. (d) HCD MS/MS output of a R242-R254 U1-70K LC1/BAD peptide precursor $[M+5H]^{5+}$. HCD results in a strong neutral loss (-98 Da) of phosphoric acid and few fragment ions (b- and y-ions). (e) Isolation and ETD fragmentation of the same precursor peptide reveals an information-rich MS/MS spectrum (c- and z-ions), while preserving the phosphorylated serine, labile by HCD. Charged-reduced precursors (•) are also annotated in the MS/MS spectrum (f) ETD fragmentation produces the vast majority of unique peptide matches to U1-70K LC1/BAD as compared with HCD. (g) ETD fragmentation facilitates the identification of increased number of post-translationally modified peptides matched to U1-70K LC1/BAD as compared with HCD. (h) Protein coverage generated by HCD fragmentation compared with ETD fragmentation across all trypsin incubation lengths (15, 30, 60, 120, 240 minutes, and overnight). Peptides are colored according to whether each residue is unmodified (green), mono-methylated at an Arginine/Lysine (yellow), di-methylated at Arginine/Lysine (red) or phosphorylated at Serine/Threonine/Tyrosine (turquoise).

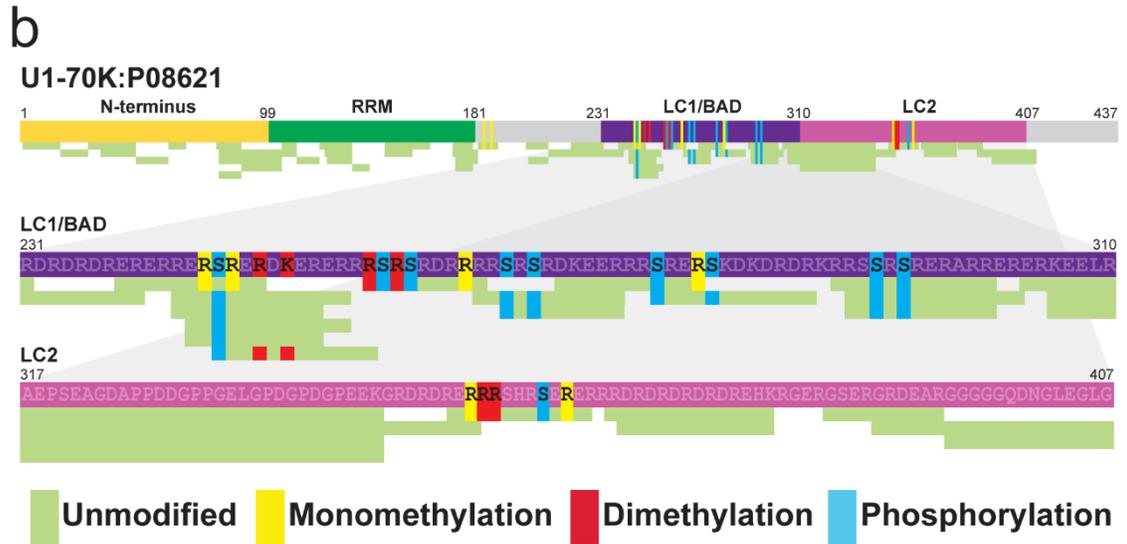
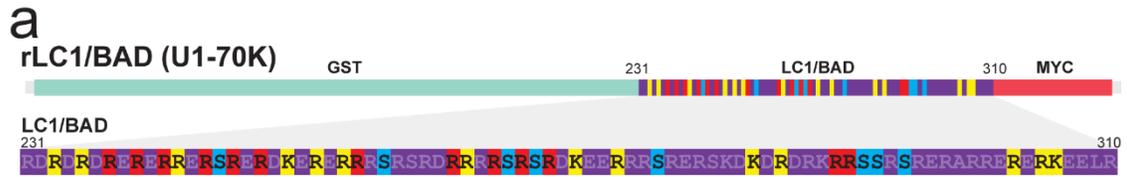


Figure 2.2. Middle-down ETD MS strategy provides enhanced sequence coverage of U1-70K arginine-rich low complexity domains. (a) Summation of modified peptide spectral matches (PSMs) of all trypsin digestion times and ETD/HCD dissociation methods on the LC1/BAD domain of the purified recombinant LC1/BAD protein. Peptides are colored according to whether each residue is unmodified (*green*), monomethylated at an Arginine/Lysine (*yellow*), dimethylated at Arginine/Lysine (*red*) or phosphorylated at Serine/Threonine/Tyrosine (*turquoise*). Using middle-down ETD, near complete coverage of the LC1/BAD domain was achieved, identifying 8 phosphorylation sites and 23 methylation sites (12 mono-methylated arginines, 13 di-methylated arginines, and four mono-methylated lysines) in total. (b) Amino acids 231-310 of the HEK-293T-native U1-70K LC1/BAD domain was almost completely sequenced (76/80 residues, 95%), identifying 9 phosphorylation sites with 8 total methylation sites. By comparison, the U1-70K LC2 domain, without BAD or RS motifs but similar arginine content, contains less methylation and phosphorylation PTM.

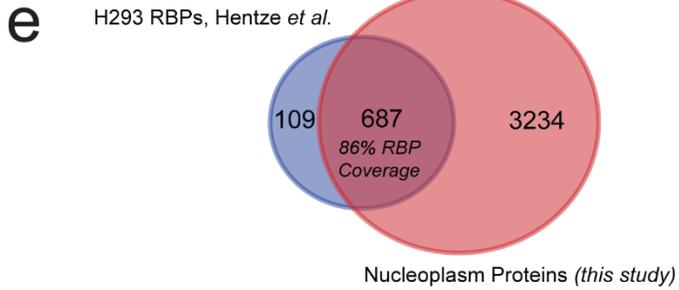
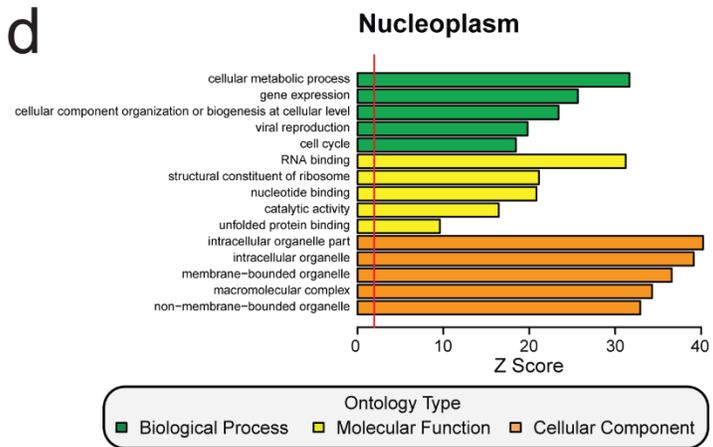
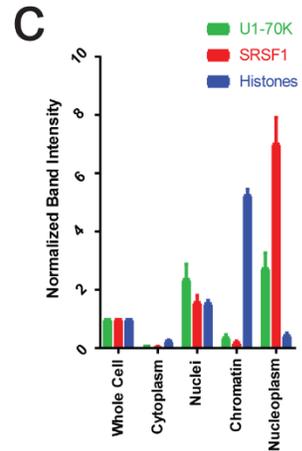
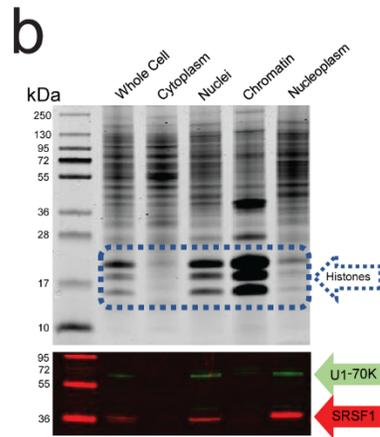
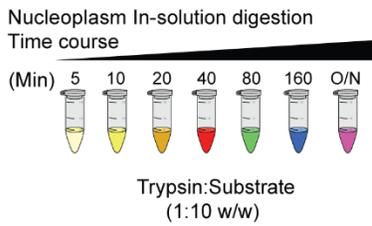
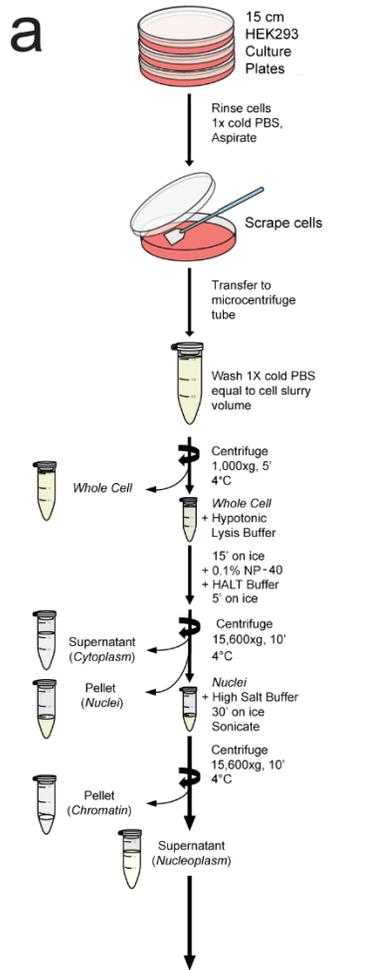


Figure 2.3. Isolation of nucleoplasm by cellular fractionation enriches for RNA-binding proteins. (a) Schematic of the HEK-293T cellular fractionation protocol. (b) At the top of the panel, the individual fractions were resolved by SDS-PAGE and stained with Coomassie Blue to visualize protein. Histone proteins at the indicated molecular weight span approximately 15-22 kDa and are outlined (*dashed blue box*). At the bottom of the panel, immunoblot analysis for U1-70K (*green*) and SRSF1 (*red*) in cellular fractions. (c) Densitometry quantification of nuclear histone and RBP proteins across cellular fractions. Histone proteins show a 5.26-fold enrichment in the chromatin fraction, and 0.45-fold depletion in the nuclear extract fraction relative to the whole cell homogenate. U1-70K shows a 2.75-fold enrichment in the nucleoplasm as compared to levels in the whole cell homogenate. Similarly, SRSF1 shows a 7.02-fold enrichment in the nucleoplasm compared to whole cell homogenate. (d) Gene ontology (GO) analysis of nucleoplasm proteins identified by middle-down proteomics. Significant over-representation of an ontology term within the nucleoplasm sample compared to the background proteome is reflected with Z score greater than 1.96, equivalent to $p < 0.05$ (to the right of the red line). (e) When comparing to 796 RNA-binding proteins previously identified in HEK-293T cells⁶⁴, our nucleoplasm extract dataset matched to 687 of these (86%), as well as identified 3234 other nuclear proteins.

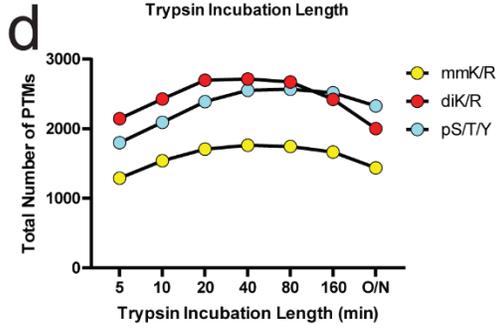
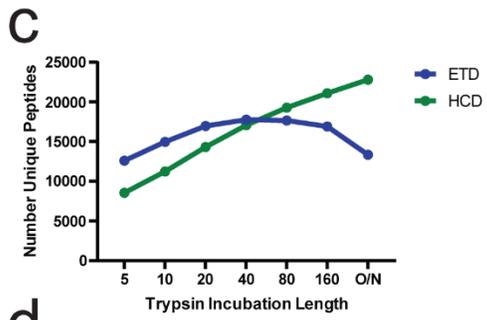
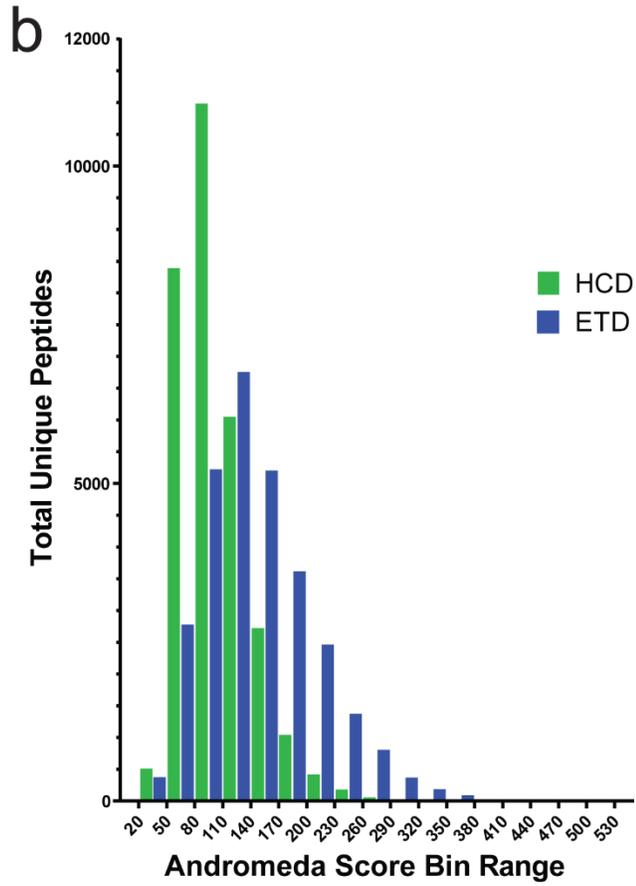


Figure 2.4. Middle-down ETD analysis of nucleoplasm contributes to greater coverage of Arg-rich RNA-binding proteins. (a) Venn diagram showing overlap between ETD and HCD fragmentation strategies of 57,871 unique peptides sequenced from nucleoplasm extract. (b) Andromeda scores of peptides fragmented by ETD (*blue bars*) versus HCD (*green bars*). (c) The number of peptides fragmented by either ETD (*blue*) or HCD (*green*) at each trypsin incubation time length. (d) The total number of PTM across increasing trypsin incubation time lengths (monomethylation of Lysine/Arginine = mmK/R, *yellow*; dimethylation of Lysine/Arginine = dmK/R, *red*; phosphorylation of Serine/Threonine/Tyrosine = pS/T/Y, *turquoise*).

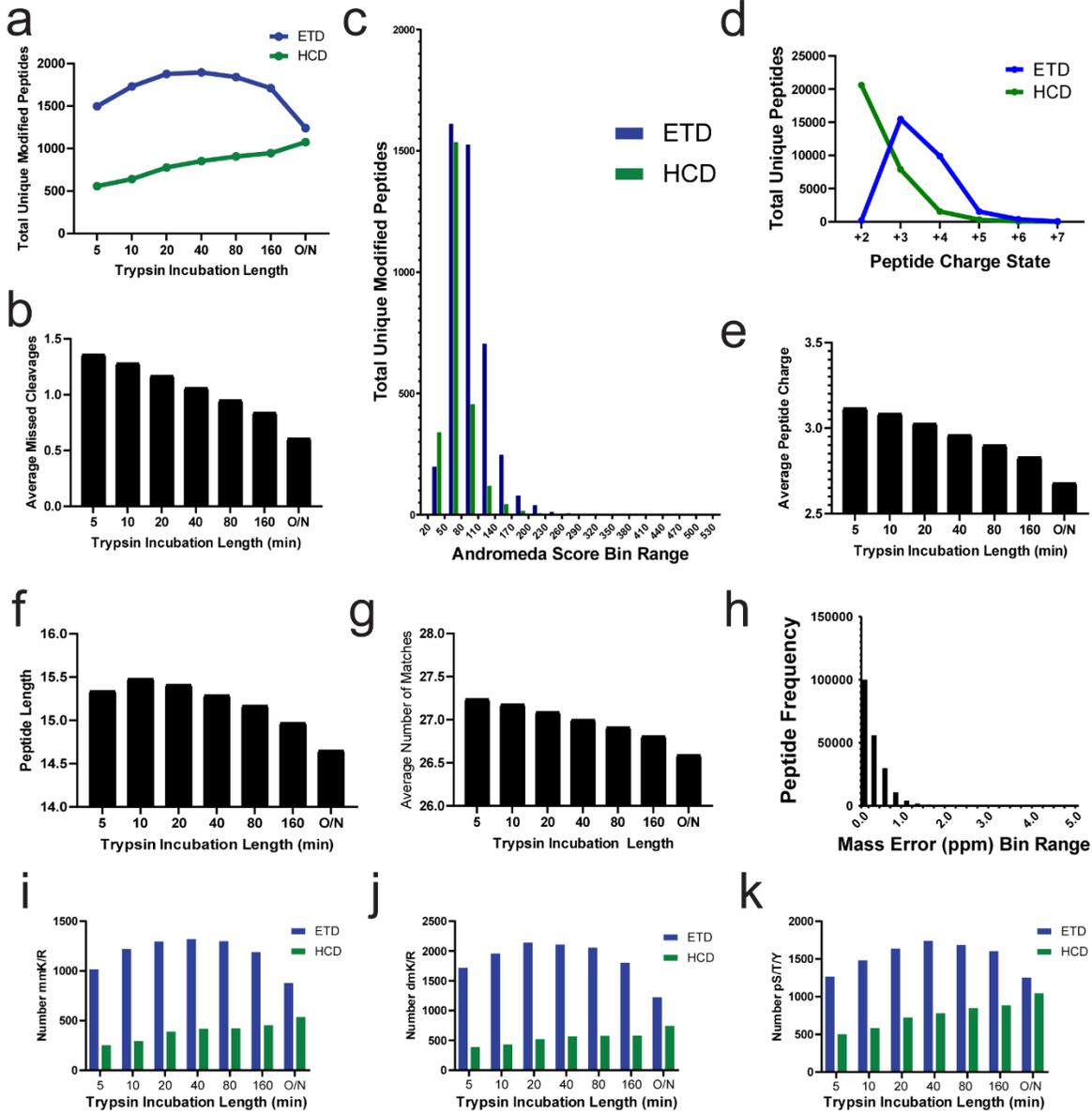


Figure 2.5. HCD and ETD fragmentation of nucleoplasmic sample enhances identification of unique peptides. (a) The number of unique modified peptides fragmented by ETD and HCD at all trypsin digest time points. (b) Average missed cleavages events decrease as trypsin incubation times increase. (c) Andromeda scores of unique modified peptides fragmented by ETD (*blue bars*) versus HCD (*green bars*). (d) As determined by the decision tree approach ¹⁹⁰, peptides with charges greater than +2 are preferentially sequenced by ETD. (e) The average charge of a peptide decreases with increasing trypsin incubation time. (f) Peptide length decreases with increasing trypsin incubation time. (g) The average number of matches per peptide across trypsin incubation time lengths. (h) Most assigned peptides (97%) matched to the measured mass with under 1.0 parts per million (ppm) error. (i) The number of peptides fragmented by ETD (*blue bars*) versus HCD (*green bars*) containing monomethylation at Lysine/Arginine (mmK/R) across trypsin incubation time lengths. (j) The number of peptides fragmented by ETD (*blue bars*) versus HCD (*green bars*) containing dimethylation at Lysine/Arginine (dmK/R) across trypsin incubation time lengths. (k) The number of peptides fragmented by ETD (*blue bars*) versus HCD (*green bars*) containing phosphorylation at Serine/Threonine/Tyrosine (pS/T/Y) across trypsin incubation time lengths.

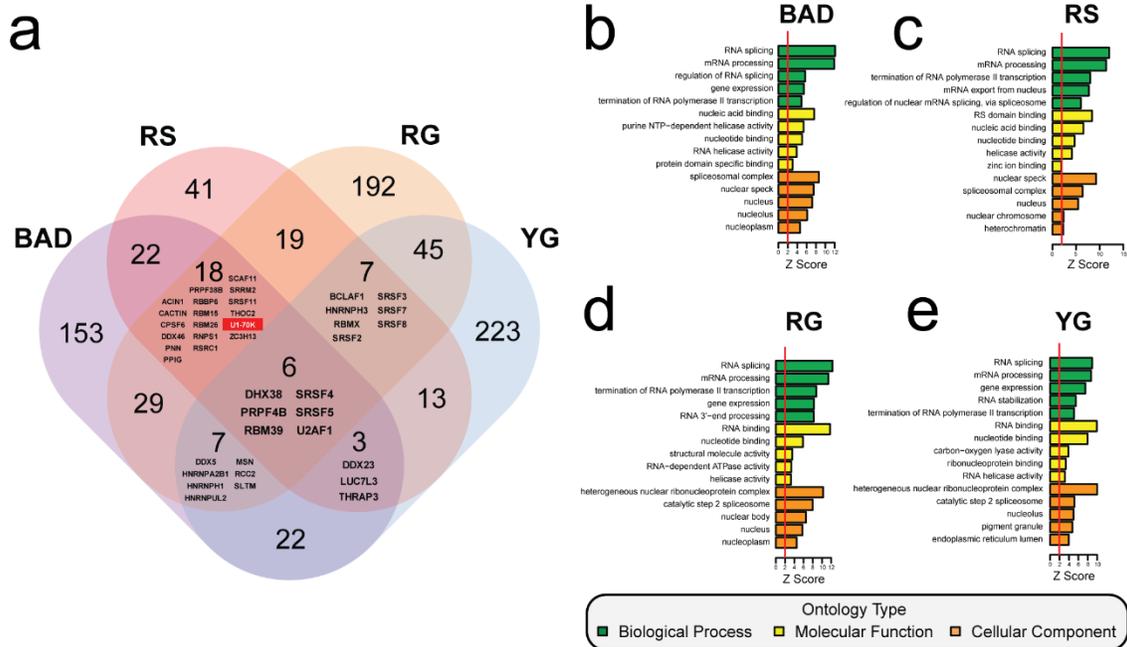


Figure 2.6. Motif algorithms resolve RNA-binding protein subgroups with distinct biological properties. (a) A scoring algorithm was developed (see methods) to select unique peptides that contain BAD-/ RS-/RG-/YG-patterned primary sequences, and the group membership of proteins containing these motifs by Venn diagram. The U1 snRNP member U1-70K (*red*) contains BAD, RS and RG motifs. (b-e) Gene ontology (GO) analysis of BAD and RS proteins identified by our algorithm. Significant over-representation of an ontology term is reflected with a Z score greater than 1.96, equivalent to $p < 0.05$ (to the right of the red line). (b) The BAD proteins (237) have the expected biological processes ('RNA splicing' and 'mRNA processing'), and subcellular localization ('spliceosomal complex' and 'nucleus'). (c) RS proteins (117) retain similar processes and localization ('nuclear speckles' and 'spliceosomal complex'). (d) RG proteins (323) localized to the 'heterogeneous nuclear ribonucleoprotein complex' and 'catalytic step 2 spliceosome', and are involved in 'RNA splicing' and 'RNA binding'. (e) YG proteins (326) also localized to the 'heterogeneous nuclear ribonucleoprotein complex' and 'catalytic step 2 spliceosome', and the primary group biological process and molecular function were 'RNA splicing' and 'RNA binding', respectively.

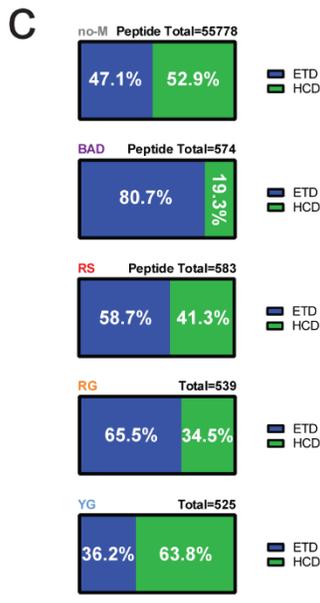
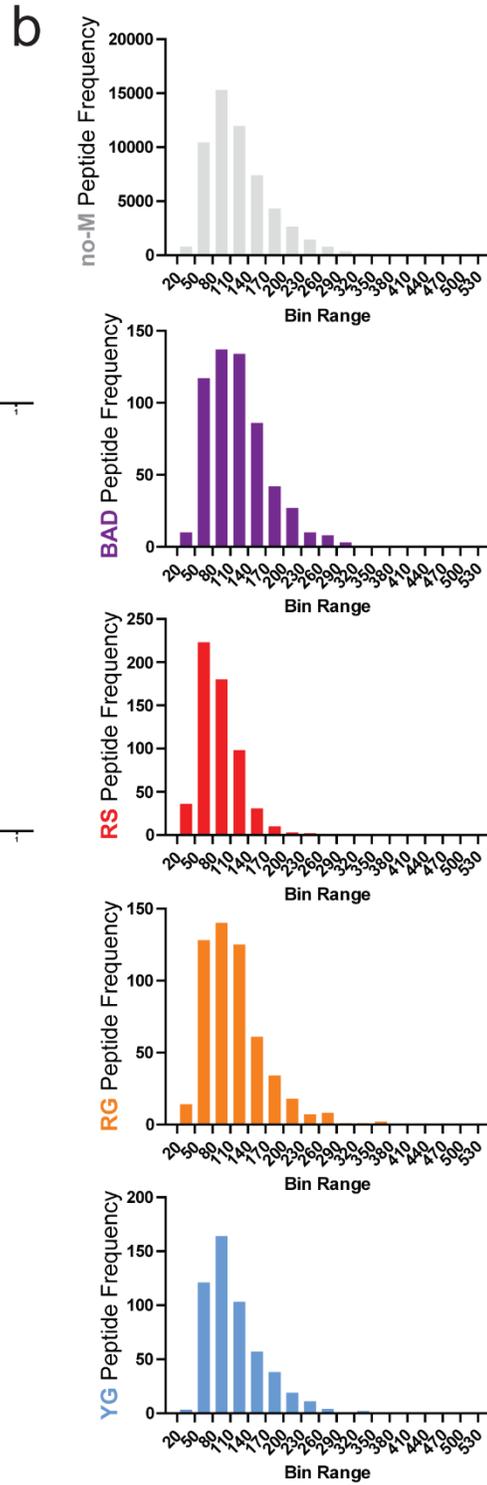
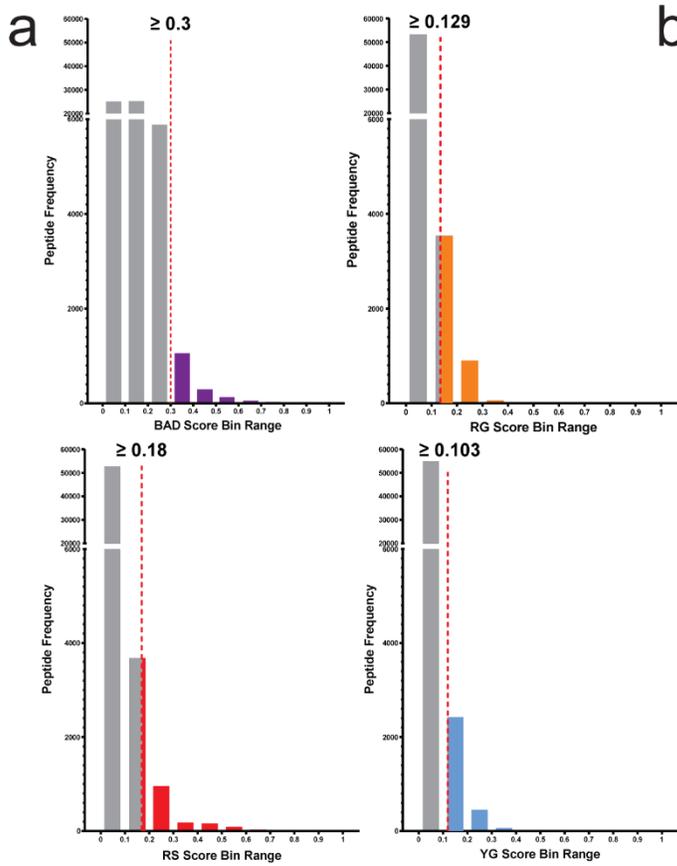
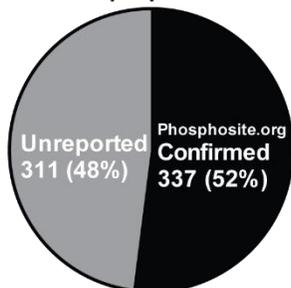


Figure 2.7. A motif algorithm stringently selects high-confidence peptides. (a) A scoring algorithm was developed (see methods) to select identified unique peptides that contain BAD-/RS-/RG-/YG-patterned primary sequence. A total of 57,871 unique peptides were scored using the scoring algorithms, and a conservative score cutoffs of 0.3 or greater (BAD), 0.18 or greater (RS), 0.129 or greater (RG) or 0.103 or greater (YG) were used for each algorithm. (b) Histogram analysis of the MaxQUANT Andromeda Scores of no-M (*grey*), BAD (*purple*), RS (*red*), RG (*orange*) and YG (*blue*) peptides. (c) Frequency of ETD (*blue*) or HCD (*green*) fragmentation strategies yielding successfully-identified unique peptides to no-M, BAD, RS, RG and YG peptide subgroups.

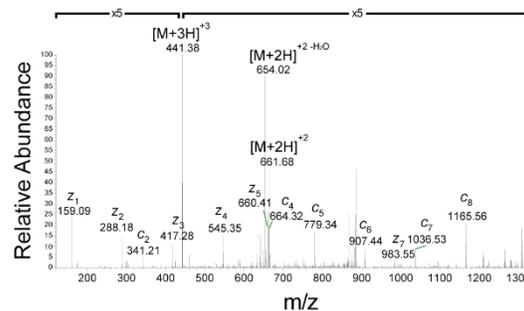
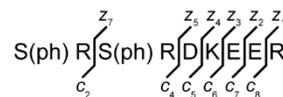
a

648 Unique pS/T/Y Sites



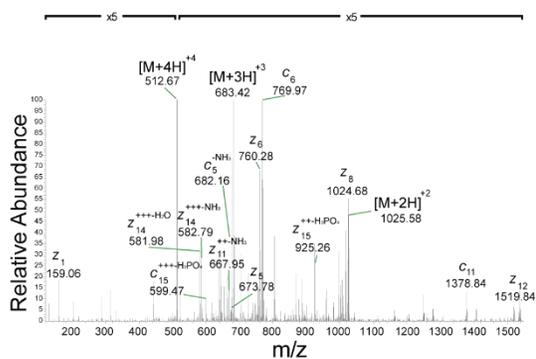
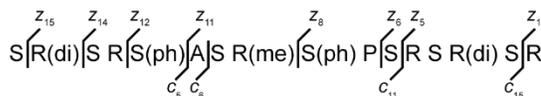
b

U1-70K [S266-R274]; 20'; Scan #2811; ETD



c

SRSF4 [S474-R489]; 160'; Scan #17701; ETD MS/MS



d

LUC7L3 [S333-R341]; 160'; Scan #782; ETD MS/MS

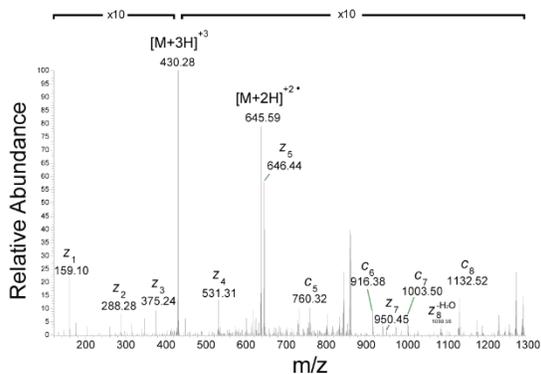
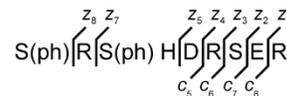


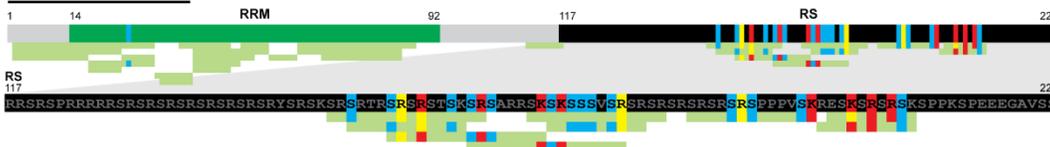
Figure 2.8. PTM validation by phosphosite.org database comparison and manual inspection of spectra. (a) Phosphorylation sites at Serine/Threonine/Tyrosine (pS/T/Y) observed within the nucleoplasm extract dataset were compared with sites listed on the proteomic PTM database phosphosite.org²⁰³, and either validated (confirmed), or represent new PTM sites unreported by phosphosite.org. (b) ETD-fragmented S266-274 captures a doubly-phosphorylated peptide within the LC1/BAD domain of U1-70K. Both the $[M+2H]^{2+}$ and $[M+3H]^{3+}$ precursor ions are readily assigned, along with a series of c- and z-type fragment ions. (c) ETD-fragmented MS/MS spectra of S474-R489 peptide mapping to SRSF4. The $[M+4H]^{4+}$ precursor is the most abundant ion in a spectra which includes a collection of c- and z-ions (d) An ETD-fragmented MS/MS spectra of LUC7L3 spanning residues S333-R341 capturing phosphorylation at S333 and S334 within the BAD/RS domain of LUC7L3. A near-complete z-series of fragment ions is assigned, along with the $[M+3H]^{3+}$ precursor ion.

a

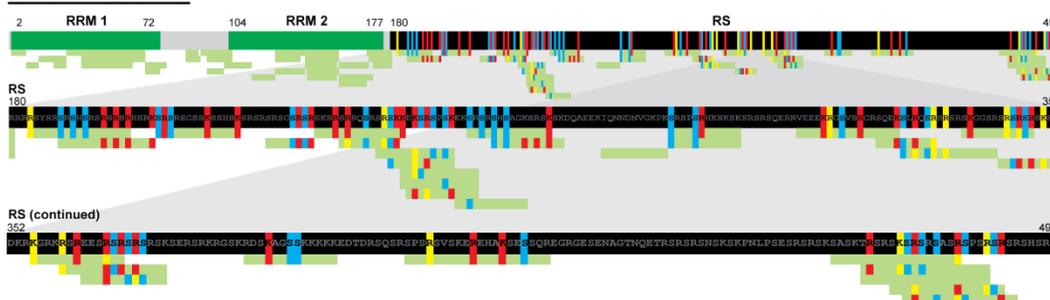
LUC7L2:Q9Y383



SRSF2:Q01130

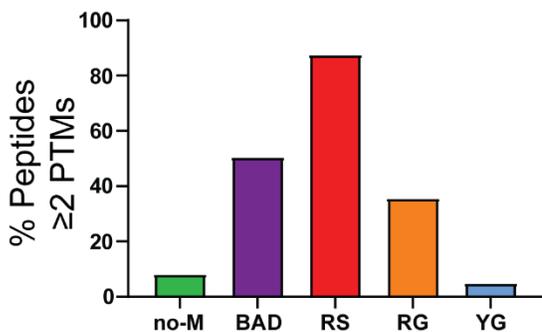


SRSF4:Q08170



● Unmodified ● Monomethylation ● Dimethylation ● Phosphorylation

b



c

# PTMs per peptide	no-M	BAD	RS	RG	YG
0	49923	227	23	294	464
1	1274	57	49	53	35
2	1255	86	127	66	14
3	1528	95	151	57	4
4	1170	71	137	48	8
5	603	36	95	20	0
6	25	2	1	1	0

d

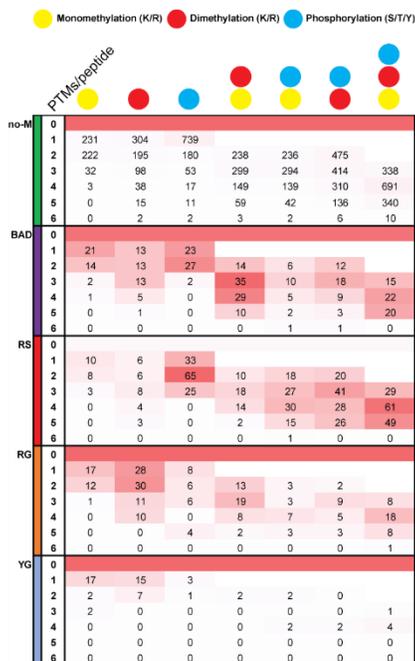
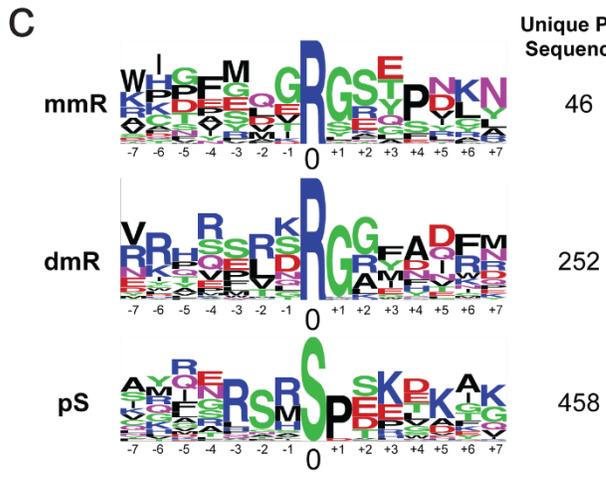
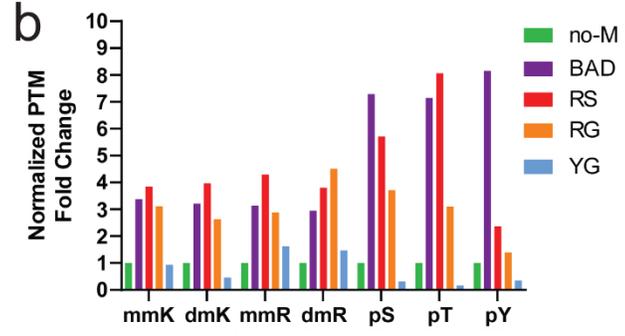
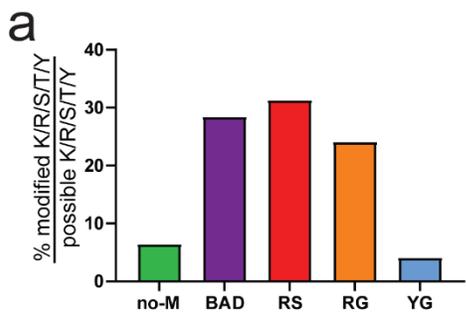


Figure 2.9. Arginine-rich RNA-binding proteins are enriched for PTMs. (a) Unique peptides were mapped to arginine-rich proteins LUC7L2, SRSF2, and SRSF4 (Uniprot ID following semicolon). Peptides are colored according to whether each residue is unmodified (*green*), monomethylated at an Arginine/Lysine (*yellow*), dimethylated at Arginine/Lysine (*red*) or phosphorylated at Serine/Threonine/Tyrosine (*turquoise*). High modification density was observed in the BAD and/or RS low-complexity (LC) domains of each of the three proteins. Sequences used to build this illustration are listed in **Supplemental Table S5**. (b) The percentage of peptides containing two or more PTMs was calculated for no-Motif (no-M), BAD, RS, RG and YG motif-containing peptides. (c) The number of peptides containing no (0) to increasingly-modified states of PTM (1-6) were tallied and compared, using a heat map to illustrate occupancies of each subgroup. (d) The number of peptides containing uniform to double- or triple-combinations of PTM subtypes were calculated and compared between peptide groups (no-M vs BAD vs RS vs RG vs YG), using a heat map to illustrate the frequency of each (monomethylation of Arginine/Lysine (*yellow*), dimethylation of Arginine/Lysine (*red*) or phosphorylated of Serine/Threonine/Tyrosine (*turquoise*)).



d

Sequence	Percent of Total Instances
RSRS	17.6
RSRS	16.6
RSRS	12.1
RSRS	9.4
RSRS	7.3
RSRS	6.9
RSRS	6.8
RSRS	4.2
RSRS	4.1
RSRS	4.1
RSRS	3.2
RSRS	2.6
RSRS	2.4
RSRS	1.8
RSRS	0.8

Figure 2.10. Arginine methylation and serine phosphorylation are enriched in RSRS motifs. (a) The density of modified residues (K/R monomethylation, K/R dimethylation, S/T/Y phosphorylation) within a peptide divided by the density of the total residues (K/R/S/T/Y) within the same peptide was averaged for peptide subgroups (no-M, BAD, RS, RG and YG) and presented as an averaged percentage. Approximately 31% of all residues in RS peptides are post-translationally modified, compared to 6.5% of no-M. In comparison, 28% of BAD peptide residues, 24% of RG peptide residues are post-translationally modified. Approximately 4% of residues within YG peptides are modified. (b) Fold enrichment of PTM frequency for lysine, arginine, and serine were observed in BAD, RS and RG peptides, normalized to the PTM frequency observed in no-M peptides. (c) Motif Logos illustrating residue frequencies adjacent to a center modified residue was performed for monomethylated arginine (mmR), dimethylated arginine (dmR) and phospho-serine (pS). (d) Fifteen combinations of arginine methylation and serine phosphorylation observed within RSRS sequences were normalized compared by a percent of total RSRS motif occurrences calculation and a heat map illustration.

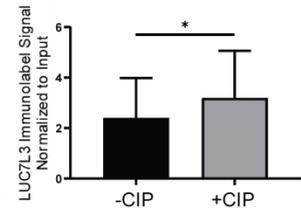
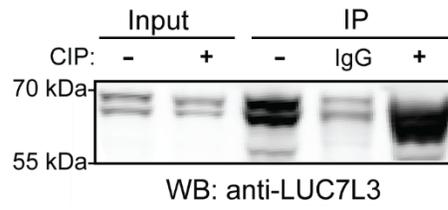
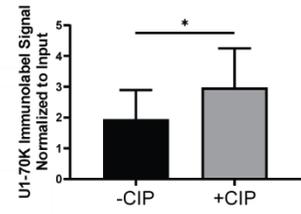
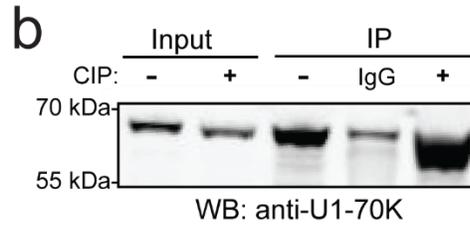
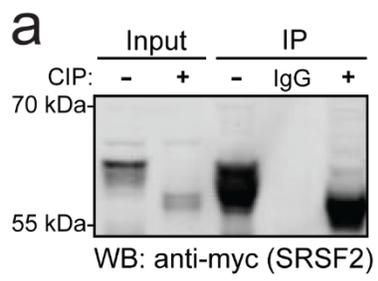
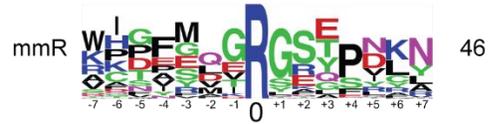
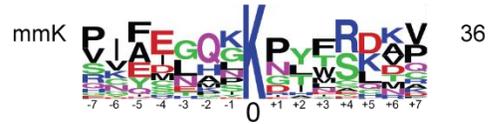


Figure 2.11. Dephosphorylation increases protein interactions between SRSF2 and the BAD RNA-binding proteins U1-70K and LUC7L3. (a) Nucleoplasm fractions of HEK-293T cells transfected with Myc-tagged recombinant SRSF2 were treated with water (-CIP) or calf intestinal alkaline phosphatase (+CIP) for 60 minutes at 37°C. Fractions were analyzed by SDS-PAGE and immunoblotted for Myc. Differences in electrophoretic mobility were observed due to the loss of phosphorylation as a result of phosphatase treatment. (b) Immunoprecipitation of Myc-tagged SRSF2 in -CIP and +CIP treated nucleoplasm fractions followed by immunoblotting for BAD proteins (U1-70K and LUC7L3). Paired t-test (n=3) showed significant enrichment of U1-70K (p=0.0347) and LUC7L3 (p=0.0422) in the co-immunoprecipitate complex following CIP treatment.

Unique PTM Sequences



Unique PTM Sequences



Figure 2.12. Motif analysis of nucleoplasmic PTM-modified peptide sequences. Unique PTM-containing sequences were searched using the Motif Logo tool on PhosphoSite.org to illustrate the frequencies of residues with increased occupancies at positions relative to a center modified residue (mmR=monomethyl arginine, dmR=dimethyl arginine, mmK=monomethyl lysine, dmK=dimethyl lysine, pS=phospho-serine, pT=phospho-threonine, pY=phospho-tyrosine).

CHAPTER 3.0 : Phosphorylation regulates arginine-rich RNA-binding protein solubility and oligomerization

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

*These findings were published in the *Journal of Biological Chemistry* on October 13, 2021 .
Supplemental tables can be found online at: [https://www.jbc.org/article/S0021-9258\(21\)01112-1/fulltext](https://www.jbc.org/article/S0021-9258(21)01112-1/fulltext)

3.1 ABSTRACT

Post-translational modifications (PTMs) such as phosphorylation of RNA-binding proteins (RBPs) regulate several critical steps in RNA metabolism, including spliceosome assembly, alternative splicing, and mRNA export. Notably, serine-/arginine- (SR) rich RBPs are densely phosphorylated compared with the remainder of the proteome. Previously, we showed that dephosphorylation of the splicing factor SRSF2 regulated increased interactions with similar arginine-rich RBPs U1-70K and LUC7L3. However, the large-scale functional and structural impact of these modifications on RBPs remains unclear. 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 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. We determined that 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 decreased high molecular weight SRSF2 species formation. Furthermore, upon pharmacological inhibition of SRPKs in mammalian cells, we observed cytoplasmic mislocalization and increased formation of cytoplasmic granules as well as cytoplasmic SRSF2 tubular structures that associated with microtubules by immunocytochemical staining. Collectively, these findings demonstrate that phosphorylation may be a critical modification that prevents arginine-rich RBP insolubility and oligomerization.

3.2 INTRODUCTION

RNA-binding proteins (RBPs) cooperatively engage both RNA and protein²⁶⁴. 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 molecules⁶⁴. Unbiased RNA interactome studies have identified many RBPs containing low complexity (LC) domains that participate in both RNA and protein interactions^{55,64-66}. LC domains are typically composed of a select few residues out of the entire amino acid code, giving rise to protein domains that are intrinsically disordered²⁶⁵. 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 fibrils²⁶⁶ *in vitro* and *in vivo*²⁶⁷. These assembly states are believed to be influenced in large part by RNA molecules^{70,268} and post-translational modifications (PTMs)^{140,146}. Although we are beginning to decipher a “molecular grammar” regulating LLPS⁶⁸, 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^{52,54,77,96,100,269-271}. Interestingly, LC domains are necessary for RBP LLPS and fibrillization^{100,270,272,273}, processes found to be regulated by PTM. LC RBPs are commonly modified by reversible PTM in the physiological milieu^{68,274}, yet in neurodegenerative disease phosphorylation PTMs increasingly occupy RBPs such as TDP-43^{134,275,276}. 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 phosphorylation²⁷⁷. One such group of arginine-rich RBPs with high levels of phosphorylation is the serine-/arginine-rich (SR) splicing factor family of RBPs¹¹⁹. This twelve member RBP family is known to contain at least one RRM RNA-binding domain²⁷⁸ 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 man⁵⁵. The most extensively studied regulator of SR protein function is phosphorylation, primarily catalyzed by nuclear cdc2-like kinases (CLKs)²⁷⁹ and cytoplasmic SR protein kinases (SRPKs)^{124,280,281}. Phosphorylation regulates nearly every facet of SR protein function, including splicing¹²⁶, coupling to sites of active transcription^{282,283}, subcellular localization^{125,284,285}, nuclear speckle compartmentalization^{124,279,286} and binding partner selection and affinity^{172,242,278,284}. Importantly, it is not fully understood whether excessive, or rather, insufficient phosphorylation alters the stability of SR proteins.

Our group²⁷⁷ and others^{220,285,287,288} 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 conditions²⁸⁹. 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 proteins²⁹⁰. 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.

3.3 MATERIALS AND METHODS

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), rabbit polyclonal anti-RBM25 (Abcam ab72237), 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-panQKI (generated by UC Davis/NIH NeuroMab Ab_10671658), rabbit polyclonal anti-ZC3H18 (ThermoFisher PA5-59322), rabbit polyclonal anti-SRRM1 (ab221061), mouse anti-GAPDH (Abcam ab8245), 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 Scotto²⁹¹ (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 Net Charge Per Residue (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 pH²⁹². 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 $\sim +0.013$, 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 hours²⁹³ 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 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 group¹⁹⁸ 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

μ L 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 μ g/mL Aprotinin, 10 μ g/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 μ L 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 μ g/mL Aprotinin, 10 μ g/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 μ g) 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 μ g) were incubated with either distilled water (mock) or 50 Units of calf intestinal alkaline phosphatase (QuickCIP, NEB M0525L) and brought up to 100 μ L with PBS. The samples were heated at 37°C for 1 hour.

Sedimentation Assay

Following QuickCIP dephosphorylation, 50 μ L of the 100 μ L sample was added to polycarbonate ultracentrifuge tubes. Samples were spun at 100,000xg for 1 hour at 4°C. The supernatant (vol=50 μ L) was transferred to a LoBind tube (Eppendorf 0030108442) and the insoluble pellet was resuspended in 50 μ L 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 μ L of total (pre-spin) fractions were added and 15 μ L of soluble and pellet fractions were added. For mass spectrometry analysis 30 μ L of soluble and pellet fractions were used.

Western Blotting

Western Blotting was performed according to standard protocol as previously described²⁷⁷. 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

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 µg of Lys-C (Wako 121-05063). Following Lys-C digestion, samples were diluted to 1M urea and digested overnight with 1 µg 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 µm Dr. Maisch, Germany) fused silica column (20 cm × 75 µm 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 Andromeda²⁹⁴ 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 search²⁰⁰. 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 ± 20 ppm was applied prior to mass accuracy calibration and ± 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.

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 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 WGCNA⁸⁸. The weighted protein co-expression network was built using the log₂-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 mediods 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^{89,295}. 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 group²⁹⁶:

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

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).

SR protein similarity and insolubility enrichment assessment

The RS domain (AA117-221) of SRSF2 was searched September 15, 2021 for similarity using the Uniprot pBLAST feature (<http://www.uniprot.org/blast/>) using the following parameters: Target Database Human, E-threshold: 10, Matrix: Auto, Filtering: None, Gapped: Yes, Hits: 1000. Similar proteins were then filtered to remove unreviewed protein entries, leaving 193 proteins with E-values less than 0.005 and similarity to the SRSF2 RS domain of greater than 20% (**Supplemental Table S6**). Protein alignment was performed using Clustal Omega multiple sequence alignment (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The list of 193 proteins was then compared to the list of proteins with increased insolubility following dephosphorylation ($n=734$) using a hypergeometric Fisher's exact test overlap using R.

SRPK2 *in vitro* kinase reaction

Human recombinant SRPK2 (0.4 μg /reaction; EMD Millipore 14-666; Glu46-end) and SRSF2 (0.8 μg /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). Different concentrations (10, 50, 100 μM) of SRPK inhibitor SRPIN340 was added previous to SRSF2 substrate. Either water (mock) or ATP was added to a final concentration of 100 μM to a final volume of 25 μL 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).

Radioactive kinase assays were performed as previously described²⁹⁷. In short, kinase reactions were set up in Kinase Buffer I (25 mM MOPS pH7.2, 12.5 mM β -glycerol phosphate, 25 mM MgCl₂, 5 mM EGTA, 2 mM EDTA, 0.25 mM DTT; Abcam ab189135) with 1 μ M “cold” ATP and 10 μ Ci γ -[³²P]ATP or only 10 μ Ci γ -[³²P]ATP totaling 10 μ L. Reactions were incubated for 10 min at 30°C and stopped by the addition of SDS loading buffer and analyzed by both SDS-PAGE and autoradiography.

Blue Native Gel Electrophoresis

Blue native PAGE was carried out as described previously¹⁰⁰. 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 (ThermoFisher 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).

Immunocytochemistry

Immunocytochemical staining was carried out as previously described²⁹⁸. 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.

Immunoprecipitation

Immunoprecipitations were performed essentially as described²⁷⁷. Briefly, HEK-293 cells were transiently transfected with either myc-tagged SRSF2 plasmid or a mock control pUC19 vector. Harvested cells were fractionated to isolate the nucleoplasm and cytoplasm as described above. Protein concentrations were determined using a bicinchoninic acid (BCA) assay (ThermoFisher 23225). Approximately 20 μ L of Protein A Sepharose 4B beads (catalog no. 101042, Invitrogen) were washed twice in immunoprecipitation (IP) buffer [50 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 0.5% (v/v) NP-40, 0.5% (v/v) CHAPS, HALT phosphatase inhibitor cocktail (1:100, ThermoFisher 87786) then blocked in 0.1 mg/ml bovine serum albumin (ThermoFisher 23209), then washed an additional three times in IP buffer. Anti-myc (4 μ g) mouse

monoclonal antibody (Cell Signaling 2276) or mouse isotype IgG control (4 μ g, BD Pharmingen 550339) were incubated with the bead slurry in IP buffer and rotated for 90 minutes to conjugate antibody to beads. The beads were then washed three times in IP buffer, then nucleoplasm lysates were added to beads (0.2 mg per IP) and incubated rotating overnight at 4°C. The beads were then washed three times in IP wash buffer (IP buffer lacking glycerol and CHAPS) then resuspended in IP wash buffer. Preceding the final wash, the bead slurry was transferred to a new microcentrifuge tube to limit contamination. The bead slurry was then centrifuged at 500 x g for 5 min at 4°C and the pelleted beads were resuspended with 1X Laemmli sample buffer (8% glycerol, 2% SDS, 50 mM Tris pH 6.8, 3.25% β -mercaptoethanol) and run by SDS-PAGE. Nucleoplasm fractions (10 μ g) were included as input loading controls. Three independent biological replicates were performed and a representative image was shown.

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository²⁹⁹ with the dataset identifier PXD026894 on 6/23/2021.

3.4 RESULTS

3.4.1 Phosphorylation prevents SRSF2 aggregation

Here, we use SRSF2 as a paradigm to study the regulation of arginine-rich RBP solubility, structure and morphology by phosphorylation. In SRSF2, the arginine-/serine-rich (RS) domain is highly phosphorylated²⁷⁷, a region with high probability of intrinsic disorder (**Fig.3.1a**)^{300,301}. We incubated lysates containing recombinant SRSF2-myc, a known phosphoprotein, with calf intestinal alkaline phosphatase (CIP) which corresponded to a faster migrating SRSF2 band

(**Fig.3.1b**) by SDS-PAGE, suggesting substantial dephosphorylation of SRSF2^{244,302}. 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)³⁰³⁻³⁰⁵ (**Fig.3.1c**). 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.3.2a**). We resolved equal amounts of total (T), soluble (S) and pelleted (P) fractions by SDS-PAGE and immunoblotted for SRSF2-myc (**Fig.3.2b**). Labeling of each fraction with the *hypo*SRSF2 antibody confirmed that the pellet fraction was enriched with *hypo*SRSF2 (**Fig.3.1d**). 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.3.2c**). 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.

3.4.2 Proteomics reveals RBPs that aggregate following dephosphorylation

Following dephosphorylation, the soluble and insoluble (i.e., pellet) samples were analyzed by label-free quantitative proteomics, using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) in biological quadruplicate (**Fig.3.2a, Online 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

(Online Supplemental Table S2). Following database search and removal of proteins with >50% missing values, we identified 4,120 unique proteins (**Online Supplemental Table S3**).

To discover proteins with the largest changes in solubility following dephosphorylation, we calculated the \log_2 fold differences of fraction insoluble values between phosphatase and mock treatments and visualized this as a volcano plot (**Fig.3.2d, Online 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. To ask whether SR proteins were enriched within this group, we performed a homology search of the RS domain of SRSF2 using the protein BLAST tool. We identified many proteins with high homology to the RS domain of SRSF2 ($n=193$), including other SR proteins (SRSF3/4/5/6/7), as well as the SR-like proteins SRRM1 and LUC7L3 (**Fig.3.3a, Online Supplemental Table S6**). Using a one-dimensional hypergeometric Fisher's exact test (FET) analysis, we concluded that the SR/SR-like group was significantly enriched to the list of proteins that experienced significantly decreased solubility following dephosphorylation (BH-corrected p-value = 0.0132) (**Fig.3.3b**). 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.

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

Phosphorylation significantly alters the net charge of a protein, adding a -2 charge with each phosphorylated residue at physiological pH²⁹². We hypothesized that proteins with highly positive net charge (high densities of arginine/lysine) may be predisposed to aggregate when not sufficiently phosphorylated. We plotted the distribution of net charge per residue (NCPR) of the nucleoplasm proteome sequenced and highlighted those proteins that were two standard deviations below (< -0.083 ; *yellow*) or above ($> +0.099$; *purple*) the mean NPCR (**Fig.3.2e**). Interestingly, 11/12 members of the SR protein family surpass the positive NCPR 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 (**Online Supplemental Table S8**).

Using the aforementioned NCPR criteria to split our proteome into groups of low, middle and high NCPR, we asked whether higher intrinsic NCPR values conferred increased susceptibility to protein aggregation following dephosphorylation. Indeed, we saw a stepwise increase in average fraction insolubility values with increasing NCPR value (**Fig.3.2f**). When compared with the group of proteins with low NCPR values ($n=69$, < -0.083), middle-charged proteins ($n=3796$, $-0.083 < x < +0.099$) had a significantly increased average \log_2 fraction insolubility value (0.89 versus 0.436, respectively). Moreover, the high NCPR group ($n=133$, $> +0.099$) had the highest fraction insolubility value (1.56), which was significantly increased when compared with both the middle-charged and low-charged groups. These data suggest proteins with a high NCPR ($> +0.099$) are highly susceptible to increased aggregation following dephosphorylation. SRSF2, as well as SRSF3/4/5/6/7, all surpassed the $+0.099$ NCPR threshold and also experienced significantly aggregation following dephosphorylation. We conclude 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.

3.4.4 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 with shared biology that may co-aggregate when not sufficiently phosphorylated. To test this, we performed Weighted Gene Correlation Network Analysis (WGCNA)⁸⁸ 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 (**Fig.3.4**). The network reduced our proteome into 27 modules [rank ordered by size, M1 (largest) – M27 (smallest)] each assigned a representative color (**Online 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.3.5a**). 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.3.5b-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 insoluble module M9 (mRNA splicing),

both unchanged in solubility profile after dephosphorylation (**Fig.3.5b**). 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.3.5c**). 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 (**Online 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.

3.4.5 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.3.6a**) and compared with immunoblot signals of select endogenous proteins from total, soluble and pelleted fractions (**Fig.3.6b**). 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. Other nuclear RBPs in HEK293 nucleoplasm lysates including TDP-43 (M5), RBM25 (M2) and ZC3H18 (M9), however, showed no significant solubility changes following dephosphorylation. It should be noted that some RBPs harbor relatively few phosphorylation PTMs in the absence of stress conditions (e.g., TDP-43), demonstrated by the lack of mobility change following dephosphorylation³⁰⁶. Nevertheless, 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.

3.4.6 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 (**Online Supplemental Table S8**). In a theoretical exercise of how phosphorylation changes the net charge of the SRSF2, we calculated the average SRSF2 NCPR (**Fig.3.7a**) and local charge density in a 21 residue sliding window range (**Fig.3.7b**) assuming three separate states: no phosphorylation (*hypo*SRSF2), MS-observed phosphorylation (pSRSF2+)²⁷⁷, and full phosphorylation (*hyper*pSRSF2) (**Online 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.3.7c-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 interactions³⁰⁷. The mock-treated

SRSF2-myc sample had a recognizable monomer species band at ~ 37 kDa, whereas dephosphorylated SRSF2 exhibited monomer as well as high molecular weight oligomeric species (**Fig.3.7d**). These data suggest that dephosphorylation leads to high molecular weight oligomeric SRSF2 structures, although in complex cellular lysate we could not determine whether the high molecular weight species are due to phosphorylation-dependent changes alone on SRSF2 or in part by protein-protein interactions as well.

To directly demonstrate how SRSF2 structure changes with phosphorylation, we performed an *in vitro* kinase reaction using purified SR protein kinase 2 (SRPK2) and SRSF2 substrate and analyzed reactions by non-denaturing native PAGE (**Fig.3.7e-f**)¹²³. As proof of principle, we confirmed SRPK2 robustly phosphorylates SRSF2 substrate *in vitro* using radiolabeled ATP (**Fig.3.8**). 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^{308,309}. 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 smear between 480-1,000 kDa (**Fig.3.7f**). In contrast, SRPK2-phosphorylated SRSF2 exhibited a decreased molecular weight range of oligomeric species, as well as a prominent monomer band at ~50 kDa. Reactions incubated with increasing concentrations of SRPIN340, an ATP-competitive selective SRPK inhibitor³¹⁰, yielded a dose-dependent recovery of high molecular weight oligomer species along with a depletion of SRSF2 monomer, suggesting that phosphorylation directly regulates the oligomerization state of SRSF2.

3.4.7 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. To validate SRPIN340 we analyzed cell lysates by western blot using mAb104¹¹⁷, an antibody that labels phosphoepitopes on numerous SR proteins³¹¹ (**Fig.3.9a**). 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.3.9b**). Independently, we compared pSR protein labeling with that of phosphoTDP-43 (pTDP-43)²⁷⁵. We found that while pTDP-43 did not quantitatively change with SRPIN340 treatment, pSRSF2 signal was significantly decreased at 50 μ M SRPIN340. These results demonstrate that inhibition of SRPKs using SRPIN340 successfully reduces pSRSF2 levels, as well as overall SR protein phosphorylation.

Next, we asked whether SR protein phosphorylation dysregulation could fundamentally alter the localization of the RBP SRSF2. Using a nuclear-cytoplasmic fractionation procedure (**Fig.3.9c**)^{198,306}, we isolated cytoplasmic and nuclear fractions from cells that underwent either vehicle- or SRPIN340-treatment (**Fig.3.9d**). In HEK293 cells, SRSF2 appeared to occupy cytoplasmic and nuclear extracts at similar levels. However, in SRPIN340-treated cells SRSF2 was primarily cytoplasmic. Quantification of nuclear fraction abundances (percent nuclear SRSF2) demonstrates that SRSF2 is significantly less nuclear in SRPIN340-treated cells (~18%) versus vehicle-treated cells (~40%) (**Fig.3.9e**). Therefore, we demonstrate that SRSF2 is enriched in cytoplasmic fractions when its phosphorylation state is decreased. The cytoplasmic mislocalization of SRSF2 following pharmacological inhibition of SRPKs was also demonstrated with an antibody raised against a non-phospho epitope of SRSF2 (**Fig. 3.10**).

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 proteinopathies³¹². Immunocytochemical (ICC) staining using the nuclear speckle antibody SC35 demonstrated canonical nuclear speckle morphology of pSRSF2 in vehicle-treated cells (**Fig.3.9f**). In cells incubated with SRPIN340, we observed unusual cytoplasmic granules (**Fig.9g**). We performed quantification and observed a significant increase in the average number of cytoplasmic granules per cell in the SRPIN340 condition (**Fig.3.9h**). 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.3.9i**). 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.3.9j-k**). Cytoplasmic mislocalization was independently validated with an antibody raised against a non-phospho epitope of SRSF2 (**Fig.3.10**). 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.3.9j-k**).

3.4.8 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.3.11a**). Interestingly, several proteins within this module are components of 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.3.11b**). Furthermore, ICC staining of SRSF2 with either TUBA1A or TUBB8 demonstrates that dephosphorylated SRSF2 preferentially associates with microtubule structures (**Fig.3.11c-d**). Thus, network-based proteomics revealed an association between arginine-rich RBPs like SRSF2 and microtubules, confirmed by both biochemical and cell imaging approaches.

3.5 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 more frequently exists as a monomer species *in vitro*. Inhibition of SR protein kinases within mammalian cells decreased SR protein phosphorylation, resulting in increased cytoplasmic SRSF2 localization, granule and tubule formation. We also observed SRSF2 to form unusual cytoplasmic tubule SRSF2 structures which co-localize with cytoskeletal proteins. Collectively, these data suggest that phosphorylation tunes SRSF2 net charge, solubility and structure by virtue of multimer disassembly *in vitro* and *in vivo*.

SRSF2 is a well-studied classical SR protein with integral roles in constitutive and alternative splicing²⁷⁸. This study highlights numerous functions of SRSF2 that are regulated by

phosphorylation and are associated with RBP dysregulation in neurodegenerative disease, including decreased solubility, high molecular weight oligomer formation, and cytoplasmic mislocalization and granule and tubule-like morphology formation. Another phenotype of dysregulated RBPs in neurodegenerative disease is the accumulation of splicing errors^{82,101,102} which could possibly be influenced by SR protein phosphorylation status. Given the substantial solubility and localization changes of SRSF2 we observed upon SRPK inhibition using SRPIN340, we hypothesize that many genes would be alternatively spliced as a result of SR protein hypophosphorylation. Use of the broad SRPK inhibitor SRPIN340 may provide a stronger effect of SR protein phosphorylation suppression than siRNA-mediated knockdown alone³⁰³. Indeed, several groups have discovered altered splicing as a result of SRPK inhibition using SRPIN340 in mammalian cells^{313,314}. Future studies investigating the regulation of the SR protein splicing network by phosphorylation may help to resolve specific genes susceptible to splicing defects upon kinase dysregulation in disease.

Although mutations in SRSF2 are frequently observed in individuals with myelodysplastic syndromes (MDS) or chronic myelomonocytic leukemia (CMML)³¹⁵⁻³¹⁹, 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^{320,321}. 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^{54,96,204,322}. As kinases are dysregulated in AD¹³⁴, 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 tangles³²³. Our group has not identified SRSF2 as a protein with significantly increased insolubility in AD³²⁴. 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.

3.6 ACKNOWLEDGMENTS

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

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.

3.7 FIGURES

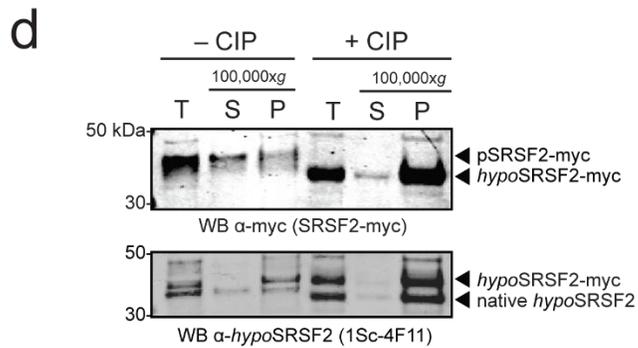
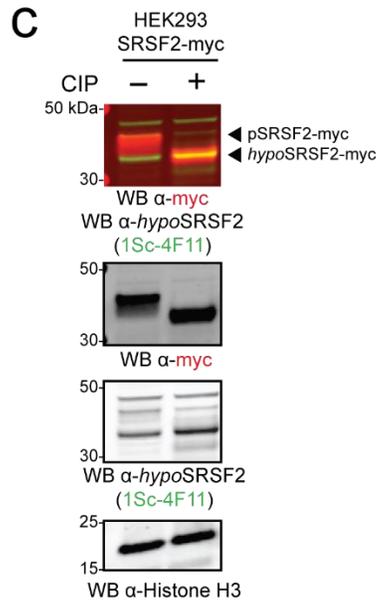
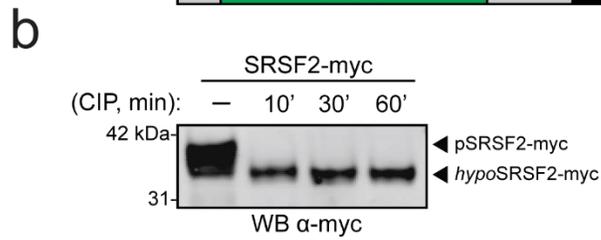
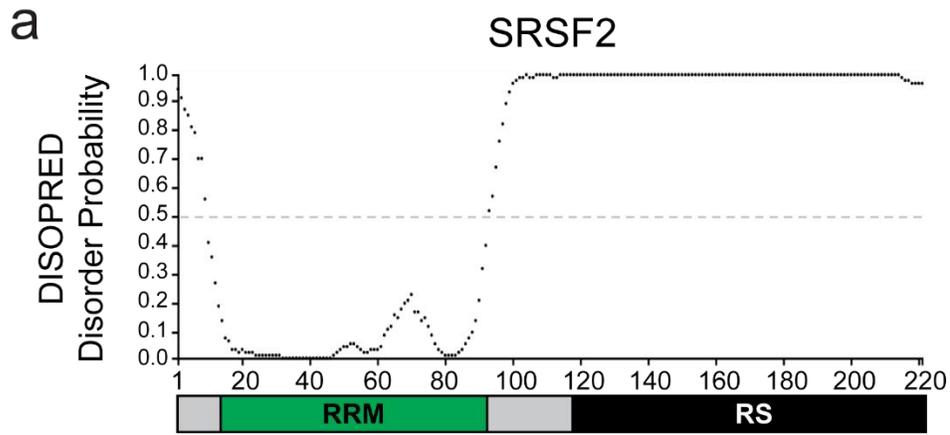


Figure 3.1. The arginine-/serine-rich (RS) domain of SRSF2 is predicted to be highly disordered. The DISOPRED 3.0 algorithm predicts intrinsically disordered regions of proteins based on primary sequence alone. The protein architecture of SRSF2 below is sized to match the x-axis of the disorder plot. 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).

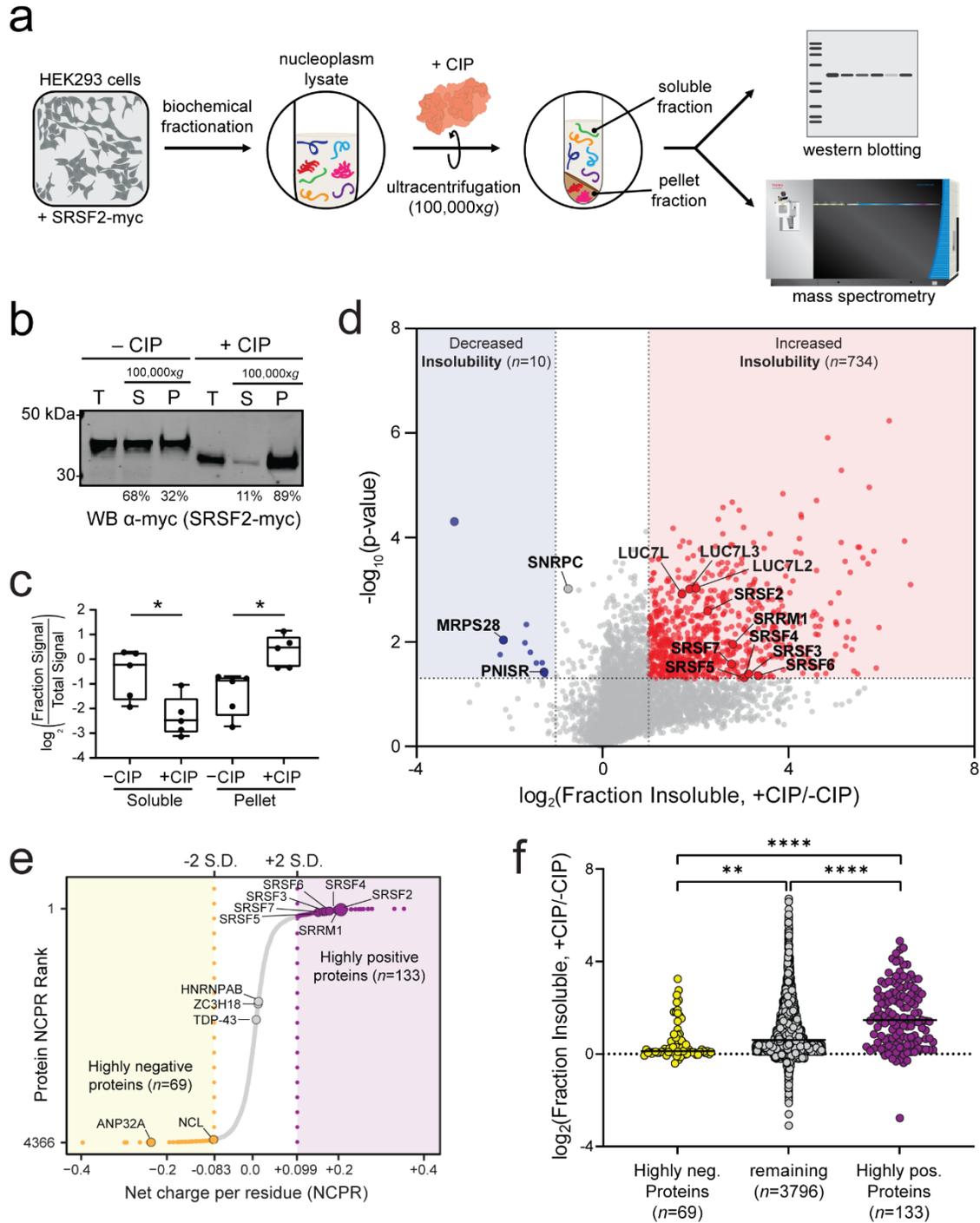


Figure 3.2. Phosphorylation regulates arginine-rich RNA-binding protein solubility. (a) Sample preparation and proteomic workflow. Nucleoplasm extracts of HEK293 cells expressing recombinant SRSF2-myc protein were incubated with either calf intestinal phosphatase (+CIP) or distilled water (-CIP) at 37°C 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*.

a

SRSF2 124 RRRRSRS--RSTRSRSTRSRSTRSRKSR-SRT 153
 LUC7L3 272 R-RRE-EEEREKERARDREERKRSTRSR|ERR 323
 RBM39 40 --SRSRHERK--RS-----KSK---ERK 56
 SRRM1 182 PRRRSSPVRREERKRSHSRSPRHRTKSRSP|SRS 282
 LUC7L2 270 R-SRSREHRRHRSRSMRERKRRTSRKSR-HRH 303

SRSF2 154 RSRSRSTSKSRSAARRSKKSSSVSRSTRSRSR 186
 LUC7L3 324 RSRSDRRRSRSHDRSERK----HRSRSDRRR 352
 RBM39 57 RSRDRERKKSKSREKRSRKERRRSRSTRDR 89
 SRRM1 283 KSRSTRSRSPSHTRPRR-----HRSRSTRSP 311
 LUC7L2 304 RSRSSRSRSHQRSRHS----SRDRSER-- 330



b

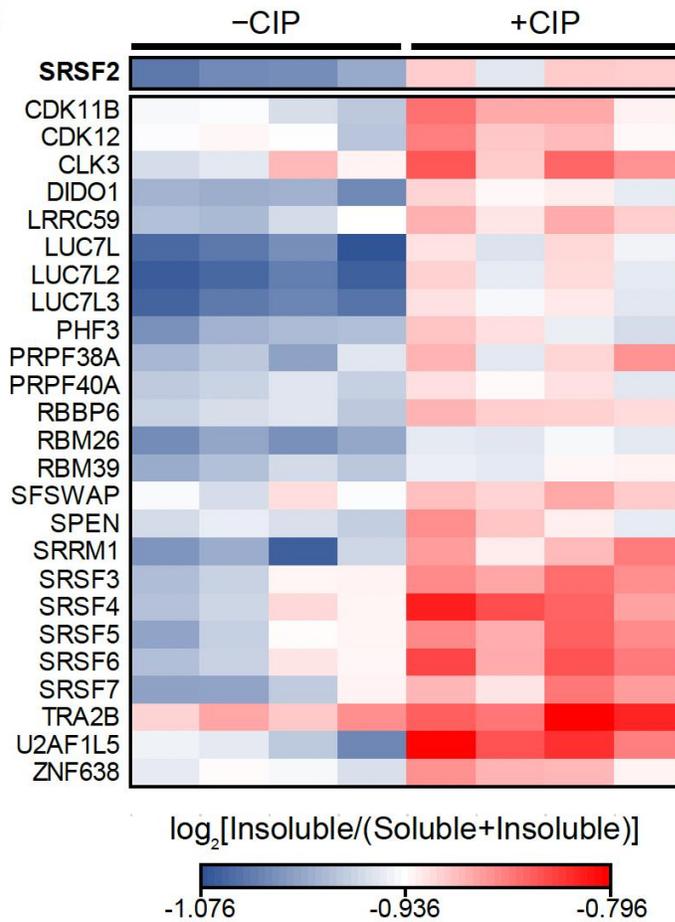


Figure 3.3. SR and SR-like proteins are enriched in detergent insoluble pellet when dephosphorylated. (a) Protein BLAST of the RS domain of SRSF2 identified 193 SR and SR-like proteins. Protein sequence alignment with residues colored according to charge. Sequences were aligned and compared according to the residue charge status (basic=purple, acid=red, acidic when phosphorylated=orange, nonpolar=black). (b) The group of SR and SR-like proteins with significantly altered fraction insoluble values following dephosphorylation.

Cluster Dendrogram

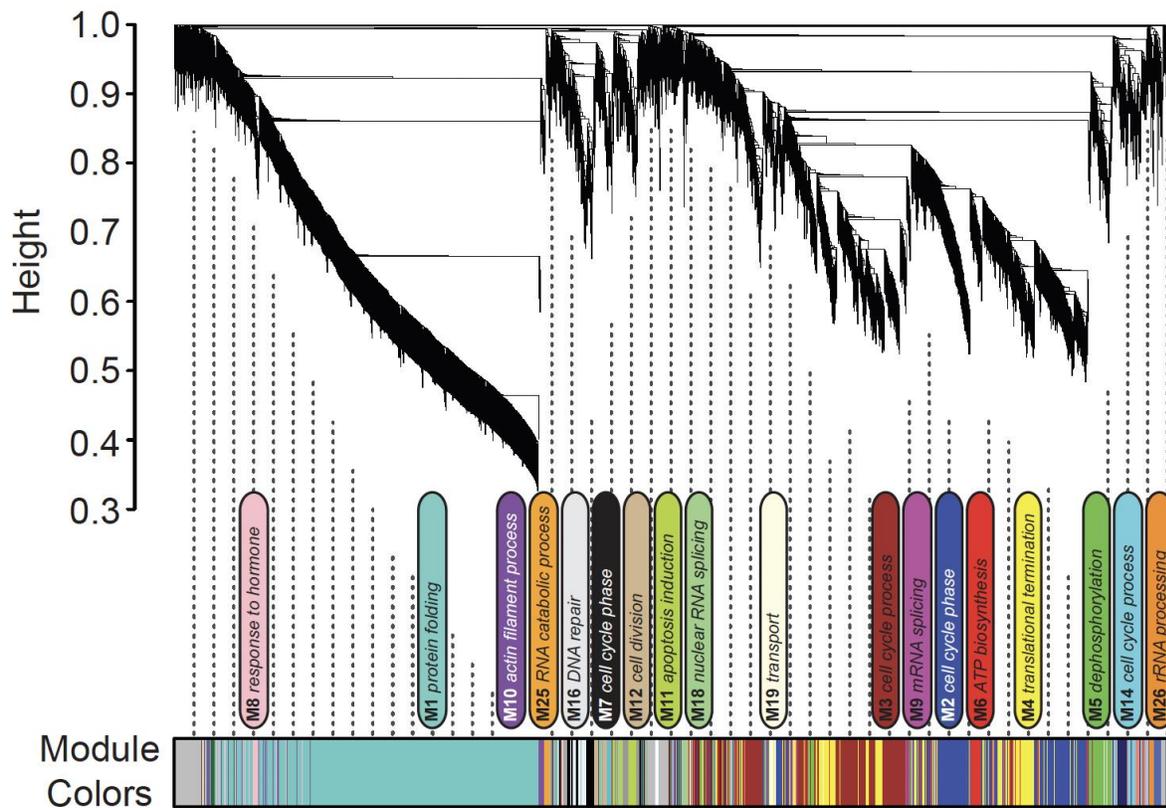
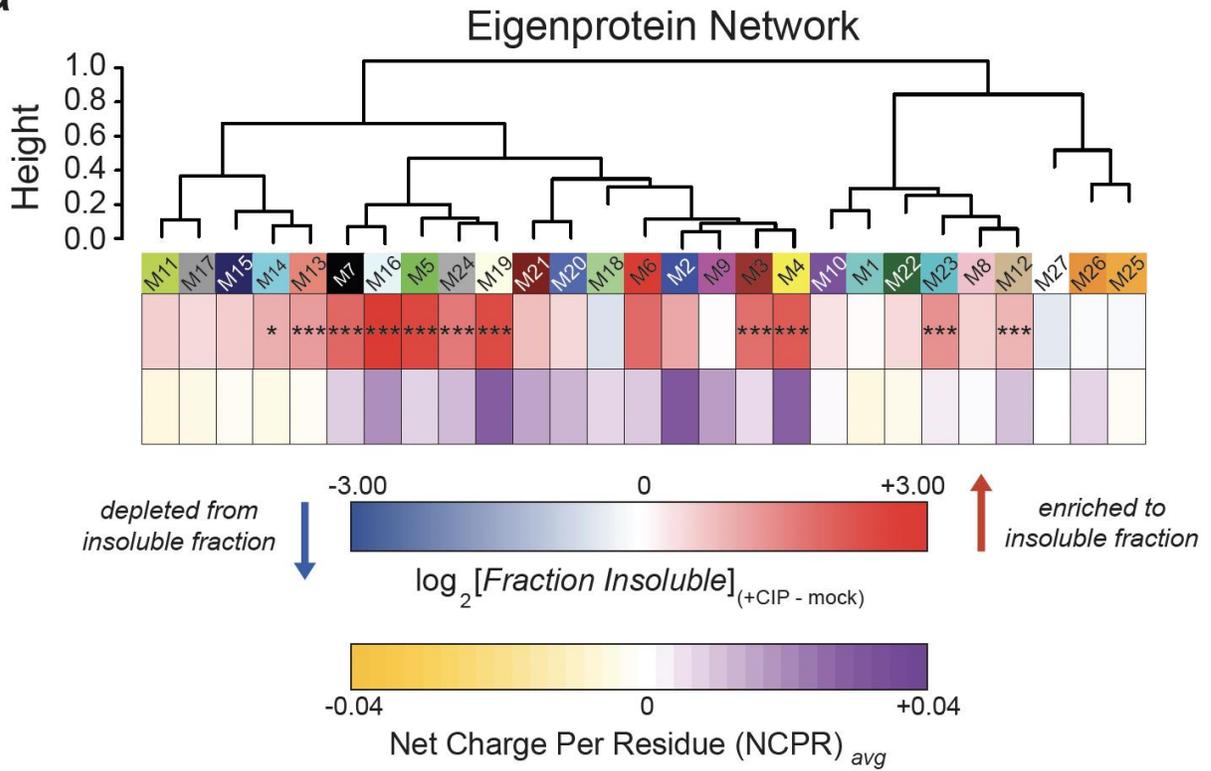
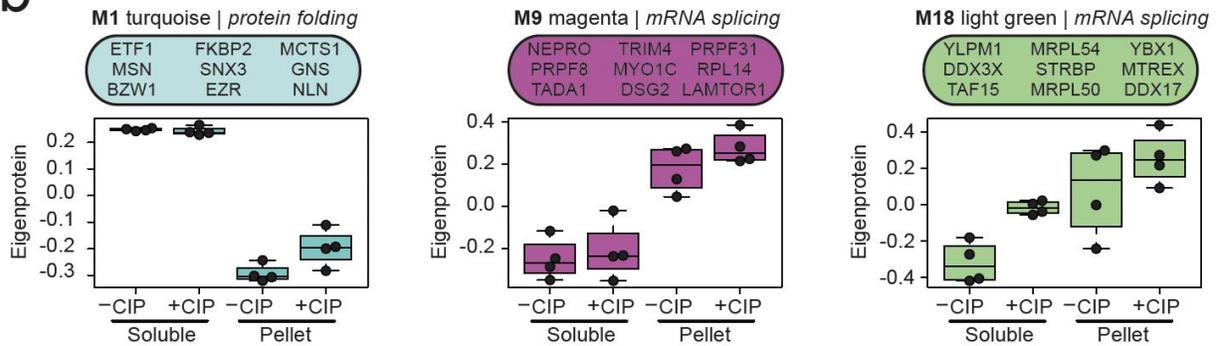


Figure 3.4. The nuclear proteome was separated into discrete groups by weighted gene correlation network analyses. Weighted Gene Correlation Network Analysis (WGCNA) cluster dendrogram groups all proteins ($n=4,366$) measured by hierarchical clustering into 27 different protein modules (M1-M27). The top generalizable biological process gene ontology term was given as a title to each module.

a



b



c

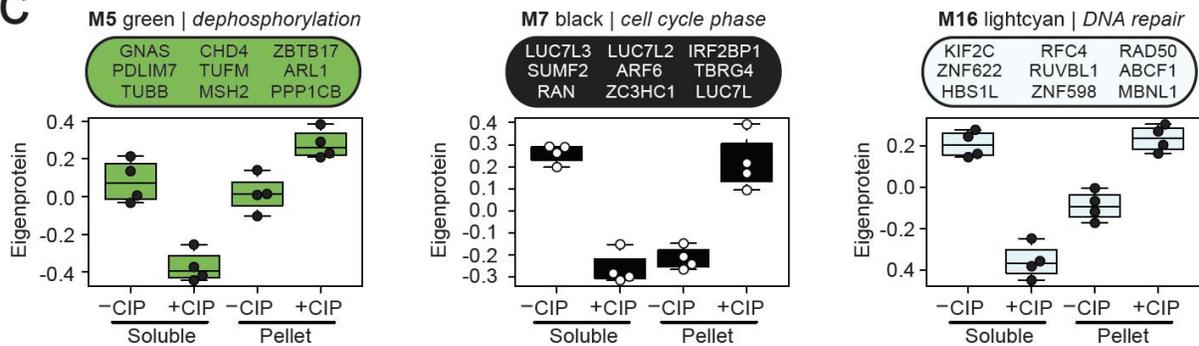
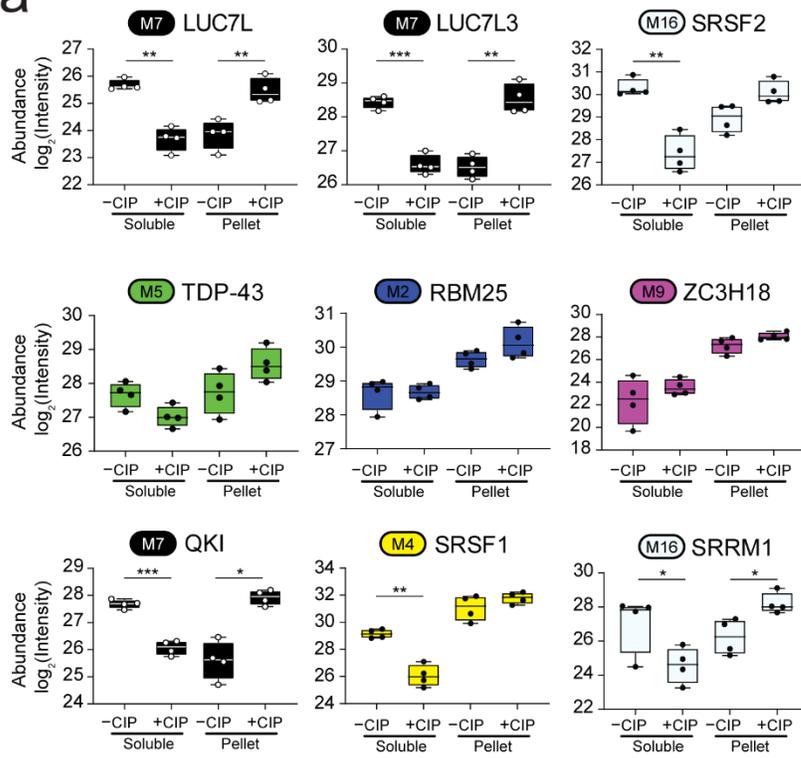


Figure 3.5. A correlation network approach groups proteins into modules with discrete gene ontologies and solubility patterns in response to dephosphorylation. (a) Protein modules were clustered to assess module relatedness based on correlation with abundances in detergent-soluble and -insoluble pellet 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.

a



b

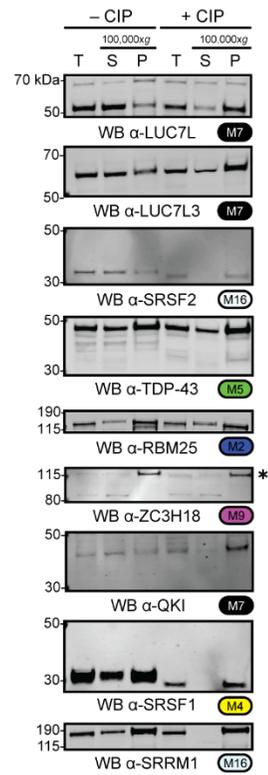


Figure 3.6. RNA-binding proteins have variable abundance patterns in soluble and pellet fractions following dephosphorylation. (a) Box and whisker plots of mass spectrometry abundance measurements ($n=4$) of RNA-binding proteins in soluble and pellet fractions in -CIP and +CIP conditions. (two-tailed 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.

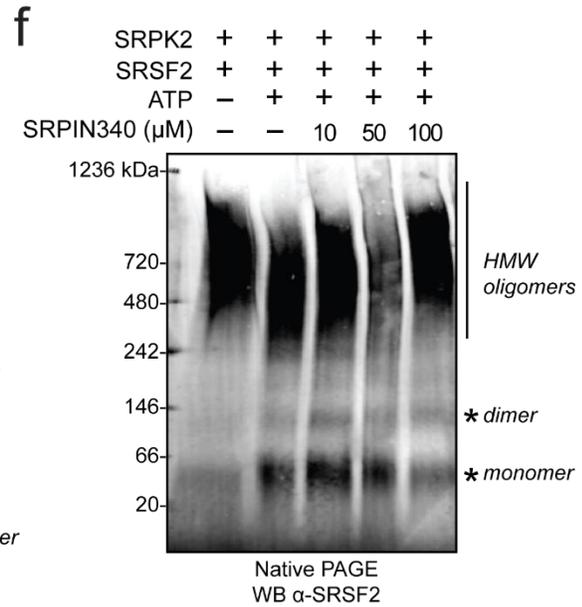
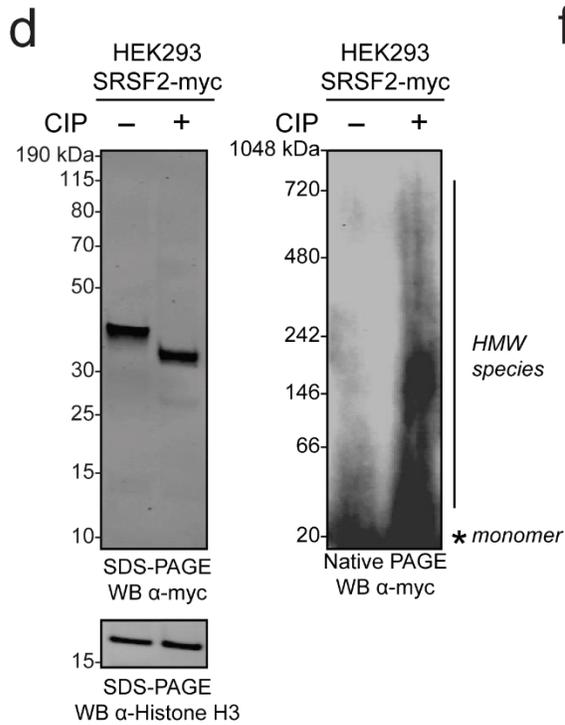
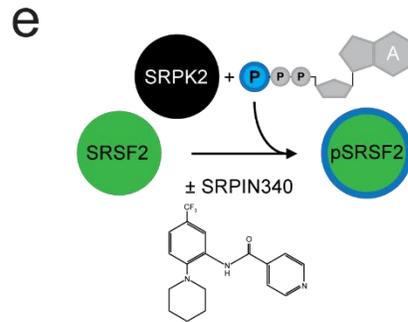
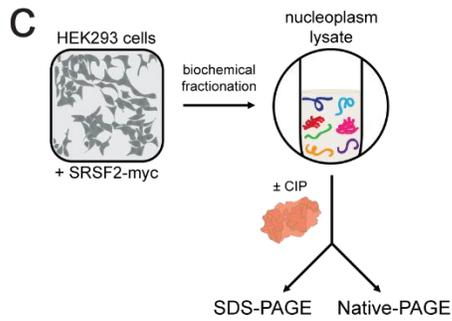
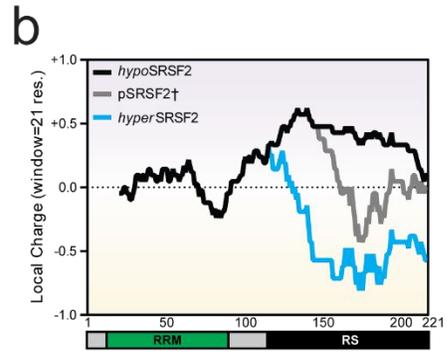
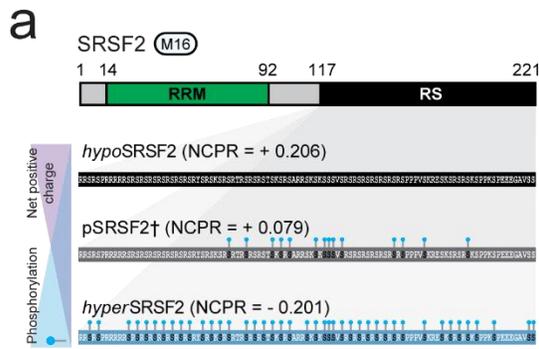


Figure 3.7. SRSF2 net charge and insolubility, respectively, increases substantially with dephosphorylation. (a) SRSF2 net charge per residue (NCPR) calculated according to phosphorylation state, i.e., no phosphorylation (*hypo*SRSF2), phosphorylation sites previously observed by middle-down mass spectrometry (pSRSF2[†], Kundinger & 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*, Kundinger & 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.

SRPK2	-	+	-	+
SRSF2	+	+	+	+
"Cold" ATP	+	+	-	-
γ -[32P] ATP	+	+	+	+

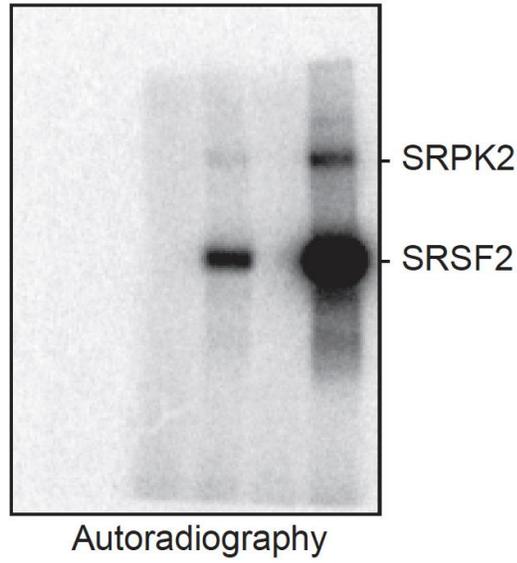
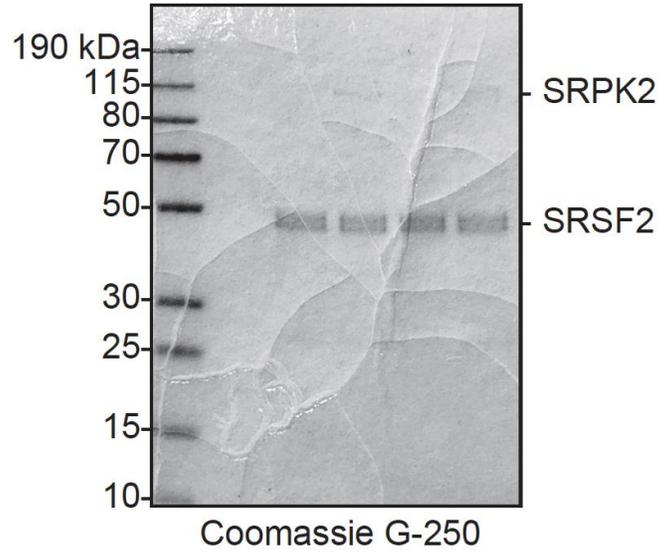


Figure 3.8. Validation of *in vitro* SRPK2-SRSF2 kinase reactions. (a-b) Serine-/arginine protein kinase 2 (SRPK2) and SRSF2, expressed and purified from *E. coli* were mixed in the presence or lack thereof (-/+, respectively) ATP and incubated at 30°C for 30 min and run by denaturing SDS-PAGE or isoelectric 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.

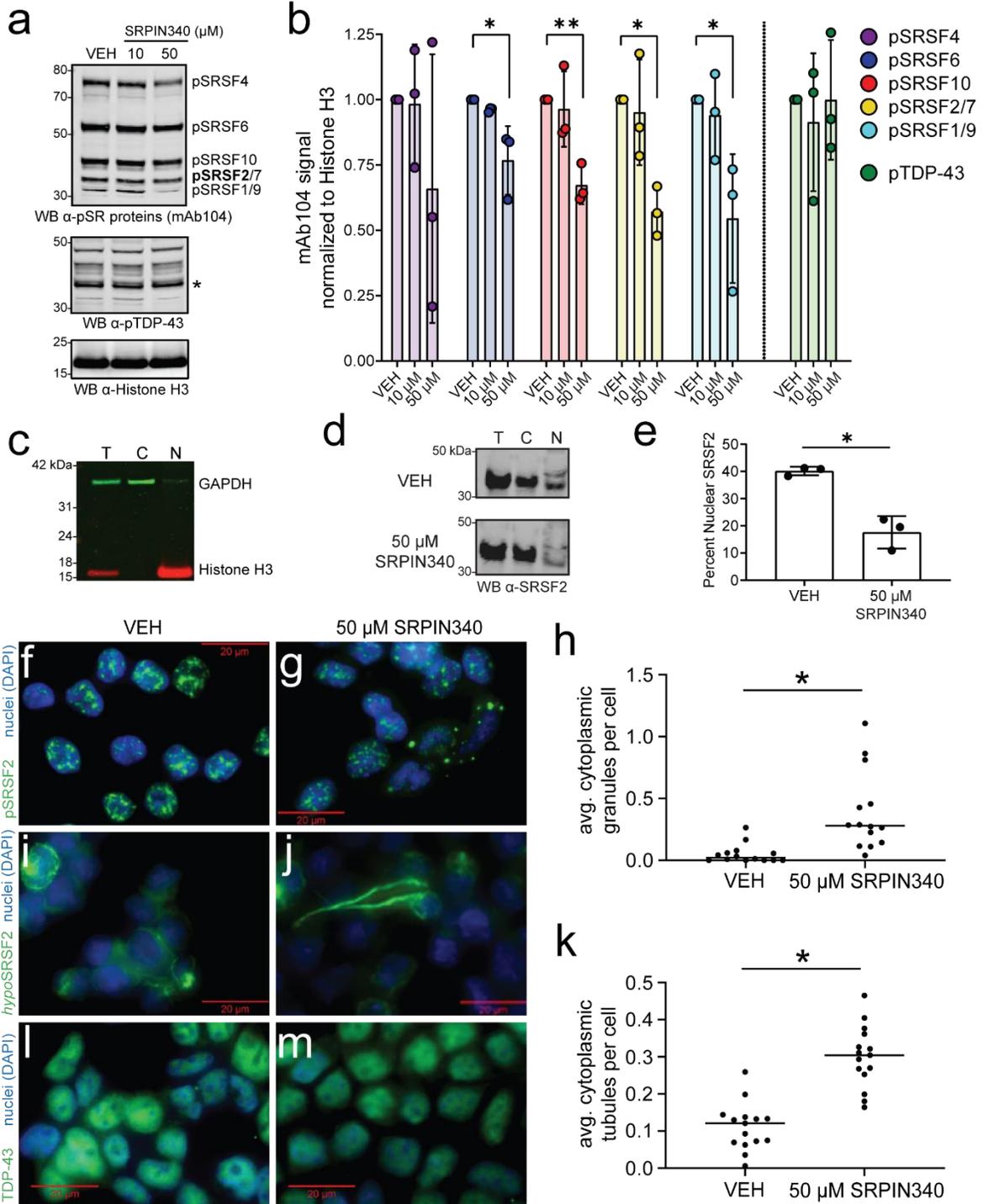


Figure 3.9. Inhibiting SRPKs decreases SR protein phosphorylation and increases cells harboring cytoplasmic SRSF2 granule and tubule structures in HEK293 cells. (a) HEK293 cells were incubated with either DMSO (vehicle, VEH) or increasing concentrations of SRPK inhibitor SRPIN340 for a length of 12 hours. 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 (hypoSRSF2) (*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-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.

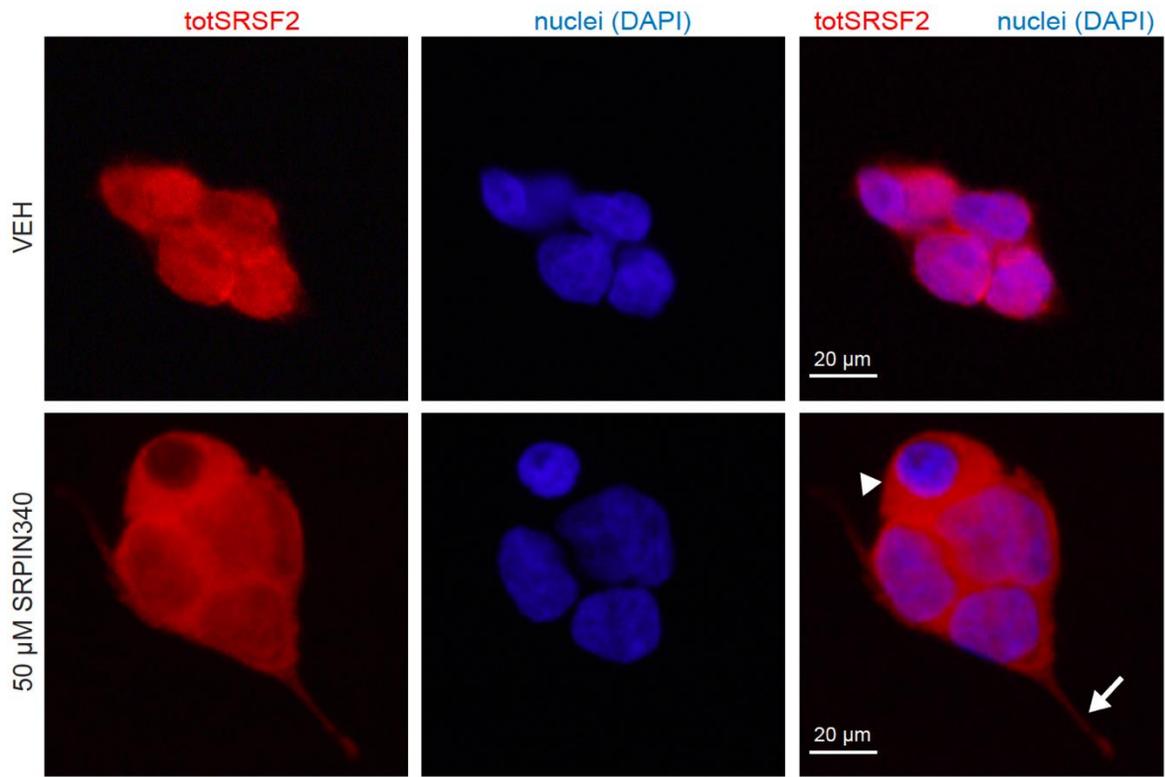


Figure 3.10. Hypophosphorylated SRSF2 exhibits nuclear mislocalization phenotypes. ICC staining of vehicle (DMSO) and 50 μ M SRPIN340-treated HEK293 cells for total SRSF2 (totSRSF2, red) and DAPI (nuclei; blue). We observed cells with SRSF2 cytoplasmic mislocalization (arrowhead) and tubule formation (arrow).

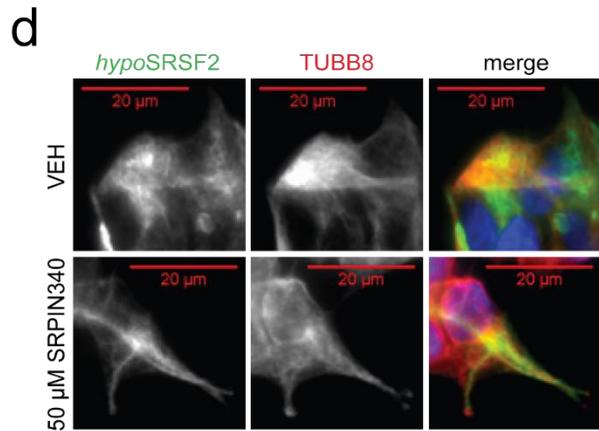
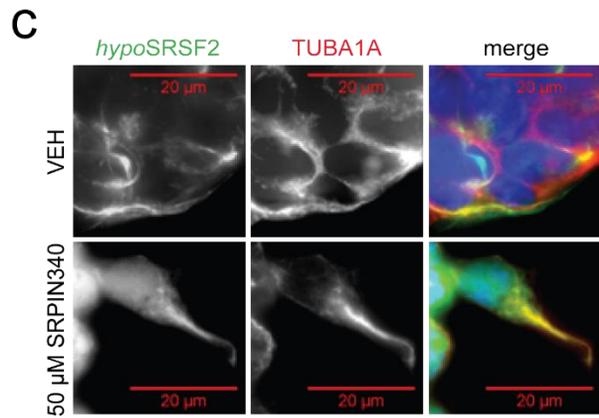
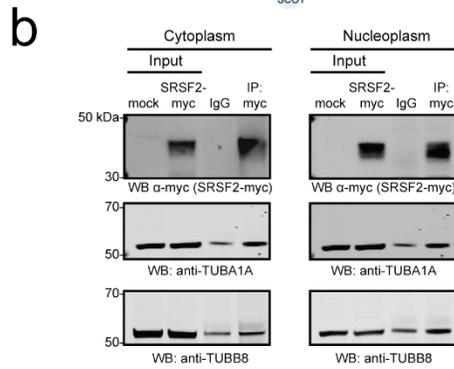
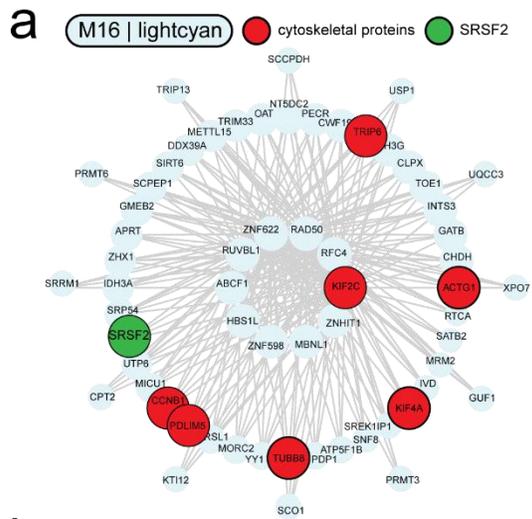


Figure 3.11. Association of SRSF2 with microtubule proteins. (a) I-graph of module 16 (M16) representing hub proteins and corresponding gene symbols as nodes. Node size and edges (gray) are reflective of the degree of intramodular connectivity in WGCNA (kME). SRSF2 (*green*) and cytoskeletal-associated 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.

CHAPTER 4.0 : DISCUSSION.

4.1 Summary

Post-translational modifications (PTMs), including phosphorylation of arginine-rich RNA-binding proteins (RBPs), regulate RNA metabolism through RBP interactions, spliceosome assembly, alternative splicing and mRNA export. Here we use a middle-down (partial protein digestion) proteomic approach coupled with electron transfer dissociation (ETD) fragmentation to significantly improve protein coverage and PTM identification in mammalian cell extracts by mass spectrometry³⁰⁶. We discover a substantial number of PTMs, particularly phosphorylation, in arginine-rich low complexity (LC) domains of RBPs. We demonstrate that removing phosphate increases interactions between structurally similar RBPs. We next dephosphorylated nuclear extracts *in vitro* and performed ultracentrifugation, analyzing equal amounts of detergent-soluble and -insoluble fractions by mass spectrometry-based proteomics. Using correlational network analyses, we identified modules and groups of RBPs that aggregate following dephosphorylation, including the SR protein family and SR-like proteins. However, increased aggregation was not observed for all nuclear RBPs let alone all nuclear proteins. We demonstrate that phosphorylation regulates the structure and oligomerization of the canonical arginine-rich RBP SRSF2. Inhibition of kinases specific to arginine-rich RBPs resulted in a decrease in SR protein phosphorylation and increased SRSF2 cytoplasmic mislocalization, granule and fibril formation. Taken together, these results demonstrate that phosphorylation occurs within arginine-rich LC domains of RBPs at levels not observed elsewhere in the proteome, and this PTM subtype regulates protein interactions to decrease the mislocalization and aggregation of SRSF2 and other arginine-rich RBPs.

4.2 Future directions of mass spectrometry to improve sequencing of arginine-rich RNA-binding proteins

Several efforts have attempted to apply alternative proteomic approaches to sequence arginine-rich protein sequences and improve PTM identifications within these sequences. One such approach includes “Confetti”³²⁵, which is a method of using multiple proteases to sequence similar protein sequences and/or proteomes. Analyzing PTMs from the same sample when digested with different proteases increases confidence that the peptide or protein identified is a true identification not an erroneous one simply due to noise. Specific types of PTMs can also be targeted by immunoaffinity enrichment approaches combined with novel proteomic methods. Maron *et al.* used a label-free ETD/HCD proteomic methodology to identify arginine methylated peptides enriched using immunoaffinity based PTMScan methods³²⁶. Despite these advances, additional approaches are needed to reliably map PTMs with these highly repetitive but functionally important regions.

Importantly, middle-down proteomic approaches which utilize incomplete trypsin digestion methods and result in longer peptides (which increase the search space needed to correctly match spectra to peptides) need vastly expanded computational power. Improved search engines, including MSFragger²⁶⁰, MetaMorpheus³²⁷ and TagGraph²⁶¹, have recently been developed and applied in order to overcome these limitations. MSFragger is exclusive among these search engines in that it has superior PTM localization algorithms to support more sensitive PTM identification hits with higher confidence in terms of localization³²⁸. Another limitation of studying site-specific PTMs is that a PTM may not be abundant enough to be sequenced in every sample, as only a fraction of any protein in a sample is modified at any given site. To combat this, a Bioconductor package may be employed called Differential Enrichment Analysis of Proteomics (DEP) data which uses a k-nearest neighbor calculation to generate imputed values for abundances missing in random samples and a minimum imputed value to abundances missing at non-random

(i.e., in experimental conditions)³²⁹. However, more innovative methods need to be continually applied to enhance our confidence that the ions we collect and spectra we sequence are true identifications.

4.3 Use of middle-down proteomic approaches to sequence post-translational modification sites in arginine-rich RNA-binding proteins in Alzheimer's disease

In Chapter 2, we describe a nuclear-cytoplasmic fractionation procedure that isolated nucleoplasmic extracts which were enriched with RBPs, including arginine-rich RBPs. To sequence arginine-rich RBPs, which enrich to insoluble fractions of AD brain extract³³⁰, a similar method to isolate nuclei may be used to enrich RBPs from AD brain extract will be important to address whether the phosphorylation status of arginine-rich RBPs is dysregulated in AD. However, brain homogenate has high cellular heterogeneity and methods used to isolate neurons, let alone neuronal nuclei, are technically difficult³³¹. Recent advances that attempt to address this problem include using fluorescence activated cell sorting (FACS) for NeuN positive nuclei of neurons extracted from human postmortem frontal cortex brain tissue³³². Using a middle-down ETD approach to sequencing FACS-sorted NeuN+ neuronal nuclei populations in Control and AD samples may reveal important PTM differences that could underlie the altered splicing observed in disease^{82,101,102}. However, given the localization and morphological changes observed for SRSF2 upon dephosphorylation described in Chapter 3, it may be that cytoplasmic pools of SRSF2 undergo the largest alterations in phosphorylation status in disease. Therefore, a comprehensive examination of nuclear and cytoplasmic pools separately may yield important insights into any alterations of SRSF2 and other arginine-rich RBPs that may occur in disease.

4.4 Recent discoveries of RNA-binding protein functions regulated by post-translational modifications

We describe the initial discovery of PTMs, with a special emphasis on the phosphorylation, on arginine-rich RBPs and SR proteins in the Introduction. While it may be that large changes in phosphorylation status trigger certain functions, residue-specific PTM changes can confer fundamental alterations in RBP behavior. For example, the nuclear SR-protein kinase CLK1 reorganizes splicing factor U1-70K for early spliceosome protein assembly by phosphorylating S226 within U1-70K, a site N-terminal to the arginine-rich low complexity LC1/BAD domain³³³. A major finding of Chapter 2 is that methylation and phosphorylation frequently co-modify RSRS motifs in mammalian cells. Support for this finding come from studies in *S. cerevisiae*, showing the co-occurrence of methylation and phosphorylation in condensate-associated proteins³³⁴. A large-scale proteomic investigation of substrates of protein arginine methyltransferases (PRMTs) revealed that arginine methylation is not only more abundant than we previously thought, but the transcriptome and proteome are significantly changed upon inhibiting specific classes of PRMTs³²⁶. The group also found that PRMTs differentially methylate similar RBPs (FUS and TAF15) and can modulate binding interactions between these proteins. Arginine methylation is increasingly being understood as a core modifier of RBPs which can regulate RNA metabolism^{274,335,336}. PRMT7-mediated arginine methylation regulates gene expression and virulence in the single-cell eukaryote *Leishmania major*³³⁷. However, we are still at an early stage of applying next-generation sequencing tools to routinely identify and measure co-occurring PTMs to study how these modifications change in different biological conditions. Combining next-generation proteomic sequencing methods with RNA-seq and/or cryo-EM technologies may allow us to define discover a ‘splicing PTM code’ which may dictate how the spliceosome is built and what pre-mRNA substrates are spliced under different conditions.

4.5 Is phosphorylation a trigger for or a consequence of protein aggregation?

A fundamental question of this thesis is whether post-translational modifications trigger, or are rather simply ‘blaze marks’, of earlier protein aggregation events. Numerous studies highlight certain PTMs that speed up aggregation kinetics while others indicate removing PTMs achieve the same effect^{140,143,144,146,147}. While it is immediately clear that this question is context dependent (i.e., which protein is modified), there remains a large amount of investigation to discover underlying rules regarding the PTM regulation of protein aggregation and the process of LLPS, the liquid de-mixing of proteins driven by multivalent interactions, in the broader context⁷⁵. In Chapter 3, we make the argument that phosphorylation prevents the insolubility, oligomerization and aggregation of the arginine-rich RBP SRSF2. Therefore, we and others⁷⁵ speculate that phosphorylation prevents the aggregation of low complexity proteins (e.g., arginine-rich RBPs) and that any phosphorylation PTMs found on aggregates in post-mortem samples may be a response of the cell to buffer liquid-to-solid transitions of certain low complexity proteins in the cell.

We hypothesize that insufficient phosphorylation may cause many protein aggregation events in neurodegenerative disease, and aggregate protein hyperphosphorylation discovered in post-mortem tissue may be the result of an earlier cellular response to reverse aggregation. An example which exemplifies this theory is the prevailing literature of AD Tau, which is hyperphosphorylated in post-mortem brain samples^{13,134}. Although it has long been assumed that phosphorylation of Tau promotes aggregation because of this correlation, several recent studies indicate that phosphorylation may instead mitigate Tau aggregation. The Kosik group identified phosphomimetic substitutions in Tau that suppress Tau aggregation in a *in vivo* Cry2 optogenetic model¹⁴². In support of this, another group recently found phosphorylation of Tau *in vitro* inhibits aggregation, seeding activity and microtubule binding/polymerization¹⁴³. Another recent,

independent example that supports this theory is the study of the hyperphosphorylated TDP-43 which was found to reduce phase separation and aggregation of TDP-43, rendering the protein more liquid-like *in vitro*¹⁴⁴. An important consideration remains that soluble disease-associated proteins may be more toxic than aggregates³³⁸, in which aggregated proteins may be better degraded and cleared. Additional investigation of other proteins may reveal more RBPs that are regulated in a similar manner.

4.6 Does phosphorylation regulate the core splicing function of arginine-rich RNA-binding proteins?

A striking fact that underpins our understanding of arginine-rich RBPs, and SR proteins in particular, is that the regulation of RBP behavior is not based on primary protein sequence alone but also to a substantial degree on the types and number of PTMs¹³⁰. Dynamic states of arginine-rich RBP phosphorylation are achieved by counterbalancing the activity of both kinases and phosphatases²⁸⁵. We hypothesize that attempts to alter the extent of hyperphosphorylation and hypophosphorylation of SR proteins in cells, achieved through kinase inhibition, results in fundamental change in RBP modification and function in key post-transcriptional processes such as splicing. In Chapter 3 we highlight gross solubility, structural and localization changes of SRSF2 that occur when phosphorylation levels are altered by SRPK inhibition. Other groups have discovered that treatment of HeLa cells with the ATP-competitive SRPK inhibitor SRPIN340 causes alternative splicing events in *VEGF*^{311,313} and *MAP2K1* transcripts³¹⁴. We further hypothesize that the removal of SRSF2 and other SR proteins from solution due to insufficient phosphorylation would be a surrogate for functional depletion. Experiments in cells with reduced SRSF2 expression reveal altered splicing programs^{339,340} and deficits in transcription initiation and elongation³⁴¹, as splicing is inextricably coupled to transcription³⁴². Future global studies of RNA

expression and splicing patterns in SRPIN340-treated cells compared with controls may elucidate scores of individual genes that rely on correct SR protein phosphorylation for proper splicing (or other steps in post-transcriptional processing), helping to sketch out a framework of the splicing code under PTM control.

4.7 The *in vivo* roles of phosphatases

A caveat of the study of the regulation of arginine-rich RBPs by phosphorylation is that the majority of studies have been conducted on the role of kinases (SRPKs, CLKs) in the absence of consideration of phosphatases. It is generally assumed that phosphatases have relatively little substrate specificity³⁴³. For example, dephosphorylation motifs are not nearly as well described as counterpart kinase motifs, partly due to the complexity conferred by the formation of holoenzyme phosphatase structures *in vivo*, which complicates the investigation into the motif specificity of individual phosphatases³⁴³. This means that there could be an exponential number of holoenzymes formed by different combinations of phosphatases. Moreover, the two major phosphatases, PP1 and PP2A, together are responsible for a large body of substrates, making the identification of specific protein substrate motifs technically challenging³⁴³⁻³⁴⁶. Interestingly, a large number of PP1 mammalian substrates are intrinsically disordered³⁴⁵, a property widely shared by arginine-rich RBPs. A recent study used MS-based proteomic approaches to identify phosphatase motifs after incubating PTM peptide libraries with either PP1 or PP2A³⁴³. This study identified a phosphatase motif, RXXpS, that was differentially phosphorylated when incubated with the phosphatases PP1 versus PP2A. Importantly, the researchers identified numerous RxxpS sites in SR proteins that were preferentially dephosphorylated by PP1 but not PP2A, suggesting that not only that SR proteins are dephosphorylated by PP1 but also that we can broadly apply these methods to detect phosphatase substrate motif preferences.

In support of the finding of PP1 phosphatase specificity to RS motifs in SR proteins, recent work has elucidated PP1 phosphatase binding motifs in the N-terminal domains of SR proteins that regulate the phosphorylation status of the RS domain, nuclear speckle formation and splicing^{242,285}. Moreover, other groups have further strengthened the link between dephosphorylated arginine-rich RBPs and alterations in cellular binding patterns, localization and morphology, as dephosphorylated SRSF2 binds more strongly to mature poly(A)+ mRNAs than phosphorylated SRSF2 does¹²⁵. Instead of a classical nuclear export signal (NES), SRSF2 export (and thereby subcellular localization) is regulated by dephosphorylation of nuclear SRSF2, which increases SRSF2 interaction with the nuclear export factor NXF1¹²⁵. Therefore, increased attention and innovative approaches must be applied to comprehensively describe the role of phosphatases and specific sites of dephosphorylation in regulating arginine-rich RBP behavior.

4.8 Implications of phosphorylation dysregulation in AD

Mass spectrometry-based proteomic investigations of the sarkosyl insoluble fractions of brain extract identified numerous arginine-rich splicing RBPs that aggregate in both early onset familial and late onset sporadic AD^{54,95,96}. Furthermore, our group demonstrated arginine-rich RBP protein insolubility in AsymAD brains that harbor little to no Tau aggregation, suggesting that the aggregation of arginine-rich RBPs may be a critical event that can possibly seed later Tau aggregation⁹⁶. Immunohistochemical staining revealed arginine-rich RBPs including U1-70K form fibril aggregates in AD brain that co-localize with Tau^{54,96,98}. In summary, arginine-rich RBPs aggregate and associate with Tau in AD due to some unknown mechanism.

Proteinopathies are believed to be promoted by altered PTM status of aggregating proteins⁷⁵. Here, we show that reduced phosphorylation states promotes the insolubility, oligomerization and cytoplasmic mislocalization and granule and tubule formation of arginine-

rich RBPs. Each of the aforementioned events are pathological RBP phenotypes associated with neurodegenerative diseases including AD. Importantly, in AD several kinases have altered steady-state expression levels¹³⁴. This suggests that a kinase specific to arginine-rich RBPs could be altered in terms of expression or activity that may help to explain the several pathological phenotypes we observe with arginine-rich RBPs in AD. Indeed, examination of large brain cohorts by RNA-seq³⁴⁷ and mass spectrometry²⁷¹ at Mt. Sinai and Emory, respectively, demonstrate that the brain-specific kinase SRPK2²⁸⁰, which preferentially phosphorylates arginine-rich RBPs, has decreased steady-state levels of both RNA and protein (**Fig.4.1a-b**). It is unknown whether SRPK2 downregulation occurs as a result of A β deposition. Decreased SRPK2 transcript and protein in AD may result in chronic, insufficient phosphorylation of arginine-rich RBPs and could help to explain the cytoplasmic mislocalization, fibrillization and splicing alterations observed in AD (**Fig.4.1c**)^{54,100,101}. Continued examination of arginine-rich RBP aggregation in AD through the framework of altered phosphorylation may reveal new therapeutic targets and options to treat AD in the future that may importantly influence the Amyloid cascade.

4.9 Future Directions

Phosphorylation of low complexity domains is often considered non-specific events due to off-target promiscuous kinases³⁴⁸. Here, we use innovative MS-based proteomic approaches to provide evidence that phosphorylation is enriched within arginine-rich low complexity RBPs and identify modified residues specifically. We then use a series of biochemical approaches to demonstrate that phosphorylation regulates the binding interactions, solubility, oligomerization, localization and morphology of an arginine-rich RBP, SRSF2. Based on our global proteomic study of solubility changes correlated with changes in phosphorylation, we hypothesize that arginine-rich RBPs as a whole function in a manner similar to SRSF2. These findings provide a

framework for understanding the role of phosphorylation in preventing numerous potentially detrimental functions of arginine-rich RBPs *in vivo*.

The evidence that I present herein suggest a much more nuanced understanding of PTMs and low complexity RBPs is warranted. Applying the approaches described in Chapter 3 to neurodegenerative disease-associated RBPs may help to reveal mechanisms underlying protein aggregation and may even serve to guide identification of RBPs that are susceptible to protein aggregation in diseases or disorders with kinase and/or phosphatase dysregulation³⁴⁹, including AD. Although our group has not discovered SRSF2 as a protein that was enriched in detergent insoluble pellets of AD brain extract, we have identified numerous arginine-rich RBPs including U1-70K, LUC7L3 and SRSF11⁹⁵. Future studies that include soluble fractions alongside insoluble fractions analyzed in the manner described in Chapter 3, may help to identify novel proteins with altered stabilities in AD. Interestingly, the Wisniewski group recently used laser microdissection to isolate Tau neurofibrillary tangles (NFTs) from AD brain sections and discovered that SRSF2 and other SR proteins preferentially associate with phosphorylated AD Tau³²³. This exciting result, in combination with the results described herein, argue that more in-depth examinations of SRSF2 such as immunohistochemistry in AD brain is warranted. Combining these techniques with innovative MS proteomic methods may identify specific residues and avenues for therapies to prevent phosphorylation dysregulation in AD.

We can also use the proteome analyzed in Chapter 3 to develop hypotheses as to why certain arginine-rich RBPs do not undergo major solubility changes upon dephosphorylation. For example, we did not observe a change in U1-70K solubility upon incubation with phosphatase. Possibly, the LC1/BAD domain of U1-70K may be an example of a ‘phosphomimetic’ version of an RS domain, with aspartate and glutamate residues substituting for serine and effectively

mimicking a phosphorylated RS domain. U1-70K may also be a particularly phosphatase resistant, although this has not been established. The lack of solubility changes upon dephosphorylation suggests that the aggregation of U1-70K, notably evident in Alzheimer's disease^{54,95,96}, is not initiated by changes in phosphorylation status. Instead, other PTMs including methylation could influence this process.

Finally, understanding how phosphorylation alters the splicing of specific transcripts is critical. This question is particularly important in our understanding of AD, as it was recently discovered that altered splicing is a *bona fide* hallmark of AD pathology^{82,101,102}. We are currently examining a next-generation RNA-seq dataset in mammalian HEK293 samples treated with the SRPK inhibitor SRPIN340 versus vehicle control to explore whether global changes in splicing occur as a result of inhibition of SR protein phosphorylation (**Fig.3.9a-b**). We also plan to compare any splicing changes with those observed under conditions of SRSF2 knockdown³³⁹ and those found in AD¹⁰¹ to determine whether phosphorylation-regulated splicing events are conserved. Unambiguously linking specific sites of phosphorylation to complex biological functions and pathologies of RBPs now appears as an attainable goal in the modern age of proteomics.

This work highlights phosphorylation as a PTM that prevents the oligomerization, insolubility and mislocalization of arginine-rich RBPs. Whether the phosphorylation status of arginine-rich RBPs in AD is altered is now a framework which we can begin to examine using next generation proteomic sequencing methods. If altered phosphorylation of arginine-rich LC domains of RBPs is confirmed, interventions that promote proper phosphorylation of these proteins can be tested in cell and animal models of AD. The role of altered phosphorylation in regulating arginine-rich RBP aggregation may reveal new therapeutic avenues to treat AD in the future.

4.10 FIGURES

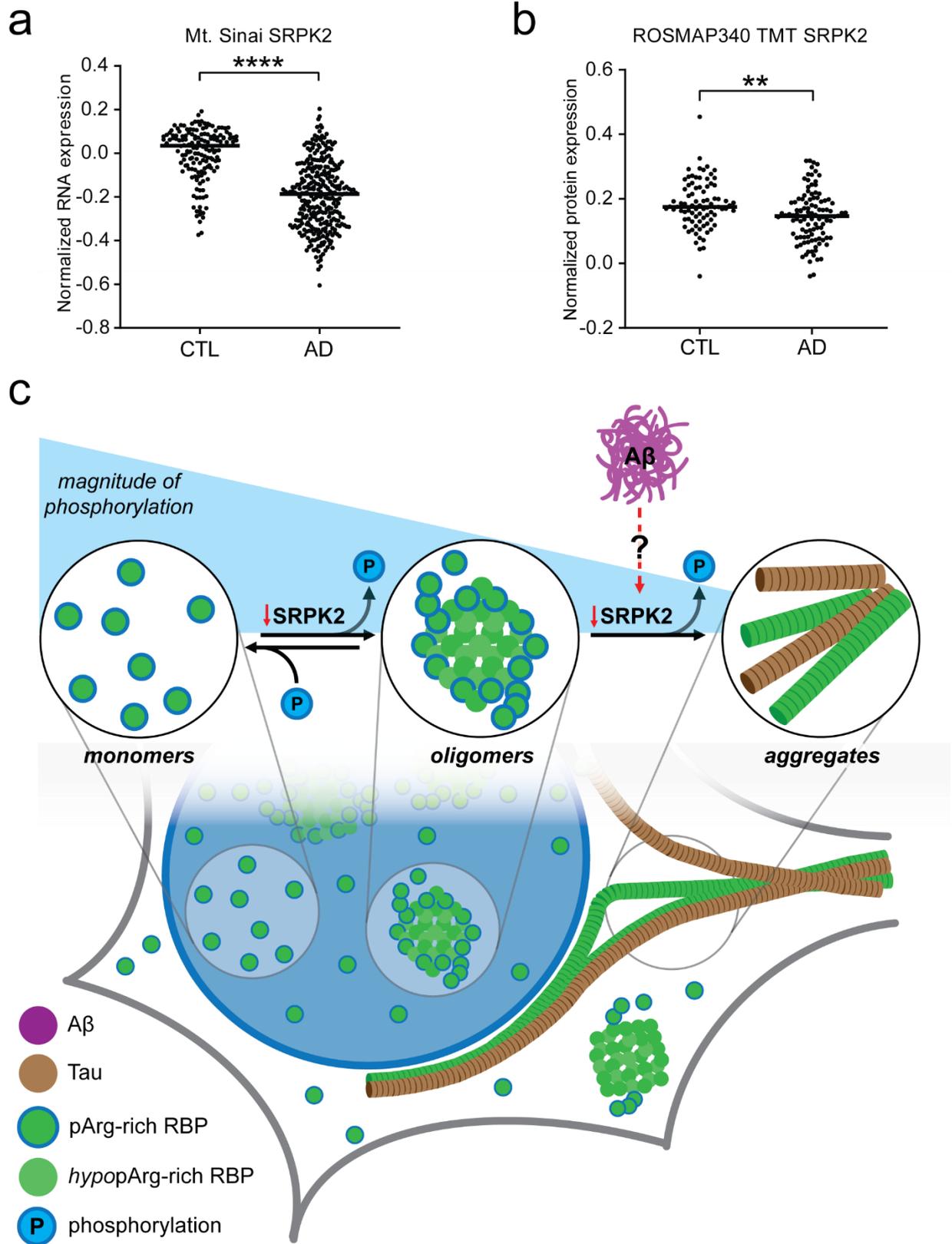


Figure 4.1. Proposed model of arginine-rich RNA-binding protein solubility and oligomerization changes regulated by phosphorylation. (a) Levels of SRPK2 RNA transcript as measured by RNA-sequencing in control (CTL) and Alzheimer's disease (AD) brain (two-tailed t-test, *p < 0.05, **p < 0.01, ***p < 0.001). (b) Levels of SRPK2 protein as measured by mass spectrometry in CTL and AD brain (two-tailed t-test, *p < 0.05, **p < 0.01, ***p < 0.001). (c) Equilibrium of structural states (monomer, oligomer, aggregate) of arginine-rich (Arg-rich) RNA-binding proteins (*green*) in varying states of phosphorylation (*dark turquoise border*). The decreasing magnitude of phosphorylation (*turquoise triangle*) is illustrated to correlate to decreasing SRPK2 levels. Phosphorylation regulates higher order SRSF2 solubility, oligomerization and structure formation.

CHAPTER 5.0: REFERENCES

1. 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2021;17(3):327-406. doi: <https://doi.org/10.1002/alz.12328>.
2. Knopman DS, Amieva H, Petersen RC, Chételat G, Holtzman DM, Hyman BT, Nixon RA, Jones DT. Alzheimer disease. *Nature Reviews Disease Primers*. 2021;7(1):33. doi: 10.1038/s41572-021-00269-y.
3. Myshko D. Some Blues Plans Won't Cover New Alzheimer's Therapy Aduhelm2021. Available from: <https://www.formularywatch.com/view/fda-expands-eua-for-lilly-s-covid-19-monoclonal-antibody>.
4. Levy R. Alzheimer—The Life of a Physician and the Career of a Disease. Konrad Maurer and Ulrike Maurer. Translated by Neil Levi with Alistair Burns, 2003. Columbia University Press, Columbia, OH. Pages: 270. \$49.50. *International Journal of Geriatric Psychiatry*. 2004;19(5):498-9. doi: <https://doi.org/10.1002/gps.1036>.
5. Whitehouse PJ, Maurer K, Ballenger JF. *Concepts of Alzheimer Disease. Biological, Clinical, and Cultural Perspectives*: Johns Hopkins University Press; 1999.
6. Graeber MB, Kösel S, Grasbon-Frodl E, Möller HJ, Mehraein P. Histopathology and APOE genotype of the first Alzheimer disease patient, Auguste D. *Neurogenetics*. 1998;1(3):223-8. Epub 2000/03/29. doi: 10.1007/s100480050033. PubMed PMID: 10737127.
7. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br J Psychiatry*. 1968;114(512):797-811. Epub 1968/07/01. doi: 10.1192/bjp.114.512.797. PubMed PMID: 5662937.
8. Terry RD, Gonatas NK, Weiss M. The ultrastructure of the cerebral cortex in Alzheimer's disease. *Trans Am Neurol Assoc*. 1964;89:12. Epub 1964/01/01. PubMed PMID: 5828485.
9. Glenner GG, Wong CW. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem Biophys Res Commun*. 1984;120(3):885-90. Epub 1984/05/16. doi: 10.1016/s0006-291x(84)80190-4. PubMed PMID: 6375662.
10. Masters CL, Simms G, Weinman NA, Multhaup G, McDonald BL, Beyreuther K. Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc Natl Acad Sci U S A*. 1985;82(12):4245-9. Epub 1985/06/01. doi: 10.1073/pnas.82.12.4245. PubMed PMID: 3159021; PMCID: PMC397973.
11. Weingarten MD, Lockwood AH, Hwo SY, Kirschner MW. A protein factor essential for microtubule assembly. *Proc Natl Acad Sci U S A*. 1975;72(5):1858-62. Epub 1975/05/01. doi: 10.1073/pnas.72.5.1858. PubMed PMID: 1057175; PMCID: PMC432646.
12. Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, Vogel FS, Hughes JP, van Belle G, Berg L. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*. 1991;41(4):479-86. Epub 1991/04/01. doi: 10.1212/wnl.41.4.479. PubMed PMID: 2011243.
13. Goedert M, Jakes R, Vanmechelen E. Monoclonal antibody AT8 recognises tau protein phosphorylated at both serine 202 and threonine 205. *Neurosci Lett*. 1995;189(3):167-9. Epub 1995/04/21. doi: 10.1016/0304-3940(95)11484-e. PubMed PMID: 7624036.

14. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta neuropathologica*. 1991;82(4):239-59. Epub 1991/01/01. doi: 10.1007/bf00308809. PubMed PMID: 1759558.
15. Haass C, Kaether C, Thinakaran G, Sisodia S. Trafficking and proteolytic processing of APP. *Cold Spring Harbor perspectives in medicine*. 2012;2(5):a006270.
16. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science (New York, NY)*. 1992;256(5054):184-5. Epub 1992/04/10. doi: 10.1126/science.1566067. PubMed PMID: 1566067.
17. Bishop NA, Lu T, Yankner BA. Neural mechanisms of ageing and cognitive decline. *Nature*. 2010;464(7288):529-35. doi: 10.1038/nature08983.
18. Bekris LM, Yu CE, Bird TD, Tsuang DW. Genetics of Alzheimer disease. *J Geriatr Psychiatry Neurol*. 2010;23(4):213-27. Epub 2010/11/04. doi: 10.1177/0891988710383571. PubMed PMID: 21045163; PMCID: PMC3044597.
19. Tanzi RE. The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med*. 2012;2(10). Epub 2012/10/03. doi: 10.1101/cshperspect.a006296. PubMed PMID: 23028126; PMCID: PMC3475404.
20. Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, Grzeschik KH, Multhaup G, Beyreuther K, Müller-Hill B. The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature*. 1987;325(6106):733-6. Epub 1987/02/19. doi: 10.1038/325733a0. PubMed PMID: 2881207.
21. Goate A, Chartier-Harlin M-C, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*. 1991;349(6311):704-6.
22. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, Tsuda T, Mar L, Foncin JF, Bruni AC, Montesi MP, Sorbi S, Rainero I, Pinessi L, Nee L, Chumakov I, Pollen D, Brookes A, Sanseau P, Polinsky RJ, Wasco W, Da Silva HA, Haines JL, Pericak-Vance MA, Tanzi RE, Roses AD, Fraser PE, Rommens JM, St George-Hyslop PH. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*. 1995;375(6534):754-60. Epub 1995/06/29. doi: 10.1038/375754a0. PubMed PMID: 7596406.
23. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu C-e, Jondro PD, Schmidt SD, Wang K, Crowley AC, Fu Y-H, Guenette SY, Galas D, Nemens E, Wijsman EM, Bird TD, Schellenberg GD, Tanzi RE. Candidate Gene for the Chromosome 1 Familial Alzheimer's Disease Locus. *Science (New York, NY)*. 1995;269(5226):973-7. doi: 10.1126/science.7638622.
24. Oakley H, Cole SL, Logan S, Maus E, Shao P, Craft J, Guillozet-Bongaarts A, Ohno M, Disterhoft J, Van Eldik L, Berry R, Vassar R. Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J Neurosci*. 2006;26(40):10129-40. Epub 2006/10/06. doi: 10.1523/jneurosci.1202-06.2006. PubMed PMID: 17021169; PMCID: PMC6674618.
25. Oddo S, Caccamo A, Kitazawa M, Tseng BP, LaFerla FM. Amyloid deposition precedes tangle formation in a triple transgenic model of Alzheimer's disease. *Neurobiol Aging*. 2003;24(8):1063-70. Epub 2003/12/04. doi: 10.1016/j.neurobiolaging.2003.08.012. PubMed PMID: 14643377.

26. Bertram L, Tanzi RE. Alzheimer's disease: one disorder, too many genes? *Hum Mol Genet.* 2004;13 Spec No 1:R135-41. Epub 2004/02/07. doi: 10.1093/hmg/ddh077. PubMed PMID: 14764623.
27. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM. Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease: A Meta-analysis. *JAMA.* 1997;278(16):1349-56. doi: 10.1001/jama.1997.03550160069041.
28. Ignatius MJ, Gebicke-Härter PJ, Skene JH, Schilling JW, Weisgraber KH, Mahley RW, Shooter EM. Expression of apolipoprotein E during nerve degeneration and regeneration. *Proc Natl Acad Sci U S A.* 1986;83(4):1125-9. Epub 1986/02/01. doi: 10.1073/pnas.83.4.1125. PubMed PMID: 2419900; PMCID: PMC323024.
29. Masliah E, Mallory M, Ge N, Alford M, Veinbergs I, Roses AD. Neurodegeneration in the central nervous system of apoE-deficient mice. *Exp Neurol.* 1995;136(2):107-22. Epub 1995/12/01. doi: 10.1006/exnr.1995.1088. PubMed PMID: 7498401.
30. Belloy ME, Napolioni V, Greicius MD. A Quarter Century of APOE and Alzheimer's Disease: Progress to Date and the Path Forward. *Neuron.* 2019;101(5):820-38. Epub 2019/03/08. doi: 10.1016/j.neuron.2019.01.056. PubMed PMID: 30844401; PMCID: PMC6407643.
31. Zannis VI, Breslow JL, Utermann G, Mahley RW, Weisgraber KH, Havel RJ, Goldstein JL, Brown MS, Schonfeld G, Hazzard WR, Blum C. Proposed nomenclature of apoE isoproteins, apoE genotypes, and phenotypes. *J Lipid Res.* 1982;23(6):911-4. Epub 1982/08/01. PubMed PMID: 7130859.
32. Verghese PB, Castellano JM, Garai K, Wang Y, Jiang H, Shah A, Bu G, Frieden C, Holtzman DM. ApoE influences amyloid- β (A β) clearance despite minimal apoE/A β association in physiological conditions. *Proceedings of the National Academy of Sciences.* 2013;110(19):E1807-E116. doi: 10.1073/pnas.1220484110.
33. Verghese PB, Castellano JM, Holtzman DM. Apolipoprotein E in Alzheimer's disease and other neurological disorders. *Lancet Neurol.* 2011;10(3):241-52. Epub 2011/02/26. doi: 10.1016/s1474-4422(10)70325-2. PubMed PMID: 21349439; PMCID: PMC3132088.
34. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science (New York, NY).* 1993;261(5123):921-3. Epub 1993/08/13. doi: 10.1126/science.8346443. PubMed PMID: 8346443.
35. Arboleda-Velasquez JF, Lopera F, O'Hare M, Delgado-Tirado S, Marino C, Chmielewska N, Saez-Torres KL, Amarnani D, Schultz AP, Sperling RA, Leyton-Cifuentes D, Chen K, Baena A, Aguillon D, Rios-Romenets S, Giraldo M, Guzmán-Vélez E, Norton DJ, Pardilla-Delgado E, Artola A, Sanchez JS, Acosta-Urbe J, Lalli M, Kosik KS, Huentelman MJ, Zetterberg H, Blennow K, Reiman RA, Luo J, Chen Y, Thiyyagura P, Su Y, Jun GR, Naymik M, Gai X, Bootwalla M, Ji J, Shen L, Miller JB, Kim LA, Tariot PN, Johnson KA, Reiman EM, Quiroz YT. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. *Nature Medicine.* 2019;25(11):1680-3. doi: 10.1038/s41591-019-0611-3.
36. Thambisetty M, An Y, Tanaka T. Alzheimer's disease risk genes and the age-at-onset phenotype. *Neurobiology of aging.* 2013;34(11):2696. e1-. e5.
37. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, Bellenguez C, Frizatti A, Chouraki V, Martin ER,

Slegers K, Badarinarayan N, Jakobsdottir J, Hamilton-Nelson KL, Moreno-Grau S, Olaso R, Raybould R, Chen Y, Kuzma AB, Hiltunen M, Morgan T, Ahmad S, Vardarajan BN, Epelbaum J, Hoffmann P, Boada M, Beecham GW, Garnier J-G, Harold D, Fitzpatrick AL, Valladares O, Moutet M-L, Gerrish A, Smith AV, Qu L, Bacq D, Denning N, Jian X, Zhao Y, Del Zompo M, Fox NC, Choi S-H, Mateo I, Hughes JT, Adams HH, Malamon J, Sanchez-Garcia F, Patel Y, Brody JA, Dombroski BA, Naranjo MCD, Daniilidou M, Eiriksdottir G, Mukherjee S, Wallon D, Uphill J, Aspelund T, Cantwell LB, Garzia F, Galimberti D, Hofer E, Butkiewicz M, Fin B, Scarpini E, Sarnowski C, Bush WS, Meslage S, Kornhuber J, White CC, Song Y, Barber RC, Engelborghs S, Sordon S, Voijnovic D, Adams PM, Vandenberghe R, Mayhaus M, Cupples LA, Albert MS, De Deyn PP, Gu W, Himali JJ, Beekly D, Squassina A, Hartmann AM, Orellana A, Blacker D, Rodriguez-Rodriguez E, Lovestone S, Garcia ME, Doody RS, Munoz-Fernandez C, Sussams R, Lin H, Fairchild TJ, Benito YA, Holmes C, Karamujić-Comić H, Frosch MP, Thonberg H, Maier W, Roshchupkin G, Ghetti B, Giedraitis V, Kawalia A, Li S, Huebinger RM, Kilander L, Moebus S, Hernández I, Kamboh MI, Brundin R, Turton J, Yang Q, Katz MJ, Concaro L, Lord J, Beiser AS, Keene CD, Helisalmi S, Kloszewska I, Kukull WA, Koivisto AM, Lynch A, Tarraga L, Larson EB, Haapasalo A, Lawlor B, Mosley TH, Lipton RB, Solfrizzi V, Gill M, Longstreth WT, Montine TJ, Frisardi V, Diez-Fairen M, Rivadeneira F, Petersen RC, Deramecourt V, Alvarez I, Salani F, Ciarrella A, Boerwinkle E, Reiman EM, Fievet N, Rotter JJ, Reisch JS, Hanon O, Cupidi C, Andre Uitterlinden AG, Royall DR, Dufouil C, Maletta RG, de Rojas I, Sano M, Brice A, Cecchetti R, George-Hyslop PS, Ritchie K, Tsolaki M, Tsuang DW, Dubois B, Craig D, Wu C-K, Soininen H, Avramidou D, Albin RL, Fratiglioni L, Germanou A, Apostolova LG, Keller L, Koutroumani M, Arnold SE, Panza F, Gkatzima O, Asthana S, Hannequin D, Whitehead P, Atwood CS, Caffarra P, Hampel H, Quintela I, Carracedo Á, Lannfelt L, Rubinsztein DC, Barnes LL, Pasquier F, Frölich L, Barral S, McGuinness B, Beach TG, Johnston JA, Becker JT, Passmore P, Bigio EH, Schott JM, Bird TD, Warren JD, Boeve BF, Lupton MK, Bowen JD, Proitsi P, Boxer A, Powell JF, Burke JR, Kauwe JSK, Burns JM, Mancuso M, Buxbaum JD, Bonuccelli U, Cairns NJ, McQuillin A, Cao C, Livingston G, Carlson CS, Bass NJ, Carlsson CM, Hardy J, Carney RM, Bras J, Carrasquillo MM, Guerreiro R, Allen M, Chui HC, Fisher E, Masullo C, Crocco EA, DeCarli C, Bisceglia G, Dick M, Ma L, Duara R, Graff-Radford NR, Evans DA, Hodges A, Faber KM, Scherer M, Fallon KB, Riemenschneider M, Fardo DW, Heun R, Farlow MR, Kölsch H, Ferris S, Leber M, Foroud TM, Heuser I, Galasko DR, Giegling I, Gearing M, Hüll M, Geschwind DH, Gilbert JR, Morris J, Green RC, Mayo K, Growdon JH, Feulner T, Hamilton RL, Harrell LE, Drichel D, Honig LS, Cushion TD, Huentelman MJ, Hollingworth P, Hulette CM, Hyman BT, Marshall R, Jarvik GP, Meggy A, Abner E, Menzies GE, Jin L-W, Leonenko G, Real LM, Jun GR, Baldwin CT, Grozeva D, Karydas A, Russo G, Kaye JA, Kim R, Jessen F, Kowall NW, Vellas B, Kramer JH, Vardy E, LaFerla FM, Jöckel K-H, Lah JJ, Dichgans M, Leverenz JB, Mann D, Levey AI, Pickering-Brown S, Lieberman AP. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immunity and lipid processing. *Nature genetics*. 2019;51(3):414-30. doi: 10.1038/s41588-019-0358-2.

38. Schwartzenuber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, Young AMH, Franklin RJM, Johnson T, Estrada K, Gaffney DJ, Beltrao P, Bassett A. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nature genetics*. 2021;53(3):392-402. Epub 2021/02/17. doi: 10.1038/s41588-020-00776-w. PubMed PMID: 33589840; PMCID: PMC7610386.

39. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hägg S, Athanasiu L, Voyle N, Proitsi P, Witoelar A, Stringer S, Aarsland D, Almdahl IS, Andersen F, Bergh S, Bettella F, Bjornsson S, Brækhus A, Bråthen G, de Leeuw C, Desikan RS, Djurovic S, Dumitrescu L, Fladby T, Hohman TJ, Jonsson PV, Kiddle SJ, Rongve A, Saltvedt I, Sando SB, Selbæk G, Shoai M, Skene NG, Snaedal J, Stordal E, Ulstein ID, Wang Y, White LR, Hardy J, Hjerling-Leffler J, Sullivan PF, van der Flier WM, Dobson R, Davis LK, Stefansson H, Stefansson K, Pedersen NL, Ripke S, Andreassen OA, Posthuma D. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics*. 2019;51(3):404-13. doi: 10.1038/s41588-018-0311-9.
40. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*. 2019;20(8):467-84. doi: 10.1038/s41576-019-0127-1.
41. Müller U, Winter P, Graeber MB. A presenilin 1 mutation in the first case of Alzheimer's disease. *Lancet Neurol*. 2013;12(2):129-30. Epub 2012/12/19. doi: 10.1016/s1474-4422(12)70307-1. PubMed PMID: 23246540.
42. Mutations in APP: ALZFORUM; 2021 [cited 2021]. Available from: <https://www.alzforum.org/mutations/app>.
43. Goedert M, Spillantini MG. Tau mutations in frontotemporal dementia FTDP-17 and their relevance for Alzheimer's disease. *Biochimica et biophysica acta*. 2000;1502(1):110-21. Epub 2000/07/19. doi: 10.1016/s0925-4439(00)00037-5. PubMed PMID: 10899436.
44. Umeda T, Maekawa S, Kimura T, Takashima A, Tomiyama T, Mori H. Neurofibrillary tangle formation by introducing wild-type human tau into APP transgenic mice. *Acta neuropathologica*. 2014;127(5):685-98. Epub 2014/02/18. doi: 10.1007/s00401-014-1259-1. PubMed PMID: 24531886.
45. Drummond E, Wisniewski T. Alzheimer's disease: experimental models and reality. *Acta neuropathologica*. 2017;133(2):155-75. Epub 2016/12/28. doi: 10.1007/s00401-016-1662-x. PubMed PMID: 28025715; PMCID: PMC5253109.
46. Price JL, Morris JC. Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Ann Neurol*. 1999;45(3):358-68. Epub 1999/03/11. doi: 10.1002/1531-8249(199903)45:3<358::aid-ana12>3.0.co;2-x. PubMed PMID: 10072051.
47. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR, Jr., Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):280-92. Epub 2011/04/26. doi: 10.1016/j.jalz.2011.03.003. PubMed PMID: 21514248; PMCID: PMC3220946.
48. Götz J, Bodea L-G, Goedert M. Rodent models for Alzheimer disease. *Nature Reviews Neuroscience*. 2018;19(10):583-98. doi: 10.1038/s41583-018-0054-8.
49. AMYLOID-RELATED THERAPEUTICS [Internet]2021. Available from: https://www.alzforum.org/therapeutics/search?fda_statuses=&target_types%5B%5D=170&therapy_types=&conditions%5B%5D=145&keywords-entry=&keywords=#results.
50. Sevigny J, Chiao P, Bussièrè T, Weinreb PH, Williams L, Maier M, Dunstan R, Salloway S, Chen T, Ling Y, O'Gorman J, Qian F, Arastu M, Li M, Chollate S, Brennan MS, Quintero-Monzon O, Scannevin RH, Arnold HM, Engber T, Rhodes K, Ferrero J, Hang Y, Mikulskis A, Grimm J, Hock C, Nitsch RM, Sandrock A. The antibody aducanumab reduces A β plaques in

Alzheimer's disease. *Nature*. 2016;537(7618):50-6. Epub 2016/09/02. doi: 10.1038/nature19323. PubMed PMID: 27582220.

51. Knopman DS, Jones DT, Greicius MD. Failure to demonstrate efficacy of aducanumab: An analysis of the EMERGE and ENGAGE trials as reported by Biogen, December 2019. *Alzheimers Dement*. 2021;17(4):696-701. Epub 2020/11/03. doi: 10.1002/alz.12213. PubMed PMID: 33135381.

52. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, Masliah E, Mackenzie IR, Feldman H, Feiden W, Kretzschmar HA, Trojanowski JQ, Lee VM. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (New York, NY)*. 2006;314(5796):130-3. Epub 2006/10/07. doi: 10.1126/science.1134108. PubMed PMID: 17023659.

53. Brettschneider J, Arai K, Del Tredici K, Toledo JB, Robinson JL, Lee EB, Kuwabara S, Shibuya K, Irwin DJ, Fang L, Van Deerlin VM, Elman L, McCluskey L, Ludolph AC, Lee VM, Braak H, Trojanowski JQ. TDP-43 pathology and neuronal loss in amyotrophic lateral sclerosis spinal cord. *Acta Neuropathol*. 2014;128(3):423-37. Epub 2014/06/12. doi: 10.1007/s00401-014-1299-6. PubMed PMID: 24916269; PMCID: PMC4384652.

54. Bai B, Hales CM, Chen PC, Gozal Y, Dammer EB, Fritz JJ, Wang X, Xia Q, Duong DM, Street C, Cantero G, Cheng D, Jones DR, Wu Z, Li Y, Diner I, Heilman CJ, Rees HD, Wu H, Lin L, Szulwach KE, Gearing M, Mufson EJ, Bennett DA, Montine TJ, Seyfried NT, Wingo TS, Sun YE, Jin P, Hanfelt J, Willcock DM, Levey A, Lah JJ, Peng J. U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proc Natl Acad Sci U S A*. 2013;110(41):16562-7. Epub 2013/09/12. doi: 10.1073/pnas.1310249110. PubMed PMID: 24023061; PMCID: PMC3799305.

55. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. 2012;149(6):1393-406. Epub 2012/06/05. doi: 10.1016/j.cell.2012.04.031. PubMed PMID: 22658674.

56. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*. 2008;582(14):1977-86. Epub 2008/03/18. doi: 10.1016/j.febslet.2008.03.004. PubMed PMID: 18342629; PMCID: PMC2858862.

57. Wolozin B, Apicco D. RNA binding proteins and the genesis of neurodegenerative diseases. *Adv Exp Med Biol*. 2015;822:11-5. Epub 2014/11/25. doi: 10.1007/978-3-319-08927-0_3. PubMed PMID: 25416971; PMCID: PMC4694570.

58. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, Mann D, Tsuchiya K, Yoshida M, Hashizume Y, Oda T. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun*. 2006;351(3):602-11. Epub 2006/11/07. doi: 10.1016/j.bbrc.2006.10.093. PubMed PMID: 17084815.

59. Kwiatkowski TJ, Jr., Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, Davis A, Gilchrist J, Kasarskis EJ, Munsat T, Valdmanis P, Rouleau GA, Hosler BA, Cortelli P, de Jong PJ, Yoshinaga Y, Haines JL, Pericak-Vance MA, Yan J, Ticozzi N, Siddique T, McKenna-Yasek D, Sapp PC, Horvitz HR, Landers JE, Brown RH, Jr. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science (New York, NY)*. 2009;323(5918):1205-8. Epub 2009/03/03. doi: 10.1126/science.1166066. PubMed PMID: 19251627.

60. Taylor JP, Brown RH, Jr., Cleveland DW. Decoding ALS: from genes to mechanism. *Nature*. 2016;539(7628):197-206. Epub 2016/11/11. doi: 10.1038/nature20413. PubMed PMID: 27830784; PMCID: PMC5585017.
61. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P, Ganesalingam J, Williams KL, Tripathi V, Al-Saraj S, Al-Chalabi A, Leigh PN, Blair IP, Nicholson G, de Belleruche J, Gallo JM, Miller CC, Shaw CE. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science (New York, NY)*. 2009;323(5918):1208-11. Epub 2009/03/03. doi: 10.1126/science.1165942. PubMed PMID: 19251628; PMCID: PMC4516382.
62. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Vande Velde C, Bouchard JP, Lacomblez L, Pochigaeva K, Salachas F, Pradat PF, Camu W, Meininger V, Dupre N, Rouleau GA. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature genetics*. 2008;40(5):572-4. Epub 2008/04/01. doi: 10.1038/ng.132. PubMed PMID: 18372902.
63. Van Deerlin VM, Leverenz JB, Bekris LM, Bird TD, Yuan W, Elman LB, Clay D, Wood EM, Chen-Plotkin AS, Martinez-Lage M, Steinbart E, McCluskey L, Grossman M, Neumann M, Wu IL, Yang WS, Kalb R, Galasko DR, Montine TJ, Trojanowski JQ, Lee VM, Schellenberg GD, Yu CE. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol*. 2008;7(5):409-16. Epub 2008/04/09. doi: 10.1016/s1474-4422(08)70071-1. PubMed PMID: 18396105; PMCID: PMC3546119.
64. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol*. 2018;19(5):327-41. Epub 2018/01/18. doi: 10.1038/nrm.2017.130. PubMed PMID: 29339797.
65. Baltz AG, 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, Landthaler M. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell*. 2012;46(5):674-90. Epub 2012/06/12. doi: 10.1016/j.molcel.2012.05.021. PubMed PMID: 22681889.
66. Castello A, Fischer B, Frese CK, Horos R, Alleaume AM, Foehr S, Curk T, Krijgsveld J, Hentze MW. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell*. 2016;63(4):696-710. Epub 2016/07/28. doi: 10.1016/j.molcel.2016.06.029. PubMed PMID: 27453046; PMCID: PMC5003815.
67. Posey AE, Holehouse AS, Pappu RV. Phase Separation of Intrinsically Disordered Proteins. *Methods Enzymol*. 2018;611:1-30. Epub 2018/11/26. doi: 10.1016/bs.mie.2018.09.035. PubMed PMID: 30471685.
68. Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahn M, Maharana S, Lemaitre R, Pozniakovskiy A, Drechsel D, Poser I, Pappu RV, Alberti S, Hyman AA. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*. 2018;174(3):688-99 e16. Epub 2018/07/03. doi: 10.1016/j.cell.2018.06.006. PubMed PMID: 29961577; PMCID: PMC6063760.
69. Wheeler RJ, Hyman AA. Controlling compartmentalization by non-membrane-bound organelles. *Philos Trans R Soc Lond B Biol Sci*. 2018;373(1747). Epub 2018/04/11. doi: 10.1098/rstb.2017.0193. PubMed PMID: 29632271; PMCID: PMC5904305.
70. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, Taylor JP. Phase separation by low complexity domains promotes stress granule assembly and drives

- pathological fibrillization. *Cell*. 2015;163(1):123-33. Epub 2015/09/26. doi: 10.1016/j.cell.2015.09.015. PubMed PMID: 26406374; PMCID: PMC5149108.
71. Ganassi M, Mateju D, Bigi I, Mediani L, Poser I, Lee HO, Seguin SJ, Morelli FF, Vinet J, Leo G, Pansarasa O, Cereda C, Poletti A, Alberti S, Carra S. A Surveillance Function of the HSPB8-BAG3-HSP70 Chaperone Complex Ensures Stress Granule Integrity and Dynamism. *Molecular Cell*. 2016;63(5):796-810. doi: <https://doi.org/10.1016/j.molcel.2016.07.021>.
72. Alberti S, Mateju D, Mediani L, Carra S. Granulostasis: Protein Quality Control of RNP Granules. *Frontiers in Molecular Neuroscience*. 2017;10(84). doi: 10.3389/fnmol.2017.00084.
73. Xue S, Gong R, He F, Li Y, Wang Y, Tan T, Luo SZ. Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic-acidic motifs. *Sci Adv*. 2019;5(11):eaax5349. Epub 2019/11/15. doi: 10.1126/sciadv.aax5349. PubMed PMID: 31723601; PMCID: PMC6834393.
74. Gebauer F, Schwarzl T, Valcárcel J, Hentze MW. RNA-binding proteins in human genetic disease. *Nat Rev Genet*. 2021;22(3):185-98. Epub 2020/11/26. doi: 10.1038/s41576-020-00302-y. PubMed PMID: 33235359.
75. Sternburg EL, Gruijs da Silva LA, Dormann D. Post-translational modifications on RNA-binding proteins: accelerators, brakes, or passengers in neurodegeneration? *Trends in Biochemical Sciences*. 2021. doi: <https://doi.org/10.1016/j.tibs.2021.07.004>.
76. Josephs KA, Whitwell JL, Tosakulwong N, Weigand SD, Murray ME, Liesinger AM, Petrucelli L, Senjem ML, Ivnik RJ, Parisi JE, Petersen RC, Dickson DW. TAR DNA-binding protein 43 and pathological subtype of Alzheimer's disease impact clinical features. *Ann Neurol*. 2015;78(5):697-709. Epub 2015/08/01. doi: 10.1002/ana.24493. PubMed PMID: 26224156; PMCID: PMC4623932.
77. Vanderweyde T, Yu H, Varnum M, Liu-Yesucevitz L, Citro A, Ikezu T, Duff K, Wolozin B. Contrasting pathology of the stress granule proteins TIA-1 and G3BP in tauopathies. *J Neurosci*. 2012;32(24):8270-83. Epub 2012/06/16. doi: 10.1523/JNEUROSCI.1592-12.2012. PubMed PMID: 22699908; PMCID: PMC3402380.
78. Jiang L, Lin W, Zhang C, Ash PEA, Verma M, Kwan J, van Vliet E, Yang Z, Cruz AL, Boudeau S, Maziuk BF, Lei S, Song J, Alvarez VE, Hovde S, Abisambra JF, Kuo MH, Kanaan N, Murray ME, Crary JF, Zhao J, Cheng JX, Petrucelli L, Li H, Emili A, Wolozin B. Interaction of tau with HNRNPA2B1 and N(6)-methyladenosine RNA mediates the progression of tauopathy. *Mol Cell*. 2021. Epub 2021/08/29. doi: 10.1016/j.molcel.2021.07.038. PubMed PMID: 34453888.
79. Maziuk BF, Apicco DJ, Cruz AL, Jiang L, Ash PEA, da Rocha EL, Zhang C, Yu WH, Leszyk J, Abisambra JF, Li H, Wolozin B. RNA binding proteins co-localize with small tau inclusions in tauopathy. *Acta Neuropathol Commun*. 2018;6(1):71. Epub 2018/08/03. doi: 10.1186/s40478-018-0574-5. PubMed PMID: 30068389; PMCID: PMC6069705.
80. Apicco DJ, Ash PEA, Maziuk B, LeBlang C, Medalla M, Al Abdullatif A, Ferragud A, Botelho E, Ballance HI, Dhawan U, Boudeau S, Cruz AL, Kashy D, Wong A, Goldberg LR, Yazdani N, Zhang C, Ung CY, Tripodis Y, Kanaan NM, Ikezu T, Cottone P, Leszyk J, Li H, Luebke J, Bryant CD, Wolozin B. Reducing the RNA binding protein TIA1 protects against tau-mediated neurodegeneration in vivo. *Nat Neurosci*. 2018;21(1):72-80. Epub 2017/12/24. doi: 10.1038/s41593-017-0022-z. PubMed PMID: 29273772; PMCID: PMC5745051.
81. McMillan PJ, Strovast TJ, Baum M, Mitchell BK, Eck RJ, Hendricks N, Wheeler JM, Latimer CS, Keene CD, Kraemer BC. Pathological tau drives ectopic nuclear speckle scaffold

- protein SRRM2 accumulation in neuron cytoplasm in Alzheimer's disease. *Acta Neuropathologica Communications*. 2021;9(1):117. doi: 10.1186/s40478-021-01219-1.
82. Hsieh Y-C, Guo C, Yalamanchili HK, Abreha M, Al-Ouran R, Li Y, Dammer EB, Lah JJ, Levey AI, Bennett DA, De Jager PL, Seyfried NT, Liu Z, Shulman JM. Tau-mediated Disruption of the Spliceosome Triggers Cryptic RNA-splicing and Neurodegeneration in Alzheimer's Disease. *bioRxiv*. 2019:514927. doi: 10.1101/514927.
83. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207. doi: 10.1038/nature01511.
84. Coon JJ, Syka JEP, Shabanowitz J, Hunt DF. Tandem Mass Spectrometry for Peptide and Protein Sequence Analysis. *BioTechniques*. 2005;38(4):519-23. doi: 10.2144/05384te01. PubMed PMID: 15884666.
85. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng*. 2009;11:49-79. Epub 2009/04/30. doi: 10.1146/annurev-bioeng-061008-124934. PubMed PMID: 19400705.
86. Bagranyan K, Barash JR, Arnon SS, Kalkum M. Attomolar detection of botulinum toxin type A in complex biological matrices. *PLoS One*. 2008;3(4):e2041. Epub 2008/05/01. doi: 10.1371/journal.pone.0002041. PubMed PMID: 18446228; PMCID: PMC2323579.
87. Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW, Smith RD. Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem Soc Rev*. 2012;41(10):3912-28. Epub 2012/04/14. doi: 10.1039/c2cs15331a. PubMed PMID: 22498958; PMCID: PMC3375054.
88. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. doi: 10.1186/1471-2105-9-559.
89. Seyfried NT, Dammer EB, Swarup V, Nandakumar D, Duong DM, Yin L, Deng Q, Nguyen T, Hales CM, Wingo T, Glass J, Gearing M, Thambisetty M, Troncoso JC, Geschwind DH, Lah JJ, Levey AI. A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Syst*. 2017;4(1):60-72 e4. Epub 2016/12/19. doi: 10.1016/j.cels.2016.11.006. PubMed PMID: 27989508; PMCID: PMC5269514.
90. Greenberg SG, Davies P. A preparation of Alzheimer paired helical filaments that displays distinct tau proteins by polyacrylamide gel electrophoresis. *Proc Natl Acad Sci U S A*. 1990;87(15):5827-31. Epub 1990/08/01. doi: 10.1073/pnas.87.15.5827. PubMed PMID: 2116006; PMCID: PMC54421.
91. Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc Natl Acad Sci U S A*. 1998;95(11):6469-73. Epub 1998/05/30. doi: 10.1073/pnas.95.11.6469. PubMed PMID: 9600990; PMCID: PMC27806.
92. Hasegawa M, Fujiwara H, Nonaka T, Wakabayashi K, Takahashi H, Lee VM, Trojanowski JQ, Mann D, Iwatsubo T. Phosphorylated alpha-synuclein is ubiquitinated in alpha-synucleinopathy lesions. *J Biol Chem*. 2002;277(50):49071-6. Epub 2002/10/16. doi: 10.1074/jbc.M208046200. PubMed PMID: 12377775.
93. Pérez M, Valpuesta JM, de Garcini EM, Quintana C, Arrasate M, López Carrascosa JL, Rábano A, García de Yébenes J, Avila J. Ferritin is associated with the aberrant tau filaments present in progressive supranuclear palsy. *Am J Pathol*. 1998;152(6):1531-9. Epub 1998/06/17. PubMed PMID: 9626057; PMCID: PMC1858430.
94. Forman MS, Zhukareva V, Bergeron C, Chin SS, Grossman M, Clark C, Lee VM, Trojanowski JQ. Signature tau neuropathology in gray and white matter of corticobasal

- degeneration. *Am J Pathol.* 2002;160(6):2045-53. Epub 2002/06/12. doi: 10.1016/s0002-9440(10)61154-6. PubMed PMID: 12057909; PMCID: PMC1850831.
95. Cherry JD, Zeineddin A, Dammer EB, Webster JA, Duong D, Seyfried NT, Levey AI, Alvarez VE, Huber BR, Stein TD, Kiernan PT, McKee AC, Lah JJ, Hales CM. Characterization of Detergent Insoluble Proteome in Chronic Traumatic Encephalopathy. *J Neuropathol Exp Neurol.* 2018;77(1):40-9. Epub 2017/11/18. doi: 10.1093/jnen/nlx100. PubMed PMID: 29145658; PMCID: PMC5939631.
96. Hales CM, Dammer EB, Deng Q, Duong DM, Gearing M, Troncoso JC, Thambisetty M, Lah JJ, Shulman JM, Levey AI, Seyfried NT. Changes in the detergent-insoluble brain proteome linked to amyloid and tau in Alzheimer's Disease progression. *Proteomics.* 2016;16(23):3042-53. Epub 2016/10/09. doi: 10.1002/pmic.201600057. PubMed PMID: 27718298; PMCID: PMC5462625.
97. Guo Q, Dammer EB, Zhou M, Kundinger SR, Gearing M, Lah JJ, Levey AI, Shulman JM, Seyfried NT. Targeted Quantification of Detergent-Insoluble RNA-Binding Proteins in Human Brain Reveals Stage and Disease Specific Co-aggregation in Alzheimer's Disease. *Frontiers in Molecular Neuroscience.* 2021;14(29). doi: 10.3389/fnmol.2021.623659.
98. Diner I, Hales CM, Bishof I, Rabenold L, Duong DM, Yi H, Laur O, Gearing M, Troncoso J, Thambisetty M, Lah JJ, Levey AI, Seyfried NT. Aggregation properties of the small nuclear ribonucleoprotein U1-70K in Alzheimer disease. *J Biol Chem.* 2014;289(51):35296-313. Epub 2014/10/31. doi: 10.1074/jbc.M114.562959. PubMed PMID: 25355317; PMCID: PMC4271217.
99. Hales CM, Seyfried NT, Dammer EB, Duong D, Yi H, Gearing M, Troncoso JC, Mufson EJ, Thambisetty M, Levey AI, Lah JJ. U1 small nuclear ribonucleoproteins (snRNPs) aggregate in Alzheimer's disease due to autosomal dominant genetic mutations and trisomy 21. *Mol Neurodegener.* 2014;9:15. Epub 2014/04/30. doi: 10.1186/1750-1326-9-15. PubMed PMID: 24773620; PMCID: PMC4022210.
100. Bishof I, Dammer EB, Duong DM, Kundinger SR, Gearing M, Lah JJ, Levey AI, Seyfried NT. RNA-binding proteins with basic-acidic dipeptide (BAD) domains self-assemble and aggregate in Alzheimer's disease. *J Biol Chem.* 2018;293(28):11047-66. Epub 2018/05/29. doi: 10.1074/jbc.RA118.001747. PubMed PMID: 29802200; PMCID: PMC6052236.
101. Raj T, Li YI, Wong G, Humphrey J, Wang M, Ramdhani S, Wang YC, Ng B, Gupta I, Haroutunian V, Schadt EE, Young-Pearse T, Mostafavi S, Zhang B, Sklar P, Bennett DA, De Jager PL. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat Genet.* 2018;50(11):1584-92. Epub 2018/10/10. doi: 10.1038/s41588-018-0238-1. PubMed PMID: 30297968.
102. Johnson ECB, Dammer EB, Duong DM, Yin L, Thambisetty M, Troncoso JC, Lah JJ, Levey AI, Seyfried NT. Deep proteomic network analysis of Alzheimer's disease brain reveals alterations in RNA binding proteins and RNA splicing associated with disease. *Mol Neurodegener.* 2018;13(1):52. Epub 2018/10/06. doi: 10.1186/s13024-018-0282-4. PubMed PMID: 30286791; PMCID: PMC6172707.
103. Rybak-Wolf A, Plass M. RNA Dynamics in Alzheimer's Disease. *Molecules.* 2021;26(17):5113. PubMed PMID: doi:10.3390/molecules26175113.
104. Matera AG, Wang Z. A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology.* 2014;15(2):108-21. doi: 10.1038/nrm3742.

105. Wang Y, Liu J, Huang BO, Xu YM, Li J, Huang LF, Lin J, Zhang J, Min QH, Yang WM, Wang XZ. Mechanism of alternative splicing and its regulation. *Biomed Rep.* 2015;3(2):152-8. Epub 2015/03/24. doi: 10.3892/br.2014.407. PubMed PMID: 25798239; PMCID: PMC4360811.
106. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008;40(12):1413-5. Epub 2008/11/04. doi: 10.1038/ng.259. PubMed PMID: 18978789.
107. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008;456(7221):470-6. doi: 10.1038/nature07509.
108. Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature.* 2009;458(7237):475-80. Epub 2009/03/28. doi: 10.1038/nature07851. PubMed PMID: 19325628; PMCID: PMC2673513.
109. Philipps D, Celotto AM, Wang QQ, Tarng RS, Graveley BR. Arginine/serine repeats are sufficient to constitute a splicing activation domain. *Nucleic Acids Res.* 2003;31(22):6502-8. Epub 2003/11/07. doi: 10.1093/nar/gkg845. PubMed PMID: 14602908; PMCID: PMC275541.
110. Coulter LR, Landree MA, Cooper TA. Identification of a new class of exonic splicing enhancers by in vivo selection. *Mol Cell Biol.* 1997;17(4):2143-50. Epub 1997/04/01. doi: 10.1128/mcb.17.4.2143. PubMed PMID: 9121463; PMCID: PMC232062.
111. Baker BS. Sex in flies: the splice of life. *Nature.* 1989;340(6234):521-4. Epub 1989/08/17. doi: 10.1038/340521a0. PubMed PMID: 2505080.
112. Theissen H, Etzerodt M, Reuter R, Schneider C, Lottspeich F, Argos P, Lührmann R, Philipson L. Cloning of the human cDNA for the U1 RNA-associated 70K protein. *Embo j.* 1986;5(12):3209-17. Epub 1986/12/01. PubMed PMID: 3028775; PMCID: PMC1167314.
113. Spritz RA, Strunk K, Surowy CS, Hoch SO, Barton DE, Francke U. The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding. *Nucleic Acids Res.* 1987;15(24):10373-91. Epub 1987/12/23. doi: 10.1093/nar/15.24.10373. PubMed PMID: 2447561; PMCID: PMC339950.
114. Krainer AR, Mayeda A, Kozak D, Binns G. Functional expression of cloned human splicing factor SF2: homology to rna-binding proteins, U1 70K, and drosophila splicing regulators. *Cell.* 1991;66(2):383-94. doi: [https://doi.org/10.1016/0092-8674\(91\)90627-B](https://doi.org/10.1016/0092-8674(91)90627-B).
115. Ge H, Zuo P, Manley JL. Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell.* 1991;66(2):373-82. Epub 1991/07/26. doi: 10.1016/0092-8674(91)90626-a. PubMed PMID: 1855257.
116. Fu XD, Maniatis T. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc Natl Acad Sci U S A.* 1992;89(5):1725-9. Epub 1992/03/01. doi: 10.1073/pnas.89.5.1725. PubMed PMID: 1531875; PMCID: PMC48525.
117. Roth MB, Murphy C, Gall JG. 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.* 1990;111(6 Pt 1):2217-23. Epub 1990/12/01. doi: 10.1083/jcb.111.6.2217. PubMed PMID: 1703534; PMCID: PMC2116404.
118. Zahler AM, Lane WS, Stolk JA, Roth MB. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes & development.* 1992;6(5):837-47. Epub 1992/05/01. PubMed PMID: 1577277.
119. Fu XD. The superfamily of arginine/serine-rich splicing factors. *Rna.* 1995;1(7):663-80. Epub 1995/09/01. PubMed PMID: 7585252; PMCID: PMC1369309.

120. Manley JL, Krainer AR. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* 2010;24(11):1073-4. Epub 2010/06/03. doi: 10.1101/gad.1934910. PubMed PMID: 20516191; PMCID: PMC2878644.
121. Siebel CW, Guthrie C. The essential yeast RNA binding protein Npl3p is methylated. *Proceedings of the National Academy of Sciences.* 1996;93(24):13641-6. doi: 10.1073/pnas.93.24.13641.
122. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science.* 2009;325(5942):834-40. Epub 2009/07/18. doi: 10.1126/science.1175371. PubMed PMID: 19608861.
123. Edmond V, Moysan E, Khochbin S, Matthias P, Brambilla C, Brambilla E, Gazzeri S, Eymin B. Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. *Embo j.* 2011;30(3):510-23. Epub 2010/12/16. doi: 10.1038/emboj.2010.333. PubMed PMID: 21157427; PMCID: PMC3034009.
124. Gui JF, Lane WS, Fu XD. A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature.* 1994;369(6482):678-82. Epub 1994/06/23. doi: 10.1038/369678a0. PubMed PMID: 8208298.
125. Botti V, McNicoll F, Steiner MC, Richter FM, Solovyeva A, Wegener M, Schwich OD, Poser I, Zarnack K, Wittig I, Neugebauer KM, Muller-McNicoll M. Cellular differentiation state modulates the mRNA export activity of SR proteins. *J Cell Biol.* 2017;216(7):1993-2009. Epub 2017/06/09. doi: 10.1083/jcb.201610051. PubMed PMID: 28592444; PMCID: PMC5496613.
126. Mermoud JE, Cohen P, Lamond AI. Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res.* 1992;20(20):5263-9. Epub 1992/10/25. doi: 10.1093/nar/20.20.5263. PubMed PMID: 1331983; PMCID: PMC334330.
127. Mermoud JE, Cohen PT, Lamond AI. Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *Embo j.* 1994;13(23):5679-88. Epub 1994/12/01. PubMed PMID: 7988565; PMCID: PMC395533.
128. Stojdl DF, Bell JC. SR protein kinases: the splice of life. *Biochem Cell Biol.* 1999;77(4):293-8. Epub 1999/11/05. PubMed PMID: 10546892.
129. Mathew R, Hartmuth K, Möhlmann S, Urlaub H, Ficner R, Lührmann R. Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat Struct Mol Biol.* 2008;15(5):435-43. Epub 2008/04/22. doi: 10.1038/nsmb.1415. PubMed PMID: 18425142.
130. Zhou Z, Fu XD. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma.* 2013;122(3):191-207. Epub 2013/03/26. doi: 10.1007/s00412-013-0407-z. PubMed PMID: 23525660; PMCID: PMC3660409.
131. Schaffert L-N, Carter WG. Do Post-Translational Modifications Influence Protein Aggregation in Neurodegenerative Diseases: A Systematic Review. *Brain Sciences.* 2020;10(4):232. PubMed PMID: doi:10.3390/brainsci10040232.
132. Joshi G, Wang Y. Golgi defects enhance APP amyloidogenic processing in Alzheimer's disease. *Bioessays.* 2015;37(3):240-7. Epub 2014/12/30. doi: 10.1002/bies.201400116. PubMed PMID: 25546412; PMCID: PMC4407201.
133. Chen G-f, Xu T-h, Yan Y, Zhou Y-r, Jiang Y, Melcher K, Xu HE. Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica.* 2017;38(9):1205-35. doi: 10.1038/aps.2017.28.

134. Ping L, Kundinger SR, Duong DM, Yin L, Gearing M, Lah JJ, Levey AI, Seyfried NT. Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci Data*. 2020;7(1):315. Epub 2020/09/29. doi: 10.1038/s41597-020-00650-8. PubMed PMID: 32985496; PMCID: PMC7522715.
135. Abreha MH, Dammer EB, Ping L, Zhang T, Duong DM, Gearing M, Lah JJ, Levey AI, Seyfried NT. Quantitative Analysis of the Brain Ubiquitylome in Alzheimer's Disease. *Proteomics*. 2018;18(20):e1800108. Epub 2018/09/20. doi: 10.1002/pmic.201800108. PubMed PMID: 30230243; PMCID: PMC6283072.
136. Cohen TJ, Guo JL, Hurtado DE, Kwong LK, Mills IP, Trojanowski JQ, Lee VM. The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nat Commun*. 2011;2:252. Epub 2011/03/24. doi: 10.1038/ncomms1255. PubMed PMID: 21427723; PMCID: PMC3120096.
137. Huseby CJ, Hoffman CN, Cooper GL, Cocuron JC, Alonso AP, Thomas SN, Yang AJ, Kuret J. Quantification of Tau Protein Lysine Methylation in Aging and Alzheimer's Disease. *J Alzheimers Dis*. 2019;71(3):979-91. Epub 2019/08/28. doi: 10.3233/jad-190604. PubMed PMID: 31450505; PMCID: PMC6844542.
138. Eftekhazadeh B, Hyman BT, Wegmann S. Structural studies on the mechanism of protein aggregation in age related neurodegenerative diseases. *Mech Ageing Dev*. 2016;156:1-13. Epub 2016/03/24. doi: 10.1016/j.mad.2016.03.001. PubMed PMID: 27005270.
139. Hofweber M, Dormann D. Friend or foe-Post-translational modifications as regulators of phase separation and RNP granule dynamics. *J Biol Chem*. 2019;294(18):7137-50. Epub 2018/12/28. doi: 10.1074/jbc.TM118.001189. PubMed PMID: 30587571; PMCID: PMC6509508.
140. Ambadipudi S, Biernat J, Riedel D, Mandelkow E, Zweckstetter M. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature communications*. 2017;8(1). doi: 10.1038/s41467-017-00480-0.
141. Wegmann S, Eftekhazadeh B, Tepper K, Zoltowska KM, Bennett RE, Dujardin S, Laskowski PR, MacKenzie D, Kamath T, Commins C, Vanderburg C, Roe AD, Fan Z, Molliex AM, Hernandez-Vega A, Muller D, Hyman AA, Mandelkow E, Taylor JP, Hyman BT. Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J*. 2018;37(7). Epub 2018/02/24. doi: 10.15252/embj.201798049. PubMed PMID: 29472250; PMCID: PMC5881631.
142. Zhang X, Vigers M, McCarty J, Rauch JN, Fredrickson GH, Wilson MZ, Shea JE, Han S, Kosik KS. The proline-rich domain promotes Tau liquid-liquid phase separation in cells. *J Cell Biol*. 2020;219(11). Epub 2020/10/01. doi: 10.1083/jcb.202006054. PubMed PMID: 32997736; PMCID: PMC7594490.
143. Haj-Yahya M, Gopinath P, Rajasekhar K, Mirbaha H, Diamond MI, Lashuel HA. Site-Specific Hyperphosphorylation Inhibits, Rather than Promotes, Tau Fibrillization, Seeding Capacity, and Its Microtubule Binding. *Angewandte Chemie International Edition*. 2020;59(10):4059-67. doi: <https://doi.org/10.1002/anie.201913001>.
144. da Silva LG, Simonetti F, Hutten S, Riemenschneider H, Sternburg EL, Pietrek LM, Gebel J, Dötsch V, Edbauer D, Hummer G, Stelzl LS, Dormann D. Disease-linked TDP-43 hyperphosphorylation suppresses TDP-43 condensation and aggregation. *bioRxiv*. 2021:2021.04.30.442163. doi: 10.1101/2021.04.30.442163.
145. Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, Tycko R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-

- Complexity Domains. *Cell*. 2017;171(3):615-27.e16. Epub 2017/09/26. doi: 10.1016/j.cell.2017.08.048. PubMed PMID: 28942918; PMCID: PMC5650524.
146. Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, O'Meally R, Dignon GL, Conicella AE, Zheng W, Best RB, Cole RN, Mittal J, Shewmaker F, Fawzi NL. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *The EMBO journal*. 2017;36(20):2951-67. Epub 2017/08/10. doi: 10.15252/embj.201696394. PubMed PMID: 28790177; PMCID: PMC5641905.
147. Qamar S, Wang G, Randle SJ, Ruggeri FS, Varela JA, Lin JQ, Phillips EC, Miyashita A, Williams D, Ströhl F, Meadows W, Ferry R, Dardov VJ, Tartaglia GG, Farrer LA, Kaminski Schierle GS, Kaminski CF, Holt CE, Fraser PE, Schmitt-Ulms G, Klenerman D, Knowles T, Vendruscolo M, St George-Hyslop P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell*. 2018;173(3):720-34.e15. Epub 2018/04/21. doi: 10.1016/j.cell.2018.03.056. PubMed PMID: 29677515; PMCID: PMC5927716.
148. Hofweber M, Hutten S, Bourgeois B, Spreitzer E, Niedner-Boblentz A, Schifferer M, Ruepp MD, Simons M, Niessing D, Madl T, Dormann D. Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell*. 2018;173(3):706-19.e13. Epub 2018/04/21. doi: 10.1016/j.cell.2018.03.004. PubMed PMID: 29677514.
149. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowitz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell*. 2015;57(5):936-47. Epub 2015/03/10. doi: 10.1016/j.molcel.2015.01.013. PubMed PMID: 25747659; PMCID: PMC4352761.
150. Ryan VH, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, Amaya J, Burke KA, Mittal J, Fawzi NL. Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol Cell*. 2018;69(3):465-79.e7. Epub 2018/01/24. doi: 10.1016/j.molcel.2017.12.022. PubMed PMID: 29358076; PMCID: PMC5801700.
151. Jack CR, Jr., Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol*. 2010;9(1):119-28. Epub 2010/01/20. doi: 10.1016/s1474-4422(09)70299-6. PubMed PMID: 20083042; PMCID: PMC2819840.
152. Zahler AM, Lane WS, Stolk JA, Roth MB. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes & Development*. 1992;6(5):837-47. doi: 10.1101/gad.6.5.837.
153. Shepard PJ, Hertel KJ. The SR protein family. *Genome Biol*. 2009;10(10):242. Epub 2009/10/28. doi: 10.1186/gb-2009-10-10-242. PubMed PMID: 19857271; PMCID: PMC2784316.
154. Bishof I, Dammer EB, Duong DM, Kundinger S, Gearing M, Lah JJ, Levey AI, Seyfried NT. RNA-binding proteins with basic-acidic dipeptide (BAD) domains self-assemble and aggregate in Alzheimer's disease. *Journal of Biological Chemistry*. 2018. doi: 10.1074/jbc.RA118.001747.
155. Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV, Jedd G. Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. *bioRxiv*. 2019:771592. doi: 10.1101/771592.
156. Tripathi V, Song DY, Zong X, Shevtsov SP, Hearn S, Fu XD, Dundr M, Prasanth KV. SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol Biol*

- Cell. 2012;23(18):3694-706. Epub 2012/08/03. doi: 10.1091/mbc.E12-03-0206. PubMed PMID: 22855529; PMCID: Pmc3442416.
157. Phelan MM, Goult BT, Clayton JC, Hautbergue GM, Wilson SA, Lian LY. The structure and selectivity of the SR protein SRSF2 RRM domain with RNA. *Nucleic Acids Res.* 2012;40(7):3232-44. Epub 2011/12/06. doi: 10.1093/nar/gkr1164. PubMed PMID: 22140111; PMCID: Pmc3326313.
158. Saitoh N, Sakamoto C, Hagiwara M, Agredano-Moreno LT, Jimenez-Garcia LF, Nakao M. The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. *Mol Biol Cell.* 2012;23(6):1115-28. Epub 2012/01/21. doi: 10.1091/mbc.E11-09-0783. PubMed PMID: 22262462; PMCID: Pmc3302738.
159. Graveley BR. Sorting out the complexity of SR protein functions. *Rna.* 2000;6(9):1197-211. Epub 2000/09/22. doi: 10.1017/s1355838200000960. PubMed PMID: 10999598; PMCID: PMC1369994.
160. Tazi J, Kornstadt U, Rossi F, Jeanteur P, Cathala G, Brunel C, Luhrmann R. Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature.* 1993;363(6426):283-6. Epub 1993/05/20. doi: 10.1038/363283a0. PubMed PMID: 8387646.
161. Shen H, Green MR. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell.* 2004;16(3):363-73. Epub 2004/11/05. doi: 10.1016/j.molcel.2004.10.021. PubMed PMID: 15525510.
162. Shen H, Kan JL, Green MR. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell.* 2004;13(3):367-76. Epub 2004/02/18. doi: 10.1016/s1097-2765(04)00025-5. PubMed PMID: 14967144.
163. Wu JY, Maniatis T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell.* 1993;75(6):1061-70. Epub 1993/12/17. doi: 10.1016/0092-8674(93)90316-i. PubMed PMID: 8261509.
164. Kohtz JD, Jamison SF, Will CL, Zuo P, Luhrmann R, Garcia-Blanco MA, Manley JL. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature.* 1994;368(6467):119-24. Epub 1994/03/10. doi: 10.1038/368119a0. PubMed PMID: 8139654.
165. Cao W, Garcia-Blanco MA. A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. *J Biol Chem.* 1998;273(32):20629-35. Epub 1998/08/01. PubMed PMID: 9685421.
166. Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *Embo j.* 1992;11(7):2655-64. Epub 1992/07/01. PubMed PMID: 1628625; PMCID: PMC556741.
167. Hanakahi LA, Sun H, Maizels N. High affinity interactions of nucleolin with G-G-paired rDNA. *J Biol Chem.* 1999;274(22):15908-12. Epub 1999/05/21. doi: 10.1074/jbc.274.22.15908. PubMed PMID: 10336496.
168. Brahm H, Meheus L, de Brabandere V, Fischer U, Luhrmann R. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA (New York, NY).* 2001;7(11):1531-42. Epub 2001/11/27. PubMed PMID: 11720283; PMCID: PMC1370196.
169. Zhang R, So BR, Li P, Yong J, Glisovic T, Wan L, Dreyfuss G. Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell.* 2011;146(3):384-95. Epub 2011/08/06. doi: 10.1016/j.cell.2011.06.043. PubMed PMID: 21816274; PMCID: PMC3160754.

170. Liddy KA, White MY, Cordwell SJ. Functional decorations: post-translational modifications and heart disease delineated by targeted proteomics. *Genome Med.* 2013;5(2):20. Epub 2013/03/01. doi: 10.1186/gm424. PubMed PMID: 23445784; PMCID: PMC3706772.
171. Misteli T. RNA splicing: What has phosphorylation got to do with it? *Curr Biol.* 1999;9(6):R198-200. Epub 1999/04/21. PubMed PMID: 10209090.
172. Xiao SH, Manley JL. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & development.* 1997;11(3):334-44. Epub 1997/02/01. doi: 10.1101/gad.11.3.334. PubMed PMID: 9030686.
173. Sanford JR, Bruzik JP. Developmental regulation of SR protein phosphorylation and activity. *Genes Dev.* 1999;13(12):1513-8. Epub 1999/07/01. PubMed PMID: 10385619; PMCID: PMC316808.
174. Larsen SC, Sylvestersen KB, Mund A, Lyon D, Mullari M, Madsen MV, Daniel JA, Jensen LJ, Nielsen ML. Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. *Sci Signal.* 2016;9(443):rs9. Epub 2016/09/01. doi: 10.1126/scisignal.aaf7329. PubMed PMID: 27577262.
175. Woppmann A, Patschinsky T, Bringmann P, Godt F, Luhrmann R. Characterisation of human and murine snRNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants. *Nucleic Acids Res.* 1990;18(15):4427-38. Epub 1990/08/11. doi: 10.1093/nar/18.15.4427. PubMed PMID: 2143816; PMCID: PMC331261.
176. Keil-Dlouha VV, Zylber N, Imhoff J, Tong N, Keil B. Proteolytic activity of pseudotrypsin. *FEBS Lett.* 1971;16(4):291-5. Epub 1971/09/01. PubMed PMID: 11945964.
177. Keil-Dlouha VV, Zylber N, Tong N, Keil B. Cleavage of glucagon by alpha- and beta-trypsin. *FEBS Lett.* 1971;16(4):287-90. Epub 1971/09/01. PubMed PMID: 11945963.
178. Good DM, Wirtala M, McAlister GC, Coon JJ. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics.* 2007;6(11):1942-51. Epub 2007/08/04. doi: 10.1074/mcp.M700073-MCP200. PubMed PMID: 17673454.
179. Liu J, McLuckey SA. Electron Transfer Dissociation: Effects of Cation Charge State on Product Partitioning in Ion/Ion Electron Transfer to Multiply Protonated Polypeptides. *Int J Mass Spectrom.* 2012;330-332:174-81. Epub 2012/12/25. doi: 10.1016/j.ijms.2012.07.013. PubMed PMID: 23264749; PMCID: PMC3525064.
180. Pitteri SJ, Chrisman PA, Hogan JM, McLuckey SA. Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO₂*. *Anal Chem.* 2005;77(6):1831-9. Epub 2005/03/15. doi: 10.1021/ac0483872. PubMed PMID: 15762593; PMCID: PMC1564063.
181. Dongré AR, Jones JL, Somogyi Á, Wysocki VH. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J Am Chem Soc.* 1996;118(35):8365-74. doi: 10.1021/ja9542193.
182. Tabb DL, Huang Y, Wysocki VH, Yates JR, 3rd. Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal Chem.* 2004;76(5):1243-8. Epub 2004/02/28. doi: 10.1021/ac0351163. PubMed PMID: 14987077; PMCID: PMC2813199.
183. Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, Walsh CT. Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem.* 1999;71(19):4250-3. Epub 1999/10/12. PubMed PMID: 10517147.

184. Shi SD, Hemling ME, Carr SA, Horn DM, Lindh I, McLafferty FW. Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal Chem.* 2001;73(1):19-22. Epub 2001/02/24. PubMed PMID: 11195502.
185. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A.* 2004;101(26):9528-33. Epub 2004/06/24. doi: 10.1073/pnas.0402700101. PubMed PMID: 15210983; PMCID: PMC470779.
186. Mikesch LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta.* 2006;1764(12):1811-22. Epub 2006/11/23. doi: 10.1016/j.bbapap.2006.10.003. PubMed PMID: 17118725; PMCID: PMC1853258.
187. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JE, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A.* 2007;104(7):2193-8. Epub 2007/02/09. doi: 10.1073/pnas.0607084104. PubMed PMID: 17287358; PMCID: PMC1892997.
188. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A.* 2007;104(7):2199-204. Epub 2007/02/09. doi: 10.1073/pnas.0611217104. PubMed PMID: 17287340; PMCID: PMC1794346.
189. Khidekel N, Ficarro SB, Clark PM, Bryan MC, Swaney DL, Rexach JE, Sun YE, Coon JJ, Peters EC, Hsieh-Wilson LC. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol.* 2007;3(6):339-48. Epub 2007/05/15. doi: 10.1038/nchembio881. PubMed PMID: 17496889.
190. Swaney DL, McAlister GC, Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods.* 2008;5(11):959-64. Epub 2008/10/22. doi: 10.1038/nmeth.1260. PubMed PMID: 18931669; PMCID: PMC2597439.
191. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JE, Zabrouskov V, Coon JJ. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J Proteome Res.* 2008;7(8):3127-36. Epub 2008/07/11. doi: 10.1021/pr800264t. PubMed PMID: 18613715; PMCID: PMC2601597.
192. Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, Garcia BA. High throughput characterization of combinatorial histone codes. *Mol Cell Proteomics.* 2009;8(10):2266-84. Epub 2009/08/06. doi: 10.1074/mcp.M900238-MCP200. PubMed PMID: 19654425; PMCID: PMC2758755.
193. Molden RC, Garcia BA. Middle-Down and Top-Down Mass Spectrometric Analysis of Co-occurring Histone Modifications. *Curr Protoc Protein Sci.* 2014;77:23.7.1-7.8. Epub 2014/08/02. doi: 10.1002/0471140864.ps2307s77. PubMed PMID: 25081742; PMCID: PMC4160057.
194. Sidoli S, Garcia BA. Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert review of proteomics.* 2017;14(7):617-26. Epub 2017/06/28. doi: 10.1080/14789450.2017.1345632. PubMed PMID: 28649883.
195. Wang H, Straubinger RM, Aletta JM, Cao J, Duan X, Yu H, Qu J. Accurate localization and relative quantification of arginine methylation using nanoflow liquid chromatography coupled to electron transfer dissociation and orbitrap mass spectrometry. *J Am Soc Mass*

- Spectrom. 2009;20(3):507-19. Epub 2008/12/27. doi: 10.1016/j.jasms.2008.11.008. PubMed PMID: 19110445; PMCID: PMC3351756.
196. Earley L, Anderson LC, Bai DL, Mullen C, Syka JE, English AM, Dunyach JJ, Stafford GC, Jr., Shabanowitz J, Hunt DF, Compton PD. Front-end electron transfer dissociation: a new ionization source. *Anal Chem.* 2013;85(17):8385-90. Epub 2013/08/06. doi: 10.1021/ac401783f. PubMed PMID: 23909443; PMCID: PMC3822909.
197. Diner I, Hales CM, Rabenold L, Bishof I, Duong DM, Yi H, Laur O, Gearing M, Troncoso J, Thambisetty M, Lah JJ, Levey AI, Seyfried NT. Aggregation Properties of the Small Nuclear Ribonucleoprotein U1-70K in Alzheimer Disease. *J Biol Chem.* 2014. Epub 2014/10/31. doi: 10.1074/jbc.M114.562959. PubMed PMID: 25355317.
198. O. XG. Small Scale Nuclear Protein Extraction Without the Use of a Detergent 2005 [updated August 2005; cited 2019]. Available from: <https://gozani-lab-website.github.io/website/protocols.html>.
199. Swaney DL, McAlister GC, Coon JJ. Decision tree–driven tandem mass spectrometry for shotgun proteomics. *Nat Methods.* 2008;5:959. doi: 10.1038/nmeth.1260
<https://www.nature.com/articles/nmeth.1260#supplementary-information>.
200. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research.* 2011;10(4):1794-805. Epub 2011/01/25. doi: 10.1021/pr101065j. PubMed PMID: 21254760.
201. Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, Griffin TJ. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics.* 2013;13(8):1352-7. Epub 2013/02/16. doi: 10.1002/pmic.201200352. PubMed PMID: 23412978; PMCID: PMC3633484.
202. Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, Griffin TJ. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics.* 2013;13(8):1352-7. Epub 2013/03/15. doi: 10.1002/pmic.201200352. PubMed PMID: 23412978.
203. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 2015;43(Database issue):D512-20. Epub 2014/12/18. doi: 10.1093/nar/gku1267. PubMed PMID: 25514926; PMCID: PMC4383998.
204. Hales CM, Dammer EB, Diner I, Yi H, Seyfried NT, Gearing M, Glass JD, Montine TJ, Levey AI, Lah JJ. Aggregates of small nuclear ribonucleic acids (snRNAs) in Alzheimer's disease. *Brain Pathol.* 2014;24(4):344-51. Epub 2014/02/28. doi: 10.1111/bpa.12133. PubMed PMID: 24571648; PMCID: PMC4096308.
205. Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, Aderem A, Boyle R, Brunner E, Donohoe S, Fausto N, Hafen E, Hood L, Katze MG, Kennedy KA, Kregenow F, Lee H, Lin B, Martin D, Ranish JA, Rawlings DJ, Samelson LE, Shio Y, Watts JD, Wollscheid B, Wright ME, Yan W, Yang L, Yi EC, Zhang H, Aebersold R. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol.* 2005;6(1):R9. Epub 2005/01/12. doi: 10.1186/gb-2004-6-1-r9. PubMed PMID: 15642101; PMCID: PMC549070.
206. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, Edes J, Loevenich SN, Aebersold R. The PeptideAtlas project. *Nucleic Acids Res.* 2006;34(Database issue):D655-D8. Epub 2005/12/28. doi: 10.1093/nar/gkj040. PubMed PMID: 16381952.

207. 2018 [cited 2018 December 19, 2018]. Available from: https://db.systemsbio.net/sbeams/cgi/PeptideAtlas/GetProtein?atlas_build_id=472&protein_name=P08621&action=QUERY.
208. Chan WY, Chan TW, O'Connor PB. Electron transfer dissociation with supplemental activation to differentiate aspartic and isoaspartic residues in doubly charged peptide cations. *J Am Soc Mass Spectrom.* 2010;21(6):1012-5. Epub 2010/03/23. doi: 10.1016/j.jasms.2010.02.002. PubMed PMID: 20304674; PMCID: PMC3114307.
209. O'Connor PB, Cournoyer JJ, Pitteri SJ, Chrisman PA, McLuckey SA. Differentiation of aspartic and isoaspartic acids using electron transfer dissociation. *J Am Soc Mass Spectrom.* 2006;17(1):15-9. Epub 2005/12/13. doi: 10.1016/j.jasms.2005.08.019. PubMed PMID: 16338146.
210. Zhang X, Herring CJ, Romano PR, Szczepanowska J, Brzeska H, Hinnebusch AG, Qin J. Identification of phosphorylation sites in proteins separated by polyacrylamide gel electrophoresis. *Anal Chem.* 1998;70(10):2050-9. Epub 1998/06/03. PubMed PMID: 9608844.
211. Szczepanowska J, Zhang X, Herring CJ, Qin J, Korn ED, Brzeska H. Identification by mass spectrometry of the phosphorylated residue responsible for activation of the catalytic domain of myosin I heavy chain kinase, a member of the PAK/STE20 family. *Proc Natl Acad Sci U S A.* 1997;94(16):8503-8. Epub 1997/08/05. PubMed PMID: 9238006; PMCID: PMC22975.
212. Gillece-Castro BLS, J. T., editor. Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics. ASMS Conference on Mass Spectrometry and Allied Topics 1995; Atlanta, GA.
213. Haller IM, U. A.; Chait, B. T., editor. Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics. ASMS Conference on Mass Spectrometry and Allied Topics; 1997; Atlanta, GA.
214. Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. *Nature.* 2002;419(6903):182-5. doi: 10.1038/nature01031.
215. Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 2002;12(8):1231-45. Epub 2002/08/15. doi: 10.1101/gr.473902. PubMed PMID: 12176931; PMCID: PMC186633.
216. Trendel J, Schwarzl T, Horos R, Prakash A, Bateman A, Hentze MW, Krijgsveld J. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell.* 2019;176(1-2):391-403.e19. Epub 2018/12/12. doi: 10.1016/j.cell.2018.11.004. PubMed PMID: 30528433.
217. He C, Sidoli S, Warneford-Thomson R, Tatomer DC, Wilusz JE, Garcia BA, Bonasio R. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol Cell.* 2016;64(2):416-30. Epub 2016/10/22. doi: 10.1016/j.molcel.2016.09.034. PubMed PMID: 27768875; PMCID: PMC5222606.
218. Queiroz RML, Smith T, Villanueva E, Marti-Solano M, Monti M, Pizzinga M, Mirea DM, Ramakrishna M, Harvey RF, Dezi V, Thomas GH, Willis AE, Lilley KS. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat Biotechnol.* 2019;37(2):169-78. Epub 2019/01/05. doi: 10.1038/s41587-018-0001-2. PubMed PMID: 30607034; PMCID: PMC6591131.
219. Patel A, Lee HO, Jawerth L, Maharana S, Janel M, Hein MY, Stoykov S, Mahamid J, Saha S, Franzmann TM, Pozniakovski A, Poser I, Maghelli N, Royer LA, Weigert M, Myers EW, Grill S, Drechsel D, Hyman AA, Alberti S. A Liquid-to-Solid Phase Transition of the ALS

- Protein FUS Accelerated by Disease Mutation. *Cell*. 2015;162(5):1066-77. Epub 2015/09/01. doi: 10.1016/j.cell.2015.07.047. PubMed PMID: 26317470.
220. Cho S, Hoang A, Sinha R, Zhong XY, Fu XD, Krainer AR, Ghosh G. 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*. 2011;108(20):8233-8. Epub 2011/05/04. doi: 10.1073/pnas.1017700108. PubMed PMID: 21536904; PMCID: PMC3100968.
221. Bedford MT, Clarke SG. Protein arginine methylation in mammals: who, what, and why. *Mol Cell*. 2009;33(1):1-13. Epub 2009/01/20. doi: 10.1016/j.molcel.2008.12.013. PubMed PMID: 19150423; PMCID: PMC3372459.
222. Bedford MT, Richard S. Arginine methylation an emerging regulator of protein function. *Mol Cell*. 2005;18(3):263-72. Epub 2005/05/04. doi: 10.1016/j.molcel.2005.04.003. PubMed PMID: 15866169.
223. Mai S, Qu X, Li P, Ma Q, Cao C, Liu X. Global regulation of alternative RNA splicing by the SR-rich protein RBM39. *Biochim Biophys Acta*. 2016;1859(8):1014-24. Epub 2016/06/30. doi: 10.1016/j.bbagr.2016.06.007. PubMed PMID: 27354116.
224. Mayeda A, Krainer AR. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*. 1992;68(2):365-75. Epub 1992/01/24. doi: 10.1016/0092-8674(92)90477-t. PubMed PMID: 1531115.
225. Caceres JF, Stamm S, Helfman DM, Krainer AR. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science*. 1994;265(5179):1706-9. Epub 1994/09/16. doi: 10.1126/science.8085156. PubMed PMID: 8085156.
226. Yang X, Bani MR, Lu SJ, Rowan S, Ben-David Y, Chabot B. The A1 and A1B proteins of heterogeneous nuclear ribonucleoproteins modulate 5' splice site selection in vivo. *Proc Natl Acad Sci U S A*. 1994;91(15):6924-8. Epub 1994/07/19. doi: 10.1073/pnas.91.15.6924. PubMed PMID: 8041722; PMCID: PMC44310.
227. Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell*. 2001;8(6):1351-61. Epub 2002/01/10. doi: 10.1016/s1097-2765(01)00409-9. PubMed PMID: 11779509.
228. Konarska MM, Vilardell J, Query CC. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell*. 2006;21(4):543-53. Epub 2006/02/18. doi: 10.1016/j.molcel.2006.01.017. PubMed PMID: 16483935.
229. Caceres JF, Sreaton GR, Krainer AR. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev*. 1998;12(1):55-66. Epub 1998/02/21. doi: 10.1101/gad.12.1.55. PubMed PMID: 9420331; PMCID: PMC316398.
230. Huang Y, Gattoni R, Stevenin J, Steitz JA. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell*. 2003;11(3):837-43. Epub 2003/04/02. doi: 10.1016/s1097-2765(03)00089-3. PubMed PMID: 12667464.
231. Huang Y, Steitz JA. Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell*. 2001;7(4):899-905. Epub 2001/05/05. doi: 10.1016/s1097-2765(01)00233-7. PubMed PMID: 11336712.
232. Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J, Lian LY, Wilson SA, Allain FH. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *Embo j*. 2006;25(21):5126-37. Epub 2006/10/13. doi: 10.1038/sj.emboj.7601385. PubMed PMID: 17036044; PMCID: PMC1630407.

233. Wiese H, Kuhlmann K, Wiese S, Stoepel NS, Pawlas M, Meyer HE, Stephan C, Eisenacher M, Drepper F, Warscheid B. Comparison of alternative MS/MS and bioinformatics approaches for confident phosphorylation site localization. *J Proteome Res.* 2014;13(2):1128-37. Epub 2013/12/25. doi: 10.1021/pr400402s. PubMed PMID: 24364495.
234. Ma CT, Hagopian JC, Ghosh G, Fu XD, Adams JA. Regiospecific phosphorylation control of the SR protein ASF/SF2 by SRPK1. *Journal of molecular biology.* 2009;390(4):618-34. Epub 2009/05/30. doi: 10.1016/j.jmb.2009.05.060. PubMed PMID: 19477182; PMCID: PMC2741142.
235. Xiang S, Gapsys V, Kim HY, Bessonov S, Hsiao HH, Mohlmann S, Klaukien V, Ficner R, Becker S, Urlaub H, Luhrmann R, de Groot B, Zweckstetter M. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure.* 2013;21(12):2162-74. Epub 2013/11/05. doi: 10.1016/j.str.2013.09.014. PubMed PMID: 24183573.
236. Serrano P, Aubol BE, Keshwani MM, Forli S, Ma CT, Dutta SK, Geralt M, Wuthrich K, Adams JA. Directional Phosphorylation and Nuclear Transport of the Splicing Factor SRSF1 Is Regulated by an RNA Recognition Motif. *J Mol Biol.* 2016;428(11):2430-45. Epub 2016/04/20. doi: 10.1016/j.jmb.2016.04.009. PubMed PMID: 27091468; PMCID: PMC4884534.
237. Long Y, Sou WH, Yung KKY, Liu H, Wan SWC, Li Q, Zeng C, Law COK, Chan GHC, Lau TCK, Ngo JCK. Distinct mechanisms govern the phosphorylation of different SR protein splicing factors. *J Biol Chem.* 2019;294(4):1312-27. Epub 2018/11/28. doi: 10.1074/jbc.RA118.003392. PubMed PMID: 30478176; PMCID: PMC6349114.
238. Aubol BE, Serrano P, Fattet L, Wuthrich K, Adams JA. Molecular Interactions Connecting the Function of the Serine-Arginine-rich protein SRSF1 to Protein Phosphatase 1. *J Biol Chem.* 2018. Epub 2018/09/07. doi: 10.1074/jbc.RA118.004587. PubMed PMID: 30185622.
239. Lischwe MA, Cook RG, Ahn YS, Yeoman LC, Busch H. Clustering of glycine and NG,NG-dimethylarginine in nucleolar protein C23. *Biochemistry.* 1985;24(22):6025-8. Epub 1985/10/22. doi: 10.1021/bi00343a001. PubMed PMID: 4084504.
240. Corley SM, Gready JE. Identification of the RGG box motif in Shadoo: RNA-binding and signaling roles? *Bioinform Biol Insights.* 2008;2:383-400. Epub 2008/01/01. PubMed PMID: 19812790; PMCID: PMC2735946.
241. Amanchy R, Periaswamy B, Mathivanan S, Reddy R, Tattikota SG, Pandey A. A curated compendium of phosphorylation motifs. *Nat Biotechnol.* 2007;25(3):285-6. Epub 2007/03/09. doi: 10.1038/nbt0307-285. PubMed PMID: 17344875.
242. Aubol BE, Hailey KL, Fattet L, Jennings PA, Adams JA. Redirecting SR Protein Nuclear Trafficking through an Allosteric Platform. *Journal of molecular biology.* 2017;429(14):2178-91. Epub 2017/06/04. doi: 10.1016/j.jmb.2017.05.022. PubMed PMID: 28576472; PMCID: PMC5536852.
243. Ngo JC, Chakrabarti S, Ding JH, Velazquez-Dones A, Nolen B, Aubol BE, Adams JA, Fu XD, Ghosh G. Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol Cell.* 2005;20(1):77-89. Epub 2005/10/08. doi: 10.1016/j.molcel.2005.08.025. PubMed PMID: 16209947.
244. Keshwani MM, Aubol BE, Fattet L, Ma CT, Qiu J, Jennings PA, Fu XD, Adams JA. Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *Biochem J.* 2015;466(2):311-22. Epub 2014/12/23. doi: 10.1042/bj20141373. PubMed PMID: 25529026; PMCID: PMC5053020.

245. Colwill K, Feng LL, Yeakley JM, Gish GD, Caceres JF, Pawson T, Fu XD. SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *J Biol Chem*. 1996;271(40):24569-75. Epub 1996/10/04. PubMed PMID: 8798720.
246. Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, Bergeron J, Borchers CH, Corthals GL, Costello CE, Deutsch EW, Domon B, Hancock W, He F, Hochstrasser D, Marko-Varga G, Salekdeh GH, Sechi S, Snyder M, Srivastava S, Uhlen M, Wu CH, Yamamoto T, Paik YK, Omenn GS. The human proteome project: current state and future direction. *Mol Cell Proteomics*. 2011;10(7):M111.009993. Epub 2011/07/12. doi: 10.1074/mcp.M111.009993. PubMed PMID: 21742803; PMCID: PMC3134076.
247. Garcia BA, Mollah S, Ueberheide BM, Busby SA, Muratore TL, Shabanowitz J, Hunt DF. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc*. 2007;2(4):933-8. doi: 10.1038/nprot.2007.106. PubMed PMID: 17446892.
248. Boisvert FM, Chenard CA, Richard S. Protein interfaces in signaling regulated by arginine methylation. *Sci STKE*. 2005;2005(271):re2. Epub 2005/02/17. doi: 10.1126/stke.2712005re2. PubMed PMID: 15713950.
249. Meister G, Eggert C, Buhler D, Brahms H, Kambach C, Fischer U. Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Current biology : CB*. 2001;11(24):1990-4. Epub 2001/12/19. PubMed PMID: 11747828.
250. Musiani D, Bok J, Massignani E, Wu L, Tabaglio T, Ippolito MR, Cuomo A, Ozbek U, Zorgati H, Ghoshdastider U, Robinson RC, Guccione E, Bonaldi T. Proteomics profiling of arginine methylation defines PRMT5 substrate specificity. *Sci Signal*. 2019;12(575). Epub 2019/04/04. doi: 10.1126/scisignal.aat8388. PubMed PMID: 30940768.
251. Zhang X, Gan L, Pan H, Guo S, He X, Olson ST, Mesecar A, Adam S, Unterman TG. Phosphorylation of serine 256 suppresses transactivation by FKHR (FOXO1) by multiple mechanisms. Direct and indirect effects on nuclear/cytoplasmic shuttling and DNA binding. *J Biol Chem*. 2002;277(47):45276-84. Epub 2002/09/14. doi: 10.1074/jbc.M208063200. PubMed PMID: 12228231.
252. Sims RJ, 3rd, Rojas LA, Beck DB, Bonasio R, Schuller R, Drury WJ, 3rd, Eick D, Reinberg D. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science (New York, NY)*. 2011;332(6025):99-103. Epub 2011/04/02. doi: 10.1126/science.1202663. PubMed PMID: 21454787; PMCID: PMC3773223.
253. Zhao DY, Gish G, Braunschweig U, Li Y, Ni Z, Schmitges FW, Zhong G, Liu K, Li W, Moffat J, Vedadi M, Min J, Pawson TJ, Blencowe BJ, Greenblatt JF. SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*. 2016;529(7584):48-53. Epub 2015/12/25. doi: 10.1038/nature16469. PubMed PMID: 26700805.
254. Fong JY, Pignata L, Goy PA, Kawabata KC, Lee SC, Koh CM, Musiani D, Massignani E, Kotini AG, Penson A, Wun CM, Shen Y, Schwarz M, Low DH, Rialdi A, Ki M, Wollmann H, Mzoughi S, Gay F, Thompson C, Hart T, Barbash O, Luciani GM, Szewczyk MM, Wouters BJ, Delwel R, Papapetrou EP, Baryte-Lovejoy D, Arrowsmith CH, Minden MD, Jin J, Melnick A, Bonaldi T, Abdel-Wahab O, Guccione E. Therapeutic Targeting of RNA Splicing Catalysis through Inhibition of Protein Arginine Methylation. *Cancer Cell*. 2019;36(2):194-209 e9. Epub 2019/08/14. doi: 10.1016/j.ccell.2019.07.003. PubMed PMID: 31408619.
255. Yang Y, Hadjikyriacou A, Xia Z, Gayatri S, Kim D, Zurita-Lopez C, Kelly R, Guo A, Li W, Clarke SG, Bedford MT. PRMT9 is a type II methyltransferase that methylates the splicing factor SAP145. *Nature communications*. 2015;6:6428. Epub 2015/03/05. doi: 10.1038/ncomms7428. PubMed PMID: 25737013; PMCID: PMC4351962.

256. Lamond AI, Spector DL. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol.* 2003;4(8):605-12. Epub 2003/08/19. doi: 10.1038/nrm1172. PubMed PMID: 12923522.
257. Charenton C, Wilkinson ME, Nagai K. Mechanism of 5' splice site transfer for human spliceosome activation. *Science.* 2019;364(6438):362-7. Epub 2019/04/13. doi: 10.1126/science.aax3289. PubMed PMID: 30975767; PMCID: PMC6525098.
258. Maertens GN, Cook NJ, Wang W, Hare S, Gupta SS, Oztop I, Lee K, Pye VE, Cosnefroy O, Snijders AP, KewalRamani VN, Fassati A, Engelman A, Cherepanov P. Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc Natl Acad Sci U S A.* 2014;111(7):2728-33. Epub 2014/01/23. doi: 10.1073/pnas.1320755111. PubMed PMID: 24449914; PMCID: PMC3932936.
259. Daubner GM, Clery A, Jayne S, Stevenin J, Allain FH. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *Embo j.* 2012;31(1):162-74. Epub 2011/10/18. doi: 10.1038/emboj.2011.367. PubMed PMID: 22002536; PMCID: PMC3252578.
260. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods.* 2017;14(5):513-20. Epub 2017/04/11. doi: 10.1038/nmeth.4256. PubMed PMID: 28394336; PMCID: PMC5409104.
261. Devabhaktuni A, Lin S, Zhang L, Swaminathan K, Gonzalez CG, Olsson N, Pearlman SM, Rawson K, Elias JE. TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat Biotechnol.* 2019;37(4):469-79. Epub 2019/04/03. doi: 10.1038/s41587-019-0067-5. PubMed PMID: 30936560; PMCID: PMC6447449.
262. Frese CK, Altelaar AF, van den Toorn H, Nolting D, Griep-Raming J, Heck AJ, Mohammed S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem.* 2012;84(22):9668-73. Epub 2012/10/31. doi: 10.1021/ac3025366. PubMed PMID: 23106539.
263. Frese CK, Zhou H, Taus T, Altelaar AF, Mechtler K, Heck AJ, Mohammed S. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res.* 2013;12(3):1520-5. Epub 2013/01/26. doi: 10.1021/pr301130k. PubMed PMID: 23347405; PMCID: PMC3588588.
264. Gebauer F, Schwarzl T, Valcárcel J, Hentze MW. RNA-binding proteins in human genetic disease. *Nature Reviews Genetics.* 2021;22(3):185-98. doi: 10.1038/s41576-020-00302-y.
265. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 2002;11(4):739-56. Epub 2002/03/23. doi: 10.1110/ps.4210102. PubMed PMID: 11910019; PMCID: PMC2373528.
266. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell.* 2012;149(4):753-67. Epub 2012/05/15. doi: 10.1016/j.cell.2012.04.017. PubMed PMID: 22579281.
267. St George-Hyslop P, Lin JQ, Miyashita A, Phillips EC, Qamar S, Randle SJ, Wang G. 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*. 2018;1693:11-23. doi: <https://doi.org/10.1016/j.brainres.2018.04.036>.
268. Schwartz Jacob C, Wang X, Podell Elaine R, Cech Thomas R. RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Reports*. 2013;5(4):918-25. doi: <https://doi.org/10.1016/j.celrep.2013.11.017>.
269. Neumann M, Rademakers R, Roeber S, Baker M, Kretzschmar HA, Mackenzie IR. A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain : a journal of neurology*. 2009;132(Pt 11):2922-31. Epub 2009/08/14. doi: 10.1093/brain/awp214. PubMed PMID: 19674978; PMCID: PMC2768659.
270. Liu-Yesucevitz L, Lin AY, Ebata A, Boon JY, Reid W, Xu YF, Kobrin K, Murphy GJ, Petrucelli L, Wolozin B. ALS-linked mutations enlarge TDP-43-enriched neuronal RNA granules in the dendritic arbor. *J Neurosci*. 2014;34(12):4167-74. Epub 2014/03/22. doi: 10.1523/jneurosci.2350-13.2014. PubMed PMID: 24647938; PMCID: PMC3960463.
271. Johnson ECB, Carter EK, Dammer EB, Duong DM, Gerasimov ES, Liu Y, Liu J, Betarbet R, Ping L, Yin L, Serrano GE, Beach TG, Peng J, De Jager PL, Gaiteri C, Bennett DA, Gearing M, Wingo TS, Wingo AP, Lah JJ, Levey AI, Seyfried NT. Large-Scale Deep Multi-Layer Analysis of Alzheimer's Disease Brain Reveals Strong Proteomic Disease-Related Changes Not Observed at the RNA Level. *bioRxiv*. 2021:2021.04.05.438450. doi: 10.1101/2021.04.05.438450.
272. Furukawa Y, Kaneko K, Matsumoto G, Kurosawa M, Nukina N. Cross-seeding fibrillation of Q/N-rich proteins offers new pathomechanism of polyglutamine diseases. *J Neurosci*. 2009;29(16):5153-62. Epub 2009/04/24. doi: 10.1523/jneurosci.0783-09.2009. PubMed PMID: 19386911; PMCID: PMC6665470.
273. Waelter S, Boeddrich A, Lurz R, Scherzinger E, Lueder G, Lehrach H, Wanker EE. Accumulation of mutant huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation. *Mol Biol Cell*. 2001;12(5):1393-407. Epub 2001/05/22. doi: 10.1091/mbc.12.5.1393. PubMed PMID: 11359930; PMCID: PMC34592.
274. Liu Q, Dreyfuss G. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol*. 1995;15(5):2800-8. Epub 1995/05/01. doi: 10.1128/mcb.15.5.2800. PubMed PMID: 7739561; PMCID: PMC230511.
275. Hasegawa M, Arai T, Nonaka T, Kametani F, Yoshida M, Hashizume Y, Beach TG, Buratti E, Baralle F, Morita M, Nakano I, Oda T, Tsuchiya K, Akiyama H. Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann Neurol*. 2008;64(1):60-70. Epub 2008/06/12. doi: 10.1002/ana.21425. PubMed PMID: 18546284; PMCID: PMC2674108.
276. Neumann M, Kwong LK, Lee EB, Kremmer E, Flatley A, Xu Y, Forman MS, Troost D, Kretzschmar HA, Trojanowski JQ, Lee VM. Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. *Acta Neuropathol*. 2009;117(2):137-49. Epub 2009/01/07. doi: 10.1007/s00401-008-0477-9. PubMed PMID: 19125255; PMCID: PMC2693625.
277. Kundinger SR, Bishof I, Dammer EB, Duong DM, Seyfried NT. Middle-Down Proteomics Reveals Dense Sites of Methylation and Phosphorylation in Arginine-Rich RNA-Binding Proteins. *J Proteome Res*. 2020. Epub 2020/01/30. doi: 10.1021/acs.jproteome.9b00633. PubMed PMID: 31994892.

278. Tacke R, Manley JL. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO journal*. 1995;14(14):3540-51. Epub 1995/07/17. PubMed PMID: 7543047; PMCID: PMC394422.
279. Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *Embo j*. 1996;15(2):265-75. Epub 1996/01/15. PubMed PMID: 8617202; PMCID: PMC449941.
280. Wang HY, Lin W, Dyck JA, Yeakley JM, Songyang Z, Cantley LC, Fu XD. SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol*. 1998;140(4):737-50. Epub 1998/03/21. doi: 10.1083/jcb.140.4.737. PubMed PMID: 9472028; PMCID: PMC2141757.
281. Kannan N, Neuwald AF. Evolutionary constraints associated with functional specificity of the CMGC protein kinases MAPK, CDK, GSK, SRPK, DYRK, and CK2alpha. *Protein Sci*. 2004;13(8):2059-77. Epub 2004/07/27. doi: 10.1110/ps.04637904. PubMed PMID: 15273306; PMCID: PMC2279817.
282. Misteli T, Cáceres JF, Clement JQ, Krainer AR, Wilkinson MF, Spector DL. Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *The Journal of cell biology*. 1998;143(2):297-307. Epub 1998/10/24. doi: 10.1083/jcb.143.2.297. PubMed PMID: 9786943; PMCID: PMC2132840.
283. Misteli T, Cáceres JF, Spector DL. The dynamics of a pre-mRNA splicing factor in living cells. *Nature*. 1997;387(6632):523-7. Epub 1997/05/29. doi: 10.1038/387523a0. PubMed PMID: 9168118.
284. Yeakley JM, Tronchère H, Olesen J, Dyck JA, Wang HY, Fu XD. Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *The Journal of cell biology*. 1999;145(3):447-55. Epub 1999/05/04. doi: 10.1083/jcb.145.3.447. PubMed PMID: 10225947; PMCID: PMC2185075.
285. Aubol BE, Serrano P, Fattet L, Wuthrich K, Adams JA. Molecular interactions connecting the function of the serine-arginine-rich protein SRSF1 to protein phosphatase 1. *J Biol Chem*. 2018;293(43):16751-60. Epub 2018/09/07. doi: 10.1074/jbc.RA118.004587. PubMed PMID: 30185622; PMCID: PMC6204901.
286. Aubol BE, Keshwani MM, Fattet L, Adams JA. Mobilization of a splicing factor through a nuclear kinase-kinase complex. *Biochem J*. 2018;475(3):677-90. Epub 2018/01/18. doi: 10.1042/bcj20170672. PubMed PMID: 29335301; PMCID: PMC6293969.
287. Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV, Jedd G. Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol Cell*. 2020. Epub 2020/02/13. doi: 10.1016/j.molcel.2020.01.025. PubMed PMID: 32048997.
288. Peng TY, Lee KR, Tarn WY. 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*. 2008;275(16):4152-63. Epub 2008/07/18. doi: 10.1111/j.1742-4658.2008.06564.x. PubMed PMID: 18631359; PMCID: PMC7164085.
289. Nikolakaki E, Drosou V, Sanidas I, Peidis P, Papamarcaki T, Iakoucheva LM, Giannakouros T. RNA association or phosphorylation of the RS domain prevents aggregation of RS domain-containing proteins. *Biochim Biophys Acta*. 2008;1780(2):214-25. Epub 2007/11/21. doi: 10.1016/j.bbagen.2007.10.014. PubMed PMID: 18022399.

290. Robison R. The Possible Significance of Hexosephosphoric Esters in Ossification. *The Biochemical journal*. 1923;17(2):286-93. Epub 1923/01/01. doi: 10.1042/bj0170286. PubMed PMID: 16743183; PMCID: PMC1259346.
291. Shi J, Hu Z, Pabon K, Scotto KW. Caffeine regulates alternative splicing in a subset of cancer-associated genes: a role for SC35. *Mol Cell Biol*. 2008;28(2):883-95. Epub 2007/11/21. doi: 10.1128/mcb.01345-07. PubMed PMID: 18025108; PMCID: PMC2223418.
292. Cooper JA, Sefton BM, Hunter T. Detection and quantification of phosphotyrosine in proteins. *Methods Enzymol*. 1983;99:387-402. Epub 1983/01/01. doi: 10.1016/0076-6879(83)99075-4. PubMed PMID: 6196603.
293. Heaton BE, Trimarco JD, Hamele CE, Harding AT, Tata A, Zhu X, Tata PR, Smith CM, Heaton NS. SRSF protein kinases 1 and 2 are essential host factors for human coronaviruses including SARS-CoV-2. *bioRxiv*. 2020. Epub 2020/08/21. doi: 10.1101/2020.08.14.251207. PubMed PMID: 32817937; PMCID: PMC7430567.
294. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-72. Epub 2008/11/26. doi: 10.1038/nbt.1511. PubMed PMID: 19029910.
295. Zambon AC, Gaj S, Ho I, Hanspers K, Vranizan K, Evelo CT, Conklin BR, Pico AR, Salomonis N. GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics*. 2012;28(16):2209-10. Epub 2012/06/30. doi: 10.1093/bioinformatics/bts366. PubMed PMID: 22743224; PMCID: PMC3413395.
296. Seyfried NT, Gozal YM, Dammer EB, Xia Q, Duong DM, Cheng D, Lah JJ, Levey AI, Peng J. Multiplex SILAC analysis of a cellular TDP-43 proteinopathy model reveals protein inclusions associated with SUMOylation and diverse polyubiquitin chains. *Mol Cell Proteomics*. 2010;9(4):705-18. Epub 2010/01/06. doi: 10.1074/mcp.M800390-MCP200. PubMed PMID: 20047951; PMCID: PMC2860236.
297. Ghalei H, Schaub FX, Doherty JR, Noguchi Y, Roush WR, Cleveland JL, Stroupe ME, Karbstein K. Hrr25/CK1 δ -directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *Journal of Cell Biology*. 2015;208(6):745-59. doi: 10.1083/jcb.201409056.
298. Dammer EB, Fallini C, Gozal YM, Duong DM, Rossoll W, Xu P, Lah JJ, Levey AI, Peng J, Bassell GJ, Seyfried NT. 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*. 2012;7(6):e38658. Epub 2012/07/05. doi: 10.1371/journal.pone.0038658. PubMed PMID: 22761693; PMCID: PMC3380899.
299. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, 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 AF, Ternent T, Brazma A, Vizcaíno JA. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*. 2019;47(D1):D442-d50. Epub 2018/11/06. doi: 10.1093/nar/gky1106. PubMed PMID: 30395289; PMCID: PMC6323896.
300. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015;31(6):857-63. Epub 2014/11/14. doi: 10.1093/bioinformatics/btu744. PubMed PMID: 25391399; PMCID: PMC4380029.
301. Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res*. 2006;34(1):305-12. Epub 2006/01/13. doi: 10.1093/nar/gkj424. PubMed PMID: 16407336; PMCID: PMC1326245.

302. Aubol BE, Wu G, Keshwani MM, Movassat M, Fattet L, Hertel KJ, Fu XD, Adams JA. Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol Cell*. 2016;63(2):218-28. Epub 2016/07/12. doi: 10.1016/j.molcel.2016.05.034. PubMed PMID: 27397683; PMCID: PMC4941815.
303. Saitoh N, Sakamoto C, Hagiwara M, Agredano-Moreno LT, Jiménez-García LF, Nakao M. The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. *Molecular Biology of the Cell*. 2012;23(6):1115-28. doi: 10.1091/mbc.e11-09-0783. PubMed PMID: 22262462.
304. Cavaloc Y, Bourgeois CF, Kister L, Stévenin J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *Rna*. 1999;5(3):468-83. Epub 1999/03/27. doi: 10.1017/s1355838299981967. PubMed PMID: 10094314; PMCID: PMC1369774.
305. Kadri F, Pacifici M, Wilk A, Parker-Struckhoff A, Del Valle L, Hauser KF, Knapp PE, Parsons C, Jeansonne D, Lassak A, Peruzzi F. HIV-1-Tat Protein Inhibits SC35-mediated Tau Exon 10 Inclusion through Up-regulation of DYRK1A Kinase. *J Biol Chem*. 2015;290(52):30931-46. Epub 2015/11/05. doi: 10.1074/jbc.M115.675751. PubMed PMID: 26534959; PMCID: PMC4692221.
306. Kundinger SR, Bishof I, Dammer EB, Duong DM, Seyfried NT. Middle-Down Proteomics Reveals Dense Sites of Methylation and Phosphorylation in Arginine-Rich RNA-Binding Proteins. *Journal of proteome research*. 2020;19(4):1574-91. Epub 2020/01/30. doi: 10.1021/acs.jproteome.9b00633. PubMed PMID: 31994892.
307. Wittig I, Braun HP, Schägger H. Blue native PAGE. *Nat Protoc*. 2006;1(1):418-28. Epub 2007/04/05. doi: 10.1038/nprot.2006.62. PubMed PMID: 17406264.
308. Enami M, Ishihama A. Protein phosphorylation in *Escherichia coli* and purification of a protein kinase. *J Biol Chem*. 1984;259(1):526-33. Epub 1984/01/10. PubMed PMID: 6368541.
309. Macek B, Gnad F, Soufi B, Kumar C, Olsen JV, Mijakovic I, Mann M. Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol Cell Proteomics*. 2008;7(2):299-307. Epub 2007/10/17. doi: 10.1074/mcp.M700311-MCP200. PubMed PMID: 17938405.
310. Fukuhara T, Hosoya T, Shimizu S, Sumi K, Oshiro T, Yoshinaka Y, Suzuki M, Yamamoto N, Herzenberg LA, Herzenberg LA, Hagiwara M. Utilization of host SR protein kinases and RNA-splicing machinery during viral replication. *Proc Natl Acad Sci U S A*. 2006;103(30):11329-33. Epub 2006/07/15. doi: 10.1073/pnas.0604616103. PubMed PMID: 16840555; PMCID: PMC1544086.
311. Hatcher JM, Wu G, Zeng C, Zhu J, Meng F, Patel S, Wang W, Ficarro SB, Leggett AL, Powell CE, Marto JA, Zhang K, Ki Ngo JC, Fu XD, Zhang T, Gray NS. SRPKIN-1: A Covalent SRPK1/2 Inhibitor that Potently Converts VEGF from Pro-angiogenic to Anti-angiogenic Isoform. *Cell Chem Biol*. 2018;25(4):460-70.e6. Epub 2018/02/27. doi: 10.1016/j.chembiol.2018.01.013. PubMed PMID: 29478907; PMCID: PMC5973797.
312. Conlon EG, Manley JL. RNA-binding proteins in neurodegeneration: mechanisms in aggregate. *Genes Dev*. 2017;31(15):1509-28. Epub 2017/09/16. doi: 10.1101/gad.304055.117. PubMed PMID: 28912172; PMCID: PMC5630017.
313. Gammons MV, Dick AD, Harper SJ, Bates DO. SRPK1 inhibition modulates VEGF splicing to reduce pathological neovascularization in a rat model of retinopathy of prematurity. *Invest Ophthalmol Vis Sci*. 2013;54(8):5797-806. Epub 2013/06/14. doi: 10.1167/iovs.13-11634. PubMed PMID: 23761094; PMCID: PMC6485497.

314. Siqueira RP, Barbosa Éde A, Polêto MD, Righetto GL, Seraphim TV, Salgado RL, Ferreira JG, Barros MV, de Oliveira LL, Laranjeira AB, Almeida MR, Júnior AS, Fietto JL, Kobarg J, de Oliveira EB, Teixeira RR, Borges JC, Yunes JA, Bressan GC. Potential Antileukemia Effect and Structural Analyses of SRPK Inhibition by N-(2-(Piperidin-1-yl)-5-(Trifluoromethyl)Phenyl)Isonicotinamide (SRPIN340). *PLoS One*. 2015;10(8):e0134882. Epub 2015/08/06. doi: 10.1371/journal.pone.0134882. PubMed PMID: 26244849; PMCID: PMC4526641.
315. 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 WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64-9. Epub 2011/09/13. doi: 10.1038/nature10496. PubMed PMID: 21909114.
316. 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 HP, Klein HU, Dugas M, Aburatani H, Kohlmann A, Miyano S, Haferlach C, Kern W, Ogawa S. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28(2):241-7. Epub 2013/11/14. doi: 10.1038/leu.2013.336. PubMed PMID: 24220272; PMCID: PMC3918868.
317. Thol F, Kade S, Schlarman C, Löffel P, Morgan M, Krauter J, Wlodarski MW, Kölking B, Wichmann M, Görlich K, Göhring G, Bug G, Ottmann O, Niemeyer CM, Hofmann WK, Schlegelberger B, Ganser A, Heuser M. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 2012;119(15):3578-84. Epub 2012/03/06. doi: 10.1182/blood-2011-12-399337. PubMed PMID: 22389253.
318. Wu SJ, Kuo YY, Hou HA, Li LY, Tseng MH, Huang CF, Lee FY, Liu MC, Liu CW, Lin CT, Chen CY, Chou WC, Yao M, Huang SY, Ko BS, Tang JL, Tsay W, Tien HF. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood*. 2012;120(15):3106-11. Epub 2012/08/31. doi: 10.1182/blood-2012-02-412296. PubMed PMID: 22932795.
319. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, Yoon CJ, Ellis P, Wedge DC, Pellagatti A, Shlien A, Groves MJ, Forbes SA, Raine K, Hinton J, Mudie LJ, McLaren S, Hardy C, Latimer C, Della Porta MG, O'Meara S, Ambaglio I, Galli A, Butler AP, Walldin G, Teague JW, Quek L, Sternberg A, Gambacorti-Passerini C, Cross NC, Green AR, Boulwood J, Vyas P, Hellstrom-Lindberg E, Bowen D, Cazzola M, Stratton MR, Campbell PJ. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-27; quiz 99. Epub 2013/09/14. doi: 10.1182/blood-2013-08-518886. PubMed PMID: 24030381; PMCID: PMC3837510.
320. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, Kim J, Yun J, Xie Y, McKnight SL. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science (New York, NY)*. 2014;345(6201):1139-45. Epub 2014/08/02. doi: 10.1126/science.1254917. PubMed PMID: 25081482; PMCID: PMC4459787.
321. Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, Spille JH, Afeyan LK, Zamudio AV, Shrinivas K, Abraham BJ, Boija A, Decker TM, Rimel JK, Fant CB, Lee TI, Cisse, II, Sharp PA, Taatjes DJ, Young RA. Pol II phosphorylation regulates a switch

- between transcriptional and splicing condensates. *Nature*. 2019;572(7770):543-8. Epub 2019/08/09. doi: 10.1038/s41586-019-1464-0. PubMed PMID: 31391587; PMCID: PMC6706314.
322. Lester E, Ooi FK, Bakkar N, Ayers J, Woerman AL, Wheeler J, Bowser R, Carlson GA, Prusiner SB, Parker R. Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. *Neuron*. 2021. Epub 2021/04/14. doi: 10.1016/j.neuron.2021.03.026. PubMed PMID: 33848474.
323. Drummond E, Pires G, MacMurray C, Askenazi M, Nayak S, Bourdon M, Safar J, Ueberheide B, Wisniewski T. Phosphorylated tau interactome in the human Alzheimer's disease brain. *Brain*. 2020;143(9):2803-17. doi: 10.1093/brain/awaa223.
324. Guo Q, Dammer EB, Zhou M, Kundinger SR, Gearing M, Lah JJ, Levey AI, Shulman JM, Seyfried NT. 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*. 2021;14:623659. Epub 2021/04/06. doi: 10.3389/fnmol.2021.623659. PubMed PMID: 33815056; PMCID: PMC8014091.
325. Guo X, Trudgian DC, Lemoff A, Yadavalli S, Mirzaei H. Confetti: a multiprotease map of the HeLa proteome for comprehensive proteomics. *Mol Cell Proteomics*. 2014;13(6):1573-84. Epub 2014/04/04. doi: 10.1074/mcp.M113.035170. PubMed PMID: 24696503; PMCID: PMC4047476.
326. Maron MI, Lehman SM, Gayatri S, DeAngelo JD, Hegde S, Lorton BM, Sun Y, Bai DL, Sidoli S, Gupta V, Marunde MR, Bone JR, Sun Z-W, Bedford MT, Shabanowitz J, Chen H, Hunt DF, Shechter D. Independent transcriptomic and proteomic regulation by type I and II protein arginine methyltransferases. *iScience*. 2021;24(9):102971. doi: <https://doi.org/10.1016/j.isci.2021.102971>.
327. Solntsev SK, Shortreed MR, Frey BL, Smith LM. Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *J Proteome Res*. 2018;17(5):1844-51. Epub 2018/03/27. doi: 10.1021/acs.jproteome.7b00873. PubMed PMID: 29578715.
328. Yu F, Teo GC, Kong AT, Haynes SE, Avtonomov DM, Geiszler DJ, Nesvizhskii AI. Identification of modified peptides using localization-aware open search. *Nature Communications*. 2020;11(1):4065. doi: 10.1038/s41467-020-17921-y.
329. Zhang X, Smits AH, van Tilburg GB, Ovaa H, Huber W, Vermeulen M. Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat Protoc*. 2018;13(3):530-50. Epub 2018/02/16. doi: 10.1038/nprot.2017.147. PubMed PMID: 29446774.
330. Bishof I, Dammer EB, Duong DM, Gearing M, Lah JJ, Levey AI, Seyfried NT. RNA-binding proteins with mixed charge domains self-assemble and aggregate in Alzheimer's Disease. *bioRxiv*. 2018. doi: 10.1101/243014.
331. Matevossian A, Akbarian S. Neuronal nuclei isolation from human postmortem brain tissue. *J Vis Exp*. 2008(20). Epub 2008/12/17. doi: 10.3791/914. PubMed PMID: 19078943; PMCID: PMC3233860.
332. Dammer EB, Duong DM, Diner I, Gearing M, Feng Y, Lah JJ, Levey AI, Seyfried NT. Neuron enriched nuclear proteome isolated from human brain. *Journal of proteome research*. 2013;12(7):3193-206. Epub 2013/06/19. doi: 10.1021/pr400246t. PubMed PMID: 23768213; PMCID: PMC3734798.
333. Aubol BE, Wozniak JM, Fattet L, Gonzalez DJ, Adams JA. CLK1 reorganizes the splicing factor U1-70K for early spliceosomal protein assembly. *Proc Natl Acad Sci U S A*.

- 2021;118(14). Epub 2021/04/04. doi: 10.1073/pnas.2018251118. PubMed PMID: 33811140; PMCID: PMC8040622.
334. Hamey JJ, Nguyen A, Wilkins MR. Discovery of Arginine Methylation, Phosphorylation, and Their Co-occurrence in Condensate-Associated Proteins in *Saccharomyces cerevisiae*. *J Proteome Res.* 2021;20(5):2420-34. Epub 2021/04/16. doi: 10.1021/acs.jproteome.0c00927. PubMed PMID: 33856219.
335. Giuliani V, Miller MA, Liu C-Y, Hartono SR, Class CA, Bristow CA, Suzuki E, Sanz LA, Gao G, Gay JP, Feng N, Rose JL, Tomihara H, Daniele JR, Peoples MD, Bardenhagen JP, Geck Do MK, Chang QE, Vangamudi B, Vellano C, Ying H, Deem AK, Do K-A, Genovese G, Marszalek JR, Kovacs JJ, Kim M, Fleming JB, Guccione E, Viale A, Maitra A, Emilia Di Francesco M, Yap TA, Jones P, Draetta G, Carugo A, Chedin F, Heffernan TP. PRMT1-dependent regulation of RNA metabolism and DNA damage response sustains pancreatic ductal adenocarcinoma. *Nature communications.* 2021;12(1):4626. doi: 10.1038/s41467-021-24798-y.
336. Pahllich S, Zakaryan RP, Gehring H. Protein arginine methylation: Cellular functions and methods of analysis. *Biochimica et biophysica acta.* 2006;1764(12):1890-903. Epub 2006/10/03. doi: 10.1016/j.bbapap.2006.08.008. PubMed PMID: 17010682.
337. Ferreira TR, Dowle AA, Parry E, Alves-Ferreira EVC, Hogg K, Kolokousi F, Larson TR, Plevin MJ, Cruz AK, Walrad PB. PRMT7 regulates RNA-binding capacity and protein stability in *Leishmania* parasites. *Nucleic Acids Research.* 2020;48(10):5511-26. doi: 10.1093/nar/gkaa211.
338. Haass C, Selkoe DJ. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β -peptide. *Nature Reviews Molecular Cell Biology.* 2007;8(2):101-12. doi: 10.1038/nrm2101.
339. Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo GW, Ares M, Jr., Fu XD. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell.* 2013;50(2):223-35. Epub 2013/04/09. doi: 10.1016/j.molcel.2013.03.001. PubMed PMID: 23562324; PMCID: PMC3640356.
340. Shiozawa Y, Malcovati L, Gallì A, Sato-Otsubo A, Kataoka K, Sato Y, Watatani Y, Suzuki H, Yoshizato T, Yoshida K, Sanada M, Makishima H, Shiraishi Y, Chiba K, Hellström-Lindberg E, Miyano S, Ogawa S, Cazzola M. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nature Communications.* 2018;9(1):3649. doi: 10.1038/s41467-018-06063-x.
341. Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD. The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol.* 2008;15(8):819-26. Epub 2008/07/22. doi: 10.1038/nsmb.1461. PubMed PMID: 18641664; PMCID: PMC2574591.
342. Zhang S, Aibara S, Vos SM, Agafonov DE, Lührmann R, Cramer P. Structure of a transcribing RNA polymerase II and U1 snRNP complex. *Science.* 2021;371(6526):305-9. doi: 10.1126/science.abf1870.
343. Hoermann B, Kokot T, Helm D, Heinzlmeir S, Chojnacki JE, Schubert T, Ludwig C, Berteotti A, Kurzawa N, Kuster B, Savitski MM, Köhn M. Dissecting the sequence determinants for dephosphorylation by the catalytic subunits of phosphatases PP1 and PP2A. *Nature Communications.* 2020;11(1):3583. doi: 10.1038/s41467-020-17334-x.
344. Brautigan DL, Shenolikar S. Protein Serine/Threonine Phosphatases: Keys to Unlocking Regulators and Substrates. *Annu Rev Biochem.* 2018;87:921-64. Epub 2018/06/22. doi: 10.1146/annurev-biochem-062917-012332. PubMed PMID: 29925267.

345. Bollen M, Peti W, Ragusa MJ, Beullens M. The extended PP1 toolkit: designed to create specificity. *Trends Biochem Sci.* 2010;35(8):450-8. Epub 2010/04/20. doi: 10.1016/j.tibs.2010.03.002. PubMed PMID: 20399103; PMCID: PMC3131691.
346. Brautigam DL. Protein Ser/Thr phosphatases--the ugly ducklings of cell signalling. *Febs j.* 2013;280(2):324-45. Epub 2012/04/24. doi: 10.1111/j.1742-4658.2012.08609.x. PubMed PMID: 22519956.
347. Wang M, Beckmann ND, Roussos P, Wang E, Zhou X, Wang Q, Ming C, Neff R, Ma W, Fullard JF, Hauberg ME, Bendl J, Peters MA, Logsdon B, Wang P, Mahajan M, Mangravite LM, Dammer EB, Duong DM, Lah JJ, Seyfried NT, Levey AI, Buxbaum JD, Ehrlich M, Gandy S, Katsel P, Haroutunian V, Schadt E, Zhang B. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Scientific data.* 2018;5(1):180185. doi: 10.1038/sdata.2018.185.
348. Holt LJ, Tuch BB, Villén J, Johnson AD, Gygi SP, Morgan DO. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science.* 2009;325(5948):1682-6. Epub 2009/09/26. doi: 10.1126/science.1172867. PubMed PMID: 19779198; PMCID: PMC2813701.
349. De Munter S, Köhn M, Bollen M. Challenges and opportunities in the development of protein phosphatase-directed therapeutics. *ACS Chem Biol.* 2013;8(1):36-45. Epub 2012/12/12. doi: 10.1021/cb300597g. PubMed PMID: 23214403.