Quantitative investigation of protein-RNA interactions and regulation by phosphorylation

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

eingereicht an der Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von:

M.Sc. Carlos Henrique Vieira e Vieira

Präsidentin der Humboldt-Universität zu Berlin

Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

Prof. Dr. Dr. Christian Ulrichs

Gutachter/innen

1. Prof. Dr. Matthias Selbach

2. Prof. Dr. Florian Heyd

3. Prof. Dr. Markus Landthaler

Tag der mündlichen Prüfung: 08.11.2021

ZUSAMMENFASSUNG

Die posttranskriptionelle Regulierung der Genexpression ist von grundlegender Bedeutung im gesunden und erkrankten Organismus. RNA-bindende Proteine (RBPs) binden RNA direkt und kontrollieren das Schicksal von RNAs in Zellen. Gleichzeitig steuern zelluläre Signalkaskaden die Funktionen von RBPs, indem sie ihre physikochemischen Eigenschaften durch posttranslationale Modifikationen wie Phosphorylierung modulieren. Obwohl heute bereits Tausende von Phosphorylierungsstellen annotiert sind, sind entsprechende funktionelle Informationen begrenzt. Dies ist zum Teil darauf zurückzuführen, dass es keine Hochdurchsatzmethoden zur Erforschung der Funktion einer Phosphorylierungsstelle gibt. Um dieser Herausforderung zu begegnen, habe ich eine auf Shotgun-Proteomik basierende Strategie zur Messung der RNA-Bindungsaktivität von RBPs und ihren phosphorylierten Proteoformen entwickelt, die 'quantitative RNA-Interactome Capture (qRIC)' genannt wird.

QRIC quantifiziert die Pull-Down-Effizienz von RBPs, die mit Oligo(dT)-Magnetbeads isoliert werden. Diese Effizienz korreliert mit der Anzahl der RNA-Bindungsstellen und der Spezifität der Motivbindung, und spiegelt so die RNA-Bindung in vivo wieder.

In einer Gegenüberstellung der Pull-Down-Effizienz verschiedener Proteoformen in unbehandelten Zellen, habe ich qRIC als unvoreingenommenes Screening von regulatorischen Phosphorylierungsstellen in RBPs eingesetzt. Für jede einzelne Phosphorylierungsstelle wurde ein Delta-Effizienzwert berechnet, der den Einfluss auf die RNA-Bindung in vivo reflektiert. Die Effizienzunterschiede spiegelten das erwartete Verhalten von RBPs während der Phasentrennung von membranlosen Organellen und die Ladungsabstoßung zwischen Phosphorylierungsstellen und Nukleotiden bei physiologischem pH-Wert wider. Mithilfe des Delta-Effizienzwertes identifizierte ich mehrere bereits bekannte regulatorische Phosphorylierungsstellen in SF3B1, UPF1 und ELAVL1, sowie neue, bisher unbekannte und möglicherweise regulatorische Phosphorylierungsstellen in SERBP1, LARP1 und RBM20. Phosphomimetische Mutationsvarianten dieser Phosphorylierungsstellen wurden analysiert, um den molekularen Einfluss auf die Regulation der RBP-Funktion zu untersuchen. Es konnte gezeigt werden, dass die Phosphorylierung bestimmter Stellen im Spleißregulator RBM20 dessen nukleo-zytoplasmatische Lokalisierung, die Assoziation mit

zytosolischen RNA-Granula und die Spleißfunktion beeinflusst. Diese Erkenntnisse könnten sich beispielsweise auf die Entwicklung neuer Behandlungsmethoden für Patienten mit dysfunktionalen RBM20-Mutationen auswirken, die zu dilatativer Kardiomyopathie führen. QRIC kann als Hochdurchsatzverfahren dazu beitragen, unser Wissen über die Regulierung von Protein-RNA-Interaktionen durch Phosphorylierung zu erweitern.

SUMMARY

Post-transcriptional regulation of gene expression is fundamental in health and disease. RNA-binding proteins (RBPs) directly bind and govern the fate of RNAs in cells. At the same time, cell signaling cascades control RBP functions by modulating their physicochemical properties through post-translational modifications, like phosphorylation. Although thousands of phosphorylation sites have been annotated, functional information is limited. This, in part, is due to the lack of high-throughput methods that measure function. To tackle this challenge I developed a shotgun proteomics-based strategy for measuring the RNA-binding activity of RBPs and their phosphorylated proteoforms, named quantitative RNA-interactome capture (qRIC). In qRIC, pull-down efficiency of RBPs isolation with oligo(dT) magnetic beads is quantified in cells at steady state and correlates with the number of RNA-binding sites and motif binding specificity, reflecting a link to RNA-binding in vivo. By contrasting pull-down efficiency of different proteoforms in the cells, I applied qRIC as an unbiased screening of regulatory phosphorylation sites in RBPs affecting pull-down efficiency. A delta efficiency score was calculated for each individual phosphorylation site to denote its influence on RNA-binding in vivo. Efficiency differences globally reflected the expected behavior of RBPs during phase separation of membraneless organelles and charge repulsion between phosphorylation sites and nucleotides in physiological pH. Using the delta efficiency score, I identified several previously known regulatory phosphorylation sites in SF3B1, UPF1 and ELAVL1, plus novel candidate regulatory sites in SERBP1, LARP1 and RBM20. Phosphomimetic mutant variants of these sites were analysed to investigate the molecular mechanism of regulation. Importantly, I show that phosphorylation of candidate sites in the splicing regulator RBM20 affects its nucleo-cytoplasmic localization, association with cytosolic RNA granules, and splicing function. These findings could have implications for the development of novel treatments based on kinase activity for patients with dysfunctional RBM20 mutations leading to congenital dilated cardiomyopathy. I anticipate that qRIC, as a high throughput approach, will help to expand our knowledge about the regulation of protein-RNA interactions and their regulation by phosphorylation.

ERKLÄRUNG

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben. Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad. Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde. Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin vom 5. März 2015. Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsberaterinnen/Promotionsberatern stattgefunden hat und dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Berlin, _____

_____

Carlos Henrique Vieira e Vieira

DECLARATION

I hereby declare that I completed the doctoral thesis independently based on the stated resources and aids. I have not applied for a doctoral degree elsewhere and do not have a corresponding doctoral degree. I have not submitted the doctoral thesis, or parts of it, to another academic institution and the thesis has not been accepted or rejected. I declare that I have acknowledged the Doctoral Degree Regulations which underlie the procedure of the Faculty of Life Sciences of Humboldt-Universität zu Berlin, as amended on 5th March 2015. Furthermore, I declare that no collaboration with commercial doctoral degree supervisors took place, and that the principles of Humboldt-Universität zu Berlin for ensuring good academic practice were abided by.

Berlin, _____

_____

Carlos Henrique Vieira e Vieira

ACKNOWLEDGEMENTS

First and foremost, I thank Matthias. More than courses or workshops, I believe we learn from the people around us and I definitely learned from you. I'll carry you high in the rank of role models for the rest of my career for your compassion, enthusiasm and care for your group. Thank you for the best professional compliment I ever received: I also had fun working with you.

A special thank you to all my colleagues, current and previous members of the lab. You made my days richer and inspiring. Thank you for every lunch, walk around the campus, dance and sing in the lab, chat by the bench, and cake (booze) break in the coffee room. Particular thanks to the ones who contributed directly to my thesis and taught me so much: Koshi, Katrina, Boggy, Henrik, Michal, Kamilla, Chris, Mirjam, Trendelina, and Teresa.

I am really grateful for all the technical support I received during my PhD. Thank you Martha for the support with the final experiments of my thesis. And a very special thank you to Christian, who has been an incredibly motivated and brilliant colleague and is always willing to help the "kids in the kindergarten". Without you the lab goes wild! Trust me, we tried it and failed during your holidays...

A big special thanks to Sabine and Petra for all the help navigating the administrative hurdles. You made my life so much easier!

I also want to acknowledge my scientific collaborators. Vita Dauksaite, Janine Frohlich and Michael Gotthardt for their enthusiasm in our RBM20 project. Roni Lahr and Andrea Berman for their work in our LARP1 project. Michael Nielsen and his team for having me in their lab in Denmark. Erika Urdaneta, Matthew Kraushar, Rainer Nikolay, Ulrike Zinnall, Jonas Peters, and Ethiraj Ravindran for the very productive exchange and sinergy. It has been a pleasure to work with you all.

Arriving in a new institute, new city and new country is very challenging but this was made much easier by having the support from the PhD office and the Welcome center. For that, I would like to thank Michaela Herzog, Annette Schledz, Tina Diamantara and Sylvia Sibilak.

Mental and emotional health is a huge part of a thesis that often goes unrecognized. The friends I made in Berlin are the cuddles, laughs and cries that re-energize me and made it not only possible but a dream come true. Special thanks to Marta, Laura, Eric and Gesa that have been with me in this journey from the very beginning. I am also very grateful for the friends who joined the Mellow circle throughout these years and became my friends-family. I love to contemplate the friendship we created together. Another special thanks to the members of our Men Circle for creating a safe and honest space in this past year.

A big thank you to my family who supported me from afar to pursue my dreams and empowered me to trust in myself. Thank you to my mom Rosana whose love is infinite, my brothers, my grandparents and all cousins, aunts and uncles!

And finally, thank you Anne. Our relationship has been the most gratifying experiment of all my thesis and that I'll take with me. By your side, I am a better person and can deliver the best of myself.

Table of Contents

Table of Figures

Abbreviations

AA: Amino acid

ABC: Ammonium bicarbonate

CLIP: Cross-linking and immunoprecipitation

DAPI: 4',6-diamidino-2-phenylindole

dFBS: Dialysed FBS

DMEM: Dulbecco's modified Eagle medium

DNA: Deoxyribonucleic Acid

DTT: Dithiothreitol

eCLIP: Enhanced CLIP

EDTA: Ethylenediaminetetraacetic acid

EGTA: Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid)

FBS: Fetal bovine serum

HEK293T: Human embryonic kidney cells 293, antigen T

HPLC: High-performance liquid chromatography

HRP: Horseradish peroxidase

iBAQ: Intensity based absolute quantification

LC-MS/MS: Liquid chromatography tandem mass spectrometry

LysC: Lysyl endopeptidase

MDC: Max Delbrück Center for Molecular Medicine in the Helmholtz Association

mRBPs: mRNA-binding proteins

mRNA: Messenger RNA

PAR-CLIP: Photoactivatable ribonucleoside-enhanced CLIP

PBS: Phosphate-Buffered Saline

PCR: Polymerase chain reaction

pSILAC: Pulsed SILAC

PTex: Phenol-toluol extraction

PTM: Post translational modification

qRIC: Quantitative RIC

qRT-PCR: Quantitative RT-PCR

RBD: RNA-binding domain

RBPs: RNA-binding proteins

RIC: RNA-interactome capture

RNA: Ribonucleic acid

RRM: RNA recognition motif

RT-PCR: Reverse transcriptase PCR

SDS: Sodium dodecyl sulfate

SILAC: Stable isotope labeling by amino acids in cell culture

StageTip: Stop-and-Go extraction tip

TFA: Trifluoroacetic acid

TOP: 5' terminal oligopyrimidine

Tris: Tris(hydroxymethyl)aminomethane

UTR: Untranslated region

UV: Ultraviolet light

WT: wild-type

XL: cross-link

1. Introduction

Regulating protein output from gene expression in the cell is essential for life. Gene expression occurs in two key steps: transcription of DNA into RNA, and translation of RNA into protein. Consequently, regulation intervenes at the level of transcription, post-transcription, translation and post-translation. In this work, I investigated how two of these regulatory steps converge, where post-translational modification of proteins influences their ability to bind and regulate RNAs post-transcriptionally. For this section, I introduce central concepts necessary for the understanding of my results and discussion. Although several of the ideas presented here also apply to other classes of RNAs, I will be referring mostly to messenger RNAs (mRNAs).

1.1. RNA-binding proteins mediate post-transcriptional regulation of gene expression

Protein levels in cells are only partially explained by the corresponding RNA abundance, suggesting that post-transcriptional regulation of gene expression has a major impact on protein output produced from mRNA (Buccitelli and Selbach, 2020). The number and complexity of steps downstream of transcription underscore the importance of post-transcriptional regulation (Figure 1.1): From the DNA template, pre-mRNA is transcribed, capped at the 5’ end, spliced, cleaved, and polyadenylated at the 3´ end; RNAs are then transported out of the nucleus into the cytosol where ribosomes translate the RNA into proteins; and at the end of RNA’s lifecycle specialized molecular machineries are responsible for degrading RNA into its minimal components for recycling (Moore, 2005; Singh *et al.*, 2015). Central to all these processes, RNA-binding proteins (RBPs) coordinate gene expression by directly binding to RNA and forming ribonucleoprotein particles (RNPs) (Müller-McNicoll and Neugebauer, 2013; Gehring, Wahle and Fischer, 2017). RNPs are dynamic, changing composition to exert distinct effects on RNAs (Mitchell and Parker, 2014; Zarnack *et al.*, 2020). The RNP composition ultimately determines RNA’s fate. In agreement to their central role in post-transcriptional gene expression, perturbations of the network of RBP-RNA interactions can lead to cellular dysfunction and disease (Castello *et al.*, 2013; Gebauer *et al.*, 2021; Park *et al.*, 2021). In summary, RBPs adorn RNA

molecules and thereby provide a code that is read by the molecular machinery in the cell to control gene expression.

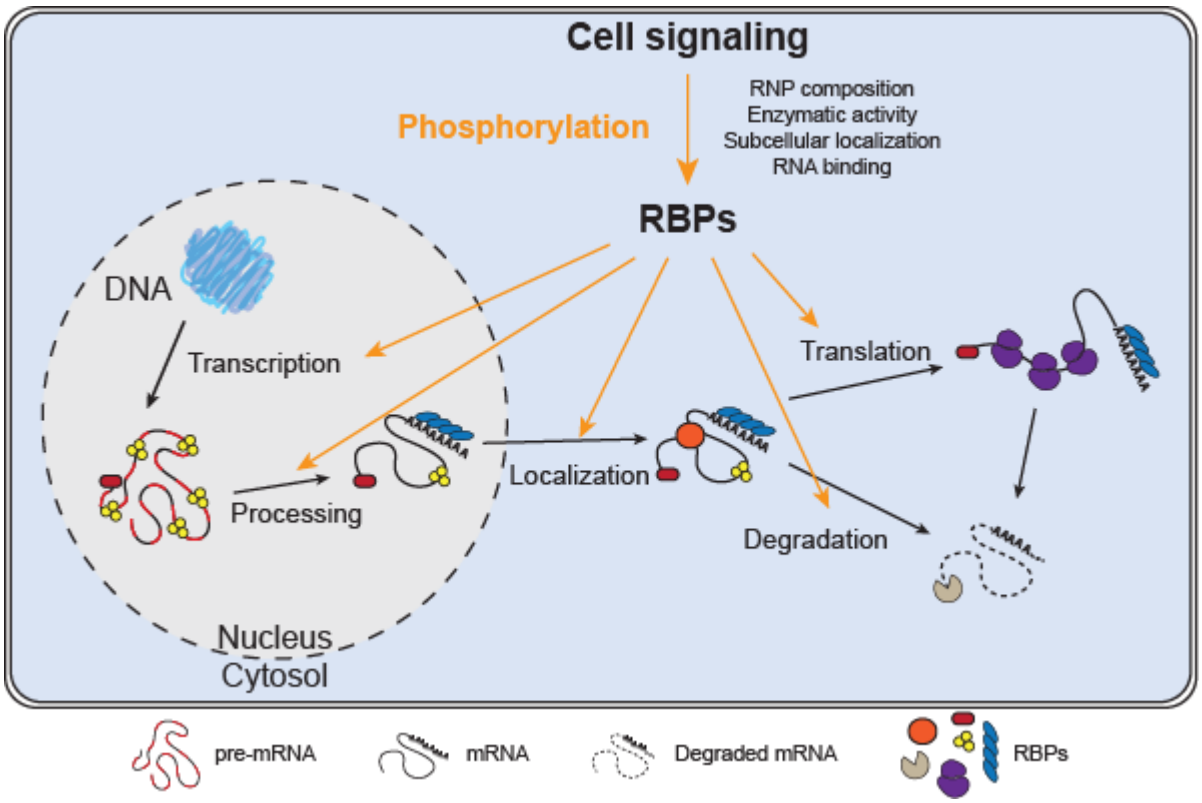


Figure 1.1 - The cellular life cycle of mRNAs from transcription to degradation, and regulation by RBPs downstream of cell signaling cascades.

1.1.1. The diverse RNA-bound proteome

RNA molecules in the cell are essentially always in association with a myriad of proteins (Khong and Parker, 2020). Historically, limited experimental conditions restricted identification of RBPs to those easier to purify due to association with virtually all mRNAs molecules in cells (Singh *et al.*, 2015). However, with modern mass spectrometry technology several hundred RBPs can be identified in a single experiment (Baltz *et al.*, 2012; Urdaneta *et al.*, 2019). As a consequence, the RNA-bound proteome has greatly expanded in recent years, with several hundred RBPs discovered and comprising at least 5 % of the human proteome, exceeding the number of transcription factors (Gerstberger, Hafner and Tuschl, 2014; Hentze *et al.*, 2018; Gebauer *et al.*, 2021).

RBPs are a diverse group of proteins often categorized by RNA-binding domains (RBDs), although only about 25 % of RBPs carry identifiable RBDs (Gerstberger, Hafner and Tuschl, 2014). One useful way to broadly classify RBPs is by their specificity in associating with RNAs (Müller-McNicoll and Neugebauer, 2013; Gehring, Wahle and Fischer, 2017): Core RBPs bind to all mRNAs in defined phases and often possessing or promoting enzymatic activity on targets to perform important general functions during mRNA metabolism; specific RBPs recognize defined sequences or structural features with high specificity and affinity and therefore target a limited set of RNAs; and non-specific generalist RBPs with little or no sequence specificity are recruited to RNAs often as effectors of RNA fate, including nucleases, helicases and RNA modifying enzymes. Finally, not only RNAs are regulated by RBPs, but also some RBPs are regulated by the interaction with RNAs. To identify the specific group of metabolic enzymes that interact with RNAs, and whose metabolic functions are likely regulated by it, such as GAPDH, the moonlight RBPs term has been coined (Castello, Hentze and Preiss, 2015).

1.1.2. RBP function is a product of subcellular localization, protein partners and sequence

Investigating the protein sequence is a powerful approach to study protein function. Sequence homology revealed that Aconitase, together with several other metabolic enzymes, are RNA-binding proteins (Hentze and Argos, 1991), a surprising finding later validated by low and high throughput methods (Castello, Hentze and Preiss, 2015). Hence, identification of RBDs by sequence homology has been historically used for annotating proteins likely to interact with RNAs *in vivo* (Corley, Burns and Yeo, 2020; Gebauer *et al.*, 2021). However, knowing that a protein carries for example a helicase domain indicates that it is involved in unwinding of the RNA secondary structure, but it does not help to identify the RNA targets nor the RNA lifecycle phase in which it acts. Therefore, analysis of RBP sequence can reveal mechanisms for protein function but is not enough to understand the cellular context.

More informative about the cellular context is knowing an RBP's protein partners, as proteins rarely work alone in the cells. This notion has been experimentally exploited for discovering novel candidate RBPs in the proteome, as any protein physically binding RNA is likely to be in close proximity with other RBPs and, consequently,

interaction with other RBPs is a strong predictor of RNA-binding (Brannan *et al.*, 2016). The RBP functional context in the cell can also be inferred from interaction with core macromolecular complexes such as polysomes, macromolecular structures formed by several actively translating ribosomes on the same mRNA molecule. Because of polysome's central role in translation, RBPs interacting with polysomes are likely involved in translational control (Imami *et al.*, 2018). Similarly, interaction with components of the spliceosome is suggestive of regulation of splicing (Maatz *et al.*, 2014). In conclusion, identifying RBP's partners in crime in the cell is a useful way to study protein function on RNA targets.

The Human Protein Cell Atlas (http://www.proteinatlas.org) shows that the more well-defined and limited the subcellular localization is, the better one can predict the function of a given protein (Thul *et al.*, 2017). Although RBPs are dispersed throughout cellular compartments, several RBPs are known to localize specifically according to their function. Furthermore, the observation that the cellular distribution of several RBPs changes according to the abundance of target RNAs suggests that subcellular localization and RNA-binding are closely linked (Khong and Parker, 2020). For example, subcellular localization and annotating RNA targets led to the identification of 42 novel RBPs regulating gene expression in the mitochondria (Van Nostrand *et al.*, 2020). Therefore, subcellular localization is intrinsically connected to RBP function in the cell and is a useful feature for studying these proteins.

In summary, the biological function of RBPs in the cell is a product of their sequence, molecular interactions, and subcellular localization. Only by understanding these three features simultaneously affecting protein function can one grasp the workings of RBPs on target RNAs.

1.1.3. RNA-binding defines RBP function

RBPs function is also defined by the set of RNA interactors. While many RBPs interact with several classes of RNAs, some RBPs selectively interact with tRNAs, snoRNAs, miRNAs and other RNA species (Mukherjee *et al.*, 2019; Van Nostrand *et al.*, 2020). This strongly suggests the involvement of RBPs in cellular events where these RNA classes are particularly relevant, and RBP cellular functions can thus be anticipated from knowing the class of bound RNAs. In addition, already the number of RNA-binding sites alone can be suggestive of RBP function: Binding to many sites might

suggest exposure to several RNA molecules, possibly due to a close relationship with core macromolecular complexes like the ribosome or the spliceosome. This correlation is also true for many proteins involved in splicing regulation, although not all proteins involved in splicing and translation show a large number of RNA-binding sites (Van Nostrand *et al.*, 2020). Therefore, knowing the class and number of target RNAs provides predictive information on the RBP function in the cell.

RBP binding strength to RNA also varies according to function. For example, helicases responsible for the unwinding of secondary structures need to access multiple nucleotide sequences and, therefore, should not strongly bind to RNA. In agreement with that, RNA-binding motifs for helicases are often poorly defined and bind with lower affinities compared to other classes of RBPs (Linder and Jankowsky, 2011; Hurt, Robertson and Burge, 2013). Conversely, it is essential for RBPs involved in alternative splicing that RNA-binding occurs at very precise locations on the pre-mRNA sequence (Ule *et al.*, 2006). Thus, knowing the RNA-binding strength and specificity can also provide hints about the function of an RBP (Li *et al.*, 2017a; Dominguez *et al.*, 2018; Mukherjee *et al.*, 2019).

Binding of RBPs to structural features in RNA can also be a good indicator for function. For example, binding to sequences at intron-exon borders is a strong indicator of involvement in splicing (Van Nostrand *et al.*, 2020). Also, the mRNA 3'UTR is a highly versatile region and interactions with this region points to regulatory roles of RBPs including RNA stability and subcellular localization (Mayr, 2017; Plass, Rasmussen and Krogh, 2017). Another interesting case is RBPs regulating translation initiation by binding to the mRNA 5´UTR, especially near the m^7Gppp cap structure essential for cap-dependent translation (Niedzwiecka *et al.*, 2002). Hence, RBPs such as LARP1 binding to the 5' terminal oligopyrimidine (TOP) motif, a motif present in mRNAs coding ribosomal genes and several other proteins that assist mRNA translation initiation in cells (called TOP mRNAs), can effectively block the access of translation initiation factors and globally repress protein translation in cells in response to nutrient deprivation (Philippe *et al.*, 2018; Berman *et al.*, 2021). Therefore, a lot can be learned about the RBP function by knowing which region and structure in the mRNA it binds.

Finally, RBPs interact in *cis* and *trans* to co-regulate RNA's lifecycle. Hence, regulatory RNA operons are coordinated by common RBPs, where expression of sets of RNAs coding for proteins in a given pathway can simultaneously be regulated by the same

RBPs (Keene, 2007). Thus, much of the leading-edge system biology research in the field dedicates to mapping such a combinatorial regulation (Li *et al.*, 2017b; Mukherjee *et al.*, 2019; Sternburg and Karginov, 2020). Together, the network of combinatorial interactions between several hundreds of RBPs and thousands of RNAs form a highly complex and fundamental component in post-transcriptional gene expression.

1.2. Phosphorylation regulates RBPs function

RBPs function can be dynamically modulated by post-translational modification (PTMs) that chemically modify amino acids structures to influence RNP composition and function, and thereby modulating gene expression (Figure 1.1) (Thapar, 2015; Zarnack *et al.*, 2020). Phosphorylation, the most prevalent PTM, is the reversible enzymatic addition and removal of a phosphoryl group to specific residues on proteins, by protein kinases and phosphatases, respectively. Most phosphorylation events happen on serine, threonine and tyrosine residues (Sharma *et al.*, 2014), although phosphorylation of other residues has also been shown with particularly high incidence in mRNA-binding proteins (Hardman *et al.*, 2019). Phosphorylation is the most common PTM in eukaryotes and over a hundred thousand sites have been identified in humans, collectively referred to as the human phosphoproteome (Ochoa *et al.*, 2020). These include over twenty thousand sites in more than 1600 RBPs, with an average of 12.6 phosphorylation sites per protein. Even though not all sites are modified simultaneously in cells, this suggests that RBPs are heavily phosphorylated, which points to the important role of RBPs as gene expression effectors of cell signaling cascades. In this section, I will discuss how post-translational modifications (PTMs), specifically phosphorylation, regulate RBP function.

1.2.1. Phosphorylation impacts RNA binding

Interaction with the RNA is central for RBP functions and can be directly regulated by phosphorylation of the RBP. Indeed, changing RBP affinity towards RNA is a common mechanism of RBP regulation by phosphorylation (Thapar, 2015; Lovci, Bengtson and Massirer, 2016). Due to the physicochemical properties of the phosphoryl group, such as its negative charge and bulky size, phosphorylation of residues often lead to order-to-disorder and disorder-to-order transitions in the protein structure (Thapar, 2015).

Because many RBPs interact with RNA through intrinsically disordered regions (Järvelin *et al.*, 2016), RNA-binding of these RBPs can then be affected by stabilization and destabilization of the protein structure mediated by phosphorylation of specific residues (Lovci, Bengtson and Massirer, 2016). Moreover, RBP-RNA binding can also be regulated locally at the RNA-binding site. RBP's amino acids contacting RNAs are phosphorylated more often than non-contacting residues (Bae *et al.*, 2020). Interestingly, while both the phosphoryl group as well as nucleotides are negatively charged in physiological conditions, suggesting charge repulsion (Ghaemi *et al.*, 2017), not all phosphorylation events lead to decreased RNA-binding (Lovci, Bengtson and Massirer, 2016; Grammatikakis, Abdelmohsen and Gorospe, 2017; Xu *et al.*, 2019). Together, phosphorylation events both close or far from the RNA contacting site in RBPs can influence the interaction strength.

1.2.2. Phosphorylation dynamically regulates RNP composition

Besides affecting the bound transcriptome, phosphorylation of RBPs also frequently regulates the exchange of protein components in RNPs (Zarnack *et al.*, 2020). For example, UPF1, an RNA-dependent helicase and ATPase, plays a central role in the nonsense-mediated mRNA decay pathway by recruiting the RNA degradation machinery upon phosphorylation (Kim and Maquat, 2019; Kurosaki, Popp and Maquat, 2019). UPF1 phosphorylation by SMG1 recruits the SMG5-SMG7 complex to the UPF1 mRNP (Okada-Katsuhata *et al.*, 2012; Kurosaki *et al.*, 2014), which in turn recruits the CCR4−NOT deadenylation complex ultimately promoting degradation of UPF1-bound mRNA (Kashima *et al.*, 2006; Kurosaki, Popp and Maquat, 2019). That way, UPF1 controls the RNA fate in the cell by recruiting other protein components into the RNP.

Interestingly, phosphorylation of RBPs can also influence RNA stability without changing the composition of RNPs. For instance, phosphorylation of NCL activates the deadenylase activity of its binding partner PARN, leading to shortening of poly-A tails in NCL-bound RNAs (Zhang *et al.*, 2018). Therefore, phosphorylation of RBPs can activate enzymatic properties of interaction partners, thereby regulating RNA structure and gene expression without affecting the RNP composition.

Also, the role of phosphorylation in maintaining RBP-rich, lipid phase separated condensates has become increasingly well studied (Rai *et al.*, 2018). Structural

changes of RBPs, such as sequence disorder-to-order transitions and RNA-affinity changes, driven by phosphorylation in several RBPs can make them more prone to aggregate and form condensates (Nosella and Forman-Kay, 2021). Several types of aggregates exist in cells and hijacking RNAs into these structures might have multiple functions depending on the condensate componentes, such as presenting RNAs to degradation machinery in processing bodies or safeguarding RNAs in stress granules during unfavourable nutritional conditions to allow faster gene expression recovery (Banani *et al.*, 2017).

1.2.3. Phosphorylation driven subcellular localization of RBPs

Phosphorylation can also influence subcellular localization of RBPs. For example, localization of ELAVL1 (HuR) in the cytosol or nucleus is controlled by phosphorylation of several sites near a nuclear localization signal (Abdelmohsen *et al.*, 2007; Doller *et al.*, 2007; Kim, Abdelmohsen, *et al.*, 2008; Kim, Yang, *et al.*, 2008; Lafarga *et al.*, 2009). Experimentally studying regulation of subcellular localization independent from other protein functions is often difficult but possible. An interesting case is that of the serine 635 site (S635) in RBM20, an RBP involved in the regulation of alternative splicing in the nucleus of muscular and cardiac cells (Guo *et al.*, 2012). Under normal conditions, S635 is heavily phosphorylated and the protein localizes in the nucleus (Murayama *et al.*, 2018; Gaertner *et al.*, 2020), where it interacts with components of the spliceosome to regulate splicing of target genes (Maatz *et al.*, 2014). Substitution of S635 for alanine (S635A) to prevent phosphorylation, renders a cytosolic form of RBM20, and therefore incapable of regulating splicing of nuclear targets (Murayama *et al.*, 2018; Gaertner *et al.*, 2020). However, the mutant phenotype can be partially rescued by adding a nuclear localization signal to the S635A mutant, forcing it to enter the nucleus, and suggesting that S635 phosphorylation specifically regulates protein subcellular localization and not the interaction with target RNA of the spliceosome (Murayama *et al.*, 2018). This shows that, in the case of RBM20 S635A mutation, splicing activity is unaltered and the phenotype is caused by protein cytosolic accumulation.

In conclusion, regulation of RBP function by phosphorylation allows cells to modulate gene expression faster without relying on *de novo* RNA synthesis. As described in this section, RBPs are often phosphorylated and phosphorylation can regulate RBP

functions in diverse ways. Due to their central role in post-transcriptional gene expression control, studying the functional consequences of phosphorylation on RBPs is important. At the same time, not all phosphorylation events are expected to be functional and this resource and time demanding process can be frustrating (Ochoa *et al.*, 2020; Watson, Schwartz and Francavilla, 2021). Therefore, it is important that functionally relevant phosphorylation sites in RBPs are identified systematically and a method to do that is still missing in the field.

1.3. Studying RBP function in the OMICS era

After introducing the state of science, I want to discuss relevant high throughput methods to study protein-RNA interactions. Dozens of methods exist, each adding features and possibilities. Here, I present the conceptualization of the approach developed in this work that allowed me to systematically identify regulatory phosphorylation sites in RBPs.

1.3.1. UV cross-linking stabilizes RBP-RNA interactions

Experimental protocols to study RBPs function *in vivo* commonly begin by stabilizing the RNA interaction (Ramanathan, Porter and Khavari, 2019; Gräwe *et al.*, 2020). That's because RBP-RNA interactions are dynamic and often short lived (Corley, Burns and Yeo, 2020; Sharma *et al.*, 2021) and therefore difficult to detect, making stabilization crucial. At the same time, stabilizing RBP-RNA interaction in the cell has the advantage that background binding during or after cell lysis can be removed by washing samples in very stringent conditions (Baltz *et al.*, 2012; Castello *et al.*, 2012).

Cross-linking with UV irradiation stands out as the most common approach to stabilize RBP-RNA interactions *in situ* (Ramanathan, Porter and Khavari, 2019). Atoms in the nucleotides can be excited with continuous wave UV light to a higher energy state for a short time period (Meisenheimer and Koch, 1997). During the microseconds that atoms are excited, nucleotides can form zero-distance covalent cross-links with amino acid residues in close proximity (Budowsky *et al.*, 1986). The fact that atoms are only excited for a short period of time grants high specificity to the cross-linking reaction but also makes it very inefficient. In fact, efficiencies lower than 5 % of interacting RBP-RNA *in vitro* have been reported (Budowsky *et al.*, 1986; Fecko *et al.*, 2007). In

addition to the overall low efficiency of UV cross-linking, cross-linking efficiency varies depending on the amino acid side chain (Meisenheimer and Koch, 1997; Kramer *et al.*, 2014; Bae *et al.*, 2020). Moreover, the molecular structure of the nucleotides involved in the cross-linking also defines the reaction efficiency. Uridines are favourably cross-linked *in vitro* (Meisenheimer and Koch, 1997) and are the only detectable cross-linking nucleotide *in vivo* (Kramer *et al.*, 2014; Bae *et al.*, 2020). Also, whether the nucleotide contacts the amino acids with the base, sugar or phosphate backbone affects cross-linking efficiency, while double stranded RNAs generally poorly cross-link (Meisenheimer and Koch, 1997).

In conclusion, UV cross-linking is an inefficient approach to stabilize RBP-RNA interactions in cells. Although alternative approaches to stabilizing RBP-RNA interactions exist, such as chemical cross-linking (Li, Song and Yi, 2014), cross-linking with UV offers several advantages that make it the most commonly used approach when studying RBPs, including easy, fast and low cost in otherwise unmodified cells or tissues.

1.3.2. Protein- and RNA-centric methods to study RBP functions

In high throughput experiments, stabilization of RBP-RNA interactions is usually followed by omics analysis of samples, such as DNA sequencing and shotgun proteomics (Ramanathan, Porter and Khavari, 2019). In the last decade the number of sequencing methods to map RNA-binding sites of RBPs both *in vivo* and *in vitro* has increased massively (Wheeler, Van Nostrand and Yeo, 2018; Lin and Miles, 2019). Although each method has its own advantages and disadvantages offering a broad range of applications, cross-linking and immunoprecipitation followed by RNA-seq (CLIP-seq, or just CLIP for short) and its variants stand out. Methods like CLIP are called protein-centric, as the RBP of interest is purified for identification of bound RNAs (Wheeler, Van Nostrand and Yeo, 2018; Lin and Miles, 2019). Bound RNAs are then sequenced and RNA-binding sites are identified by peaks of RNA sequencing reads enriched over a background control (Corcoran *et al.*, 2011; Uren *et al.*, 2012; Lovci *et al.*, 2013). By mapping sequencing peaks to the genome, not only the target's genomic identity is obtained but also single nucleotide resolution of the interacting RNA site is achieved (Hafner *et al.*, 2010; König *et al.*, 2010).

Conversely, RNA-centric methods identify the RNA-bound proteome following biochemical isolation of a single or groups of RNAs and identification of protein interaction partners via mass spectrometry-based shotgun proteomics (Gräwe *et al.*, 2020). The most routinely used method for high throughput investigation of RNA-bound proteomes was published independently by the labs of Markus Landthaler and Matthias Hentze in 2012 (Baltz *et al.*, 2012; Castello *et al.*, 2012), named RNA-interactome capture (RIC). In RIC, samples are first cross-linked using UV light, followed by cell lysis. Poly-A RNAs containing RNPs are biochemically isolated from lysates with beads attached to oligo(dT) sequences. The oligo(dT) sequences hybridize with poly-A sequences in RNAs and beads are obtained from the suspension with magnets. Isolated RNAs can be sequenced for confirmation of RNA enrichment, while the poly-A RNA bound interactome is identified *via* mass spectrometry. As most poly-A RNAs are mRNAs, RIC specifically enriches mRNA-binding proteins (mRBPs).

Protein abundance in RIC pull-downs is a direct function of RNA-binding and UV cross-linking efficiency in the cell (Hentze *et al.*, 2018). In fact, differential analysis of RBP binding to RNA across conditions has been performed by several groups with RIC, thereby identifying context-specific regulatory RBPs in mammalian tissue culture cells (Boucas *et al.*, 2015; Liepelt *et al.*, 2016; Milek *et al.*, 2017; Perez-Perri *et al.*, 2018; Garcia-Moreno *et al.*, 2019; Ignarski *et al.*, 2019; Backlund *et al.*, 2020; Hiller *et al.*, 2020; E. M. Smith *et al.*, 2021), zebrafish and fly embryos (Sysoev *et al.*, 2016; Despic *et al.*, 2017), yeast (Bresson *et al.*, 2020), and plant cells (Marondedze *et al.*, 2019). Therefore, quantifying RNA-binding in UV-based RIC experiments is useful for investigating regulatory events in RBPs. One interesting question that is tackled here is how we can use quantification of RNA-binding to identify functional phosphorylation sites in RBPs.

Isolation of ribonucleoproteins (RNPs) can also be achieved physicochemically. During the course of my doctoral work, I collaborated with Erika Urdaneta (Humboldt-University Berlin, Germany) and Benedikt Beckmann (Humboldt-University Berlin, Germany) on the development of a method to isolate RBPs from cross-linked cells based on organic phase separation, named phenol-toluene extraction (PTex) (Urdaneta *et al.*, 2019b). Organic phase separation of RNPs is possible due to the amphiphilic properties of the RNA-protein complex (Smith *et al.*, 2020). In physiological pH, nucleotides are hydrophilic and RNA dissolves in the aqueous

phase, while proteins contain both hydrophobic and hydrophilic amino acids. By dissolving the cellular content in a mixture of aqueous and organic solvents, molecules are segregated when the two phases separate in solution. RNAs separate with the aqueous phase, but cross-linked RNPs are pushed into the intermediate border between both phases. Ultimately, RNPs are precipitated from the intermediary border between phases and analysed via mass spectrometry. Importantly, RNA isolation is not limited to specific classes or RNA polyadenylation (Urdaneta *et al.*, 2019b). Two similar approaches have been developed in the laboratories of Kathrin Liley and Jeroen Krijgsveld (Queiroz *et al.*, 2019; Trendel *et al.*, 2019).

1.3.3. Conceptualizing a method to systematically study RBPs functional regulation by phosphorylation

Differential quantification of phosphorylation across experimental conditions is often used to screen for relevant sites downstream of signaling cascades (Olsen *et al.*, 2006). In such experiments, changes in the cellular signaling landscape are detected by quantifying the abundance of phosphorylated sites in trypsin digested proteomes. The key assumption is that changes in phosphorylation levels correlate with function and that important sites will be differentially phosphorylated upon perturbation. Therefore, differential RIC and phosphoproteome quantification across experimental conditions could be used to correlate changes in RNA-binding and cell signaling status of RBPs. However, to the best of my knowledge, this has not been tried yet.

One inherent difficulty of this analysis would be that an RBP with changing RNA-binding could have two opposingly acting phosphorylation sites: one increasing and one decreasing RNA-binding. Indeed, examples where sites in the same RBP show opposing functions exist (Grammatikakis, Abdelmohsen and Gorospe, 2017; Xu *et al.*, 2019). Furthermore, experiments involving differential quantification of phosphorylation across experimental conditions are generally limited to short term perturbations, as changes in the proteome and secondary phosphorylation will confound results (Olsen *et al.*, 2006; Bodenmiller *et al.*, 2010; Needham *et al.*, 2019). In conclusion, an ideal method for systematic identification of regulatory phosphorylation sites in RBPs is independent of cellular perturbations and discerns the contribution of individual sites to RNA-binding.

One way to avoid the need for perturbation-dependent experiments is to compare the modified and unmodified protein forms (proteoforms) co-existing in cells. In shotgun proteomics, proteins are first digested into peptides that are then physicochemically separated and analysed in the mass spectrometer (Figure 1.2). Peptides with exactly the same sequence stemming from different proteoforms in the cell are indistinguishable. Hence, protein quantification in shotgun proteomics is based on unmodified peptide sequences coming from all proteoforms (Cox *et al.*, 2014). At the same time, modified peptides uniquely quantify proteoforms. Hence, contrasting modified and unmodified peptides for proteoform quantification has been implemented to identify phosphorylation sites modulating protein interaction with ribosomes (Imami *et al.*, 2018), nuclear subcellular distribution (Masuda *et al.*, 2020), and thermal stability (Huang *et al.*, 2019). Therefore, a shotgun proteomics method that quantifies RNA-binding of multiple proteoforms simultaneously from lysates of cells at steady state can fulfill the requirements as an ideal method for identifying regulatory PTMs individual sites in RBPs independently of perturbations.

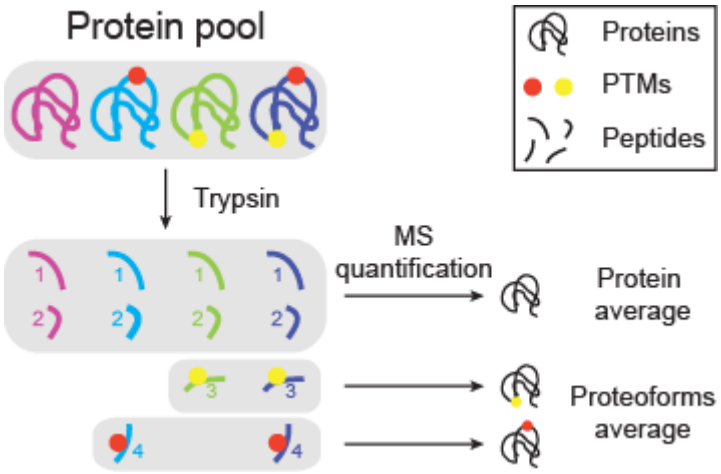


Figure 1.2 - Proteoform quantification in shotgun proteomics.

Once individual proteoforms can be quantified, the challenge is to measure the impact on RBP-RNA binding. One way to contrast proteoform-specific RBP-RNA binding is to compare the phosphorylation occupancy of RBPs, the fraction of a protein pool that is modified, in the whole cell lysate to that in the RNA-bound proteome fraction. For example, if half of the RBP pool is phosphorylated in the whole cell lysate, the same

50 % occupancy in the RNA-bound fraction is expected. Higher or lower occupancy in the RNA-bound fraction indicates positive or negative modulation of RNA-binding, respectively (Figure 1.3). Therefore, by comparing proteoform occupancy in whole cell lysates and RIC pull-down, one could identify regulatory phosphorylation sites in RBPs. The main problem with this approach is that it is difficult to quantify sample-specific phosphorylation occupancy, as the phosphorylated and unmodified peptides and total protein must be all quantified in the same sample (Sharma *et al.*, 2014).

An alternative strategy is to compare the pull-down efficiency, the fraction of protein isolated from the total cellular protein pool, of RBP proteoforms. Pull-down efficiency strongly correlates with relative RBP-RNA binding in the cell (Hentze *et al.*, 2018) Hence, if a phosphorylation site changes RBP-RNA binding, the RNA pull-down efficiency of such a proteoform will be different to the rest of the RBP pool (Figure 1.3). Importantly, pull-down efficiencies for proteoforms can be measured independently, for example in the whole cell lysates and in phosphopeptide-enriched samples. Therefore, quantification of pull-down efficiencies of mRNA-bound RBP proteoforms with shotgun proteomics fulfills the criteria for a method to quantify phosphorylation site specific changes in RBP-RNA binding at steady state conditions and was exploited in this work.

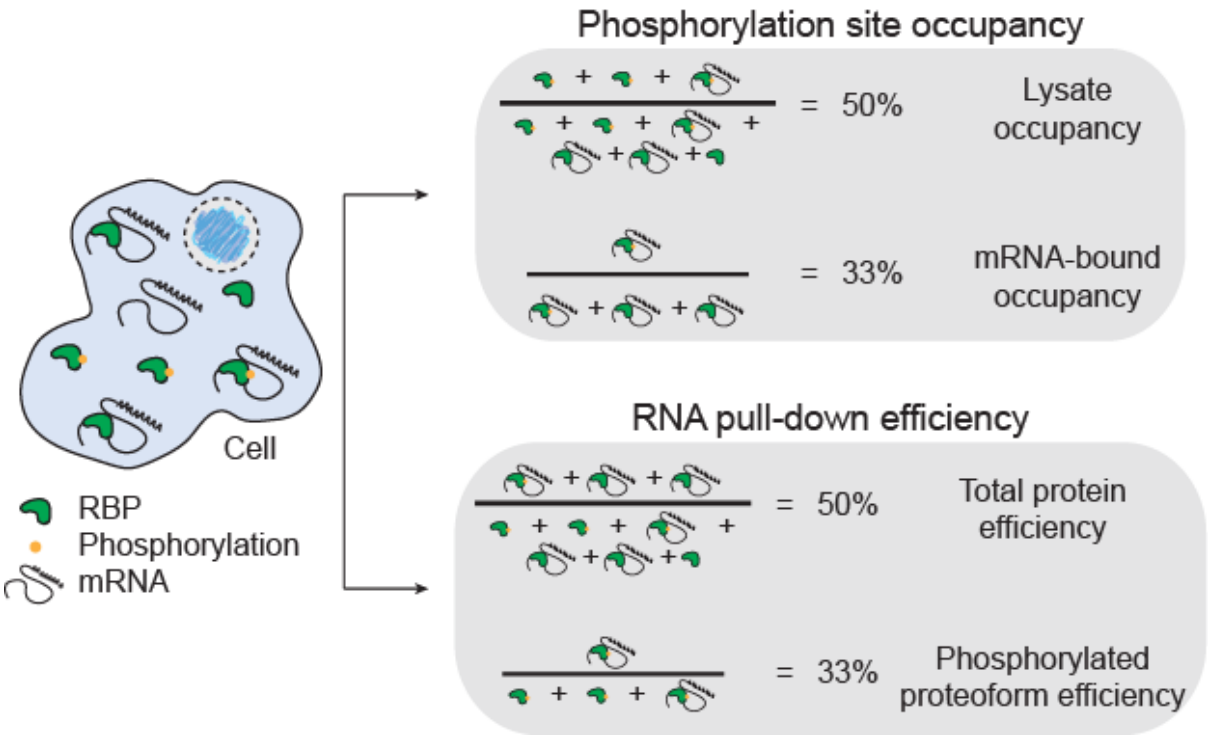


Figure 1.3 - Two shotgun proteomics-based strategies for identifying RNA-binding regulatory phosphorylation sites in RBPs.

2. Scientific aims

After almost a decade of constant development, RIC has greatly expanded the set of known RBPs and proven to be a robust method for quantitative analysis of mRNA interactomes across multiple samples and conditions. At the same time, mass spectrometry analysis of RNA interactomes opens up the possibility to simultaneously investigate post-translational modifications in RBPs and RNA-binding. Since the defining feature of RBPs is their ability to bind RNA, an attractive approach to assess the function of phosphorylation sites in RBPs is to quantify their impact on RNA-binding. Therefore, the aims of this work are:

1 - to establish a mass spectrometry-based approach for quantifying RBP-RNA interactions *in vivo*;

2 - to apply such a method in the systematic identification of regulatory phosphorylation sites in RBPs;

3 - to characterize the molecular mechanism of regulation by candidate phosphorylation sites in RBPs identified here.

3. Results

RBPs can be isolated and quantified from whole cell lysates using methods like RIC and PTex. Phosphorylated sites can also be routinely identified *via* mass spectrometry after enrichment of phosphopeptides from complex mixtures of peptides. Therefore, methods for isolating RBPs and phosphopeptide enrichment can be combined for functional investigation of phosphorylation in RBPs. The challenge is to combine these methods in a way that allows identifying regulatory phosphorylation sites in RBPs. Here, this is achieved by simultaneously quantifying pull-down efficiency of proteoforms in stable isotope amino acids in cell culture (SILAC)-labeled, UV cross-linked and otherwise untreated human embryonic kidney 293T (HEK293T) cells.

To quantify RBP pull-down efficiency, I developed quantitative RNA Interactome Capture (qRIC). The qRIC protocol starts by differentially labelling HEK293T cells with heavy or light SILAC. Protein-RNA interactions are stabilized *in vivo* by UV cross-linking, and RBPs from heavy labeled cells are isolated with RIC or PTex. RBPs are combined with the whole cell protein extract (“input”) from light labelled cells as an internal reference for accurate quantification. RBPs are then quantified in the pulled down fraction and relatively compared to a spiked-in amount of the input cell lysate. Consequently, qRIC directly associates SILAC ratios with the pull-down efficiencies, that is, the isolated fraction of the total cellular protein pool.

In addition, SILAC ratios and pull-down efficiency of phosphorylated RBPs is also quantified after enrichment of phosphopeptides by TiO2 chromatography. To compare phospho- and unmodified-proteoform pull-down efficiencies, I calculated a delta efficiency value as the ratio of SILAC ratios, representing the fold difference in pull-down efficiencies between proteoforms. Delta efficiency values higher or smaller than one suggest correlation of RBP phosphorylation with increased or decreased RNA-binding, respectively. Therefore, qRIC provides a simplified score related to the ability of phosphorylation sites to regulate RNA-binding of RBPs *in vivo*. The final qRIC experimental design is presented in Figure 3.1.

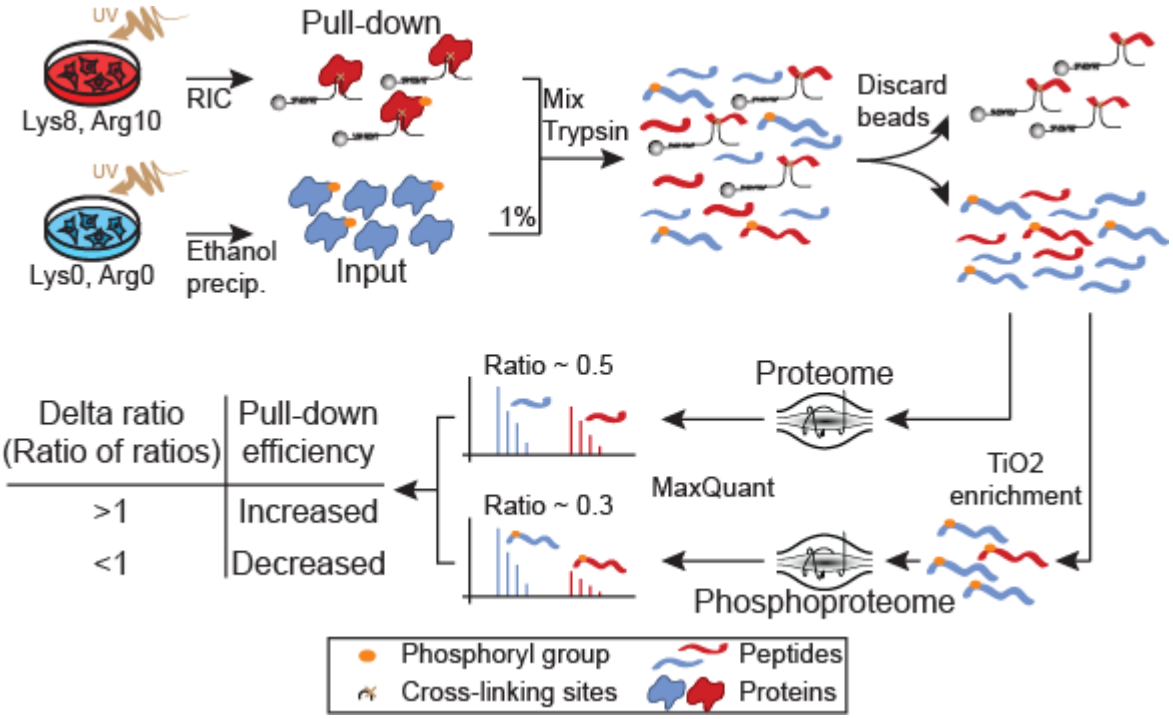


Figure 3.1 - The qRIC experimental design.

Partial results of the presented work have been submitted at Molecular Cell journal (Vieira-Vieira et al, submitted). Optimization of qRIC together with detailed discussions on the qRIC method are presented in the first section (3.1). Following section (3.2) is dedicated to my main findings on quantifying pull-down efficiency and its relationship to *in vivo* RNA-binding. In the third section (3.3), I focus on quantifying phosphorylation impact on pull-down efficiency. Finally, the last two sections (3.4 and 3.5) are where I present my results for three RBPs with candidate novel regulatory phosphorylation sites and their impact in cell biology. With a few exceptions indicated below, all experiments, data analysis and figures were performed, designed and created by me under the supervision of Matthias Selbach, and with input from the members of our lab and collaboration partners. Below is a summary of contributions from collaboration partners:

- Orsalia Hazapis in the lab of Prof. Dr. Markus Landthaler at MDC in Berlin, Germany, created the Flp-In T-Rex SERBP1_3xFLAG HEK293 cell lines expressing mutant SERBP1 variants.
- Dr. Roni Lahr in the lab of Prof. Dr. Andrea Berman at the University of Pittsburgh in Pittsburgh, United States of America, cloned, expressed and

purified full length LARP1 for EMSA. She also performed the EMSA and analysed results.

- Martha Hergeselle in the lab of Dr. Matthias Selbach at MDC in Berlin, Germany, performed the pSILAC experiment for LARP1 mutations and the immunofluorescence staining of FLAG tagged RBM20 mutants. I imaged samples and analyzed the results.
- Janine Froehlich and Dr. Vita Dauksaite in the lab of Dr. Michael Gotthardt in MDC in Berlin, Germany, performed the PCRs and luciferase measurements for the cell-based splicing reporter assays with RBM20 mutants. I analyzed the results.

3.1. Development of the quantitative RNA-Interactome Capture (qRIC)

3.1.1. Phosphorylated peptides are poorly enriched from PTex-isolated RBPs

Isolation of RBPs in PTex is not limited to RNAs containing poly-A sequence and is less time and resources consuming than RIC (Urdaneta *et al.*, 2019b). These advantages make PTex an attractive method to systematically study the impact of phosphorylation on RNA-binding. In a first experiment, I tested whether I can enrich phosphorylated peptides from RBPs isolated with PTex using TiO2 chromatography.

HEK293T cells were exposed to UV for cross-linking of RNPs *in vivo*. Cell lysate was split and 90 % used for PTex, while the remaining 10 % was ethanol precipitated and used for analysing the input whole cell lysate phosphoproteome (Figure 3.2 A). After PTex, both samples were trypsin-digested and different amounts of peptides (50, 100 or 200 µg) were used for phosphopeptides enrichment. Samples were analysed via mass spectrometry and phosphorylation sites identified by MaxQuant (Figure 3.2 B). Only singly modified peptides were considered for this analysis. Phosphorylation sites were also categorized whether they were found in annotated RBPs, mRBPs or other protein groups (Hentze *et al.*, 2018). From the phosphorylation sites identified, a larger fraction has been annotated to known RBPs in PTex enriched samples (55 %) than in input samples (30 %), suggesting successful isolation of RBPs. For both input and PTex samples, increasing the amount of peptides used for enrichment led to the identification of more phosphorylation sites, with a maximal number when 200 µg was used. While a high number of phosphorylation sites was identified in the input samples,

suggesting successful phosphopeptide enrichment, the number of identified phosphorylation sites in PTex samples was very low, with less than 200 phosphosites identified in total. Altogether, phosphopeptide enrichment from PTex isolated RBPs leads to low number of detected phosphorylation sites and is insufficient for a systemic investigation of phosphorylation on RBPs.

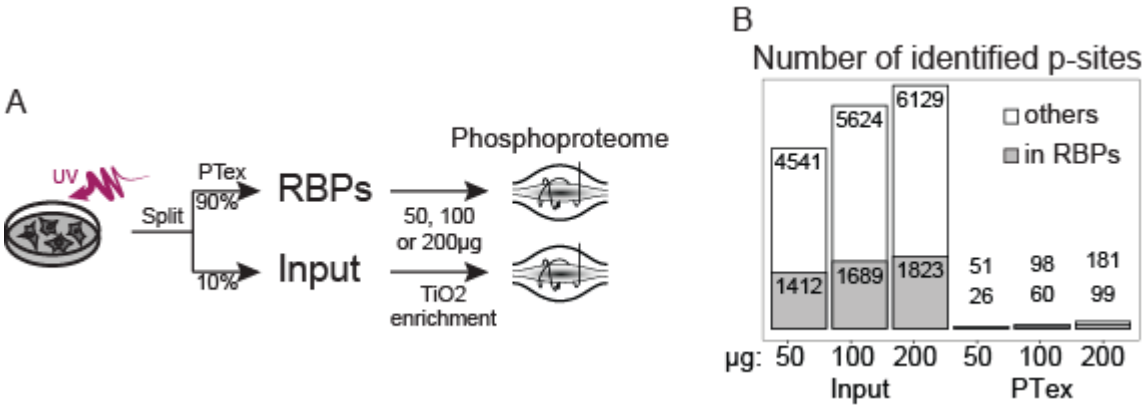


Figure 3.2 - Phosphopeptide enrichment with TiO2 chromatography in PTex-isolated RBPs.

A) Experimental design for testing phosphopeptide enrichment in PTex enriched samples.
B) Phosphosites (p-sites) identified via shotgun proteomics. Sites in RBPs were annotated accordingly (Hentze *et al.*, 2018). Numbers indicate the total number of phosphorylation sites (upper) and sites in RBPs (lower).

3.1.2. Trypsin elution in RIC improves identification of phosphosites

A key challenge in qRIC is the co-enrichment of the phosphopeptides of interest with peptides cross-linked to RNA. That is because the phosphate group in the RNA is chemically identical to that in phosphopeptides. Consequently, phosphopeptides enrichment methods based on phosphate-directed interaction with metal ions for affinity chromatography, such as the TiO2 chromatography that was used throughout this work, also isolate peptides cross-linked to RNA. In fact, methods that are nowadays used for phosphopeptide enrichment have been used in the past to isolate cross-linked protein-RNA complexes (Richter *et al.*, 2009). In conclusion, co-isolation of peptides cross-linked to RNA might explain the unsatisfactory results with PTex as even after RNA digestion many cross-linked peptides carry a moiety of nucleotides.

Phosphopeptide enrichment on peptide mixture from UV cross-linked RNPs might benefit from removing peptides cross-linked to RNA in advance. For that, I next adapted the original RIC protocol. In the original RIC protocol (Baltz *et al.*, 2012; Castello *et al.*, 2012), UV cross-linked RNPs are pulled down by hybridizing the RNA to oligo(dT) beads followed by stringent washes. For RBP elution, beads are heated up, disrupting RNA-oligo(dT) hybrids and releasing mRBP-mRNA complexes into solution. Another way to elute RBPs from the beads is to digest them with trypsin at room temperature, maintaining the mRNA-oligo(dT) hybrids. That way, digested peptides cross-linked to mRNA stay bound to beads and can be discarded, while all other tryptic peptides are released into solution. In consequence of eluting by protein digestion with trypsin, amino acid sequences between two proximal digestion sites and containing an mRNA cross-linking site are systematically removed from the downstream analysis. In fact, identifying missing peptides in digestion-eluted samples has been used to annotate mRNA-binding regions in mRBPs (Castello *et al.*, 2016; Mullari *et al.*, 2017).

To test whether the phosphopeptides enrichment is affected by the presence of peptides cross-linked to RNA moieties, I compared RIC-isolated mRBPs eluted with heat or trypsin digestion. Cells were grown in light or heavy labeled SILAC medium and cross-linked with UV (Figure 3.3 A). Medium-heavy grown cells were included as a non UV cross-linked control. Labeled cells were mixed and used for RIC while a small fraction (1 %) of the mixture was used for analysis of the input whole cell lysate. Before elution, beads were divided in two and mRBPs from each half were eluted either by heating up to 80°C for 5 min followed by overnight trypsin digestion of the supernatant, or by direct on-bead digestion with trypsin overnight at 25°C. After phosphopeptide enrichment with TiO2 chromatography, proteome and phosphoproteome were analysed (Figure 3.3 B). As expected, more phosphorylation sites were identified in both heat and trypsin eluted RIC samples than in the control group not cross-linked with UV (Figure 3.3 B). Also, only 4 and 10 % of phosphorylation sites in heat and trypsin samples were identified in proteins not previously annotated as mRBPs, suggesting successful isolation of mRBPs.

Importantly, elution of mRBPs with trypsin led to identification of 2 fold more phosphorylation sites than elution by heat (817 against 414 in both replicates together), with only 37 heat eluted sites not identified in the trypsin group (Figure 3.3

C). More phosphorylation sites might be identified in the trypsin eluted samples because more mRBPs are also generally eluted from beads. In fact, a slightly higher number of mRBPs were quantified in the trypsin eluted fractions before phosphopeptides enrichment (89 in trypsin against 69 in heat eluted samples). However, this small increase is not sufficient to explain the difference observed in the number of quantified phosphorylation sites (817 against 414 phosphosites). At the same time, quantitative analysis of the protein and phosphosite SILAC ratios over the non-cross linked control group indicates that elution was similarly efficient in both digestion and heat elution (Figure 3.3 D). Altogether, removal of peptides cross-linked to RNA before phosphopeptides enrichment improved phosphorylation sites coverage in mRBPs isolated by RIC.

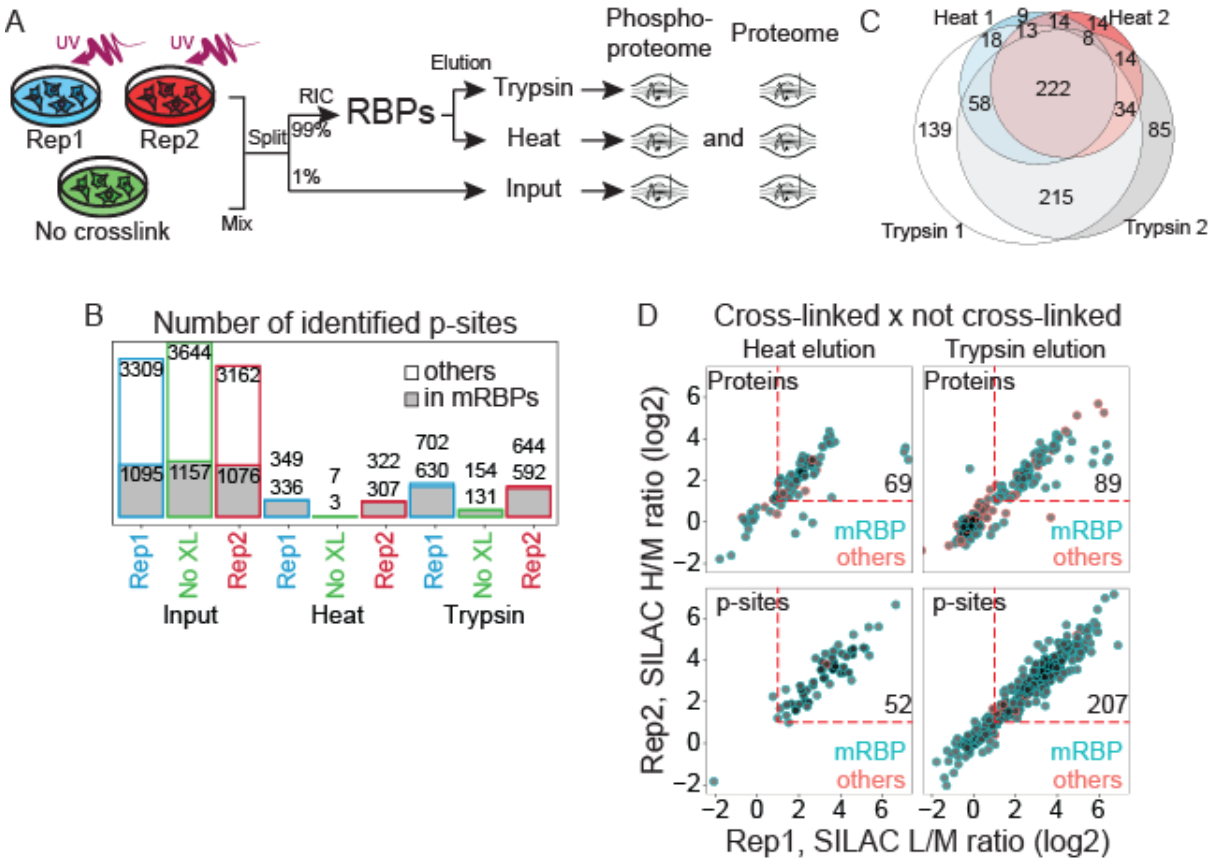


Figure 3.3 - Trypsin elution of RBPs improves phosphopeptides coverage in RIC samples.

A) Experimental design. Medium-heavy labeled cells (green) were used as a control group without UV cross-linking, while light (blue) and heavy (red) labeled cells represent two independent replicas (Rep1 and Rep2, respectively).

B) Number of phosphorylation sites (p-sites) per sample. Phosphorylation sites in RBPs were annotated accordingly (Hentze et al., 2018). XL, cross-linking.
C) Venn diagram of phosphorylation sites identified in heat- and trypsin-eluted RBPs in both experimental replicas.
D) Protein and phosphosite quantification in heat- and trypsin-eluted RBPs over the control sample without UV cross-linking. For the light-labeled group (Rep1), light-over-medium ratio is plotted (L/M) while heavy-over-medium ratio is plotted for the heavy-labeled group (Rep2). The number of proteins and phosphosites with SILAC ratio higher than two (dashed red lines) in both replicas is indicated. Previously annotated mRBPs and phosphosites in mRBPS (Hentze et al., 2018) are also indicated.

3.1.3. Ratio compression does not bias qRIC pull-down efficiency

Another challenge in qRIC relates to the quantification of the SILAC ratios. Like many other mass spectrometry based quantification methods, SILAC quantification suffers from ratio compression, a bias in which extreme ratios are underestimated due to increased noise-to-signal ratio for low abundant peptides (Lau *et al.*, 2014; Hogrebe *et al.*, 2018). SILAC allows for very accurate relative quantification but extreme ratios, higher than 10, are systematically underestimated. In qRIC, the SILAC ratio directly depends on the amount of light-labeled input spiked into the isolated mRBPs. It is then important that the amount of spike-in input is not too high or low, so it does not lead to extreme SILAC ratios. To investigate the extent of ratio compression and its consequences on pull-down efficiency estimation in qRIC, I spiking in increasing amounts of light labeled input material and measured the impact on SILAC ratios (Figure 3.4).

The same amount of cells were grown in light or heavy labeled SILAC medium and cross-linked with UV (Figure 3.4 A). While heavy labeled cells lysate was used for RIC, lysate from light labeled cells was ethanol precipitated. After RIC, oligo(dT) beads were divided in three equal parts before mRBPs elution with on-bead trypsin digestion and different amounts of ethanol-precipitated light labeled cell lysate, corresponding to 0.1, 0.5 and 2.5 % of the heavy light-labeled amount used for RIC, were added. Biological replicates were generated by swapping SILAC labels, ratios were computationally inverted before averaging with the forward label measurement for further analysis. Protein pull-down efficiencies were obtained from the SILAC ratios by multiplying it to the respective percentage of spiked-in input cell lysate. Hence, a

protein has pull-down efficiency of 1 % if the SILAC ratio is 2 fold in the sample with 0.5 % input spiked-in. The same protein with pull-down efficiency of 1 % should have a SILAC ratio of 10 fold in the sample with 0.1 % spiked-in input.

Ratio compression of extreme SILAC ratios leads to underestimation of the true pull-down efficiency in one sample, but the same ratio might be correctly quantified in one of the other two mixtures. Therefore, the average pull-down efficiency from all experiments was compared to the SILAC ratios in individual experiments (Figure 3.4 B). As expected, SILAC ratios change in relation to the average pull-down efficiency according to the amount of input material, as indicated by the horizontal shift of the data points distribution in Figure 3.4 B.

By comparing sample specific SILAC ratio and the average pull-down efficiency in all three samples, I identified under- and overestimated ratios according to where they locate in relationship to the linear trend of the data: to the left if underestimated, and to the right if overestimated. With the exception of proteins with extremely low or high SILAC ratios in samples spiked-in with 0.1 or 2.5 % input material, SILAC ratio quantification is correctly estimated for most proteins in all three mixtures compared to the average efficiency. Notably, a strong increase in the number of proteins with very low pull-down efficiency, and likely background binders to oligo(dT) beads, was observed when a higher spike-in amount (2.5 % of input material) was used. Furthermore, less than 10% of SILAC ratios are greater than 10 fold when 0.5-2.5 % spike-in is used and are not expected to strongly suffer from ratio compression. These results indicate that, although pull-down efficiencies for highly efficient mRBPs might be underestimated by SILAC ratio compression, it does not significantly impact qRIC estimates when spike-in amounts between 0.5 % and 2.5 % is used. For the rest of this thesis, I used 1 % of input material as spike-in in qRIC experiments.

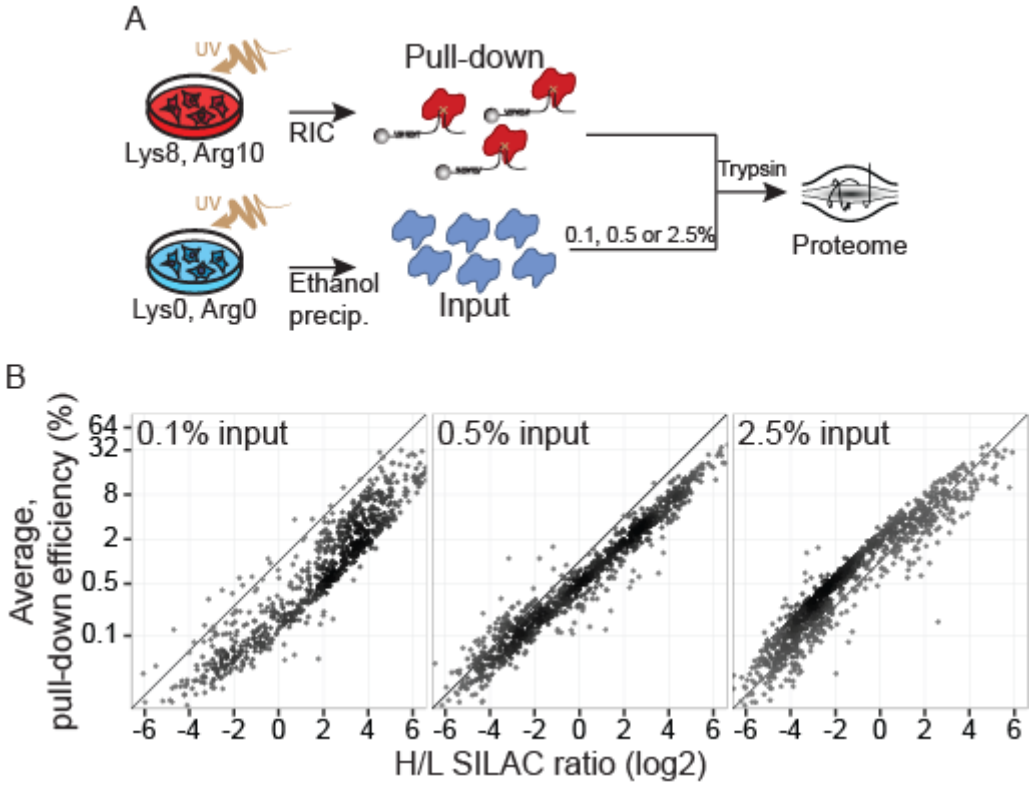


Figure 3.4 - Ratio compression impact on qRIC pull-down efficiency calculation.

A) Experimental design to evaluate ratio compression impact on pull-down efficiency estimation.
B) Average protein pull-down efficiency comparison to SILAC ratios in samples spiked-in with different amounts of input material.

3.2. Quantifying RBP-RNA interaction with qRIC

Several parts of this subsection are included in a manuscript submitted for publication to the Molecular Cell journal (Vieira-Vieira et al, submitted).

3.2.1. qRIC reproducibly quantifies pull-down efficiency from SILAC ratios

Optimization led to the final protocol (Figure 3.1) that was used for pull-down efficiency quantification of mRBPs and phosphorylation sites in HEK293T cells. Cells were fully labeled with heavy or light stable isotope amino acids and protein-RNA interactions were then stabilized *in vivo* by UV cross-linking. RNPs from heavy labeled cells were pulled down with magnetic oligo(dT) beads followed by stringent washes, and combined with ethanol-precipitated input lysate from light labelled cells, corresponding

to 1 % of the heavy-labeled lysate amount used for RIC, as an internal reference to enable accurate quantification. After protein digestion with trypsin, beads were discarded, removing mRNA-cross linked peptides from the SILAC mixture. Phosphopeptides were enriched by TiO2 chromatography, and proteome and phosphoproteome were analyzed by shotgun proteomics and MaxQuant (Cox and Mann, 2008). To determine the pull-down efficiency, heavy-over-light (H/L) SILAC ratios were multiplied by the spiked-in amount of input cell extract. A biological replica was generated by swapping SILAC labels and the whole experiment, including both biological replicates, was repeated a second time. Finally, the whole experiment was repeated a third time to include a medium-heavy labeled non-crosslinking control. For that, HEK293T cells were also labeled with medium-heavy amino acids and not irradiated with UV. Light and medium labeled cells were mixed before RIC and protocol was repeated as before.

As expected, the well-known mRBP HNRNPD was specifically pulled-down in the UV-irradiated samples but not in the non-crosslinked control (Figure 3.5 A, B, E and F). In contrast, Actin (ACTN1) was mainly detected in the input with only background signal in the UV-irradiated and non-UV-irradiated samples and therefore had low pull-down efficiency (Figure 3.5 C, D, E and F). Protein pull-down efficiencies in the forward and reverse label experiments were correlated, indicating overall good biological reproducibility (Figure 3.5 E and F).

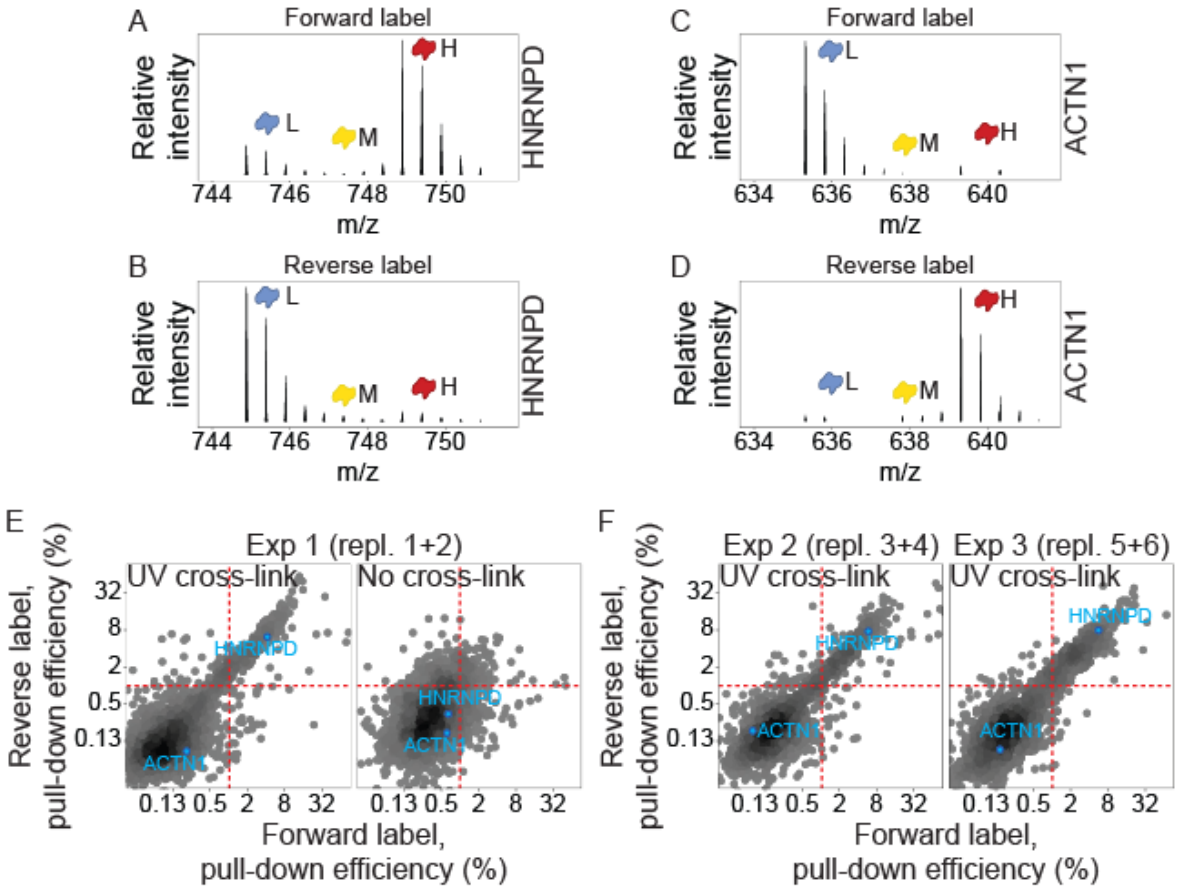


Figure 3.5 - Specific enrichment of mRBPs.

A, B, C and D) Exemplary spectra for HNRNPD and ACTN1. SILAC light (L), medium-heavy (M) and heavy (H) labeled chromatogram peaks are indicated.

E and F) Reproducibility overview of all three qRIC experiments. Dashed red lines indicate pull-down efficiency of 1 %.

In total, I identified 5,461 proteins and 2719 class I phosphorylation sites with localization probability > 0.75 in all six biological replicates combined. Protein quantification in at least one of the three forward and one of the three reverse experiments were required for further analysis and I computed the mean of all three forward and all three reverse experiments, followed by taking the mean-of-means from forward and reverse experiments to obtain a single average pull-down efficiency for 2831 proteins and 1243 class I phosphorylation sites (Figure 3.6 B).

Plotting the mean pull-down efficiencies for all quantified proteins and phosphosites showed a bimodal distribution ranging from 0 to 64 % and with one peak at about 0.2 and another one at 2-4 % (Figure 3.6 B). In control experiments without UV cross-

linking, the low efficiency peak was essentially unaltered while the higher efficiency peak was markedly reduced. Hence, the low efficiency peak appears to be largely due to non-specific contaminants, that is, proteins co-purifying with oligo(dT) beads that are not covalently cross-linked to mRNAs. Consistently, the vast majority of proteins with pull-down efficiencies greater than 1 % have been previously annotated as RNA-binding (Figure 3.6 C). Systematic benchmarking using annotated RBPs, mRBPs and cell line-specific mRBPs (Baltz *et al.*, 2012; Hentze *et al.*, 2018) revealed that qRIC can identify mRBPs with good sensitivity and specificity (Figure 3.6 D). At a cut-off of 1 % pull-down efficiency, the specificity was 90% with a sensitivity of 64 % for detecting HEK293-specific mRBPs. I also obtained pull-down efficiencies for 343 mRBPs in HEK293T cells, representing 39 % of cell-specific and 24 % of all previously annotated mRBPs, above the cut-off value. Applying the same thresholds to phosphoproteomics analysis led to quantification of 479 phosphorylation sites on 196 proteins.

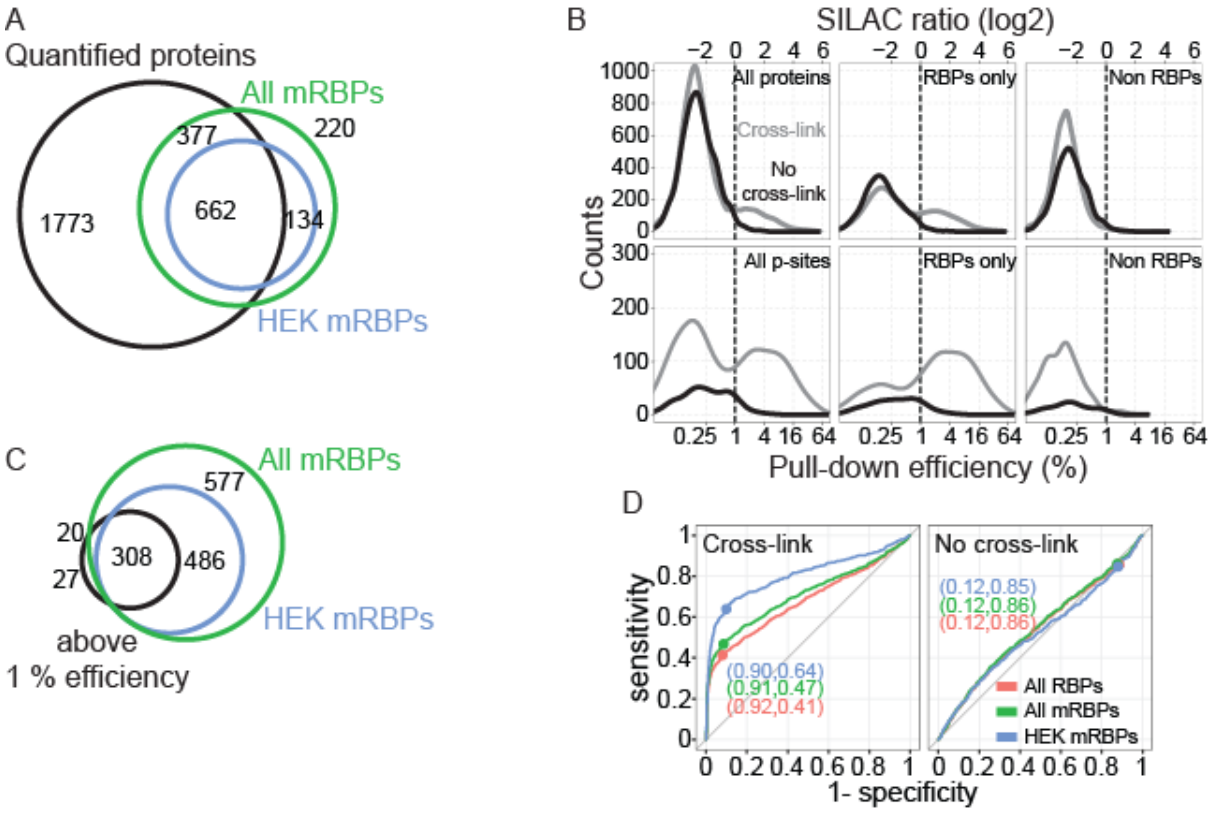


Figure 3.6 - Quantifying pull-down efficiency with qRIQ.

A) Venn diagram of RBPs quantified in qRIC, HEK293-specific mRBPs (blue) and all mRBPs identified in multiple cell lines (green).
B) Distribution of pull-down efficiencies quantified in qRIC. RBPs and phosphosites (p-sites) in RBPs were annotated accordingly.

C) Same as A, but proteins quantified with pull-down efficiency above the 1 % threshold are shown.
D) Receiver operating characteristic curves for RBP calling using three different sets of RBPs: HEK293-specific mRBPs (blue), all mRBPs identified in multiple cell lines (green), and all annotated RBPs (red). The specificity and sensitivity at cut-off of 1 % are indicated.

3.2.2. Canonical RBPs pull-down with high efficiency

Protein abundance can affect RNA-binding as highly abundant mRBPs might outcompete low abundant competitors on individual mRNA-binding sites. In that case, one would expect a positive correlation between mRBP copy number and pull-down efficiencies. In fact, I observed a small but significant positive correlation of pull-down efficiencies with absolute cellular abundance (Figure 3.7 A).

Furthermore, high affinity interaction with mRNAs in many RBPs is mediated by RNA-binding domains (RBDs) (Gerstberger, Hafner and Tuschl, 2014; Hentze *et al.*, 2018) and presence and numbers of RBDs are important factors for achieving high RNA-binding efficiencies (Lunde, Moore and Varani, 2007). To assess RBDs impact pull-down efficiencies I extracted domain annotation from the Pfam database (Mistry *et al.*, 2021) and selected domains annotated as RNA-binding (Gerstberger, Hafner and Tuschl, 2014) for proteins quantified in qRIC. As expected, the number of RBDs per protein strongly impacts pull-down efficiency (Figure 3.7 B). Pull-down efficiencies of proteins with and without annotated RBDs differed already below the 1 % threshold, confirming that our cut-off is indeed conservative. For all proteins containing only a single RBD I also compared pull-down efficiencies for different RBD types (Figure 3.7 C). Proteins with single KH_1 and RRM_1 domains showed higher pull-down efficiencies than proteins with a single DEAD box domain, consistent with the more transient nature of DEAD box helicases binding to RNA. Importantly, only a small and negative correlation with protein length was found, indicating that it is indeed the specific presence of domains rather than longer sequences that affect RNA-binding (Figure 3.7 D).

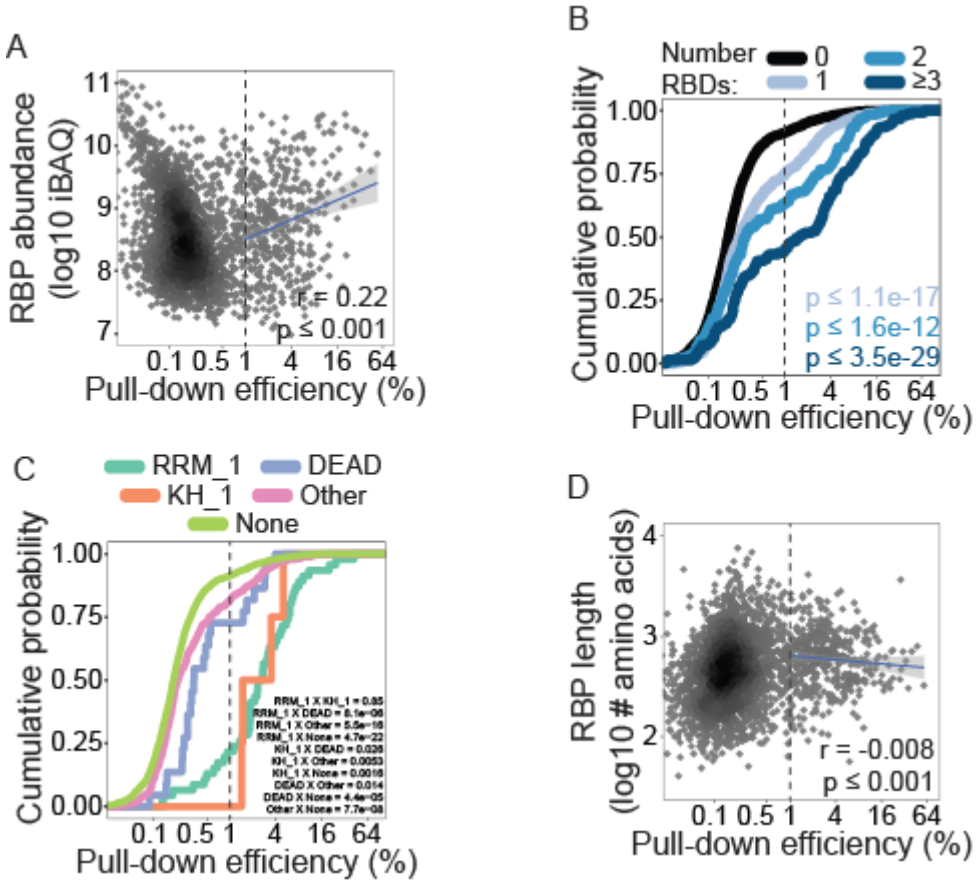


Figure 3.7 - Protein features correlating with pull-down efficiency.

A) Protein pull-down efficiency is plotted against protein abundance (iBAQ intensity) in the proteome. Pearson correlation was performed for proteins with pull-down efficiency higher than 1 %. Pearson correlation “r” value is shown. P-value was estimated from ten thousand randomized assignments of the data.

B) Pull-down efficiency cumulative probability of proteins grouped according to the number of RBDs. P-values from a two-sided Wilcoxon test for the comparison with proteins without annotated RBDs is indicated.

C) Same as B but only proteins with a single RBD are plotted and grouped according to RBD type.

D) Same as A but pull-down efficiency is plotted against protein length measured in number of amino acids.

3.2.3. *Pull-down efficiency correlates with RNA-binding* in vivo

The binding reaction between RBPs and target RNAs also depends on the availability of RNAs. Follows that mRBPs with more RNA-binding sites in the transcriptome are likely to be more occupied by RNAs and therefore present higher pull-down efficiencies in qRIC. To test this hypothesis, I obtained the number of peaks from PAR-CLIP and eCLIP data deposited in POSTAR2 and ENCODE (Zhu *et al.*, 2019; Van Nostrand *et al.*, 2020), respectively, as a proxy for the number of RNA-binding sites in the transcriptome. For PAR-CLIP, the downloaded datasets have been previously analysed with either PARalyser or Piranha peak calling tools (Corcoran *et al.*, 2011; Uren *et al.*, 2012), while eCLIP data has been analysed with CLIPPER peak caller (Lovci *et al.*, 2013). Finally, CLIP-seq data was originally produced in multiple cell lines (HEK293 and HEK293T for PAR-CLIP, and HepG2 and K562 for eCLIP) and these were analysed in parallel.

Correlation was highest for PAR-CLIP results produced in HEK293T cells (R = 0.85 and 0.69 for PARalyse and Piranha, respectively) (Figure 3.8 A). PAR-CLIP results from HEK293 cells also showed moderate to high correlation with pull-down efficiency depending on the peak calling tool (R = 0.27 and 0.47 for PARalyse and Piranha, respectively). K562 cells eCLIP data showed only moderate correlation (R = 0.30). In conclusion, CLIP-seq data produced in HEK293T, the same cell line used in qRIC, showed the highest correlation, although not the same RBPs were investigated by multiple CLIP-seq methods. Second highest correlation was found with PAR-CLIP data produced in the parental cell line HEK293. It is not clear however if the lower correlation in eCLIP-derived number of RNA-binding sites comes from the CLIP-seq method or cell lines of choice. Moreover, these results suggest strong correlation in RBPs pull-down efficiency and the number of RNA-binding sites.

Different RBPs often bind to similar RNA motifs. I therefore compared pull-down efficiencies of RBPs grouped together based on their shared binding sites via combinatorial clustering (Li *et al.*, 2017b). Interestingly, RBPs interacting with the similar motifs tend to have similar pull-down efficiencies (Figure 3.8 B). This was even true for proteins with vastly absolute abundance differences in HEK293T cells such as LIN28B and FUS. Interestingly, CPSF4 was more efficiently pulled down than CPSF1, CPSF2 and CPSF3, even though all four proteins are annotated to interact with the

same AAUAAA motif (Li *et al.*, 2017a). All four proteins are members of CPSF (cleavage and polyadenylation specificity factor), a multiprotein complex that recognizes the AAUAAA motif in the polyadenylation signals of precursor mRNAs (Keller *et al.*, 1991). While initially CPSF1 was thought to recognize and directly bind this sequence motif, more recent data demonstrated that RNA-binding occurs via CPSF4 and WDR33 instead (Keller *et al.*, 1991; Chan *et al.*, 2014; Schönemann *et al.*, 2014; Clerici *et al.*, 2018). This is consistent with my observation that CPSF4 and WDR33 but not CPSF1, CPSF2 and CPSF3 are pulled down with efficiencies above the cut-off (Figure 3.8 B).

Altogether, pull-down efficiency is strongly impacted by the RNA affinity, as indicated by the higher pull-down efficiency of mRBPs with multiple RBDs and the similar pull-down efficiency of mRBPs with the same RNA-binding motif. Also, analysis of the number of RNA-binding sites in the transcriptome indicate a linear relationship between qRIC pull-down efficiency and RNA-binding *in vivo*. I conclude that pull-down efficiencies determined via qRIC provide a meaningful read-out for comparing RNA-binding of multiple RBPs *in vivo*. It is however important to highlight that pull-down efficiency is not an absolute quantification of the fraction of the cellular mRBP pool interacting with mRNAs in the cell (mRBP occupancy), as mRNA isolation and UV cross-linking efficiencies also strongly impact it (Vieira-Vieira and Selbach, 2021). Therefore, pull-down efficiency should not be interpreted as a direct readout for the actual fraction of an RBP bound to mRNAs *in vivo*, but provide a meaningful measurement correlating to RNA-binding.

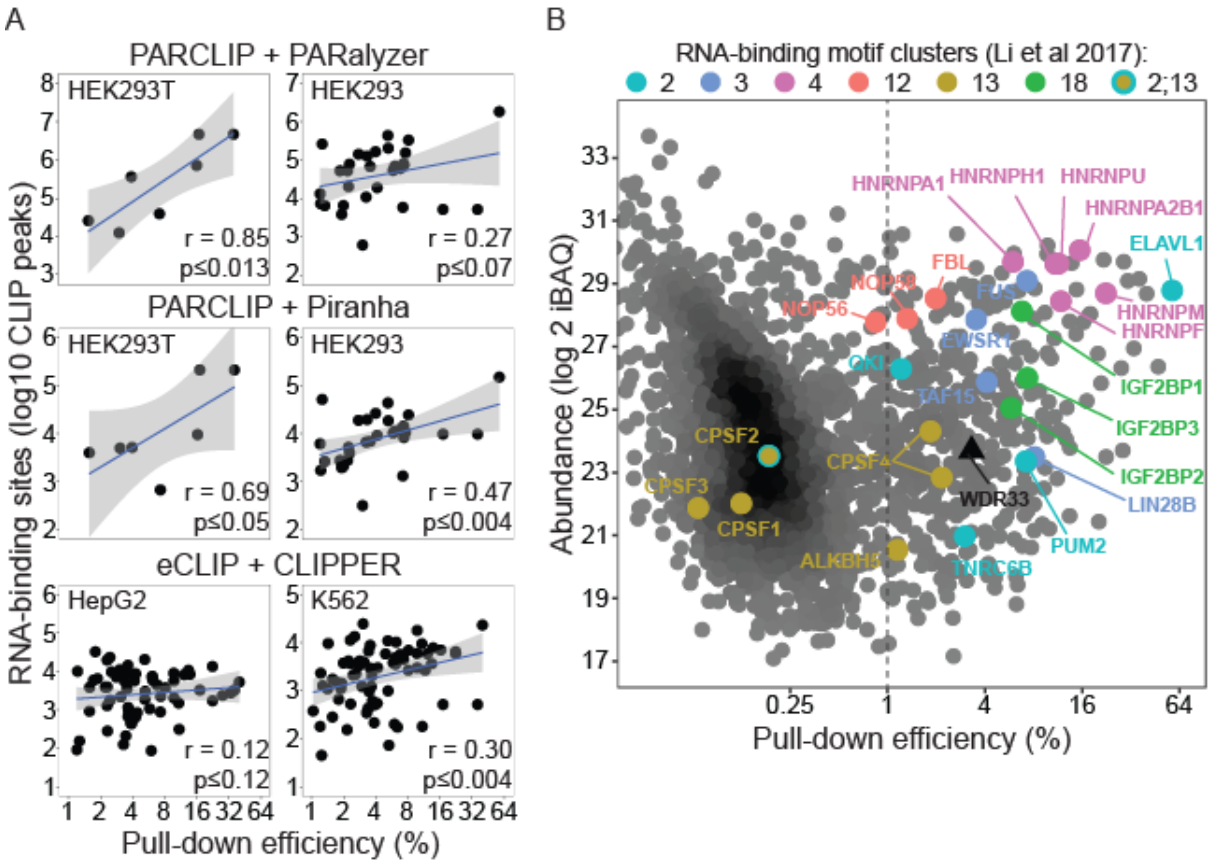


Figure 3.8 - Pull-down efficiency correlation with RNA-binding.

A) Number of peaks from CLIP data from individual RBPs produced in several cell types and analysed with different peak calling tools. The cell line from which the data was generated is indicated (HEK293T, HEK293, HepG2 or K562). Pearson correlation "r" value is shown. A p-value was calculated from 10.000 randomized samples of the data.
B) Same as Figure 3.7 A but proteins with RNA-binding motifs are highlighted. WDR33 is emphasized (see main text). The black dashed line indicates 1 % pull-down efficiency.

3.3. Identifying regulatory phosphorylation sites in RBPs

Several parts of this subsection are included in a manuscript submitted for publication to the Molecular Cell journal (Vieira-Vieira et al, submitted).

3.3.1. Calculating a delta pull-down efficiency for phosphorylated sites in mRBPs

To systematically identify phosphorylation sites with regulatory potential, I compared the pull-down efficiencies of individual sites with the efficiency of the corresponding host protein. To this end, I computed delta efficiencies as the ratio of the SILAC ratios. This yielded 395 phosphorylation sites in 166 RBPs (Figure 3.9 A). Only reproducibly quantified phosphorylation sites for which the corresponding host protein was also reproducibly quantified and either the phosphorylated peptide or the host protein passed the 1 % pull-down efficiency threshold were considered in this analysis.

The delta value indicates changes in the pull-down efficiency associated with a phosphorylation event, where positive and negative log2 delta efficiencies indicate increased and decreased pull-down, respectively. The great majority of phosphorylation sites show log2 delta efficiencies close to zero (Figure 3.9 B), suggesting that most phosphorylation events do not impact mRBP pull-down. Using a threshold at delta efficiency of 2-fold change, I found a total of 115 phosphorylation sites potentially involved in mRBP functional regulation. Interestingly, compared to the unmodified peptides from the same host proteins, the distribution of the delta efficiency values is significantly skewed to positive values for phosphorylation sites (Figure 3.9 C).

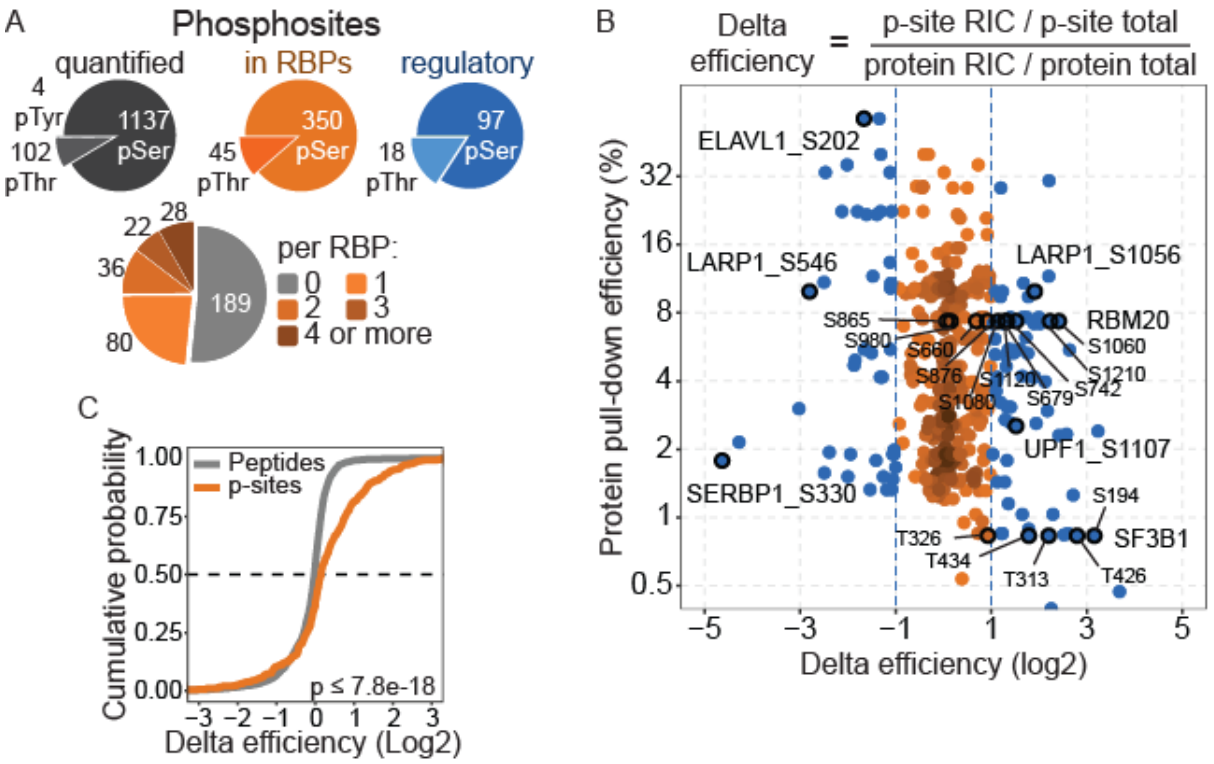


Figure 3.9 - Delta pull-down efficiency of phosphorylation sites in mRBPs.

A) Overview of quantified phosphorylation sites in qRIC.
B) Delta efficiency of individual phosphorylation sites is plotted with the pull-down efficiency of the host mRBP. Phosphorylation sites with delta efficiency higher than two (dashed blue lines) are highlighted in blue. The sites indicated in the plot are discussed in the text. The delta efficiency formula is also shown.
C) Cumulative probability of delta efficiency for unmodified peptides from mRBPs identified by at least one phosphorylation site (p-sites). Delta efficiency of phosphorylated sites (orange) was compared to unmodified peptides (grey) and the p-value from the Wilcoxon rank sum test is shown.

3.3.2. Delta efficiency is not driven by peptide misidentification

During analysis of mass spectrometry data, peptide amino acid sequences are assigned by matching the measured fragmentation spectra to the predicted fragmentation spectra of peptides in the search space, and peptides are then assigned into protein groups (Cox *et al.*, 2011). Peptide misidentification happens when a fragmentation spectra is matched to the wrong peptide sequence. Misidentified peptides have only a marginal impact on protein quantification as the wrong value is diluted away by the other assigned peptides. However, quantification of proteoform pull-down efficiency in qRIC uses peptide level information, and peptide misidentification can have drastic consequences (Bogdanow, Zauber and Selbach,

2016). For instance, a peptide originally from an RBP with high pull-down efficiency that is misidentified to a low efficiency RBP will present a drastically large delta efficiency to the incorrectly matched protein. Keeping track of peptide identification quality during data analysis is therefore crucial. The software Andromeda (incorporated in MaxQuant) generates an identification score based on the similarity of the measured and the predicted fragmentation spectra: lower scored peptides are more likely to be misidentified (Cox *et al.*, 2011).

Another type of misidentification of peptides occurs at the protein level. After identifying peptide sequences, these are used for annotating proteins by matching to the proteome. For each peptide, one of the three options will follow. The peptide sequence either matches a single protein sequence in the search space (unique match), matches two or more protein sequences (razor peptides), or does not match any protein sequences at all. While not-matching peptides are discarded during protein annotation, unique peptides are the strongest evidence of the presence of a protein in the sample. Shared peptides however pose a problem. If a given peptide sequence perfectly matches two proteins, which one is the correct one? Softwares like MaxQuant have found ways to deal with this situation using the simplicity principle, however it might still happen that a peptide is wrongly assigned to a protein. As was the case for peptide misidentification, misannotation of peptides into proteins can lead to drastic delta pull-down efficiencies in qRIC. Isoforms are a particularly problematic case in this context. For example, mRBP isoforms might show discrepant pull-down efficiencies, but without unique peptides to differentiate them the average efficiency will be assigned to both isoforms. As a consequence, individual peptides will show large delta efficiencies according to the expression level of both isoforms.

To investigate the extent to which peptide misidentification and misannotation of peptides in protein groups impacts pull-down efficiency of proteoforms in qRIC, I calculated the delta efficiency of unmodified peptides and classified them according to their Andromeda identification scores or uniqueness in protein group. Importantly, delta efficiency values are similarly distributed in low and high scoring peptides, as well as in unique peptides and non-unique peptides (Figure 3.10). For comparison, phosphorylated peptides showed a broader and skewed distribution to positive log2 delta efficiency values. In conclusion, although I can not rule out the possibility that misidentification and misannotation impact individual cases, this does not seem to be

generally the case in my data. Therefore, phosphosite delta efficiency likely reflects differential pull-down efficiency of proteoforms instead of misidentification and misannotation of peptides.

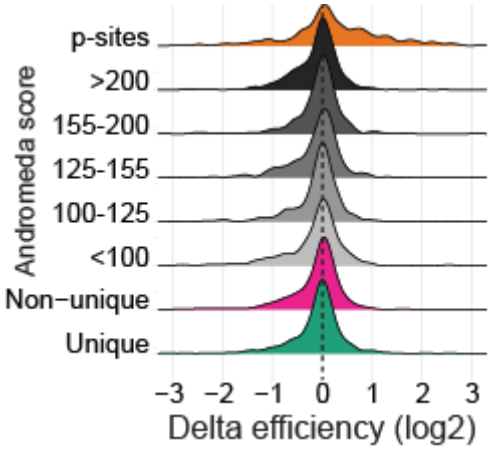


Figure 3.10 - Delta efficiency distribution of unmodified peptides in mRBPs quantified with qRIC.

3.3.3. Delta efficiency reflects expected RNA charge repulsion and phase separation by phosphorylation in RBPs

Confident that delta efficiency largely reflects biological differences in the pull-down efficiency of phosphorylated proteoforms, I assessed the performance of my method in identifying previously described regulatory phosphorylation sites in the global level and at individual RBPs.

Protein phosphoryl groups and RNAs are expected to repel each other at physiological conditions due to charge repulsion. To test the impact of charge repulsion on pull-down efficiency, I measured the linear distance expressed in the number of amino acids between phosphorylated sites and the nearest RNA-binding site (Bae *et al.*, 2020) in the host mRBP (Figure 3.11 A). If charge repulsion negatively impacts RNA-binding, I expect phosphorylated sites in closer proximity to show decreased pull-down efficiency. In fact, phosphorylation sites with negative log2 delta efficiencies are significantly closer to the nearest RNA-binding site, while those with positive log2 delta efficiency are significantly further away. The total protein length in all groups show no significant differences, indicating that my findings are specific (Figure 3.11 B). These

findings support the notion that phosphorylation negatively affects RNA-binding *in vivo*. However, the peptides within an RBP that are covalently cross-linked to the RNA are lost in the qRIC workflow and, therefore, my analysis is biased against the detection of activating phosphorylation sites near cross-linking residues.

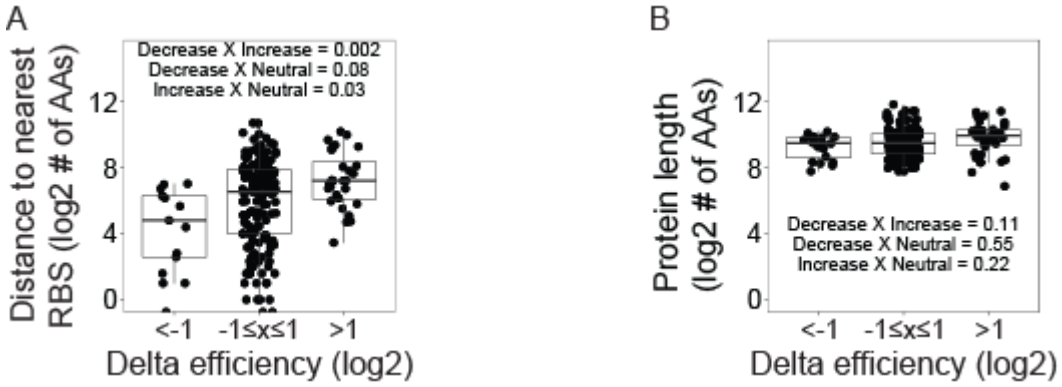


Figure 3.11 - Phosphorylation linear distance from RNA-interacting sites.

A) Distance to the nearest RNA-binding site (RBS) measured in number of amino acids for phosphosites with reduced, neutral or increased pull-down efficiency.
B) Total protein length in number of amino acids of proteins hosting phosphosites analysed in A.

Membraneless organelles dissolve during nuclear-envelope breakdown in mitosis, only to be re-formed once mitosis is complete. Recent work shows that DYRK3 acts as a central dissolvase of several types of membraneless organelle during mitosis (Rai *et al.*, 2018). Through its intrinsically disordered N-terminal domain, DYRK3 partitions in condensates where the kinase activity is required for dissolution of aggregates (Wippich *et al.*, 2013). Because RBPs are in close contact with RNA in condensates, I rationalized that DYRK3 mediated dissolution of membraneless organelles should lead to reduced RBP-RNA interaction. Accordingly, I found significantly smaller qRIC delta efficiency values for phosphorylation sites in DYRK3 interactors (Figure 3.12). In conclusion, qRIC is sensitive enough to detect global changes in RNA-binding of RBPs involved in phase separation of condensates mediated by DYRK3.

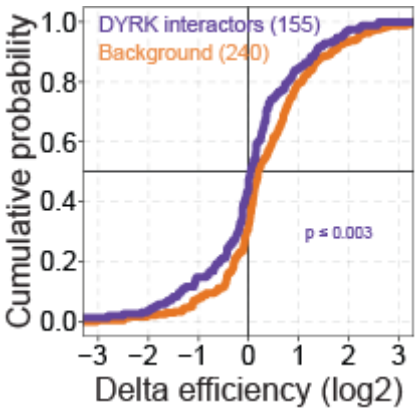


Figure 3.12 - Phosphosites in DYRK3 interactors show lower delta efficiency in qRIC.

Delta efficiency of phosphorylated sites in DYRK3 interaction partners (purple) was compared to non-interactors (orange) and the p-value from the Wilcoxon rank sum test is shown. Number of phosphorylation sites in each group is shown in parenthesis.

3.3.4. Individual regulatory phosphorylation sites are identified by qRIC

Next, I assessed the performance of qRIC in identifying regulatory phosphorylation sites in individual proteins. For that, I took a detailed look at three well-characterized mRBP with phosphorylation-dependent functions: SF3B1, UPF1 and ELAVL1.

SF3B1 (a.k.a. SF3B155 or SAP155) is a member of the SF3b multi-protein component of the U2 snRNP and directly binds the branching site in the pre-mRNA during the splicing cycle (Cretu *et al.*, 2016). Several phosphorylation sites in the threonine and proline (TP) rich region of SF3B1 have been shown to regulate protein function (Boudrez *et al.*, 2002; Girard *et al.*, 2012). In my qRIC analysis, five phosphorylation sites in SF3B1 were quantified: S194, T313, T326, T426 and T434 (Figure 3.9 B). Both the T313 and the T326 phosphorylation sites have been previously found in isolated active spliceosomes (Girard *et al.*, 2012). In particular, T313 phosphorylation in SF3B1 is a well-described marker of spliceosome activity (Boudrez *et al.*, 2002; Girard *et al.*, 2012). In qRIC, the phosphorylated proteoform containing the pT313 site shows 5.7 % pull-down efficiency, in comparison to only 0.83 % of the unmodified SF3B1 protein (Figure 3.13 A). Accordingly, phosphorylated T313 SF3B1 has log2 delta efficiency of 2.20, reflecting the much higher pull-down efficiency of the phosphorylated proteoform. The finding that phosphorylated T313 SF3B1 is pulled-down with higher efficiency was further validated using specific antibodies (Figure 3.13 B). For that, the mRNA-bound proteome of HEK293T cells was obtained with RIC and analysed on SDS-PAGE

followed by immunoblotting with antibodies against SF3B1 and phosphorylated T313 SF3B1 (Figure 3.13 B, upper panel). Although the estimated pull-down efficiency with both methods vary, in all five biological replicates of this experiment, the amount of T313 phosphorylated SF3B1 pulled down was higher than SF3B1 on average by 4.13 (+- 1.41) fold, which is very close to the estimated 4.6 delta efficiency in qRIC (Figure 3.13 B, lower panel).

In addition to T313, the other four phosphorylation sites (S194, T326, T426, and T434) in SF3B1 were also found with high delta efficiency values in my qRIC screen (Figure 3.8 B). Although the functional relevance of these four sites has not been investigated, they are located near the T313 site in the N-terminal domain of SF3B1. This intrinsically disordered region of SF3B1 is not part of the core structure of the SF3b complex, and has a proposed regulatory function during the splicing cycle (Cretu *et al.*, 2016). Taken together, these results highlight the ability of qRIC to identify previously known phosphorylation sites in a well characterized RBP involved in splicing.

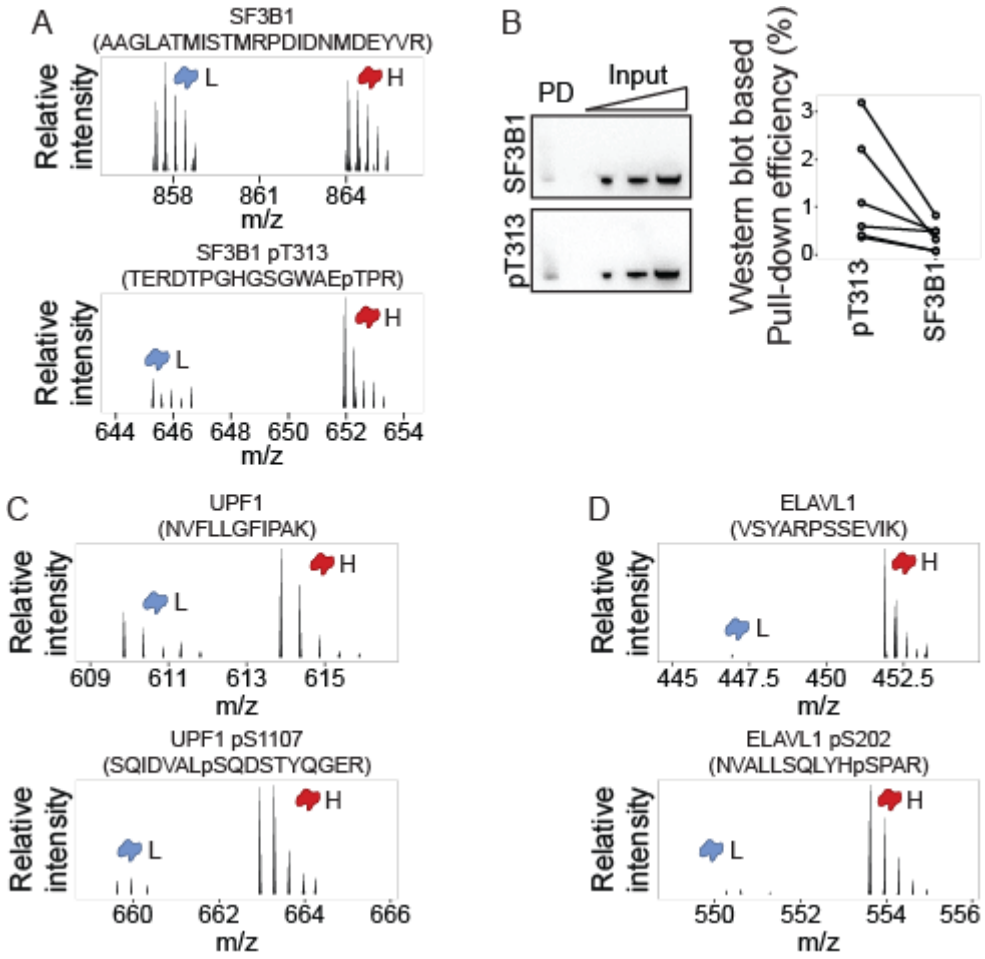


Figure 3.13 - Phosphorylation regulation of pull-down efficiency of individual sites.

A) The phosphorylated peptide spectra (bottom row) and an exemplary unmodified peptide spectra (top row) for SF3B1 T313 phosphorylation sites. SILAC light (L) and heavy (H) labeled chromatogram peaks are indicated. The peptide sequences are also indicated in parentheses.
B) Western blot based quantification of pull-down efficiency for phosphorylated T313 SF3B1 (pT313). RIC pull-downs (PD) were immunoblotted with SF3B1 and pT313 SF3B1 specific antibodies (left panel). Increasing amounts of the input whole cell lysate material (Input) were analysed in parallel. Pull-down efficiency was quantified (right panel) from the signal ratio in PD and input.
C) Same as in A, but exemplary peptides annotated to UPF1 and the phosphorylated S1107 site are shown.
D) Same as in A, but exemplary peptides annotated to ELAVL1 and the phosphorylated S202 site are shown.

In addition to SF3B1, a regulatory phosphorylation site in UPF1 was also identified in qRIC. Phosphorylation of S1107 was the only quantified site in UPF1 in my qRIC analysis (Figure 3.13 C). While S1107 phosphorylation is not essential for the degradation of target mRNAs, it times UPF1 activity and inhibition of the downstream cascade leads to accumulation of S1107 phosphorylated UPF1 on target mRNAs

(Durand, Franks and Lykke-Andersen, 2016). UPF1 is then dephosphorylated once mRNA degradation has proceeded. Therefore, unmodified UPF1 is found in cells both associated with mRNAs or free, while UPF1 is phosphorylated at S1107 when bound to mRNAs. In agreement, phosphorylated S1107 UPF1 presented higher pull-down efficiency (log2 delta efficiency of 1.5) in qRIC, suggesting that an increased fraction of the phosphorylated form is bound to mRNA.

Phosphorylation of S202 in ELAVL1 was also identified as a regulatory site in qRIC (Figure 3.13 D). ELAVL1 binds AU-enriched elements in the 3′-UTR to control splicing, localization, stability, and translation of target mRNA (Grammatikakis, Abdelmohsen and Gorospe, 2017). Central to its function, ELAVL1 shuttles in between cytosol and nucleus in cells in response to several stimuli. In fact, several modification sites in ELAVL1 have been shown to regulate its subcellular location (Grammatikakis, Abdelmohsen and Gorospe, 2017). Particularly upon phosphorylation of S202, ELAVL1 interaction with nuclear 14-3-3 is enhanced and the protein is retained in the nucleus, reducing association with cytosolic mRNA targets (Kim, Abdelmohsen, *et al.*, 2008). Consistent with these observations, phosphorylated S202 ELAVL1 shows a three-fold pull-down efficiency reduction in qRIC (log2 delta efficiency of -1.64). Hence, the reduced association of ELAVL1 phosphorylated on S202 is consistent with the known regulatory function of this phosphorylation site.

3.4. Novel candidate regulatory phosphorylation sites in mRBPs

The findings presented above show that qRIC identifies phosphorylation sites in RBPs regulating several functions, including membraneless organelles stability, RNA-binding, splicing, RNA degradation, and protein subcellular localization. These exemplify non-exhaustively the versatility of molecular mechanisms captured with qRIC. Next, I focused my attention on sites that have not yet been functionally characterized.

3.4.1. Phosphorylated S330 SERBP1 is recruited to actively translating ribosomes

In my qRIC analysis, two phosphorylation sites in SERBP1 (a.k.a. CGI-55 or PAIRBP1) were quantified: S234 and S330. While S234 phosphorylation did not affect

pull-down efficiency, S330 phosphorylation in SERBP1 tops as the most regulated site with 19 times lower pull-down efficiency than the host protein (log2 delta efficiency of -4.32) (Figure 3.14 a). This site is located near an RGG domain spanning amino acids 364 to 380 and essential for SERBP1 function (Ahn *et al.*, 2015), hinting at a regulatory role. Interestingly, S330 in SERBP1 has been shown to directly interact with RNA (Bae *et al.*, 2020). Reanalysis of previously published proteomics characterization of ribosome interacting proteins (Imami *et al.*, 2018) shows that SERBP1 phosphorylated at S330 co-sediments almost exclusively with the 40S structure, unlike the bulk SERBP1 that also co-sediments with polysomes (Figure 3.14 B). Accordingly, the strong effect on SERBP1 pull-down efficiency upon phosphorylation could then be explained by promotion of translation termination and consecutively release of the translating mRNAs from SERBP1-containing ribosomes, leading to reduced interaction of SERBP1 and mRNA.

To further investigate the functional impact of SERBP1 S330 phosphorylation, HEK293T cell lines stably overexpressing triple FLAG-tagged SERBP1 mutant variants were used. These cell lines were generated in collaboration with Orsalia Hazapis in the lab of Markus Landthaler at MDC in Berlin, Germany. Results from the SERBP1 phosphomimetic mutant carrying the single amino acid alteration of S330 to aspartic acid (S330D) were compared to the wild type protein and the non-phosphorylatable mutant with an alanine substitution (S330A). Mutant SERBP1 expression was induced with tetracycline for 48 hours before cellular lysates were layered on top of a sucrose density gradient solution and subjected to ultracentrifugation. The separation pattern of molecular complexes of different sizes on the sucrose density gradient was assessed by absorbance of 260 nm UV light (Figure 3.14 C) and 16 sample fractions were collected for analysis of protein content (Figure 3.14 D).

Similar in all samples, analysis of the sucrose separation pattern shows high amounts of proteins and RNAs in less dense portions of the sucrose density gradient where small molecular complexes are expected (fractions 1 to 4). Following the density gradient in the direction of the heavier complexes, three peaks of increasing size were visible (fractions 5 to 9) and identified according to abundance of RPL10 and RPS5: The small subunit peak (40S), large subunit peak (60S) and the peak with fully assembled ribosomes (80S). Fractions 10 to 16 contained polysome peaks with

increased molecular size corresponding to mRNA molecules with multiple translating ribosomes (polysomes). Analysis of the protein content per fraction of the gradient shows that the endogenously expressed SERBP1 is found in the 40S, 80S and polysome sedimentation fractions, similar to what has been previously shown by others (Imami *et al.*, 2018; Muto *et al.*, 2018). Surprisingly, all three mutant variants of the FLAG-tagged SERBP1 (wild-type, S330A and S330D) co-sedimented equally well with polysomes in fractions 10 to 16, suggesting that all S330 SERBP1 mutants were similarly recruited to active ribosomes. Therefore, mutating S330 to aspartic acids or alanine does not change the protein interaction with translating ribosomes, suggesting that this amino acid is not essential for SERBP1 function in translation.

On one hand, phosphomimetic mutation of S300 does not fully mimic the phosphorylation event and therefore might not lead to full regulation. On the other hand, as a hibernation factor, SERBP1 S330 phosphorylation might play a specific role during nutrient starvation and cells grown under optimal conditions might not respond (Anger *et al.*, 2013; Colleti *et al.*, 2019; Wells *et al.*, 2020). While a plausible explanation for the decreased pull-down efficiency observed in qRIC, more experiments will be necessary to understand the relationship between phosphorylated SERBP1 at S330 and actively translating ribosomes, if there is in fact one.

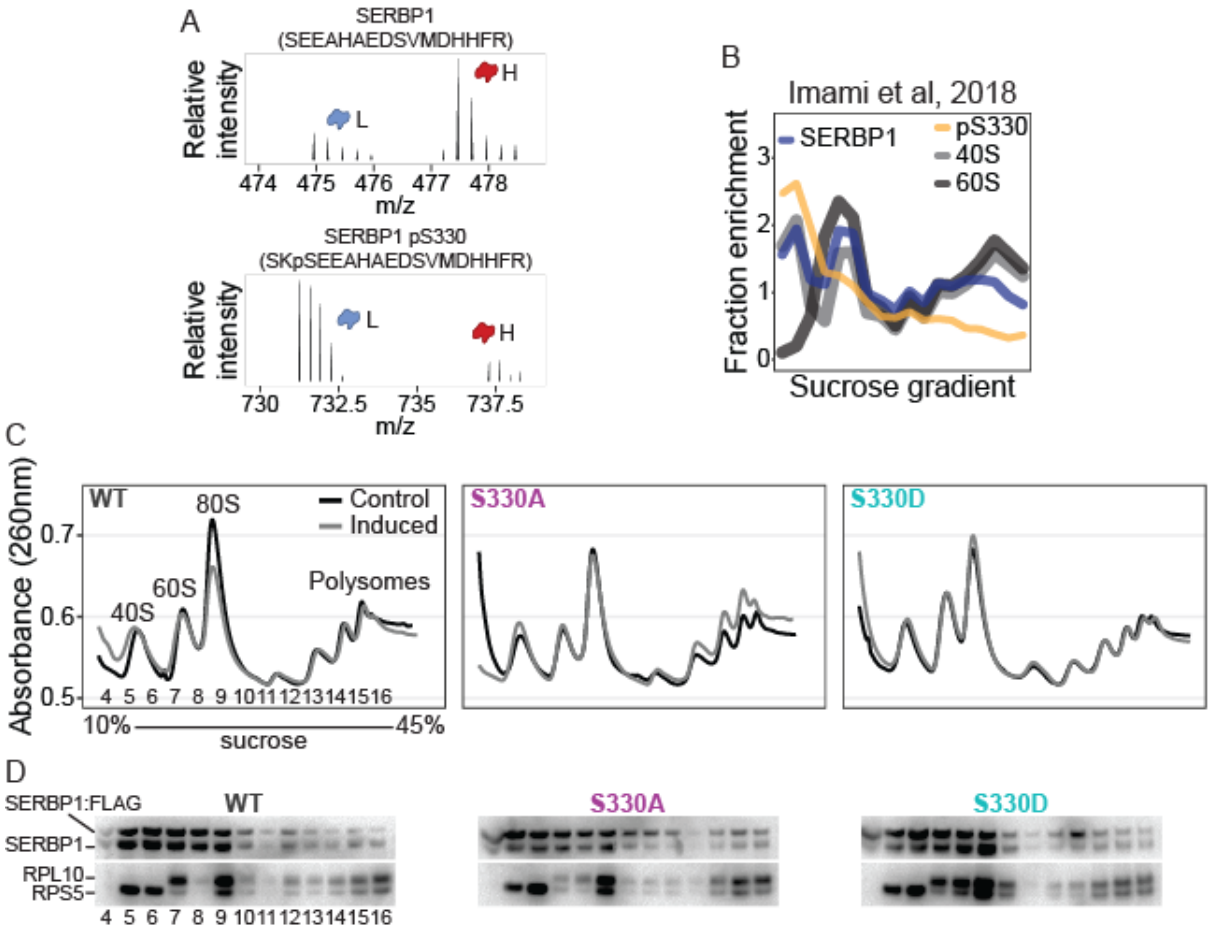


Figure 3.14 - SERBP1 S330 phosphorylation does not regulate protein interaction with polysomes.

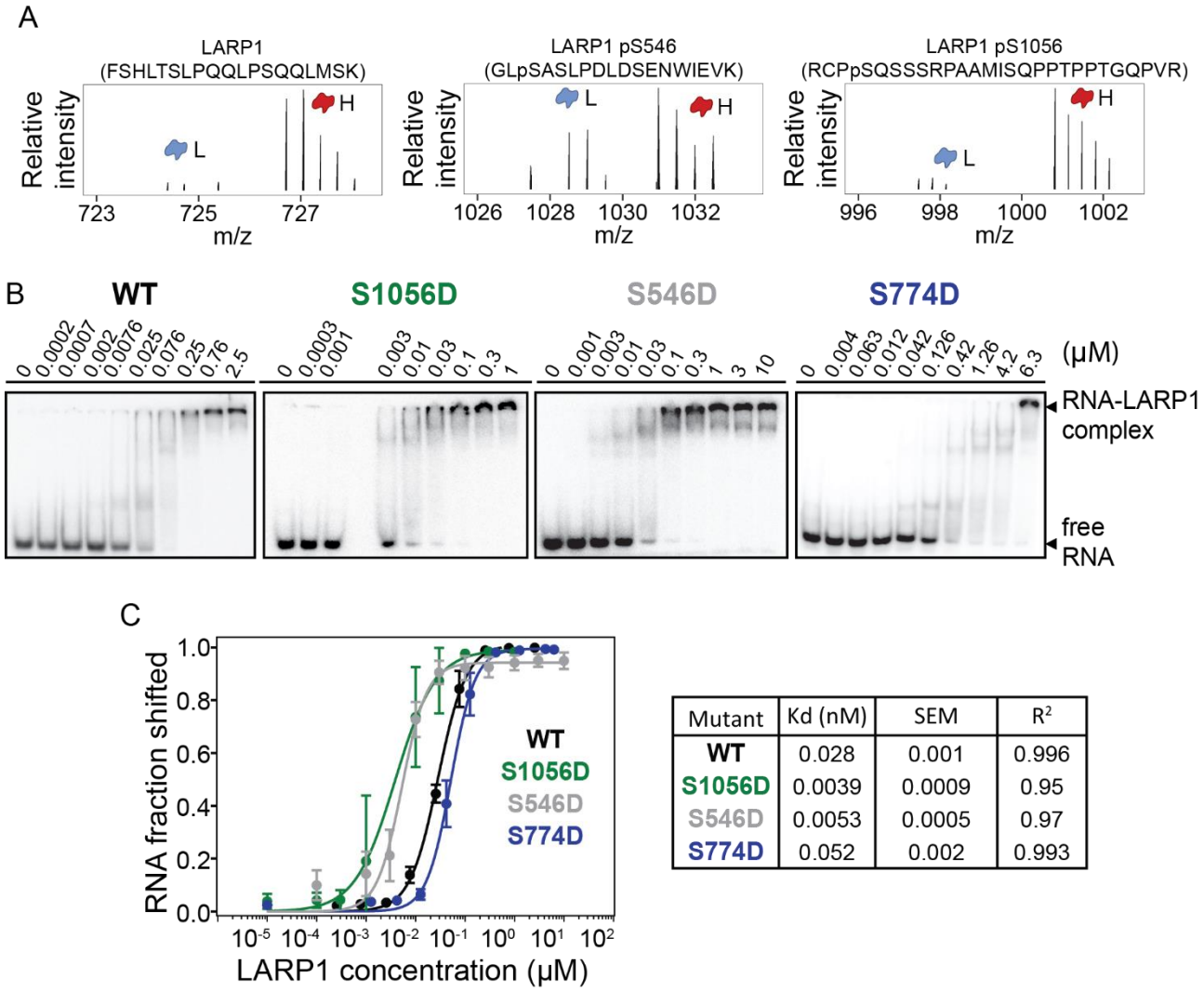
A) The phosphorylated peptide spectra (bottom) and an exemplary unmodified peptide spectra (top) for SERBP1 S330 phosphorylation sites. SILAC light (L) and heavy (H) labeled chromatogram peaks are indicated. The peptide sequences are also indicated in parentheses.
B) Sucrose fractionation enrichment of SERBP1 and the phosphorylated S330 site (pS330). Data from (Imami *et al.*, 2018). Average enrichment of all ribosomal small subunits (40S, light grey) and large subunits (60S, dark grey) are shown for reference.
C) Sucrose density gradient fractionation of cellular lysates from HEK293 cells stably expressing mutant forms of SERBP1. Cells without induced expression of SERBP1 were used as control.
D) Protein analysis of fractions from the sucrose density gradient visualized in C. Only cells with induced SERBP1 expression were analyzed.

3.4.2. LARP1 S1056 and S546 phosphorylation regulates binding affinity but not translation of TOP mRNAs

Twelve candidate regulatory phosphorylation sites were identified in LARP1. Out of these, only 2 sites showed increased efficiency (S143 and S1056, with 1.2 and 1.9

log2 delta efficiency respectively) and a single site showed decreased efficiency (S546, with -2.8 log2 delta efficiency) compared to the host protein (Figure 3.15 A). The S1056 site is an *in vitro* and *in vivo* substrate of AKT1 (Hong *et al.*, 2017) and resides just N-terminal to the DM15 domain that mediates binding to the m^7Gppp cap and the 5′TOP motif of target mRNAs (Lahr *et al.*, 2015, 2017). In comparison, S546 is nudged between the Pam2 and the RRM-L5 motifs, a region recently described to mediate simultaneous binding of LARP1 to the TOP motif and poly-A tail of target mRNAs (Al-Ashtal *et al.*, 2021), where no regulatory kinase has been identified. However, the S143 site is located in a portion of the protein not required for RNA binding and without known function. Therefore, I chose to investigate the potential regulatory impact of S1056 and S546 phosphorylation on LARP1 function.

For that, I joined forces with Roni Lahr in the lab of Andrea Berman at the University of Pittsburgh in Pittsburgh, United States of America. Full length LARP1 bearing single amino acid mutations were purified and used for EMSA (Figure 3.15 B). RNA oligos 5′-end labelled with [γ-^{32}P]-ATP covering the TOP motif of RPS6 at saturating concentration was incubated with increasing amounts of purified LARP1 mutants. At low concentrations of LARP1, the radioactive RNA migrated all the way down in native gels. However, at higher concentrations, LARP1 interacts with the RNA forming a complex of larger mass, identified by the shifted fraction of radioactive RNA in the gel (Figure 3.15 B, upper panel). By measuring the fraction of shifted RNA, we estimated the *in vitro* affinity (Kd) of LARP1 mutants to the RPS6 TOP mRNA motif (Figure 3.15 B, lower panel). Surprisingly, substitution of S546 or S1056 to aspartic acid increased TOP mRNA affinity by 5 and 7 fold, respectively. The same phosphomimetic mutation at S774 did not affect affinity. These results suggest that S1056 and S546 phosphorylation in LARP1 regulate TOP mRNA binding.



Mutant	Kd (nM)	SEM	R^2
WT	0.028	0.001	0.996
S1056D	0.0039	0.0009	0.95
S546D	0.0053	0.0005	0.97
S774D	0.052	0.002	0.993

Figure 3.15 - Phosphomimetic mutations on LARP1 affect association with the TOP motif of RPS6.

A) The phosphorylated peptide spectra for S546 and S1056 (middle and right, respectively) and an exemplary unmodified peptide spectra (left) for LARP1. SILAC light (L) and heavy (H) labeled chromatogram peaks are indicated. The peptide sequences are also indicated in parentheses.
B) Electrophoretic mobility shift assays using the TOP mRNA sequence from RPS6 and purified LARP1 phosphomimetic mutants.
C) Quantification of the fraction of RNA shifted by complex formation with LARP1 mutants. Kd values were extracted by fitting curves to the Hill equation. The mean of three independent experiments is shown with error bars indicating SEM. Kd and SEM values, and R^2 of the fitted curve are summarized in the table on the right.

To further investigate phosphorylation impact on LARP1 function *in vivo*, I generated stable HEK293 cell lines overexpressing the full length wild-type (WT) or mutant variants of LARP1 fused to BioID and a FLAG tag under the control of a tetracycline inducible promoter. Phosphomimetic LARP1 mutants were generated by substituting

phosphorylation sites by aspartic acid (S546D or S1056D), while substitution to alanine rendered a non-phosphorylatable mutant proteins (S546A or S1056A). The BioID fusion enzyme catalyzes the conversion of D-Biotin into the reactive derivative biotinoyl-5'-adenosine monophosphate (bioAMP) that promiscuously reacts with primary amines (e.g. in lysine side chains). Due to specific mutations in the enzyme, the reactive derivative diffuses from the enzyme's active core to label amines within the estimated range of approximately 10-30 nm (Kim and Roux, 2016). In this way, the fused BioID is used for biotinylation of endogenous proteins proximal to the tagged protein. Following enrichment of biotinylated proteins with streptavidin beads and mass spectrometry analysis, the bait protein interactome is mapped *in vivo* (Youn *et al.*, 2018; Go *et al.*, 2021).

Cells were grown in SILAC media and mutant LARP1 expression was induced with tetracycline for 24 hours. At this point, cells were incubated with 50 µM of D-Biotin for another 24 hours. Equal numbers of differentially labeled cells were mixed before sample preparation to minimize technical variability. SILAC ratios in BioID pull-downs reflect relative protein proximity to bait LARP1 *in vivo*. In parallel, whole proteome changes in differentially labelled cells expressing mutant LARP1 were quantified. Finally, a second biological replica of the experiment was generated by reversing the SILAC labels (reverse label), and average SILAC ratios were used for data analysis (Figure 3.16 A). LARP1 expression level and biotin pull-down was similar in all groups. To exclude the impact of proteome background, specific interactors were selected if the SILAC ratio in the BioID pull-down was 1.5 fold higher than in the whole cellular lysate. However, compared to WT LARP1, only few proteins differentially interacting with LARP1 mutants were identified (Figure 3.16 B). At the same time, none of the differential interactors have been previously described as LARP1 interactors (Berman *et al.*, 2021), nor have they been described to regulate translation initiation on TOP mRNAs. Although the differential interactors could be involved with other cellular functions of LARP1, such as mRNA stability and localization of RNAs to stress granules (Berman *et al.*, 2021), the BioID pull-down results suggest that LARP1 phosphomimetic and non-phosphorylatable mutants behave similarly to the wild-type LARP1 in cells.

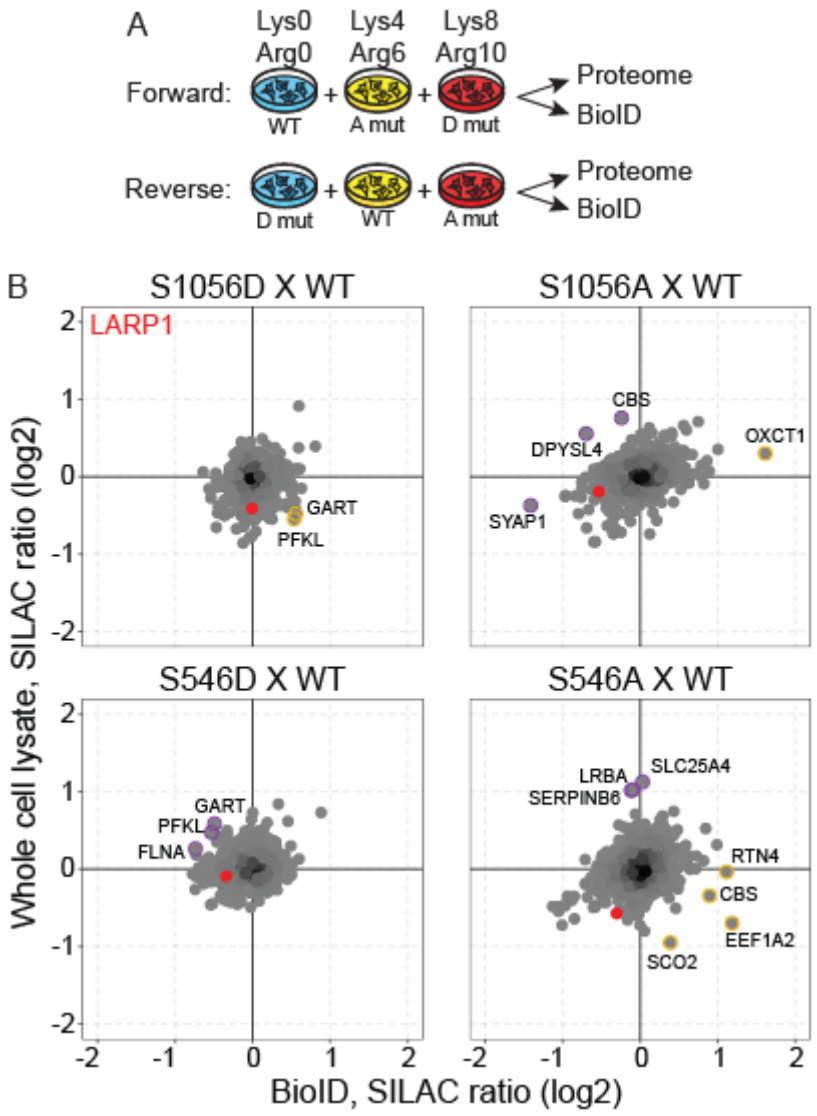


Figure 3.16 - BioID investigation of LARP1 mutant proximity partners.

A) Experimental design of BioID pull-downs with SILAC forward and reverse labeled cells.
B) Quantification of LARP1 interaction in relation to the protein expression level in the proteome. LARP1 is indicated in red while proteins with fold change of BioID over whole cell lysate difference (ratio of SILAC ratios) above the threshold are indicated by orange and purple circles.

Next, I analysed *de novo* synthesis of proteins in HEK293 cells stably overexpressing LARP1 mutants with pulsed SILAC (pSILAC) (Figure 3.17) (Schwanhäusser *et al.*, 2009). In pSILAC, cells are first cultivated in media containing light labeled amino acids and changed to media containing either medium-heavy or heavy labeled peptides upon perturbation for labeling of newly synthesized proteins, while older proteins remain labeled with light amino acids. Medium-heavy and heavy incubated cells are

mixed and protein synthesis differences between conditions is reflected on the heavy-over-medium (H/M) SILAC ratio. In my experiments, LARP1 mutant expression was induced with tetracycline for 2 hours in light media before switching to medium-heavy media for WT LARP1 expressing cells or heavy media for mutant LARP1 expressing cells, for a total pulse time of 22 hours. A second biological replica of the experiment was generated by reversing the pulse labels.

Analysis of the H/M SILAC ratios in both experimental replicas show that LARP1 was synthesized equally in all groups (Figure 3.17). Surprisingly, no proteins show H/M SILAC ratios larger than two in both experimental replicates, indicating that translation for the majority of proteins occurred at similar rates in the 22 hours that LARP1 mutants were expressed and in WT LARP1 expressing cells. Looking specifically at TOP proteins I found that their translation was also not affected by expression of LARP1 mutants (Figure 3.17). Therefore, expression of LARP1 mutants did not seem to impact protein synthesis *in vivo*. Despite the increased TOP mRNA affinity observed *in vitro*, phosphomimetic mutation of candidate regulatory phosphorylation sites in LARP1 showed little impact on TOP mRNA translation in cells.

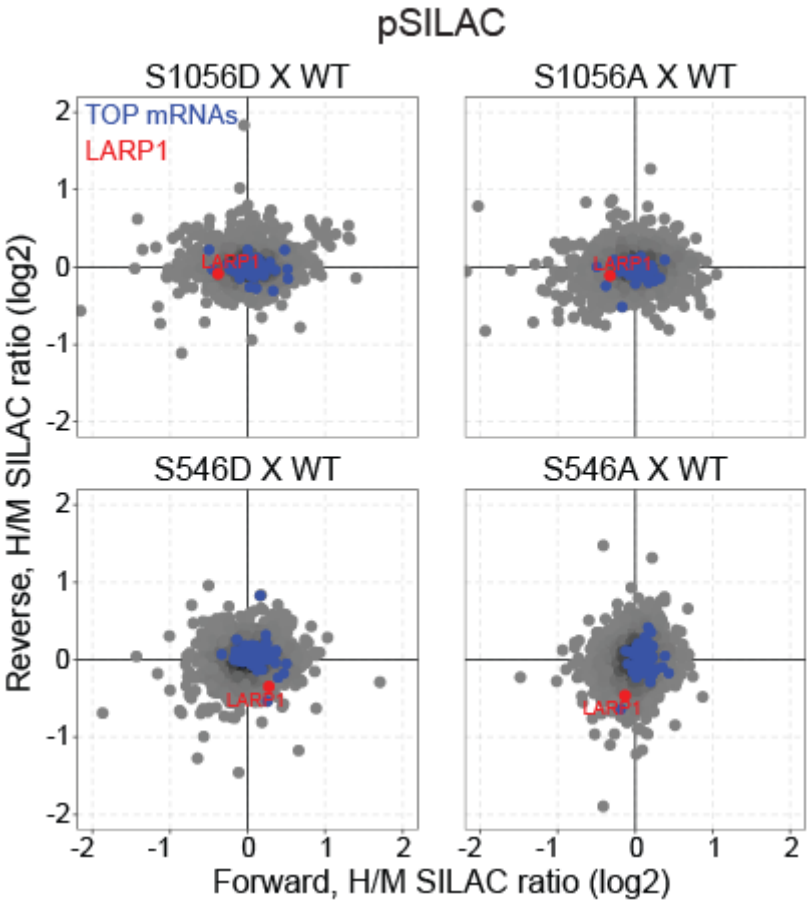


Figure 3.17 - Translation impact from expressing LARP1 mutants measured by pSILAC.

Heavy-over-medium SILAC ratios (H/M) for two biological replicates are plotted. LARP1 and TOP mRNA proteins are indicated in red and blue, respectively.

3.5. RBM20 hyper-phosphorylation regulates protein function

Several parts of this subsection are included in a manuscript submitted for publication to the Molecular Cell journal (Vieira-Vieira et al, submitted).

It caught my attention that six phosphorylation sites in the protein RBM20 were found to correlate with increased pull-down efficiency (positive log2 delta efficiencies values) (Figure 3.9). Another two sites did not pass my strict thresholds but showed similar tendencies and two other sites showed no difference. Strikingly, all ten phosphorylation sites are located in the C-terminus portion of the protein (Figure 3.18 A), a region previously shown to be essential for the splicing (Dauksaite and Gotthardt, 2018) but not for nuclear localization of RBM20 (Filippello *et al.*, 2013). In addition, I

found a proline at the position +1 in 8 out of these 10 phosphorylation sites in comparison to 4 out of 11 other annotated phosphorylated serines in RBM20 not identified in qRIC (Figure 3.18 B), indicative of phosphorylation driven by proline-directed kinases. Given the similar location in the protein and kinase motif for the candidate regulatory phosphorylation sites identified in RBM20, I reasoned that they might be similarly regulated and investigated the impact of phosphomimetic mutation on all 10 sites simultaneously.

To investigate RBM20 regulation by phosphorylation, I generated stable HEK293 cell lines overexpressing the wild-type (WT) or mutant variants of RBM20 fused to BioID and a FLAG tag under the control of a tetracycline inducible promoter. For generation of a hyper-phosphorylated phosphomimetic variant of RBM20 all 10 serines of interest were substituted by aspartic acid (D10 mutant), where substitution by alanine rendered an non-phosphorylatable RBM20 variant (A10). Cells expressing RBM20 mutants were grown in SILAC media and used for RIC and proteome analysis (Figure 3.18 C). To quantify expression and pull-down of the BioID-FLAG fused RBM20, I included an artificial protein sequence in the search database containing only the sequence of the fused part of the artificial protein. Compared to WT, both cells expressing the D10 or A10 mutants show higher levels of the RBM20 fused protein and these were similarly more abundant in RIC pull-downs. Therefore, the apparent higher pull-down of RBM20 mutant variants is explained by similarly higher expression levels in the proteome compared to WT RBM20. Unexpectedly, I conclude that D10 and A10 mutants are pulled down in RIC at similar levels to the WT RBM20 when expression differences are corrected.

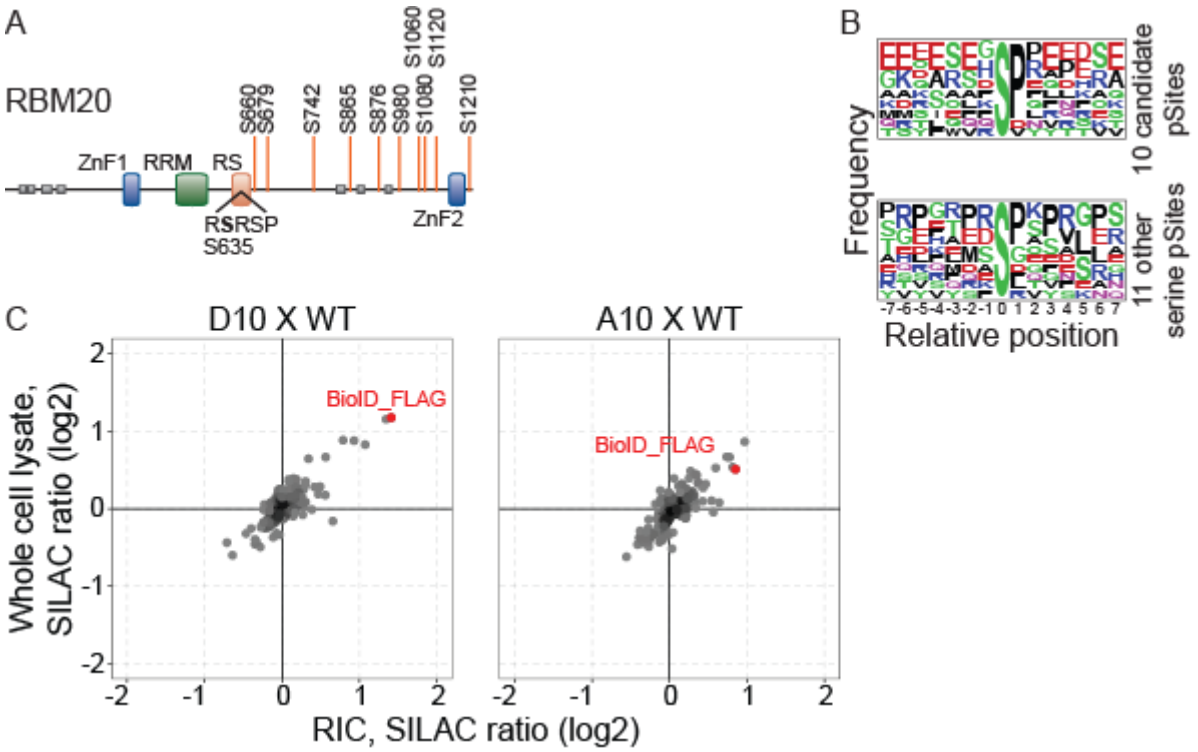


Figure 3.18 - Candidate regulatory phosphorylation sites in RBM20.

A) Scheme of the RBM20 sequence indicating the position of candidate regulatory phosphorylation sites. Other domains and regions of interest are indicated.

B) Analysis of the amino acid frequency surrounding candidate phosphorylation sites in RBM20. The 10 candidate regulatory phosphorylation sites and 11 other serine phosphorylation sites in RBM20 were aligned using WebLogo (Crooks *et al.*, 2004).

C) RIC and proteome analysis of SILAC labeled cells expressing RBM20 mutant variants.

3.5.1. Hyper-phosphorylation modulates subcellular localization and association with cytosolic stress granules

Next, I investigated the protein cellular neighbourhood of different RBM20 mutants using BioID. For this analysis, I also generated cells expressing the dysfunctional S635A mutant RBM20 variant similarly to the other variants. Cells were grown in SILAC media and mutant RBM20 expression was induced with tetracycline for 24 hours. Cells were then incubated with 50 µM of D-Biotin for another 24 hours. Similarly to the experiments described above for LARP1, differentially labeled cells were mixed and the BioID pull-down and whole proteome changes were measured with mass spectrometry by quantifying SILAC ratios (Figure 3.19 A). A second biological replica

of the experiment was generated by reversing the SILAC labels (reverse label), and the average SILAC ratios between biological replicates of the same comparison were used for data analysis.

Comparing cells expressing BirA-FLAG WT RBM20 with non-induced controls without tetracycline revealed proximity-dependent biotinylation of 95 proteins, including several splicing related factors like SF1, SF3B2, SAFB, SAFB2, U2SURP, DDX46, RBM10 and RBM17 (Figure 3.19 B). Gene ontology analysis confirmed specific enrichment of terms like “mRNA processing” and “chromatin remodeling”, consistent with the nuclear localisation of WT RBM20 and its function in splicing (Figure 3.19 E). I then investigated how RBM20 mutations affect proximity-dependent biotinylation relative to the wild-type. To avoid biases caused by differences in steady-state protein levels between cell lines, whole proteomes were quantified and I only considered proteins with proteome-normalized, proximity-dependent biotinylation changes above the 1.5 threshold (Figure 3.19 C and D). The pathogenic S635A mutant showed markedly reduced interactions with splicing factors, consistent with its impaired splicing regulatory function (Figure 3.19 D). In contrast, the D10 and A10 mutations had little impact on RMB20 interaction with splicing-related proteins. Conversely, both the S635A and D10 but not the A10 mutant showed increased interactions with proteins enriched in the GO term “cytoplasmic stress granule” (Figure 3.19 E).

To further characterize the BioID data I took advantage of a recently published protein proximity map of HEK293 cells (Go *et al.*, 2021). Projecting my WT RBM20 data onto this map revealed enrichment signals at different subcellular locations (Figure 3.19 F). When the changes of the S635A mutant relative to the WT were projected onto the map I observed a decrease in “nuclear body” and an increase in “cytoplasmic ribonucleoprotein granule” signal. This is consistent with the decreased interaction of this mutant with splicing factors and its increased localisation to cytosolic granules (Maatz *et al.*, 2014; Sun *et al.*, 2020). In contrast to the S635A mutant, the A10 and D10 mutants showed only minor changes in “nuclear body”. Importantly, however, the D10 mutant displayed an increased signal in the “cytoplasmic ribonucleoprotein granule” compartment.

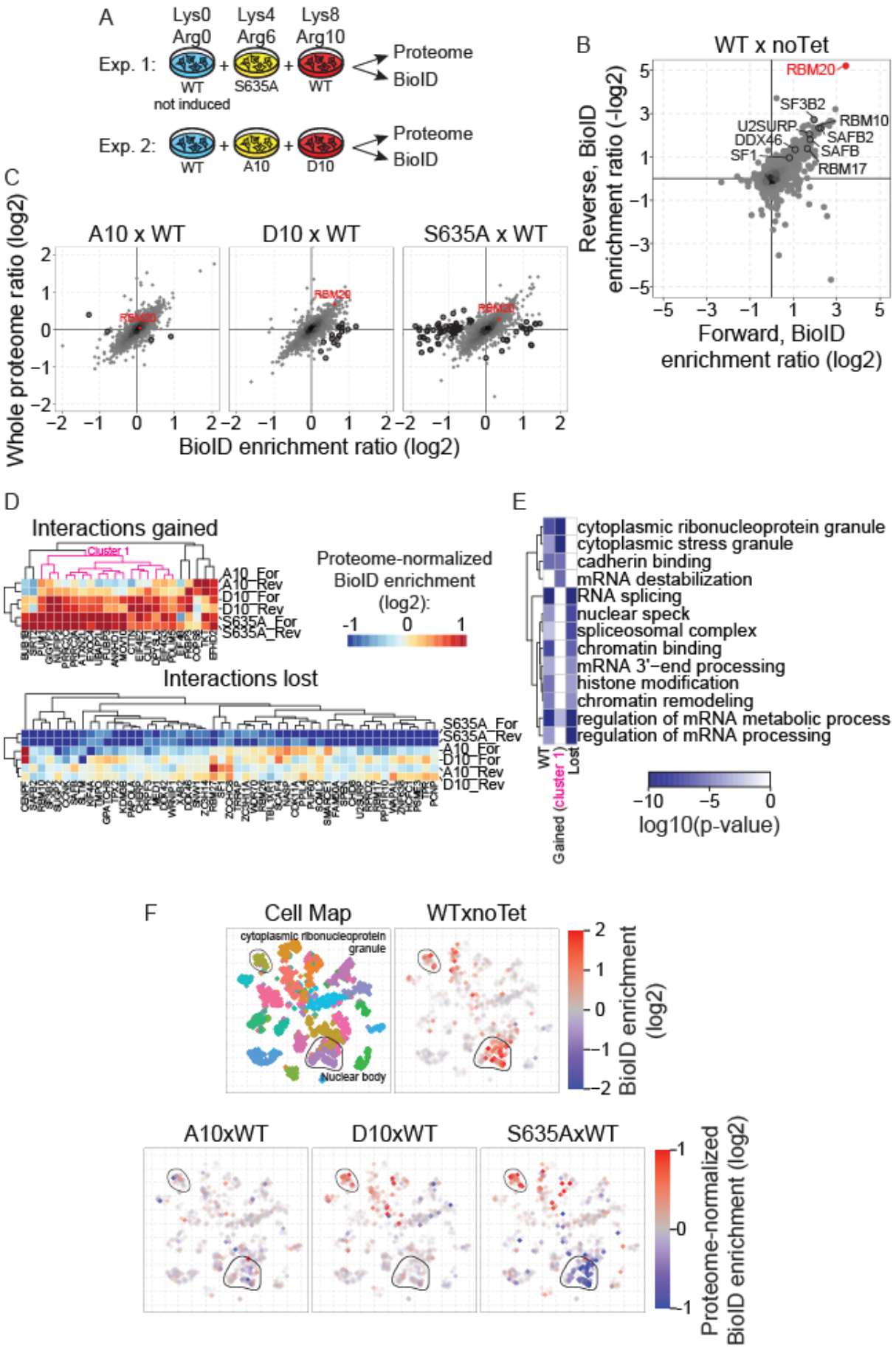

A
Lys0
Arg0
Lys4
Arg6
Lys8
Arg10
Exp. 1:
WT
not induced
S635A
WT
Proteome
BioID
Exp. 2:
WT
A10
D10
Proteome
BioID
B
WT x noTet
RBM20
SF3B2
RBM10
U2SURP
SAFB2
DDX46
SAFB
SF1
RBM17
Reverse, BioID enrichment ratio (-log2)
Forward, BioID enrichment ratio (log2)
C
A10 x WT
D10 x WT
S635A x WT
Whole proteome ratio (log2)
BioID enrichment ratio (log2)
D
Interactions gained
Cluster 1
A10_For
A10_Rev
D10_For
D10_Rev
S635A_For
S635A_Rev
Proteome-normalized BioID enrichment (log2):
Interactions lost
E
cytoplasmic ribonucleoprotein granule
cytoplasmic stress granule
cadherin binding
mRNA destabilization
RNA splicing
nuclear speck
spliceosomal complex
chromatin binding
mRNA 3'−end processing
histone modification
chromatin remodeling
regulation of mRNA metabolic process
regulation of mRNA processing
WT
Lost
Gained (cluster 1)
log10(p-value)
F
Cell Map
WTxnoTet
cytoplasmic ribonucleoprotein granule
Nuclear body
BioID enrichment (log2)
A10xWT
D10xWT
S635AxWT
Proteome-normalized BioID enrichment (log2)

Figure 3.19 - RBM20 candidate regulatory phosphorylation sites and regulation of protein proximity partners.

A) Experimental design of BioID pull-downs with SILAC forward and reverse labeled cells.
B) Reproducibility of RBM20 proximity interaction partners with BioID comparing cells expressing the WT RBM20 and non-induced cells (noTet). RBM20 and several splicing factors are highlighted.
C) Biotin enrichment and proteome changes in BioID experiments comparing mutant and WT RBM20. RBM20 and proteins preferentially enriched in mutant (over wild-type) or wild-type (over mutant) are highlighted.
D) Hierarchical clustering and heatmap visualization for RBM20 gained and lost interactions in mutant variants compared to WT. Mutant-over-WT RBM20 BioID enrichment ratios were normalized by proteome changes. Cluster 1 denotes gained interactions by both D10 and S635A but not A10.
E) Hierarchical clustering for GO terms enriched in WT RBM20 proximity interactors and interactions gained or lost by mutants.
F) Projection of BioID results onto the tSNE plot proximity map. The original proximity map is shown and compartments of interest highlighted for reference.

To validate these observations I used immunofluorescence microscopy. Stable HEK293 cell lines expressing mutant forms of RBM20 were immunostained with an fluorescent antibody against the FLAG tag fused to the N-terminal of the protein and visualized under the microscope (Figure 3.20 A). Consistent with the BioID data, both the WT and the A10 RBM20 mutant mostly co-localized with the DNA intercalator DAPI marking the nucleus. The S635A mutant was depleted in the nucleus and almost exclusively cytosolic. Interestingly, the D10 mutant (but not A10) was partially found in the cytosol while a considerable fraction of the protein was still localized to the nucleus. Quantification of cells with mostly nuclear or cytosolic FLAG (RBM20) signals confirms that the D10 phosphomimetic RBM20 accumulates in the cytosol, although less than the S635A mutant (Figure 3.20 B). Hence, phosphorylation of RBM20 appears to cause a partial relocation of the protein to the cytosol, reminiscent of the pathogenic S635A mutant.

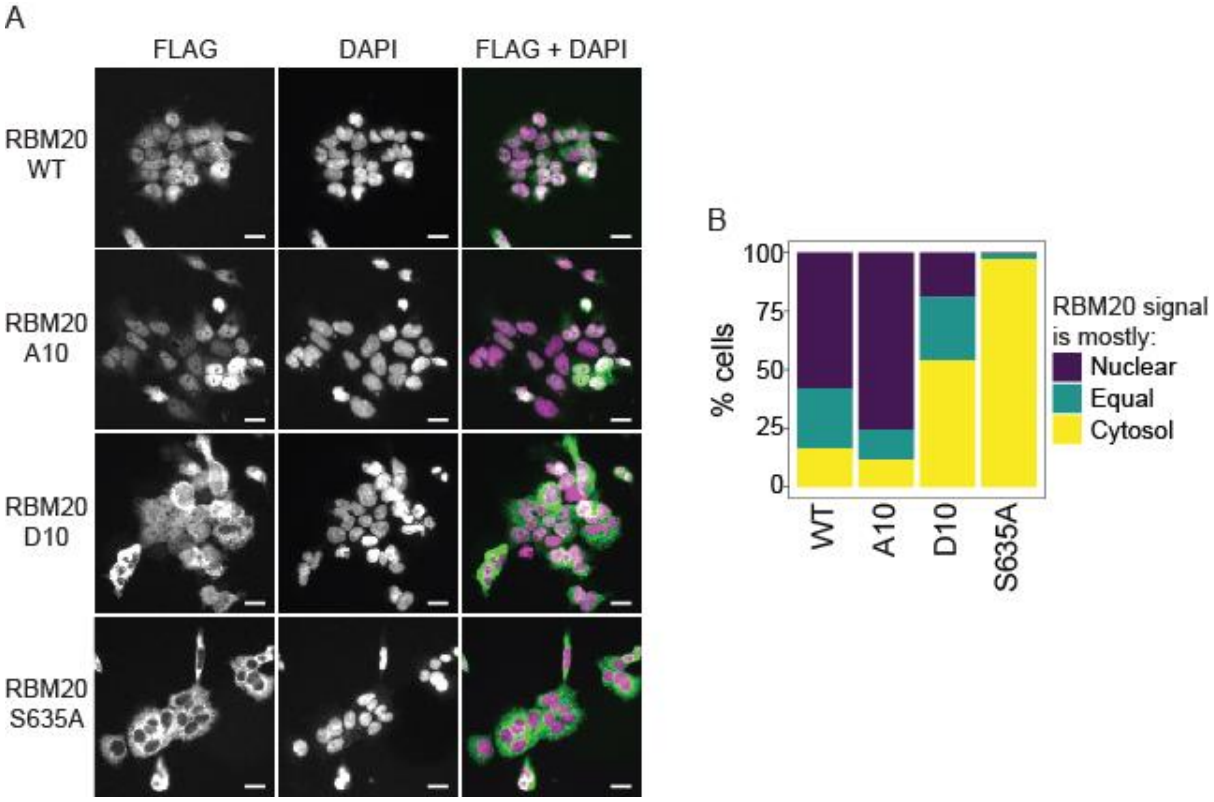


Figure 3.20 - Microscopic investigation of mutant RBM20 subcellular localization.

A) Representative immunofluorescence microscopic images of HEK293 cells stably expressing FLAG-tagged, BioID-fused RBM20 variants. RBM20 was detected by immunostaining against the FLAG tag with a specific antibody. DAPI was used to stain the nucleus.
B) Quantification of results in F. The number of cells with RBM20 content majoritarily cytosolic, nuclear or equally distributed was counted in each experimental group and divided by the total number of cells counted to obtain percentages.

In summary, the GO- (Figure 3.19 E) and the proximity map-based (Figure 3.19 F) analyses consistently indicate that the phosphomimetic D10 mutant and the pathogenic S635A mutant associate increasingly with cytosolic stress granules. While the S635A mutant also shows reduced interaction with nuclear splicing-related proteins, this is not the case for the D10 and A10 mutants. Overall, the A10 mutant behaves very similarly to wild-type RBM20. The differential subcellular localization of phosphomimetic D10 mutant was further validated with immunofluorescence microscopy that showed accumulation in the cytosol while WT RBM20 is predominantly nuclear (Figure 3.20).

3.5.2. Splicing activity is controlled by hyper-phosphorylation in RBM20

Since the pathogenic S635A mutant fails to regulate splicing of target transcripts, I asked how the phosphomimetic and non-phosphorylatable mutations affect splicing. Especifically, I investigated the impact of RBM20 phosphorylation in regulating the alternative splicing of the human Titin gene, a well known RBM20 target. To this end, I joined forces with Janine Froehlich and Vita Dauksaite in the lab of Michael Gotthardt at the MDC. Together with plasmids for transient expression of RBM20 variants, HEK293 cells were transfected for 24 hours with the plasmids encoding one of two splicing reporter systems. The first system is based on a transiently transfected Titin construct containing sequences encoding the Firefly and Renilla luciferases (Fluc and Rluc, respectively) inserted in exons 8 and 13 of the Titin PEVK region (Figure 3.21 A, upper panel). In this system, wild-type RBM20 leads to the exclusion of the Fluc-containing exon and a decreased Fluc to Rluc ratio (Guo *et al.*, 2012). Accordingly, expression of WT RBM20 dramatically decreased the normalized Firefly luciferase activity compared to the control transfected with an empty vector. While the A10 mutant repressed splicing similar to the WT RBM20, this function was mildly but significantly impaired in the D10 mutant (Figure 3.21 A, lower panel).

The other reporter system employs RT-PCR-based analysis of mRNAs derived from a transiently transfected exon/intron cassette containing exons 241, 242 and 243 of human Titin (Figure 3.21 B, upper panel) (Dauksaite and Gotthardt, 2018). In RT-PCR, the use of a pair of primers targeting the outskirts region of exon 242 yields two products of different size, corresponding to exclusion (smaller) and inclusion (larger) of exon 242. Therefore, RBM20 splicing activity can be measured by the ratio of PCR products in which the exon 242 has been included or excluded. After 24 hours of RBM20 expression, splicing of the Titin gene was visualized with RT-PCR and PCR products were quantified with qRT-PCR (Figure 3.21 B, middle and lower panels). Expression of the dysfunctional S635A and the hyperphosphorylated mutant D10 RBM20 reduced the PCR product where exon 242 was excluded (decreased splicing activity). Therefore, also in this system, I observed that the D10 mutant significantly impaired exon exclusion compared to the wild-type protein.

In conclusion, results with both cell-based splice-reporter systems support the notion that the phosphomimetic mutation D10 renders a less active RBM20 variant.

Importantly, the results were specific to the phosphomimetic variant as expression of A10 did not impact splicing. Therefore, hyper-phosphorylation of RBM20 regulates splicing activity.

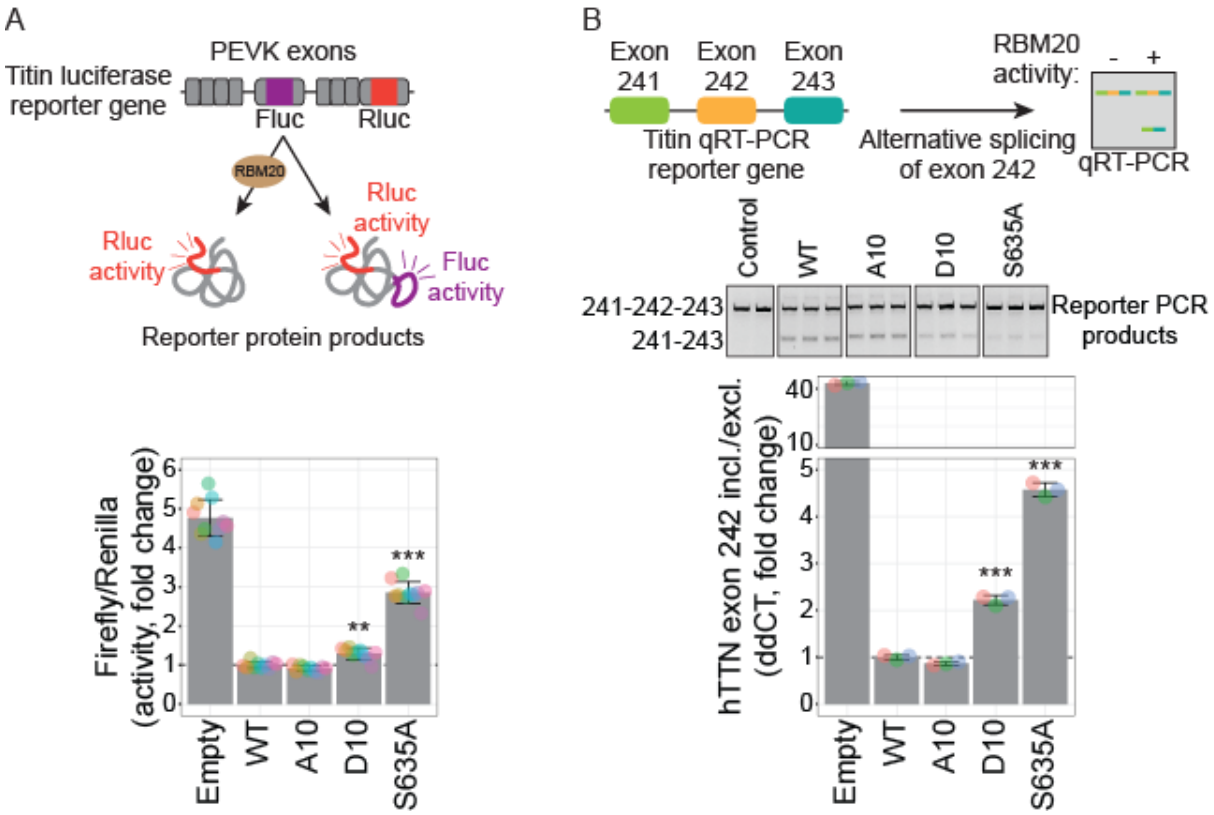


Figure 3.21 - Phosphorylation suppresses RBM20 splicing activity on Titin splicing reporters.

(A and B) Luciferase- and qRT-PCR-based splicing reporter assay for RBM20 activity on Titin exon inclusion/exclusion. RBM20-regulated Firefly (Fluc) luciferase activity relative to Renilla (Rluc) luciferase activity was quantified in A (n = 10). Inclusion and exclusion of exon 242 was quantified with qRT-PCR in B (n = 3). Data is normalized to the wild-type RBM20 level. Transfection with the empty plasmid pcDNA3.2 was used as control. Data is expressed as the mean (bars) of biological replicates (colored dots) ± standard deviation. Groups were compared by one-way ANOVA and Tukey post-test was used. P-values were considered statistically significant as follow: *p< 0.05; **p < 0.01; ***p< 0.001. P-value significance for the comparisons against the WT group are shown except for the empty vector control.

4. Discussion

During my doctoral studies, I developed quantitative RNA-Interactome Capture (qRIC), a mass spectrometry-based method for high-throughput quantification of mRBP binding to mRNAs *in vivo*. By applying qRIC in HEK293T cells at steady state, I systematically identified regulatory phosphorylation sites impacting mRBP function. Candidate regulatory phosphorylation sites in mRBPs were validated using gold-standard, low-throughput biochemical methods. In this section, I interpret these results in the context of the most up-to-date literature in the field.

4.1. Results summary

During my doctoral work, I successfully developed and optimized qRIC to quantify RBP-RNA binding *in vivo* and at steady state (Figure 3.1). Initially, I planned to isolate RNPs using PTex independently of mRNA poly-A tail hybridization with oligo(dT) beads. However, phosphopeptide enrichment with TiO2 chromatography from PTex-isolated samples proved very inefficient (Figure 3.2). Although it is possible that phosphorylation might have been lost during PTex samples preparation, my working hypothesis is that co-enrichment of peptides cross-linked to nucleotides from target RNAs outcompeted phosphopeptides during enrichment. In agreement with that, enrichment of phosphopeptides from RIC-isolated samples was drastically improved when mRBPs were eluted from oligo(dT) magnetic beads without releasing the RNA-cross linked peptides (Figure 3.3). Another important optimization step during the development of qRIC was reducing the impact of SILAC ratio compression on the pull-down efficiency estimation. This was achieved by optimizing the amount of whole cell lysate spike-in material added to isolated mRBPs in qRIC (Figure 3.4).

With a working protocol in hands (Figure 3.5), I proceeded to quantify the pull-down efficiency of the mRNA-bound proteome in HEK293T cells (Figure 3.6). Interestingly, on average, mRBPs are pulled down quite inefficiently. While efficiencies as high as 60% were observed, the great majority of mRBPs are pulled down with less than 5 % efficiency. Still, *in vivo* RNA-binding is a major contributor for pull-down efficiency variability among RBPs, as suggested by the correlation of efficiency with the number of RBDs per protein (Figure 3.7) and the number of *in vivo* RNA-binding sites, and analysis of RNA-binding motifs (Figure 3.8).

Besides quantifying RBPs, pull-down efficiency of phosphorylated proteoforms were also quantified in qRIC (Figure 3.9). Relevant to these results, I showed that misidentification and misannotation of peptides does not globally impact pull-down efficiency quantification (Figure 3.10). By calculating the delta pull-down efficiency between phosphorylated proteoforms and host proteins, I observed global patterns of regulation by phosphorylation, particularly the effect of charge repulsion in RNA-binding (Figure 3.11) and the dissolution of membraneless organelles coordinated by DYRK3 kinase (Figure 3.12). Importantly, individual phosphorylation sites involved in RBP functional regulation were identified, including sites in SF3B1, UPF1 and ELAVL1 (Figure 3.13).

One limitation of our implementation of qRIC is that the peptides UV cross-linked to RNA are systematically missed. In future experiments, chemical cleavage of RBP-RNA cross-links could be used to circumvent this bias (Bae *et al.*, 2020). Also, combining qRIC with analysis of other types of PTMs (methylation, acetylation, ubiquitination, SUMOylation etc.) could be used to assess their impact on RNA-binding. Finally, implementing qRIC with TMT labeling and deep fractionation might improve coverage of modified sites (Hogrebe *et al.*, 2018).

Among the novel 115 phosphorylation sites identified in qRIC with potential regulatory role in RBP function, sites in three RBPs were chosen for careful investigation. Phosphomimetic mutagenesis of individual sites was employed and mutant proteins analysed with low throughput methods. Preliminary results from our lab prompted me to investigate SERBP1 S330 phosphorylation role in regulating ribosome association (Imami *et al.*, 2018). However, phosphomimetic S330 SERBP1 mutation interference with the protein's ability to interact with the actively translating ribosomes was not evident (Figure 3.14). Phosphorylation of LARP1 recently emerged as a key component of cell signaling cascades governing protein translation in cells (Jia *et al.*, 2021). Investigation of TOP mRNA binding in two phosphomimetic LARP1 mutants showed increased affinity *in vitro* (Figure 3.15). Surprisingly, the increased *in vitro* affinity towards RNAs is not functional *in vivo*, as no differential interaction with the translation apparatus was found (Figure 3.16) and no translational alterations on TOP mRNAs were observed in cells expressing the phosphomimetic mutants (Figure 3.17).

Finally, regulatory phosphorylation sites in the cardiac and muscle-specific splicing factor RBM20 were validated (Figure 3.18). By analysing the proximal interaction

partners of the phosphomimetic RBM20, I found that this mutant is proximally interacting with cytosolic stress granules in cells while it does not lose interaction with nuclear components (Figure 3.19). In fact, the RBM20 phosphomimetic mutant shows altered cytosolic localization in comparison to the mostly nuclear wild-type protein (Figure 3.20). At the same time, phosphomimetic mutation of RBM20 led to loss of function in splicing (Figure 3.21).

4.2. Interpreting the pull-down efficiency values in qRIC

Several parts of this subsection have been published in a manuscript submitted to Frontiers in Molecular Biosciences (Vieira-Vieira and Selbach, 2021).

4.2.1. Use of the qRIC setup to quantify diverse protein functions and PTMs

A key concept behind qRIC is that RBPs exist simultaneously in cells in multiple proteoforms. These proteoforms can then be systematically compared without the need of experimental manipulation, like expression of mutant proteins or cellular stimulation. Thus, qRIC is conceptually similar to hotspot thermal profiling in the sense that in both cases the fraction of the total pool of proteins is relatively quantified (Huang *et al.*, 2019; I. R. Smith *et al.*, 2021). However, in contrast to assessing protein thermal stability, qRIC assesses the impact of phosphorylation on mRNA-binding and is thus a read-out for sites involved in posttranscriptional regulatory processes. Similarly, the qRIC setup could be used in the future to study several protein functions. The only requirement is that proteins can be biochemically isolated according to their function, in analogy to isolating mRBPs with RIC due to their ability to bind mRNA. Notably, several methods for enrichment of PTMs other than phosphorylation exist (Beaudette, Popp and Dittmar, 2016; Larsen *et al.*, 2016; Hendriks *et al.*, 2018) and these could also be investigated using the qRIC setup. In the future, the qRIC setup can be used for investigation of diverse protein functions and the impact by PTMs.

4.2.2. RNA occupancy of RBPs in vivo

Although my results suggest that pull-down efficiency quantified in qRIC is linked to RNA-binding, one important question is to what extent it reflects RNA occupancy *in vivo*, that is, the fraction of RBPs in the cellular pool bound to RNAs. Overall, I

observed rather low mRBP pull-down efficiencies (median 3.1 %) and consistent with previous proteome-wide observations (Castello *et al.*, 2012; Beckmann *et al.*, 2015; Urdaneta *et al.*, 2019b). If the pull-down efficiency was solely a function of mRNA occupancy at the moment RNPs were cross-linked with UV, that would suggest most mRBPs are not in association with mRNA in cells.

Importantly, in qRIC, I specifically isolated poly-A RNAs (Baltz *et al.*, 2012; Castello *et al.*, 2012). The pull-down efficiency must then reflect the RNA occupancy of RBPs by mostly mRNAs, and the low pull-down efficiency can be interpreted as indicating that mRBPs mostly associate with other classes of RNAs. This might indeed be the case for some mRBPs, as this sponge effect from other RNA classes in mRBPs has been reported (Jens and Rajewsky, 2015; Okholm *et al.*, 2020). However, several mRBPs interact almost exclusively with mRNAs (Mukherjee *et al.*, 2019; Van Nostrand *et al.*, 2020) and these do not show particularly higher pull-down efficiencies in qRIC. As I will discuss later in this section, multiple other reasons might explain the overall low pull-down efficiency of mRBPs in qRIC.

An interpretation for the average low pull-down efficiency in qRIC is that most mRBPs are free of any RNAs in cells and have low overall occupancy. While some mRBPs might indeed have low RNA occupancy, there are a couple of reasons why this is unlikely to be generally the case. First, RNAse protection assays found on average ~16 protein-protected fragments per mRNA molecule (Schueler *et al.*, 2014; Silverman *et al.*, 2014) and suggest that 3-50% of the RNA sequence is decorated by proteins (Khong and Parker, 2020). Furthermore, although proteins are more abundant than RNAs in cells (Schwanhäusser *et al.*, 2011; Feijó Delgado *et al.*, 2013), each RBP associate with hundreds to several thousands of RNA-binding sites (Mukherjee *et al.*, 2019; Van Nostrand *et al.*, 2020). Thus, it is not uncommon for RBPs to possess several RBDs (Gerstberger, Hafner and Tuschl, 2014) and form multiple contact points to RNAs targets (Bae *et al.*, 2020). Finally, protein concentration in cells vary from the high micromolar to low nanomolar range (median = 45 nM) (Wiśniewski *et al.*, 2014), close to the dissociation constant of protein-RNA interactions (median = 124 nM) (Yang *et al.*, 2013). In summary, these findings suggest that: 1) RNAs are heavily bound by RBPs in cells; 2) RBPs can interact with several RNAs and simultaneously through multiple RBDs; and 3) protein-RNA interaction is generally favoured in cells. Therefore, not only RNAs are constantly associated with proteins in

cells, due to the large availability of RNA-binding sites in the transcriptome and the favoured conditions for interaction, RBPs are also largely associated with RNAs in cells.

4.2.3. Other factors contributing to the pull-down efficiency

It is not at all surprising that pull-down efficiencies are globally low in qRIC, as two key factors independently affect pull-down efficiency besides RNA occupancy: the RNA isolation efficiency of polyA mRNAs with oligo(dT) beads and the mRNP UV cross-linking efficiency.

RNA isolation is an essential step in the RIC protocol and several RNA features can influence it, such as RNA length, subcellular localization, base composition, modifications and secondary structures. For example, mRNA is isolated in RIC through the A-T hybridization of the poly-A tail with oligo(dT) beads (Baltz *et al.*, 2012; Castello *et al.*, 2012). Therefore, the RNA isolation is directly affected by the A-T hybridization strength. Not surprisingly, a negative bias in isolating mRNAs with poly-A tails below a certain threshold length (shorter than 20 nts) has been observed (Park *et al.*, 2016), as the oligo(dT)-beads used for isolation poly-A RNAs are typically as short as 20 bases. Importantly, isolation efficiency deserves special attention when analyzing biological processes where poly-A length is regulated, like cell cycle (Park *et al.*, 2016) and maternal to zygotic transition (Despic *et al.*, 2017). For qRIC, experiments are done at steady state but RNA isolation efficiency likely still impacts delta pull-down efficiency of phosphorylation events in RBPs triggering deadenylation of mRNA targets (Zhang *et al.*, 2018) and target RNA degradation (Kurosaki *et al.*, 2014). However, particular RNA isolation biases are unlikely to drive the overall low pull-down efficiency.

Effective UV cross-linking of RNPs is essential for pull-down of RBPs in RIC. Yet, UV cross-linking efficiency is extremely low (< 5 %) (Budowsky *et al.*, 1986; Fecko *et al.*, 2007). Meaning that even if RBPs are fully occupied by mRNA in cells, only a small fraction is cross-linked and contribute to the pull-down efficiency quantification. Given the similarity of the observed average UV cross-linking efficiencies (< 5 %) and pull-down efficiencies (median 3.1 %), I expect that UV cross-linking efficiency is the main contributor for pull-down quantification in qRIC. Therefore, UV cross-linking is arguably

the most relevant factor absolutely impacting pull-down efficiency in qRIC (Sharma *et al.*, 2014; Vieira-Vieira and Selbach, 2021).

In conclusion, pull-down efficiency results from multiple biological and technical factors impacting RIC, where UV cross-linking efficiency likely plays the most important role in determining the absolute fraction of the RNA-bound proteome quantified in qRIC. Still, a strong correlation between RBPs pull-down efficiency and RNA-binding was observed, indicating that relative differences in pull-down efficiency reflects RNA-binding *in vivo* (section 3.2.2 and 3.2.3). Consequently, although RBPs are poorly pulled down in qRIC, direct comparison of proteoform pull-down efficiencies reflects the regulatory function of phosphorylation events.

4.3. Delta pull-down efficiency and functional regulation of RBPs by phosphorylation

As seen in the previous section, several factors contribute to the pull-down efficiency quantification in qRIC. Moreover, qRIC only reveals correlations between modification state and RNA-binding and does not establish causal relationships. Therefore, while it is tempting to assume that phosphorylation of sites with positive (or negative) log2 delta efficiencies causes increased (or decreased) RNA-binding, this is not the only plausible interpretation. It is also possible that changes in RNA-binding cause differences in the phosphorylation state, for example by exposing a site to kinases or phosphatases. Also, correlations between phosphorylation and RNA-binding can be indirect. For example, the phosphorylation of S202 in ELAVL1 we observed to correlate with increased mRNA-binding actually causes nuclear retention of the protein and thereby decreases mRNA-binding in the cytosol indirectly (Kim, Abdelmohsen, *et al.*, 2008; Grammatikakis, Abdelmohsen and Gorospe, 2017). Hence, while qRIC efficiently prioritizes modification sites with regulatory potential, establishing causal relationships still requires follow-up experiments. It is then instructive to more generously interpret the delta efficiency values as indicators for perturbations on RBP function rather than direct changes in RNA-binding (Vieira-Vieira and Selbach, 2021). In this section, I discuss exemplary ways by which functional regulation of RBP function by phosphorylation can impact pull-down efficiency in qRIC.

4.3.1. RNA-binding affinity

Pull-down efficiency differences can result from differential RNA-binding *in vivo*. Modulation of RNA-binding affinity is the most direct way by which phosphorylation could affect RBP-RNA interactions (Thapar, 2015). For example, phosphorylation of LARP1 in one of two sites analysed here might increase affinity towards TOP mRNAs (section 3.4.2). More generally, I also found that phosphorylation events near RNA interaction sites in RBPs correlate with reduced pull-down efficiency likely due to charge repulsion (Figure 3.10), as hydrogen bonds and hydrophobic interactions guide the RBP-RNA interaction and charge repulsion between nucleotides and phosphoryl groups negatively affect affinity (Corley, Burns and Yeo, 2020). Phosphorylation might also physically prevent interaction by blocking amino acids otherwise in direct contact with the RNA. Phosphorylation of amino acids in RBPs that directly contact nucleotides is more prevalent than in non-interacting residues (Bae *et al.*, 2020), indicating natural selection of functionally relevant modifications directly affecting RNA-binding. Therefore, regulation of the RNA affinity is a widespread mechanism of functional regulation in RBPs.

4.3.2. Subcellular localization

Protein abundance is regulated in cells in multiple ways (Buccitelli and Selbach, 2020), and phosphorylation is important for maintaining stability of many proteins (Sears *et al.*, 2000; Xu, Kim and Gumbiner, 2009). At the same time, proteins are often localized to specific subcellular compartments effectively determining their local concentration (Sundararaman *et al.*, 2016; Thul *et al.*, 2017; Van Nostrand *et al.*, 2020). Not surprisingly, several phosphorylation sites regulate RBP subcellular localization (Thapar, 2015; Murayama *et al.*, 2018). For example, phosphorylation near a nuclear localization signal in ELAVL1 induces accumulation in the cytosol (Doller *et al.*, 2007; Kim, Abdelmohsen, *et al.*, 2008; Kim, Yang, *et al.*, 2008). As a consequence, ELAVL1 differentially binds to and regulates stability and translation of cytosolic and nuclear mRNA targets (Abdelmohsen *et al.*, 2007; Doller *et al.*, 2007; Lafarga *et al.*, 2009). This example illustrates that shuttling between subcellular compartments affects interaction with RNA targets individually. Notably, I identified the hyper-phosphorylation of RBM20 as a factor determining the protein nucleus-cytosol distribution (section 3.5).

4.3.3. Protein-protein interactions and RNP composition

Another way phosphorylation regulates RBP function is by affecting protein-protein interactions and remodeling composition in RNPs (Zarnack *et al.*, 2020). In qRIC, many phosphorylation events regulating protein-protein interaction are likely to be missed as they do not directly impact RNA-binding. However, the consequences of this regulation might still be captured in the pull-down efficiency. As in the examples given throughout the text for NCL and UPF1 (section 1.2.2 and 3.3.3), phosphorylation sites involved in recruiting protein partners that drive target RNA deadenylation and degradation (Kurosaki *et al.*, 2014; Zhang *et al.*, 2018) can lead to decreased pull-down efficiency in qRIC.

Surprisingly, S1107 UPF1 phosphorylation correlates with increased pull-down efficiency of the phospho-proteoform (section 3.3.3). In that case, S1107 phosphorylation in UPF1 recruits SMG5 and SMG7 which are intermediary factors driving target RNA destabilization (Kim and Maquat, 2019; Kurosaki, Popp and Maquat, 2019). S1107 phosphorylation in UPF1 accumulates until degradation of target RNAs is activated by downstream effectors and is then removed (Durand, Franks and Lykke-Andersen, 2016). Therefore, the proteoform of UPF1 carrying the phosphorylated S1107 site exists mainly while the protein is bound to mRNA, while other UPF1 forms also exist free of mRNA. Effectively, phosphorylated UPF1 shows higher RNA occupancy than the unmodified protein even though it might be short lived. In conclusion, regulation of protein-protein interactions by phosphorylation in RBPs might impact RNA-binding and pull-down efficiency in qRIC in multiple ways.

4.3.4. Phase separation

Recent developments in the field of liquid-liquid phase separation of protein-RNA condensates put protein modifications, particularly phosphorylation, in the center (Hofweber and Dormann, 2019; Nosella and Forman-Kay, 2021). During phase separation, RBPs interact with other proteins and RNAs to form membraneless condensates. Modification of intrinsically disordered regions plays an important role in altering the phase separation propensity of RBPs and material properties of condensates (Banani *et al.*, 2017; Nosella and Forman-Kay, 2021). Importantly, intrinsically disordered regions in RBPs are rich in phosphorylation sites (Iakoucheva,

2004). Therefore, phosphorylation events affecting condensation of RBPs likely change RNA interactions. For example, targets of DYRK3 globally show decreased pull-down efficiency (Figure 3.11), in agreement with the notion that DYRK3 acts as a dissolve of multiple condensates in cells (Rai *et al.*, 2018).

4.3.5. Functionless phosphorylation events

Although the above examples show that phosphorylation can regulate RBP interaction with RNA in several ways, analysis of the delta efficiency values show that many phospho proteoforms do not impact pull-down efficiency (Figure 3.9). Regulation of RBP function might indeed not affect RNA-binding (and pull-down efficiency) in many cases, but it is also likely that many phosphorylation events are simply not functional. Quantifying phosphorylation changes across conditions in hundreds of experiments simultaneously has been used to estimate the functional likelihood of individual phosphorylation sites, and found that many might not be functional at all (Ochoa *et al.*, 2020). For example, ultra deep coverage of the phosphoproteome found extreme levels of phosphorylation during mitosis (Sharma *et al.*, 2014). Although phosphorylation is important for cell cycle progression, many of these phosphorylation events might be a side product of exposing nuclear kinases to cytosolic targets (and vice-versa). Therefore, functional regulation must not follow from protein phosphorylation, particularly because they can happen in the background of the cell signaling transduction network (Olsen *et al.*, 2006).

4.4. Novel regulatory phosphorylation sites in mRBPs

My results show that previously known and novel regulatory phosphorylation sites modulating RBP function are successfully identified in qRIC. As discussed above (section 4.3), global regulatory effects linked to charge repulsion and dissolution of membraneless condensates were encompassed in the delta pull-down efficiency. Importantly, individual regulatory phosphorylation sites in mRBPs also present identifiable delta efficiencies. Phosphorylation of T313 in SF3B1, S1107 in UPF1 and S202 in ELAVL1 are examples of previously known regulatory sites involved in RNA splicing, RNA stability and protein subcellular localization, respectively. In this section, I discuss novel regulatory phosphorylation sites in SERBP1 and LARP1.

4.4.1. Regulation of SERBP1 interaction with polysomes

SERBP1 has been implicated in multiple cellular functions (Colleti *et al.*, 2019). This pleiotropism is in part due to SERBP1 interaction with and regulation of ribosomes during mRNA translation (Anger *et al.*, 2013; Liu *et al.*, 2015; Brown *et al.*, 2018; Muto *et al.*, 2018; Wells *et al.*, 2020). Analysis of RNA-binding sites for SERBP1 indicates a sparse binding to mRNA and no recognizable binding motifs (Van Nostrand *et al.*, 2020), suggesting that SERBP1 interaction with mRNA might be secondary to binding to the polysomes. Sucrose fractionation experiments found SERBP1 to precipitate with the small ribosomal subunit (40S), the fully assembled ribosome (80S) and actively translating polysomes (Imami *et al.*, 2018; Muto *et al.*, 2018). However, when actively translating ribosomes were analysed by CryoEM, no structure containing SERBP1 was found (Behrmann *et al.*, 2015). These findings indicate that association of SERBP1 with translating ribosomes is rare and/or leads to fast disassembly of mRNA-bound ribosomes.

Interestingly, SERBP1 protein sedimentation pattern in sucrose density gradients differs from core ribosomal proteins, sedimenting majoritarily in less dense polysome fractions (Figure 3.13 A) (Imami *et al.*, 2018; Muto *et al.*, 2018). Although I can not exclude that SERBP1 interacts exclusively with ribosomes translating short length mRNAs and therefore forming smaller-sized polysomes, SERBP1 is recruited to active ribosomes to halt translation, preventing more ribosome units from associating with the same mRNA molecule (Balagopal and Parker, 2011). Furthermore, SERBP1 has been proposed to act as a hibernation factor, activated by starvation stimuli to inhibit translation while holding the idle 80S ribosome structure together and protecting it from degradation (Anger *et al.*, 2013; Wells *et al.*, 2020). Altogether, these support the notion that recruitment of SERBP1 to actively translating ribosomes where it interacts with mRNA is a regulated process likely targeted by cell signaling in response to stress.

Phosphorylated S330 SERBP1 binds to 40S but not the fully formed ribosome and polysomes (Figure 3.13 A) (Imami *et al.*, 2018). At the same time, reduced binding to polysomes could explain the strong decrease in pull-down affinity observed for phosphorylated S330, as qRIC quantifies pull-down efficiency with mRNA. Surprisingly, phosphomimetic SERBP1 is associated with polysomes at similar levels

to the wild-type protein, indicating that the strong decrease in pull-down affinity can not be explained by differential polysome interaction (section 3.4.1). Nevertheless, substitution of serine for aspartic acid only partially mimics the structure and physicochemical properties of a true phosphorylation site, while abrogating the temporal dynamic component in signaling cascades. It is then possible that incomplete mimicking of S330 phosphorylation explains the lack of effect in polysome association. Furthermore, S330 phosphorylation might be required during activation of specific signaling cascades and in response to stimuli not investigated here, for example during nutrient deprivation. In conclusion, I was not able to rule out the impact of S330 phosphorylation for SERBP1 function, as multiple biological and technical aspects complicate the analysis.

4.4.2. LARP1 binding and translation of TOP mRNAs

LARP1 is a multi-modal regulator of mRNA expression whose mechanisms of action are a topic of intense debate (Berman *et al.*, 2021). A tentative unified model to explain the mechanism of action of LARP1 points to phosphorylation-dependent binding to the 5' terminal oligopyrimidine (TOP) motif and the poly-A tail in target mRNAs simultaneously (Lahr *et al.*, 2015; Al-Ashtal *et al.*, 2021; Jia *et al.*, 2021). By interacting with the N7-methyl guanosine triphosphate (m^7Gppp) cap and TOP motif at the 5' end, LARP1 inhibits protein translation of target mRNAs by preventing binding of the eukaryotic initiation factor 4E (eIF4E) and further assembly of the eIF4F complex (Lahr *et al.*, 2017; Philippe *et al.*, 2018). Importantly, LARP1 binding to TOP motifs is regulated by phosphorylation catalysed by the mTORC1 complex and downstream kinases in response to nutrient deprivation, namely withdraw of amino acids, growth factors, and oxygen, and cell growth arrest (Fonseca *et al.*, 2015; Hong *et al.*, 2017; Philippe *et al.*, 2018; Jia *et al.*, 2021).

The observed increase in *in vitro* affinity of the S1056D phosphomimetic mutant to TOP mRNA motif is consistent with the increased pull-down efficiency in qRIC (Figure 3.14). This is not the case for S546D variant for which the increased affinity is counter-intuitive given the decreased pull-down efficiency. However, binding to mRNA targets *in vivo* is complex, as LARP1 has three RBDs: the DM15 region, where S1056 is located (interacts with TOP motif); a RRM domain, where S546 is located (interacts with TOP motif and poly-A); and the LAM domain (binds both the TOP motif and the

poly-A sequence) (Al-Ashtal *et al.*, 2021; Berman *et al.*, 2021). Therefore, testing LARP1 association with the 5'TOP motif alone misses possible effects from the simultaneous protein interaction with the poly-A tail in target mRNAs. Three critical parameters for LARP1 function were analysed: mRNA affinity, interaction with protein partners and translation regulation of TOP mRNAs.

Two critical parameters for LARP1 function were analysed: protein partners association and translation of TOP mRNAs. Notably, neither of the S546D and S1056D phosphomimetic LARP1 mutants differ in proximity interactions to the wild-type (Figure 3.15), suggesting that these proteins have a similar cellular context in cells to the wild-type LARP1. Also, expression of neither of these phosphomimetic mutants provoked changes in the translation of TOP mRNA in cells (Figure 3.16). This result is less surprising for S546 that is not essential for regulation of TOP mRNA translation in response to the mTOR kinase inhibitor Torin (Philippe *et al.*, 2018). Yet, S1056 phosphorylation is a well known Torin-sensitive site and is directly phosphorylated by Akt *in vitro* and *in vivo* downstream of mTORC1 (Hong *et al.*, 2017).

Possible explanations to the lack of translation effect in the S1056D phosphomimetic mutant include the technical limitation in partially mimicking phosphorylation with phosphomimetic mutations as outlined above (section 4.4.1). In addition, translation is a central and highly regulated process in cells. While knocking down LARP1 indeed impacts global translation (Hong *et al.*, 2017), it would be almost too simple if a single phosphorylation site could globally modulate translation in cells by regulating the core translation apparatus proteins. In fact, the mTORC1 pathway is complex and involves multiple signaling cascades (Hsu *et al.*, 2011; Sabatini, 2017). Therefore, it is likely that LARP1 phosphorylation is tightly controlled to regulate LARP1 function and this control can not be reproduced by using a single phosphomimetic mutant.

4.5. RBM20

Several parts of this subsection are included in a manuscript submitted for publication to the Molecular Cell journal (Vieira-Vieira et al, submitted).

4.5.1. RBM20 hyper-phosphorylation is a novel regulatory mechanism identified by qRIC

Our data revealed an unexpected impact of phosphorylation on the function of the splicing factor RBM20. This protein has emerged as both a key regulator of cardiac splicing and an important disease protein in dilated cardiomyopathy (Guo *et al.*, 2012). Among RBM20 targets, alternative splicing of Titin has been causally linked to cardiac function as the protein size determines passive stiffness of cardiomyocytes and diastolic function of the heart, causing dilated cardiomyopathies when dysfunctional (Methawasin *et al.*, 2014). Molecular characterization found the RRM region, the RS region and the C-term of RBM20 spanning the Zinc-finger domain 2 (ZnF2) and several conserved and unstructured regions rich in glutamine, to be essential for splicing regulation (Figure 3.18) (Dauksaite and Gotthardt, 2018; Murayama *et al.*, 2018).

In addition, RBM20 function is strongly regulated by protein abundance in the nucleus. While the single amino acid mutation (E913K) in the glutamine-rich unstructured region in the C-term has been associated with RBM20 protein instability and decreased amount of the protein in patient hearts (Beqqali *et al.*, 2016), most disease-related mutations occur in the RS region and likely regulate protein subcellular localization and lead to protein accumulation in the cytosol (Brauch *et al.*, 2009; Murayama *et al.*, 2018; Watanabe, Kimura and Kuroyanagi, 2018; Gaertner *et al.*, 2020; Lennermann, Backs and van den Hoogenhof, 2020; Schneider *et al.*, 2020). For example, the S635A mutation in the RS region was casualty linked to dilated cardiomyopathy by retaining the protein in the cytosol and reducing RBM20 splicing activity (Figure 3.19) (Guo *et al.*, 2012; Murayama *et al.*, 2018; Ihara *et al.*, 2020). Also, multiple regulatory phosphorylation sites in RBM20 have been identified, but previous studies mainly focused on phosphorylation of serine residues in the RS region of the protein (Murayama *et al.*, 2018; Sun *et al.*, 2020). This region is of special interest since it harbours most known pathogenic RBM20 mutations (Guo *et al.*, 2012; Lennermann, Backs and van den Hoogenhof, 2020).

Using qRIC, I identified several sites with regulatory potential in the disordered C-terminal half of RBM20 outside the RS region. In contrast to the presumed function of phosphorylation in the RS region, introducing phosphomimetic mutations at these

more C-terminal sites shifts the subcellular distribution of RBM20 to the cytosol (Figure 3.19 and 3.20). This highlights the important fact that PTM functions are typically site-specific (Krug *et al.*, 2019). Interestingly, we find that both the phosphomimetic D10 RBM20 mutant and the disease-causing S635A mutant show increased interactions with cytosolic stress granules (Figure 3.19). However, while the S635A mutant shows markedly reduced interaction with splicing factors and is almost exclusively cytosolic, the D10 mutant still associates with splicing factors, is partially nuclear and has a less severe impact on splicing (Section 3.5). This indicates that the subcellular distribution of RBM20 does not only depend on the RS domain. In fact, a previous study already reported that the RS region is not strictly required for nuclear localisation of RBM20 (Filippello *et al.*, 2013).

4.5.2. RBM20 subcellular localization is regulated by hyper-phosphorylation

One interesting open question emerging from my results is how hyper-phosphorylation regulates RBM20 subcellular localization. Like in other SR proteins carrying the characteristic RRM motif followed by an RS region, these sequences recognize the RNA branching point and mediate association with the spliceosome (Cho *et al.*, 2011; Thapar, 2015; Lennermann, Backs and van den Hoogenhof, 2020). At the same time, the whole amino acid sequence spanning from the RRM to the RS region has been shown to influence RBM20 nuclear localization (Filippello *et al.*, 2013) and an intact RS region sequence is also crucial (Murayama *et al.*, 2018; Ihara *et al.*, 2020). Therefore, one way by which hyper-phosphorylation could affect RBM20 subcellular localization is by interfering with RNA interaction in the nucleus. However, none of the phosphorylation sites investigated here are located in the RRM-RS region of the protein (Figure 3.18). Although allosteric regulation is possible, it seems unlikely that hyper-phosphorylation regulates nuclear localization via modulation of RNA branching point recognition and spliceosome association mediated by the RRM-RS region.

RBM20 has been recently reported to form splicing factories with the giant Titin mRNA during myoblast differentiation (Bertero *et al.*, 2019). In this structure, several RBM20 splicing targets are brought into close physical proximity for splicing regulation. Furthermore, the formation of subnuclear membraneless structures through the phase separation of other RBPs has been reported to regulate splicing (Kato *et al.*, 2012; Li *et al.*, 2020) and phosphorylation also has been shown to regulate phase separation

in diverse contexts including splicing (Nosella and Forman-Kay, 2021). Intrinsically disordered regions play an important role in phase separation and the regulatory phosphorylation sites investigated here fall in the disordered C-terminus portion of RBM20 (Erdős and Dosztányi, 2020). Therefore, I speculate that hyper-phosphorylation of RBM20 reduces its propensity to phase separate in splicing factories in the nucleus. Consequently, hyper-phosphorylated RBM20 in solution might be readily transported out of the nucleus, leading to the observed phenotype of differential cytosol-nucleus subcellular distribution the D10 phosphomimetic mutant. In the future, careful analysis of the nuclear localization of the phosphomimetic mutant could provide evidence for RBM20 phosphorylation in phase separation into splicing factories.

In the cell, proteins are synthesized in the cytosol and RBM20 must be transported into the nucleus after translation. Unlike the wild-type RBM20, the D10 RBM20 is synthesized in the cell already carrying the phosphorylation mimic, which can interfere with the interaction with auxiliary proteins and recognition by the nuclear pore complex and, therefore, with nuclear import. On the contrary, RBM20 shuttle into the nucleus could be unaffected but the hyper-phosphorylation works instead facilitating the protein exclusion back into the cytosol. However, analysis of preferential interactors of hyper-phosphorylated RBM20 does not provide evidence for altered interaction with nuclear transport factors (Figure 3.19) and this hypothesis can be neither validated nor excluded. Surprisingly, this same analysis shows that several of the gained interactors of hyper-phosphorylated RBM20 are cytosolic proteins like GIGYF2 and EIF4E2, two key components of the 4EHP-GYF2 complex that blocks translation initiation (Fu *et al.*, 2016). More generally, phosphomimetic D10 RBM20 shows enhanced interaction with cytosolic stress granules (Figure 3.19). Therefore, I speculate that hyper-phosphorylated RBM20 accumulation in the cytosol could be mediated by stronger association with cytosolic proteins. Interestingly, this also suggests that RBM20 might have a previously unknown cytosolic function in the cytosol that is regulated by phosphorylation, possibly in translation initiation.

In conclusion, the observed accumulation of hyper-phospohrylated RBM20 in the cytosol and nucleus can be caused by multiple and non-excludent mechanisms. The possibility that RBM20 might regulate translation initiation in a phosphorylation dependent manner is particularly interesting as it connects RNA processing and

translation in the cell. However, further experiments will be necessary to investigate this possibility.

4.5.3. RBM20 hyper-phosphorylation as therapeutic target in cardiomyopathies

My results suggest that phosphorylation of RBM20 partially opposes the role of the RS region in protein localization in the cell. Importantly, mutations in the RS region inhibit RBM20 function exclusively via cytosolic accumulation of the protein and splicing activity is restored if the protein is forced back into the nucleus (Murayama *et al.*, 2018). Furthermore, signaling induced by the thyroid hormone T3 and insulin affect Titin splicing in a RBM20-dependent manner, where the involvement of post-translational modifications on RBM20 has been suggested (Zhu *et al.*, 2015, 2017). Thus, manipulating RBM20 activity through modulation of phosphorylation could be physiologically and clinically relevant since cardiac splicing is known to change dynamically for example during development (Lahmers *et al.*, 2004), and diastolic heart function can be restored in animal models by increasing Titin compliance through manipulation of Titin pre-mRNA splicing (Methawasin *et al.*, 2014; Watanabe, Kimura and Kuroyanagi, 2018).

One important limitation of our data is that the phosphomimetic D10 mutant is not biochemically identical to wild-type RBM20 phosphorylated at the corresponding sites. While it is reassuring that the A10 mutant does not seem to be functionally impaired, this does not rule out issues arising from the artificial nature of phosphomimetic mutations (Dephoure *et al.*, 2013). Therefore, it would also be interesting to study the impact of RBM20 phosphorylation on its function in a more physiological model system such as cardiomyocytes (Guo *et al.*, 2012).

In conclusion, manipulation of RBM20 phosphorylation statues to control Titin splicing is a foreseeable novel therapy target for patients with congenital mutations in the RS region and diagnosed with dilated cardiomyopathies (Liss *et al.*, 2018). Inhibiting kinases, particularly proline-directed kinases such as cyclin-dependent protein kinases, mitogen-activated protein kinases, Jun N-terminal protein kinases and glycogen synthase kinase-3, or activating phosphatases that sustain RBM20 phosphorylation might restore the nuclear localization and restore the splicing of target genes, like Titin. In conclusion, as shown for RBM20, identifying regulatory

phosphorylation sites with qRIC open new opportunities for future therapies in diseases related to RBP functions.

5. Materials and methods

Several parts of this section are included in a manuscript submitted for publication to the Molecular Cell journal (Vieira-Vieira et al, submitted).

5.1. Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-FLAG Alexa 488 conjugated	Cell Signaling	Cat# 5407S; RRID:AB_1950473
anti-RBM20	Abcam	Cat# ab233147
Mouse monoclonal anti-FLAG	Sigma-Aldrich	Cat# F3165
Anti-SF3B1 (D7L5T)	Cell Signaling	Cat# 14434
Anti-Phospho-SF3B1 (Thr313) D8D8V	Cell Signaling	Cat# 25009
anti-SERBP1	Abcam	Cat# ab55993
Anti-RPL10	Sigma-Aldrich	Cat# HPA011311
Anti-RPS5	Sigma-Aldrich	Cat# HPA055878
anti-rabbit HRP conjugated secondary antibody	Amersham	Cat# NA934
anti-rabbit HRP conjugated secondary antibody	Amersham	Cat# NA931
Chemicals, Peptides, and Recombinant Proteins		
L-arginine-HCl (Arg0)	Sigma-Aldrich	Cat# A6969; CAS: 1119-34-2
L-arginine-HCl(13C6) (Arg6)	Sigma-Aldrich	Cat# 643440; CAS: 201740-91-2

L-arginine-HCl(13C6,15N4) (Arg10)	Sigma-Aldrich	Cat# 608033; CAS: 202468-25-5
L-lysine-HCl (Lys0)	Sigma-Aldrich	Cat# L8662; CAS: 657-27-2
L-lysine-2HCl(4,4,5,5-D4) (Lys4)	Cambridge Isotope Laboratories	Cat# DLM-2640-PK; CAS: 657-26-1
L-lysine-HCl(13C6,15N2) (Lys8)	Silantes	Cat# 21160410293
TRIzol Reagent	Thermo Fisher Scientific	Cat# 15596026
DNase I	Thermo Fisher Scientific	Cat# EN0521
40 kDa linear polyethylenimine (PEI40)	Polysciences	Cat# 24765-2
cOmplete™, Mini, EDTA-freier Protease-Inhibitor-Cocktail	Roche	Cat# 04693159001
phosphatase inhibitor cocktail 2	Sigma-Aldrich	Cat# P5726
phosphatase inhibitor cocktail 3	Sigma-Aldrich	Cat# P0044
Oligo d(T)25 Magnetic Beads	New England Biolabs	Cat# S1419S
Streptavidin Sepharose high capacity	Merck	Cat# GE17-5113-01
Benzonase	Merck	Cat# 101695
Critical Commercial Assays		
High-Capacity RNA-to-cDNA	Applied Biosystems	Cat# 4387406
SYBR Green master mix	Applied Biosystems	Cat# 4309155
Deposited Data		

Human UniprotKB/Swiss-Prot database (Human UniProt 2019-01)	http://www.uniprot.org/proteomes/	uniprot_human_up000005640_2019-01
qRIC proteome and phosphoproteomes datasets	This work	ProteomeXchange: PXD027137
RBM20 BioID	This work	ProteomeXchange: PXD027138
Experimental Models: Cell Lines		
Human HEK293T	DSMZ	Cat# ACC 635
Flp-In™ T-REx™ 293	Thermo Fisher Scientific	R78007
HEK Flp-In T-Rex SERBP1_WT_3xFLAG	This work	N/A
HEK Flp-In T-Rex SERBP1_S330D_3xFLAG	This work	N/A
HEK Flp-In T-Rex SERBP1_S330A_3xFLAG	This work	N/A
HEK Flp-In T-Rex LARP1_WT_BioID_FLAG	This work	N/A
HEK Flp-In T-Rex LARP1_S546D_BioID_FLAG	This work	N/A
HEK Flp-In T-Rex LARP1_S546A_BioID_FLAG	This work	N/A
HEK Flp-In T-Rex LARP1_S1056D_BioID_FLAG	This work	N/A
HEK Flp-In T-Rex LARP1_S1056A_BioID_FLAG	This work	N/A
HEK Flp-In T-Rex RBM20_WT_BioID_FLAG	This work	N/A
HEK Flp-In T-Rex RBM20_D10_BioID_FLAG	This work	N/A

HEK Flp-In T-Rex RBM20_A10_BioID_FLAG	This work	N/A
HEK Flp-In T-Rex RBM20_S635A_BioID_FLAG	This work	N/A
Recombinant DNA		
pDONR221	Thermo Fisher Scientific	Cat#12536017
pDEST_pcDNA5_BirA-FLAG_Cterm	Couzens et al. 2013	N/A
pEXPR_RBM20-WT_C-term_BirA_FLAG	This work	N/A
pEXPR_RBM20-D10_C-term_BirA_FLAG	This work	N/A
pEXPR_RBM20-A10_C-term_BirA_FLAG	This work	N/A
pEXPR_RBM20-S635A_C-term_BirA_FLAG	This work	N/A
pEXPR_SERBP1_WT_3xFLAG	This work	N/A
pEXPR_SERBP1_S330A_3xFLAG	This work	N/A
pEXPR_SERBP1_S330D_3xFLAG	This work	N/A
pEXPR_LARP1-WT_C-term_BirA_FLAG	This work	N/A
pEXPR_LARP1-S546A_C-term_BirA_FLAG	This work	N/A
pEXPR_LARP1-S546D_C-term_BirA_FLAG	This work	N/A

pEXPR_LARP1-S1056A_C-term_BirA_FLAG	This work	N/A
pEXPR_LARP1-S1056D_C-term_BirA_FLAG	This work	N/A
pcDNA3.1	Invitrogen	Cat# V79520
Software and Algorithms		
Rstudio v1.2.5033	N/A	https://www.rstudio.com
MaxQuant v1.6.0.1	Cox and Mann 2008	http://www.biochem.mpg.de/5111795/maxquant
Metascape last updated on May 2021	Zhou et al. 2019	https://metascape.org/
Fiji ImageJ v1.53c	Schindelin et al., 2012	https://imagej.net/software/fiji/

5.2. Phenol-toluene extraction (PTex)

Human Embryonic Kidney 293T (HEK293T) cells were grown in four 10-cm dishes until 80 % confluency. Cells were washed on the plate with 5 mL of ice-cold PBS and irradiated with 0.15 J/cm² of 254 nm UV light for *in vivo* cross-linking of RNPs. Cell pellets were collected in 400 µL of dPBS, followed by lysis with addition of 200 µL of Phenol, 200 µL of Toluol and 200 µL of Bromo-Chloro-Propane (BCP). Samples were vortexed for 1 min at max speed and centrifuged at 20000 g for 3 min. The aqueous upper phase containing approximately 400 µL was transferred to a fresh 2 mL tube and 300 µL of Solution D (5.85 M guanidine isothiocyanate, 31.1 mM sodium citrate, 25.6 mM N-lauroyl-sarcosine, 1 % 2-mercaptoethanol) was added vortexing the mixture for 1 min at max speed. Next, 400 µL of phenol and 200 µL of BCP were added and solution was mixed again by vortexing and centrifuged at 20000 g for 3 min. The upper aqueous and lower organic phases were discarded (600 µL each) and 400 µL of water was added to the remaining interphase, followed by 200 µL absolute ethanol.

After vortexing for 1 min at max speed, 400 µL of phenol and 200 µL of BCP was added, vortexed again and centrifuged at 20000g for 3 min. Finally, the upper aqueous and lower organic phases were again discarded (600 µL each) and the remaining interphase was precipitated with 2 mL of absolute ethanol overnight at -20°C and used for phosphoproteome and proteome analysis.

5.3. RNA Interactome Capture (RIC)

Cell pellets were resuspended in lysis buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1 % LiDS, 10 mM EDTA pH 8.0) with freshly added 5 mM of DTT, protease inhibitors (1 pill for every 50 mL final volume) (Roche, cat# 4693159001) and 1:100 dilution of phosphatase inhibitor cocktails 2 and 3 (Sigma-Aldrich, cat# P5726 and P0044). Samples were lysed and homogenized by passing it ten times through a 21 gauge needle and five times through a 26 gauge needle. Oligo(dT) magnetic beads (New England Biolabs, cat# S1419S) (0.5 mL per 0.1 g of cell pellet) were pre-washed with 3 volumes of lysis buffer before incubation with cell lysates for 3 hours at room temperature. Beads were collected with magnetic racks and washed three times with one volume of lysis buffer and two times with one volume NP-40 washing solution (50 mM Tris pH 7.5, 140 mM LiCL, 2 mM EDTA pH 8.0, 0.5 % NP-40) with freshly added 5 mM of DTT, protease inhibitors (1 pill for every 50 mL final volume) and 1:100 dilution of phosphatase inhibitor cocktails 2 and 3. Beads were further washed three times with one volume of 1 mL of 50 mM ammonium bicarbonate (ABC) solution and resuspended in ABC solution for mass spectrometry analysis or in SDS-PAGE loading buffer for Western blot analysis.

5.4. Quantitative RIC (qRIC)

Fully labeled HEK293T cells were grown in seven 15-cm dishes per label until 60 % confluency. Cells were washed on the plate with 5 mL of ice-cold PBS and irradiated with 0.15 J/cm² of 254 nm UV light for in vivo cross-linking of RNPs. After cross-linking, cells were collected in 50 mL of ice-cold dPBS by scrapping the plates. A light-labeled cells were kept as input material and heavy-labeled cells were used for RIC. For the experiment where a non-crosslinking control was included, half the number of medium-heavy labeled cells were similarly grown and the lysate was mixed with the

heavy labeled lysate before continuing with RIC. A reverse label experiment was simultaneously performed by swapping the light and heavy SILAC samples for pull-down and input sample. In total, three forward and three reverse experiments were performed, but only one included the medium-heavy labeled non-crosslinking control.

In parallel, input samples were resuspended in 100 mM Tris pH 7.5 with 0.5 % SDS and freshly added protease inhibitors (1 pill for every 50 mL final volume) and 1:100 dilution of phosphatase inhibitor cocktails 2 and 3. Input samples were lysed by boiling at 98 °C for 5 min. Excessive DNA and RNA were removed by letting the sample cool down and digestion with 1 µL of Benzonase (Merck, cat# 101695) at 37 °C for 30 min. Proteins were precipitated by addition of nine volumes of absolute ethanol and overnight incubation at -20 °C followed by 30 min centrifugation at 20000 g at 4 °C. Input protein pellets were resuspended in 200 µL of 2 M Urea, 6M Thiourea solution with freshly added 10 mM of DTT for 5 min at room temperature.

The light labeled input protein sample corresponding to 1 % of the initial input was mixed with the oligo(dT) magnetic beads from RIC with the heavy labeled cells. Proteins were alkylated in the dark with 55 mM of iodoacetamide or chloroacetamide for 20 min at 25 °C. For lysis, proteins were incubated with 10 µg of lysyl endopeptidase (Wako Chemicals, cat# 129-02541) at 25 °C for 2 hours and incubated with 10 µg of trypsin (Promega, cat# V5113) under constant agitation at 25 °C for 16 hours in the dark. Peptides were acidified with 1 % (v/v) trifluoroacetic acid and desalted with C18 Stage Tips (Rappsilber, Mann and Ishihama, 2007). A large fraction of the peptide sample (90 %) was used for enrichment of phosphopeptides. Remaining peptides were eluted with 50% acetonitrile 0.1 % formic acid, dried and resuspended in 3 % acetonitrile, 0.1 % formic acid.

For the experiment that included non-crosslinked cells in the medium-heavy channel, peptides were fractionated by cation-exchange chromatography prior to LC-MS/MS analysis to improve coverage. For that, peptides were loaded into a microcolumn tip packed with strong cation exchange Empore discs (3M and cat# 66889) and washed two times with a 0.5 % formic acid, 20 % acetonitrile aqueous solution. Three fractions were collected by eluting peptides with increasing concentration of salt: 125, 250 and 500 nM of ammonium acetate in 0.5 % formic acid, 20 % acetonitrile solution. Peptide solutions were further diluted to final 4 % acetonitrile concentration and pH acidified

with addition of formic acid to a 5 % final concentration before desalting with C18 Stage Tips for a second time.

5.4.1. Analysis of qRIC results and calculation of delta pull-down efficiency

The “proteinGroups.txt”, “peptides.txt” and “Phospho (STY)Sites.txt” tables from a single MaxQuant run including all three experiments were used for data analysis of results. Potential contaminants, reverse database hits, and peptides only identified with modification were excluded. Log2 SILAC ratios from reverse labeled experiments were mathematically inverted to reflect pull-down over input ratios. For analysis, only hits with quantified SILAC ratios in at least one forward and one reverse experiment were kept. The mean of log2 pull-down over input ratios from all three forward and reverse label experiments were taken individually, while the grand mean was taken as the mean of means in log2 space. To calculate pull-down efficiencies, the pull-down over input ratios (linear scale) were multiplied by the percentual amount of spike-in input material relative to the total amount of material used in RIC. Results in all three experiments were highly reproducible intra- and inter- experiments . As an additional quality filter, we discarded a few irreproducible measurements with a difference in the forward and inverted reverse SILAC ratios higher than 4 and 8 fold at protein and phosphopeptide level, respectively. This removed 256 proteins and 63 phosphopeptides in total but only 4 and 16, respectively, above the 1 % pull-down efficiency threshold for delta efficiency calculation.

Protein and phosphosite information was merged in a single table and the log2 delta pull-down efficiency in each experiment was calculated by subtracting the pull-down over input log2 ratio of individual phosphosite and its host protein. As a result, in log2 scale, positive (or negative) delta pull-down efficiency indicates that the phosphopeptide was pulled down more (or less) efficiently than the host protein. As before, mean delta efficiency in each experiment label was calculated and the grand mean taken. In most figures, the grand mean is plotted. Hits were deemed specific when the mean delta efficiency was above 2 fold change in either direction.

5.4.2. Sensitivity and specificity analysis of qRIC

Proteins quantified in qRIC were compared to the set of annotated mRBPs (named “all mRBPs”) obtained from Hentze et al. (2018). The subset from these that has been

identified in HEK293 cells were also selected for comparison (named “HEK mRBPs”). In addition, by uniting the set of mRBPs with the set of manually curated RBPs in Gerstberger et al (2014) we generated a full list of all RBPs (named “all RBPs”). For comparison with my qRIC data, we matched datasets by gene names. Receiver operating characteristic analysis was performed with the R package pROC (Robin *et al.*, 2011) using annotated RBPs as positives and non-annotated proteins as negatives.

5.4.3. Analysis of pull-down efficiencies correlation with specific RBP features

Protein domain annotations were obtained from the Pfam database downloaded in May 2016 (Mistry *et al.*, 2021) for all proteins quantified in qRIC. Among all domains, RBDs were selected based on the manually curated list provided in Gerstberger et al (2014). For comparison with qRIC data, proteins were matched in both datasets by Uniprot ID. For the comparison of pull-down efficiencies of RBPs in our data with corresponding CLIP-seq data, the entire POSTAR2 database of uniformly analysed PAR-CLIP experiments was downloaded in August 2019 (Zhu *et al.*, 2019). Other types of CLIP-seq data available in the POSTAR2 database were not used for the analysis as results for only few proteins exist. Similarly, the entire dataset of eCLIP experiments was downloaded from ENCODE (Davis *et al.*, 2018) in January 2021. Peaks in all datasets not mapped to human chromosomes (chromosomes 1 to 21, X, Y and mitochondrial) were excluded from analysis. CLIP-seq and qRIC datasets were merged by gene names.

5.5. Phosphopeptide enrichment with titanium oxide columns

Desalted peptides were eluted from Stage Tips in 300 µL of loading buffer (80% acetonitrile [vol/vol] and 6 % trifluoroacetic acid [vol/vol]). Phosphopeptides were enriched using a microcolumn tip packed with 1 mg of TiO2 Titansphere (GL Sciences, cat# 5020-75010). At 4 °C, the TiO2 column was equilibrated by passing through 20 µL of the loading buffer with centrifugation at 100 g. Sample solution was completely loaded on the TiO2 column via multiple consecutive steps of centrifugation at 100 g. Next, the TiO2 column was washed with 20 µL of the loading buffer, followed by 20 µL of 50 % acetonitrile (vol/vol), 0.1 % trifluoroacetic acid (vol/vol) solution. The bound

phosphopeptides were eluted using successive elution with 30 µL of 5 % NH3.H2O solution (fraction 1) followed by 30 µL with 5 % piperidine solution (fraction 2). Each fraction was collected into a fresh tube separately containing 30 µL of 20 % formic acid and further acidified with formic acid until pH smaller than 2 was obtained. The phosphopeptides were desalted with C18 Stage Tips prior to LC-MS/MS analysis.

5.6. NanoLC-MS/MS analysis of digested peptides

For LC-MS/MS analysis, desalted peptides were eluted from Stage Tips with 50 % acetonitrile 0.1 % formic acid solution, dried and resuspended in 3 % acetonitrile 0.1 % formic acid. Peptide concentration was determined based on 280 nm UV light absorbance and up to 1 µg of peptides were analysed per run. Reversed-phase liquid chromatography was performed employing an EASY nLC II 1200 (Thermo Fisher Scientific) using self-made 20 cm long C18 microcolumns packed with ReproSil-Pur C18-AQ 1.9 µm resin (Dr. Maisch, cat# r119.aq.0001) connected on-line to the electrospray ion source (Proxeon) of an HF-X Orbitrap mass spectrometer (Thermo Fisher Scientific) or a Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) in Peptide Mode and with an associated FAIMS device (Thermo Fisher Scientific) installed for pseudo ion mobility peptide separation. The mobile phases consisted of 0.1 % formic acid 5 % acetonitrile solution (Buffer A) and 0.1 % formic acid 80 % acetonitrile solution (Buffer B). Peptides were eluted at a flow rate of 250 nL/min over 44 to 214 min of increasing Buffer B concentration.

Settings for data dependent mass spectrometry analysis with HF-X mass spectrometer were as follow: positive polarity, one full scan (resolution, 60000; m/z range, 350-1800; AGC target, 3e6; max injection time, 10 ms) followed by top 20 MS/MS scans using higher-energy collisional dissociation (resolution, 15000; m/z range, 200-2000; AGC target, 1e5; max injection time, 22 ms; isolation width, 1.3 m/z; normalized collision energy, 26). Ions with an unassigned charge state, singly charged ions, and ions with charge state higher than six were rejected. Former target ions selected for MS/MS were dynamically excluded within 20 s.

Settings for data dependent mass spectrometry analysis with Exploris 480 mass spectrometer were as follow: positive polarity, one full scan (resolution, 120000; m/z range, 350-1800; normalized AGC target, 300 %; max injection time, 30 ms) followed

by top 20 MS/MS scans using higher-energy collisional dissociation (resolution, 7500; m/z range, 200-2000; normalized AGC target, 100 %; max injection time, 25 ms; isolation width, 1.3 m/z; normalized collision energy, 28). Ions with an unassigned charge state, singly charged ions, and ions with charge state higher than six were rejected. Former target ions selected for MS/MS were dynamically excluded within 20 s. Three successive rounds of a full scan followed by 20 MS/MS scans were constantly performed throughout the scanning phase with rotating correction voltages of -40, -60 or -80 V. Scans from single correction FAIMS voltages were collected into MzXML files using the FAIMS MzXML Generator tool provided by the Coon lab: https://github.com/coongroup/FAIMS-MzXML-Generator.

5.7. Processing mass spectrometry data with MaxQuant

All raw files from the same experiment were analyzed together with MaxQuant software (v1.6.0.1) (Cox and Mann, 2008) using default parameters. For increasing transparency and reproducibility of data analysis the “mqpar.xml” file generated by MaxQuant was deposited together with the raw data. Briefly, search parameters used for identification and quantification included two missed cleavage sites, cysteine carbamidomethyl as fixed modification, and the following variable modifications: methionine oxidation, protein N-terminal acetylation, and asparagine or glutamine deamidation. Up to three variable modifications per peptide were allowed. Lys0 and Arg0, Lys4 and Arg6, or Lys8 and Arg10 were set as multiplicity labels during analysis of SILAC samples. Peptide mass tolerance was 20 and 4.5 ppm for first and main search, respectively. Database search was performed with Andromeda embedded in MaxQuant against the UniProt/Swiss-Prot Human proteome (downloaded in January 2019) with common contaminant sequences provided by MaxQuant. False discovery rate was set to 1 % at peptide spectrum match and protein levels. Minimum peptide count required for protein quantification was set to two. When analysing SILAC samples, the “Requantify” option was turned on and an identical MaxQuant search but with the “Requantify” option off was performed by partial reprocessing of search post peptide searches (starting from step “Re-quantification”). The second run (with Requantify off) was used for identification and exclusion of unscrupulous ratios (defined as ratios between two requantified values). Mass spectrometry proteomics and phosphoproteomics raw data and MaxQuant output tables for the qRIC

experiments have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD027137. Similarly, raw data and MaxQuant output tables for the BioID experiment have been deposited with the identifier PXD027138.

5.8. Generation of HEK293 stable cell lines expressing mutant protein variants

For generation of plasmids for expression of proteins of interest C-terminally fused to the BirA* (a.k.a. BioID) and FLAG tag, the wild-type RBM20 coding sequence without the stop codon have been previously cloned Maatz et al (2014) while the optimized SERBP1 and LARP1 sequences were synthesized without stop codon (Twist Bioscience). Mutant sequences were synthesized by BioCat or Twist Bioscience based on the wild-type sequence by altering the corresponding serine codons to aspartic acid or alanine codons, respectively. Sequences were validated by sequencing. Following manufacturer's instruction, the synthetic sequences were cloned into the pDONR221 using Gateway Clonase II system (Thermo Fisher Scientific) and transferred to pDEST_pcDNA5_BirA-FLAG_Cterm (LARP1 and RBM20) or to pDEST_pcDNA5_Flag_Cterm (SERBP1) for protein expression (Couzens *et al.*, 2013). The resulting vectors were used to generate stable HEK293 Flp-In T-Rex cells lines overexpressing C-terminally BioID- and FLAG-tagged wild-type and mutant variants of interest. For that, HEK293 Flp-In T-Rex cells were transfected in a 6-well format by mixing 200 µL of FBS-free medium with 1 µg of the Flp-Recombinase expression vector (pOG44 plasmid, Invitrogen, cat# V600520), 0.5 µg of the destination vector containing the insert of interest, and 3.75 µL of 40 kDa linear polyethylenimine (PEI40) (Polysciences, cat# 24765-2). After a 15 min incubation, the transfection mixture was added to the cells. Cells were re-seeded into 10 cm dishes after 48 hours and allowed to attach overnight. Hygromycin (200 µg/mL) was added the next day and the cells were selected for 18 days by the addition of fresh hygromycin-containing cell culture media every 2-3 days resulting in expansion of monoclonal colonies. Monoclonal colonies were carefully transferred by pipetting cells into 6 cm dishes for expansion in media containing Hygromycin. Cells were confluent after 12 days and cultivated for further experimentation in media without the

addition of Hygromycin. Expression of the mutant proteins was validated after treatment of cells for 24 hours with 1 µg/mL of Tetracycline by immunodetection of the FLAG tag in the expected product size in an electrophoretic gel, and detection of specific peptides via shotgun proteomics when possible.

5.9. BioID investigation of proximity interactor changes in mutant protein variant

SILAC labeled cells expressing RBM20 or LARP1 variants were grown in two 15-cm dishes to 25 % confluency per experimental group and incubated for 24 hours with 1 µg/mL tetracycline to induce expression of the protein variant of interest. Light-labeled WT-expressing cells were used as control by omitting tetracycline from the media and served as an uninduced control for background binding. The quantification of SILAC ratios allowed for comparison of proteins that have been proximity labeled by the transiently expressed constructs. After the induction period all cell lines were incubated for 24 hours in the cell culture medium containing biotin. Then, cells were lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 % Triton X-100, 1 mM EDTA, 1 mM EGTA, 0.1 % SDS) with freshly added protease inhibitors (1 pill for every 10 mL final volume) and 0.5 % sodium deoxycholate. Excessive DNA and RNA were removed by digestion with 2 µL of Benzonase at 37 °C for 20 min. A small fraction of the cleared input lysate was precipitated by addition of nine volumes of absolute ethanol and overnight incubation at -20 °C followed by 30 min centrifugation at 20000 g at 4 °C. Input protein pellets were resuspended in 2 M Urea, 6 M Thiourea solution with freshly added 10 mM of DTT, alkylated in 55 mM iodoacetamide, digested with LysC for 3 hours at 25°C, diluted four times with 25 mM ammonium bicarbonate buffer and digested overnight with Trypsin at 25°C. Remaining input lysate was used for enrichment of biotinylated proteins by incubation with streptavidin-sepharose beads (Sigma-Aldrich, cat# GE17-5113-01) for 3 hours at 4 °C. Beads were washed once with lysis buffer, twice with washing buffer (50 mM HEPES-KOH pH 8.0, 100 mM KCl, 10% glycerol, 2 mM EDTA, 0.1 % NP-40) and six times with 25 mM ammonium bicarbonate buffer to completely remove detergents from the sample. Beads were eluted in ammonium bicarbonate and proteins were digested with trypsin. Beads were removed by centrifugation and peptides were acidified with 1 % (v/v) trifluoroacetic

acid and desalted with C18 Stage Tips. Finally, digested peptides were submitted to NanoLC-MS/MS analysis.

For data analysis, the “proteinGroups.txt” from a single MaxQuant run including all MzXML files were used for data analysis of results with MaxQuant. Potential contaminants, reverse database hits, and proteins only identified by modified peptides were excluded. Log2 SILAC ratios from reverse labeled experiments were mathematically inverted to reflect the induced over uninduced or the mutant over WT experiments. Only hits with quantified SILAC ratios in both forward and reverse experiments were considered for further analysis. Proteins with fold change higher than 2 in both forward and reverse experiments were considered significant when comparing induced and uninduced WT cell lines. To correct for proteome differences in cell lines expressing the different protein variants, biotin-enriched samples were normalized by subtracting log2 fold changes and the respective log2 fold change in the input proteome sample. Proteins with proteome-normalized fold changes higher than 1.5 in both forward and reverse experiments comparing mutant and WT variants were considered significant. Data presented are the mean values between forward and reverse experiments. Gene ontology enrichment of cellular components, molecular functions and biological processes was performed with Metascape (Zhou *et al.*, 2019).

5.10. RBM20 immunostaining and imaging

Cells stably expressing RBM20 variants were seeded on coverslips coated with poly-L-lysine. Variant protein expression was induced in media containing 1 µg/mL tetracycline for 24 hours before cells were fixed for 15 min with 4 % paraformaldehyde at room temperature. Fixed cells were permeabilized for 10 min with 0.5 % Triton in PBS at room temperature and nonspecific protein binding was blocked by incubation in 1.5 % BSA PBS solution for 1 hours with low shaking. Cells were immuno stained by incubation for 1 hour at room temperature with a specific antibody against FLAG (1:1000 dilution) conjugated to Alexa 488 (Cell Signaling, cat# 5407S). Nucleus was stained with DAPI (Sigma-Aldrich, cat# D9564). Images were acquired by Leica DM5000b microscope with an HCX PL FL 20x/0.50 objective. Images were further processed with Fiji ImageJ (Schindelin et al., 2012). For quantification of cells with

nuclear, cytosolic or widespread RBM20 subcellular localization at least 250 cells were counted in each group.

5.11. Cell-based Luciferase Titin splicing reporter assay

For transfection, HEK293 cells were seeded on 96-well plates and transfected with a total of 200 ng of plasmid DNA of which 1 ng was splice reporter PEVK Ex4-13 (Guo *et al.*, 2012) plus a corresponding amount of plasmid for expression of RBM20 variants or control plasmid (pcDNA3.1) in a 20x molar excess. To deliver plasmid DNA, we used PEI40 at a 1:3 ratio (DNA: PEI40). Plasmids and PEI40 in FBS-free medium were incubated for 15 min before the transfection mixture was added to the cells. Cells transfected were at a confluence of 50-60%. Each transfection experiment was repeated ten times and cell viability was measured 60 hours post-transfection using PrestoBlue (Thermo Fisher Scientific, cat# A13261). Luciferase activity was measured 60 hours post-transfection using the Dual-Luciferase® Reporter Assay System (Promega) on an Infinite® M200 Pro (TECAN) plate reader. Ratios of firefly to renilla luciferase activity were normalized to the WT RBM20 expressing cells. All data are expressed as the mean of biological replicates (n = 10) ± SEM. Group comparisons were analyzed by one-way ANOVA and Bonferroni post test. P values were considered statistically significant as follows: *P < 0.05; **P < 0.01; ***P < 0.001.

5.12. Cell-based qRT-PCR Titin splicing reporter assay

For transfection, HEK293 cells were seeded on 6-well plates and transfected with PEI40 at a 1:3 ratio (DNA:PEI40). Plasmids for the Titin exons 241-243 splicing reporter system (Dauksaite and Gotthardt, 2018) and for expression of RBM20 variants were mixed with PEI40 in FBS-free medium and incubated for 15 min before the transfection mixture was added to the cells. Each transfection experiment was performed using three technical replicates and repeated three times. Transfected cells were equally divided for RNA and protein analysis 48 hours after transfection. Similar levels of expression for all RBM20 variants was validated by immunodetection of RBM20 using a specific antibody (Abcam, cat# ab233147). RNA was isolated from cells using the TRIzol reagent following instructions from the manufacturer (Thermo Fisher Scientific, cat# 15596026). Preparations with less than 2 µg of total RNA were

treated with DNase I (Thermo Fisher Scientific, cat# EN0521) and first-strand cDNA was synthesized using High-Capacity RNA-to-cDNA kit (Applied Biosystems, cat# 4387406). Quantitative RT-PCR was performed using SYBR Green master mix (Applied Biosystems, cat# 4309155) in a 7900 HT cycler (Applied Biosystems). qRT-PCR primers are listed somewhere else (Dauksaite and Gotthardt, 2018). The quantification of the gene expression was performed using the ΔΔCT method. Relative levels of splice isoforms are presented as a ratio of mRNAs, with exon 242 included, versus mRNAs, with exon 242 excluded. The fold change in inclusion/exclusion ratio was obtained when compared to the wild type RBM20. The mean of technical replicate values was used for quantification analysis. All data are expressed as the mean of biological replicates (n = 3) ± SEM. Group comparisons were analyzed by one-way ANOVA and Bonferroni post test. P values were considered statistically significant as follows: *P < 0.05; **P < 0.01; ***P < 0.001.

5.13. LARP1 electrophoretic mobility shift assay (EMSA)

The following experiments were performed by Roni Marta Lahr in the lab of Prof. Dr. Andrea Berman in the University of Pittsburgh, Pittsburgh, U.S.A.

N-terminal His6-MBP tag followed by a Tobacco Etch Virus (TEV) protease cleavage site and glycine-linker were fused to full-length LARP1 coding sequence (accession number BC001460.2) and expressed in a modified pET28 vector. Wild-type LARP1 sequence was mutated accordingly using site-directed mutagenesis to generate the S546D, S774D and S1056D mutants. Mutations were confirmed by Sanger sequencing. Proteins were expressed in BL21(DE3) Escherichia coli cells. Cells were collected, flash frozen in liquid nitrogen, and stored at −80°C. Cell pellets were lysed by a combination of freeze-thaw and homogenization in lysis buffer (50 mM Tris–HCl, pH 7.5, 400 mM NaCl, 10 mM imidazole, 10% v/v glycerol) with Complete EDTA-free protease inhibitor tablets (Roche) and 4 mg ml−1 lysozyme. His6-MBP tagged proteins were purified by nickel agarose affinity chromatography (ThermoScientific) and eluted with 50 mM Tris–HCl, pH 7.5, 400 mM NaCl, 250 mM imidazole and 10% glycerol. TEV protease digestion was used to remove the His6-MBP tag. Nucleic acid and protein contaminants were removed by tandem HiTrap Q and HiTrap SP ion exchange (GE Healthcare Lifesciences). LARP1 was further purified by elution from a butyl HP

column (GE Healthcare Lifesciences). Purified proteins were resuspended in 50 mM Tris–HCl, pH 7.5, 250 mM NaCl, 25 % glycerol, 2 mM DTT to a final concentration of 2 mg/ml.

RNA oligos were 5′-end labelled with [γ-32P]-ATP (Perkin Elmer) using T4-polynucleotide kinase (New England Biolabs) and gel purified. Protein-RNA binding reactions were performed on ice for 30 min in 20 mM Tris–HCl, pH 8, 150 mM NaCl, 10% glycerol, 1 mM DTT, 0.5 µg tRNA (Life Technologies), 1 µg BSA and <2 nM RNA. LARP1 wild-type (WT), S546D, S774D and S1056D proteins were titrated at increasing concentrations. Protein-RNA complexes were separated on a 7–8 % polyacrylamide native gel at 120 V for 40 min at 4°C, dried, and exposed overnight. Exposed phosphor screens (GE Healthcare Lifesciences) were imaged on a Typhoon FLA plate reader (GE Healthcare Lifesciences) and quantitated using Imagequant TL (GE Healthcare Lifesciences). The fraction of shifted radio-labeled RNA was calculated by taking the ratio of the sum of the background-corrected volume intensities for all bands above the unshifted probe over total counts per lane.

5.14. Pulsed SILAC labeling (pSILAC)

Pulsed SILAC labeling of cells expressing LARP1 mutant variants was performed as described in Krausher et al (2021) (‘Protein Synthesis in the Developing Neocortex at Near-Atomic Resolution Reveals Ebp1-Mediated Neuronal Proteostasis at the 60S Tunnel Exit’, 2021). In brief, fully labeled light SILAC HEK cells stably expressing LARP1 variants were seeded in 6-well dishes at approximately 50 % confluency. On the following day, protein expression was induced by addition of final 1 µg/mL of tetracycline in medium-heavy and heavy SILAC medium to label newly synthesized proteins. After 24 hours of pulse labeling, cells were collected in ice-cold dPBS and mixed according to the experimental design. Cell pellets were directly lysed in 2M urea 6M thiourea buffer with freshly added 10 mM of DTT, alkylated in 55 mM iodoacetamide, digested with LysC for 3 hours at 25°C, diluted four times with 25 mM ammonium bicarbonate buffer and digested overnight with Trypsin at 25°C. Peptides were acidified with 1 % (v/v) trifluoroacetic acid and desalted with C18 Stage Tips. Finally, digested peptides were submitted to NanoLC-MS/MS analysis.

For data analysis, heavy-over-medium (H/M) SILAC ratios were used to quantify differences in the synthesis rates of individual proteins. The “proteinGroups.txt” from a single MaxQuant run including all raw files were used for data analysis. Potential contaminants, reverse database hits, and proteins only identified by modified peptides were excluded. Log2 H/M SILAC ratios from reverse labeled experiments were mathematically inverted to reflect the same ratio as the forward experiment. Proteins with fold change higher than 2 in both forward and reverse experiments were considered significant.

5.15. Polysome profiling by sucrose density gradient fractionation

HEK cells stably expressing SERBP1 variants were seeded in two 15-cm plates at approximately 25 % confluency in media containing 1 µg/mL of tetracycline. After two days of protein expression induction, fresh media was added for a total incubation period of 30 minutes. In the last 5 minutes of incubation, 100 µg/mL cycloheximide was added to media, cells were washed with 4 ml of ice-cold dPBS containing 100 µg/mL cycloheximide and cell pellets obtained by centrifugation at 1000g for 3 min at 4°C in dPBS with cycloheximide. For lysis, cell pellets were incubated on ice for 5 min in 350 µl of ribosome lysis buffer (20 mM Tris, 150 mM KCl, 5 mM MgCl2, 0.5 % v/v NP40 in RNAse-free water) with fresh 0.5 mM DTT, 100 µg/mL cycloheximide and 1 tablet of EDTA-free protease inhibitor (Roche) per 50 mL solution, and passed 12 times through a 21G needle. Samples were cleared by centrifugation at 17000 g for 10 min and 300 µL of the supernatant transferred to fresh tubes without carrying over the upper lipid layer.

Protein and RNA content per sample was measured by absorbance of 260 nm UV light and 300 µL per sample at 1.5 µg/µl concentration were layered on top of a 10 to 45 % sucrose density gradient prepared in ribosome lysis buffer in Beckman Coulter Ultra-Clear tubes (SELTON 13.2, 14x89 mm). Samples were separated by centrifugation at 36000 rpm for 2.5 hours at 4°C using a TH-641 rotor. For UV detection of the separation profile and fraction collection the parameters used were: AUFs, 1.0; UV wavelength, 254 nm; length of tubing, 425 mm; speed, 0.3 mm/sec; total distance, 99 mm; number of fractions, 20; automatic collector on and triggered with delay. UV signal was stabilized and normalized with water and fractions were automatically

collected from the 4rd position on, to a total of 14 fractions. Fractions were then precipitated at 4°C overnight with 150 µL (0.25 volumes) of TCA and centrifuged at 20000 g for 30 min at 4 °C. Pellets were washed with 1 mL of cold acetone, vortexed and collected by centrifugation at 16,000 g for 30 min at 4 °C.

Finally, protein pellets from individual fractions were eluted with 25 µl of SDS-PAGE buffer with 10 mM DTT and used for Western blot analysis of the protein content. Twelve µL of each fraction were separated in SDS-PAGE gel by applying 160 V for 75 minutes. Proteins were transferred from the gel to a PVDF membrane at 300mA for 60 minutes and membranes were incubated with antibodies against SERBP1, RPS5 and RPL10 at same time. Secondary antibodies conjugated with HRP were used for protein visualization and membranes were imaged on a BioRad device.

6. References

Abdelmohsen, K. *et al.* (2007) 'Phosphorylation of HuR by Chk2 Regulates SIRT1 Expression', *Molecular Cell*, pp. 543–557. doi: 10.1016/j.molcel.2007.01.011.

Ahn, J.-W. *et al.* (2015) 'SERBP1 affects homologous recombination-mediated DNA repair by regulation of CtIP translation during S phase', *Nucleic acids research*, 43(13), pp. 6321–6333.

Al-Ashtal, H. A. *et al.* (2021) 'The LARP1 La-Module recognizes both ends of TOP mRNAs', *RNA biology*, 18(2), pp. 248–258.

Anger, A. M. *et al.* (2013) 'Structures of the human and Drosophila 80S ribosome', *Nature*, 497(7447), pp. 80–85.

Backlund, M. *et al.* (2020) 'Plasticity of nuclear and cytoplasmic stress responses of RNA-binding proteins', *Nucleic acids research*, 48(9), pp. 4725–4740.

Bae, J. W. *et al.* (2020) 'Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution', *Nature structural & molecular biology*, 27(7), pp. 678–682.

Balagopal, V. and Parker, R. (2011) 'Stm1 modulates translation after 80S formation in Saccharomyces cerevisiae', *RNA* , 17(5), pp. 835–842.

Baltz, A. G. *et al.* (2012) 'The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts', *Molecular cell*, 46(5), pp. 674–690.

Banani, S. F. *et al.* (2017) 'Biomolecular condensates: organizers of cellular biochemistry', *Nature reviews. Molecular cell biology*, 18(5), pp. 285–298.

Beaudette, P., Popp, O. and Dittmar, G. (2016) 'Proteomic techniques to probe the ubiquitin landscape', *Proteomics*, 16(2), pp. 273–287.

Beckmann, B. M. *et al.* (2015) 'The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs', *Nature Communications*. doi: 10.1038/ncomms10127.

Behrmann, E. *et al.* (2015) 'Structural Snapshots of Actively Translating Human Ribosomes', *Cell*, pp. 845–857. doi: 10.1016/j.cell.2015.03.052.

Beqqali, A. *et al.* (2016) 'A mutation in the glutamate-rich region of RNA-binding motif protein 20 causes dilated cardiomyopathy through missplicing of titin and impaired Frank-Starling mechanism', *Cardiovascular research*, 112(1), pp. 452–463.

Berman, A. J. *et al.* (2021) 'Controversies around the function of LARP1', *RNA biology*, 18(2), pp. 207–217.

Bertero, A. *et al.* (2019) 'Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory', *Nature communications*, 10(1), p. 1538.

Bodenmiller, B. *et al.* (2010) 'Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast', *Science signaling*, 3(153), p. rs4.

Bogdanow, B., Zauber, H. and Selbach, M. (2016) 'Systematic Errors in Peptide and Protein Identification and Quantification by Modified Peptides', *Molecular & cellular proteomics: MCP*, 15(8), pp. 2791–2801.

Boucas, J. *et al.* (2015) 'Label-Free Protein-RNA Interactome Analysis Identifies Khsrp Signaling Downstream of the p38/Mk2 Kinase Complex as a Critical Modulator of Cell Cycle Progression', *PloS one*, 10(5), p. e0125745.

Boudrez, A. *et al.* (2002) 'Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1', *The Journal of biological chemistry*, 277(35), pp. 31834–31841.

Brannan, K. W. *et al.* (2016) 'SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes', *Molecular cell*, 64(2), pp. 282–293.

Brauch, K. M. *et al.* (2009) 'Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy', *Journal of the American College of Cardiology*, 54(10), pp. 930–941.

Bresson, S. *et al.* (2020) 'Stress-Induced Translation Inhibition through Rapid Displacement of Scanning Initiation Factors', *Molecular cell*, 80(3), pp. 470–484.e8.

Brown, A. *et al.* (2018) 'Structures of translationally inactive mammalian ribosomes', *eLife*, 7. doi: 10.7554/eLife.40486.

Buccitelli, C. and Selbach, M. (2020) 'mRNAs, proteins and the emerging principles of gene expression control', *Nature Reviews Genetics*, pp. 630–644. doi: 10.1038/s41576-020-0258-4.

Budowsky, E. I. *et al.* (1986) 'Induction of polynucleotide-protein cross-linkages by ultraviolet irradiation. Peculiarities of the high-intensity laser pulse irradiation', *European Journal of Biochemistry*, pp. 95–101. doi: 10.1111/j.1432-1033.1986.tb09837.x.

Castello, A. *et al.* (2012) 'Insights into RNA biology from an atlas of mammalian mRNA-binding proteins', *Cell*, 149(6), pp. 1393–1406.

Castello, A. *et al.* (2013) 'RNA-binding proteins in Mendelian disease', *Trends in Genetics*, pp. 318–327. doi: 10.1016/j.tig.2013.01.004.

Castello, A. *et al.* (2016) 'Comprehensive Identification of RNA-Binding Domains in Human Cells', *Molecular cell*, 63(4), pp. 696–710.

Castello, A., Hentze, M. W. and Preiss, T. (2015) 'Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins', *Trends in endocrinology and metabolism: TEM*, 26(12), pp. 746–757.

Chan, S. L. *et al.* (2014) 'CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing', *Genes & development*, 28(21), pp. 2370–2380.

Cho, S. *et al.* (2011) 'Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly', *Proceedings of the National Academy of Sciences of the United States of America*, 108(20), pp. 8233–8238.

Clerici, M. *et al.* (2018) 'Author Correction: Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex', *Nature structural & molecular biology*, 25(4), p. 355.

Colleti, C. *et al.* (2019) 'Complex interactomes and post-translational modifications of the regulatory proteins HABP4 and SERBP1 suggest pleiotropic cellular functions', *World journal of biological chemistry*, 10(3), pp. 44–64.

Corcoran, D. L. *et al.* (2011) 'PARalyzer: definition of RNA binding sites from PAR-CLIP

short-read sequence data', *Genome biology*, 12(8), p. R79.

Corley, M., Burns, M. C. and Yeo, G. W. (2020) 'How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms', *Molecular cell*, 78(1), pp. 9–29.

Couzens, A. L. *et al.* (2013) 'Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions', *Science signaling*, 6(302), p. rs15.

Cox, J. *et al.* (2011) 'Andromeda: a peptide search engine integrated into the MaxQuant environment', *Journal of proteome research*, 10(4), pp. 1794–1805.

Cox, J. *et al.* (2014) 'Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ', *Molecular & cellular proteomics: MCP*, 13(9), pp. 2513–2526.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature biotechnology*, 26(12), pp. 1367–1372.

Cretu, C. *et al.* (2016) 'Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations', *Molecular cell*, 64(2), pp. 307–319.

Crooks, G. E. *et al.* (2004) 'WebLogo: a sequence logo generator', *Genome research*, 14(6), pp. 1188–1190.

Dauksaite, V. and Gotthardt, M. (2018) 'Molecular basis of titin exon exclusion by RBM20 and the novel titin splice regulator PTB4', *Nucleic acids research*, 46(10), pp. 5227–5238.

Davis, C. A. *et al.* (2018) 'The Encyclopedia of DNA elements (ENCODE): data portal update', *Nucleic acids research*, 46(D1), pp. D794–D801.

Dephoure, N. *et al.* (2013) 'Mapping and analysis of phosphorylation sites: a quick guide for cell biologists', *Molecular Biology of the Cell*, pp. 535–542. doi: 10.1091/mbc.e12-09-0677.

Despic, V. *et al.* (2017) 'Dynamic RNA–protein interactions underlie the zebrafish maternal-to-zygotic transition', *Genome Research*, pp. 1184–1194. doi: 10.1101/gr.215954.116.

Doller, A. *et al.* (2007) 'Protein Kinase Cα-dependent Phosphorylation of the mRNA-stabilizing Factor HuR: Implications for Posttranscriptional Regulation of Cyclooxygenase-2', *Molecular Biology of the Cell*, pp. 2137–2148. doi: 10.1091/mbc.e06-09-0850.

Dominguez, D. *et al.* (2018) 'Sequence, Structure, and Context Preferences of Human RNA Binding Proteins', *Molecular cell*, 70(5), pp. 854–867.e9.

Durand, S., Franks, T. M. and Lykke-Andersen, J. (2016) 'Hyperphosphorylation amplifies UPF1 activity to resolve stalls in nonsense-mediated mRNA decay', *Nature communications*, 7, p. 12434.

Erdős, G. and Dosztányi, Z. (2020) 'Analyzing Protein Disorder with IUPred2A', *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, 70(1), p. e99.

Fecko, C. J. *et al.* (2007) 'Comparison of Femtosecond Laser and Continuous Wave UV Sources for Protein–Nucleic Acid Crosslinking', *Photochemistry and Photobiology*, pp. 1394–1404. doi: 10.1111/j.1751-1097.2007.00179.x.

Feijó Delgado, F. *et al.* (2013) 'Intracellular water exchange for measuring the dry mass, water mass and changes in chemical composition of living cells', *PloS one*, 8(7), p. e67590.

Filippello, A. *et al.* (2013) ‘Identification of nuclear retention domains in the RBM20 protein’, *FEBS letters*, 587(18), pp. 2989–2995.

Fonseca, B. D. *et al.* (2015) ‘La-related Protein 1 (LARP1) Represses Terminal Oligopyrimidine (TOP) mRNA Translation Downstream of mTOR Complex 1 (mTORC1)’, *The Journal of biological chemistry*, 290(26), pp. 15996–16020.

Fu, R. *et al.* (2016) ‘Recruitment of the 4EHP-GYF2 cap-binding complex to tetraproline motifs of tristetraprolin promotes repression and degradation of mRNAs with AU-rich elements’, *RNA* , 22(3), pp. 373–382.

Gaertner, A. *et al.* (2020) ‘Cardiomyopathy-associated mutations in the RS domain affect nuclear localization of RBM20’, *Human Mutation*, pp. 1931–1943. doi: 10.1002/humu.24096.

Garcia-Moreno, M. *et al.* (2019) ‘System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection’, *Molecular cell*, 74(1), pp. 196–211.e11.

Gebauer, F. *et al.* (2021) ‘RNA-binding proteins in human genetic disease’, *Nature Reviews Genetics*, pp. 185–198. doi: 10.1038/s41576-020-00302-y.

Gehring, N. H., Wahle, E. and Fischer, U. (2017) ‘Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation’, *Trends in biochemical sciences*, 42(5), pp. 369–382.

Gerstberger, S., Hafner, M. and Tuschl, T. (2014) ‘A census of human RNA-binding proteins’, *Nature reviews. Genetics*, 15(12), pp. 829–845.

Ghaemi, Z. *et al.* (2017) ‘Role of Electrostatics in Protein-RNA Binding: The Global vs the Local Energy Landscape’, *The journal of physical chemistry. B*, 121(36), pp. 8437–8446.

Girard, C. *et al.* (2012) ‘Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion’, *Nature communications*, 3, p. 994.

Go, C. D. *et al.* (2021) ‘A proximity-dependent biotinylation map of a human cell’, *Nature*, 595(7865), pp. 120–124.

Grammatikakis, I., Abdelmohsen, K. and Gorospe, M. (2017) ‘Posttranslational control of HuR function’, *Wiley interdisciplinary reviews. RNA*, 8(1). doi: 10.1002/wrna.1372.

Gräwe, C. *et al.* (2020) ‘RNA-Centric Methods: Toward the Interactome of Specific RNA Transcripts’, *Trends in biotechnology*. doi: 10.1016/j.tibtech.2020.11.011.

Guo, W. *et al.* (2012) ‘RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing’, *Nature medicine*, 18(5), pp. 766–773.

Hafner, M. *et al.* (2010) ‘Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP’, *Cell*, 141(1), pp. 129–141.

Hardman, G. *et al.* (2019) ‘Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation’, *The EMBO Journal*. doi: 10.15252/embj.2018100847.

Hendriks, I. A. *et al.* (2018) ‘Site-specific characterization of endogenous SUMOylation across species and organs’, *Nature communications*, 9(1), p. 2456.

Hentze, M. W. *et al.* (2018) ‘A brave new world of RNA-binding proteins’, *Nature reviews. Molecular cell biology*, 19(5), pp. 327–341.

Hentze, M. W. and Argos, P. (1991) ‘Homology between IRE-BP, a regulatory RNA-binding protein, aconitase, and isopropylmalate isomerase’, *Nucleic acids research*, 19(8), pp. 1739–1740.

Hiller, M. *et al.* (2020) ‘The mRNA Binding Proteome of Proliferating and Differentiated Muscle Cells’, *Genomics, proteomics & bioinformatics*. doi: 10.1016/j.gpb.2020.06.004.

Hofweber, M. and Dormann, D. (2019) ‘Friend or foe—Post-translational modifications as regulators of phase separation and RNP granule dynamics’, *Journal of Biological Chemistry*, pp. 7137–7150. doi: 10.1074/jbc.tm118.001189.

Hogrebe, A. *et al.* (2018) ‘Benchmarking common quantification strategies for large-scale phosphoproteomics’, *Nature communications*, 9(1), p. 1045.

Hong, S. *et al.* (2017) ‘LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs’, *eLife*, 6. doi: 10.7554/eLife.25237.

Hsu, P. P. *et al.* (2011) ‘The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling’, *Science*, 332(6035), pp. 1317–1322.

Huang, J. X. *et al.* (2019) ‘High throughput discovery of functional protein modifications by Hotspot Thermal Profiling’, *Nature methods*, 16(9), pp. 894–901.

Hurt, J. A., Robertson, A. D. and Burge, C. B. (2013) ‘Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay’, *Genome research*, 23(10), pp. 1636–1650.

Iakoucheva, L. M. (2004) ‘The importance of intrinsic disorder for protein phosphorylation’, *Nucleic Acids Research*, pp. 1037–1049. doi: 10.1093/nar/gkh253.

Ignarski, M. *et al.* (2019) ‘The RNA-Protein Interactome of Differentiated Kidney Tubular Epithelial Cells’, *Journal of the American Society of Nephrology: JASN*, 30(4), pp. 564–576.

Ihara, K. *et al.* (2020) ‘A missense mutation in the RSRSP stretch of Rbm20 causes dilated cardiomyopathy and atrial fibrillation in mice’, *Scientific reports*, 10(1), p. 17894.

Imami, K. *et al.* (2018) ‘Phosphorylation of the Ribosomal Protein RPL12/uL11 Affects Translation during Mitosis’, *Molecular cell*, 72(1), pp. 84–98.e9.

Järvelin, A. I. *et al.* (2016) ‘The new (dis)order in RNA regulation’, *Cell communication and signaling: CCS*, 14, p. 9.

Jens, M. and Rajewsky, N. (2015) ‘Competition between target sites of regulators shapes post-transcriptional gene regulation’, *Nature reviews. Genetics*, 16(2), pp. 113–126.

Jia, J.-J. *et al.* (2021) ‘mTORC1 promotes TOP mRNA translation through site-specific phosphorylation of LARP1’, *Nucleic acids research*, 49(6), pp. 3461–3489.

Kashima, I. *et al.* (2006) ‘Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay’, *Genes & development*, 20(3), pp. 355–367.

Kato, M. *et al.* (2012) ‘Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels’, *Cell*, 149(4), pp. 753–767.

Keene, J. D. (2007) ‘RNA regulons: coordination of post-transcriptional events’, *Nature*

reviews. Genetics, 8(7), pp. 533–543.

Keller, W. *et al.* (1991) 'Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA', *The EMBO Journal*, pp. 4241–4249. doi: 10.1002/j.1460-2075.1991.tb05002.x.

Khong, A. and Parker, R. (2020) 'The landscape of eukaryotic mRNPs', *RNA* , 26(3), pp. 229–239.

Kim, D. I. and Roux, K. J. (2016) 'Filling the Void: Proximity-Based Labeling of Proteins in Living Cells', *Trends in cell biology*, 26(11), pp. 804–817.

Kim, H. H., Yang, X., *et al.* (2008) 'Modification at HuR(S242) alters HuR localization and proliferative influence', *Cell cycle* , 7(21), pp. 3371–3377.

Kim, H. H., Abdelmohsen, K., *et al.* (2008) 'Nuclear HuR accumulation through phosphorylation by Cdk1', *Genes & development*, 22(13), pp. 1804–1815.

Kim, Y. K. and Maquat, L. E. (2019) 'UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond', *RNA* , 25(4), pp. 407–422.

König, J. *et al.* (2010) 'iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution', *Nature structural & molecular biology*, 17(7), pp. 909–915.

Kramer, K. *et al.* (2014) 'Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins', *Nature methods*, 11(10), pp. 1064–1070.

Krug, K. *et al.* (2019) 'A Curated Resource for Phosphosite-specific Signature Analysis', *Molecular & cellular proteomics: MCP*, 18(3), pp. 576–593.

Kurosaki, T. *et al.* (2014) 'A post-translational regulatory switch on UPF1 controls targeted mRNA degradation', *Genes & development*, 28(17), pp. 1900–1916.

Kurosaki, T., Popp, M. W. and Maquat, L. E. (2019) 'Quality and quantity control of gene expression by nonsense-mediated mRNA decay', *Nature Reviews Molecular Cell Biology*, pp. 406–420. doi: 10.1038/s41580-019-0126-2.

Lafarga, V. *et al.* (2009) 'p38 Mitogen-Activated Protein Kinase- and HuR-Dependent Stabilization of p21Cip1 mRNA Mediates the G1/S Checkpoint', *Molecular and Cellular Biology*, pp. 4341–4351. doi: 10.1128/mcb.00210-09.

Lahmers, S. *et al.* (2004) 'Developmental control of titin isoform expression and passive stiffness in fetal and neonatal myocardium', *Circulation research*, 94(4), pp. 505–513.

Lahr, R. M. *et al.* (2015) 'The La-related protein 1-specific domain repurposes HEAT-like repeats to directly bind a 5′TOP sequence', *Nucleic Acids Research*, pp. 8077–8088. doi: 10.1093/nar/gkv748.

Lahr, R. M. *et al.* (2017) 'La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs', *eLife*. doi: 10.7554/elife.24146.

Larsen, S. C. *et al.* (2016) 'Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells', *Science signaling*, 9(443), p. rs9.

Lau, H.-T. *et al.* (2014) 'Comparing SILAC- and stable isotope dimethyl-labeling approaches for quantitative proteomics', *Journal of proteome research*, 13(9), pp. 4164–4174.

Lennermann, D., Backs, J. and van den Hoogenhof, M. M. G. (2020) 'New Insights in RBM20 Cardiomyopathy', *Current heart failure reports*, 17(5), pp. 234–246.

Liepelt, A. *et al.* (2016) 'Identification of RNA-binding Proteins in Macrophages by Interactome Capture', *Molecular & cellular proteomics: MCP*, 15(8), pp. 2699–2714.

Lin, C. and Miles, W. O. (2019) 'Beyond CLIP: advances and opportunities to measure RBP-RNA and RNA-RNA interactions', *Nucleic acids research*, 47(11), pp. 5490–5501.

Linder, P. and Jankowsky, E. (2011) 'From unwinding to clamping - the DEAD box RNA helicase family', *Nature reviews. Molecular cell biology*, 12(8), pp. 505–516.

Liss, M. *et al.* (2018) 'Drug discovery with an RBM20 dependent titin splice reporter identifies cardenolides as lead structures to improve cardiac filling', *PLOS ONE*, p. e0198492. doi: 10.1371/journal.pone.0198492.

Liu, F. *et al.* (2015) 'Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry', *Nature methods*, 12(12), pp. 1179–1184.

Li, W. *et al.* (2020) 'Biophysical properties of AKAP95 protein condensates regulate splicing and tumorigenesis', *Nature cell biology*, 22(8), pp. 960–972.

Li, X., Song, J. and Yi, C. (2014) 'Genome-wide Mapping of Cellular Protein–RNA Interactions Enabled by Chemical Crosslinking', *Genomics, Proteomics & Bioinformatics*, pp. 72–78. doi: 10.1016/j.gpb.2014.03.001.

Li, Y. E. *et al.* (2017a) 'Identification of high-confidence RNA regulatory elements by combinatorial classification of RNA-protein binding sites', *Genome biology*, 18(1), p. 169.

Li, Y. E. *et al.* (2017b) 'Identification of high-confidence RNA regulatory elements by combinatorial classification of RNA-protein binding sites', *Genome biology*, 18(1), p. 169.

Lovci, M. T. *et al.* (2013) 'Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges', *Nature structural & molecular biology*, 20(12), pp. 1434–1442.

Lovci, M. T., Bengtson, M. H. and Massirer, K. B. (2016) 'Post-Translational Modifications and RNA-Binding Proteins', *Advances in experimental medicine and biology*, 907, pp. 297–317.

Lunde, B. M., Moore, C. and Varani, G. (2007) 'RNA-binding proteins: modular design for efficient function', *Nature reviews. Molecular cell biology*, 8(6), pp. 479–490.

Maatz, H. *et al.* (2014) 'RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing', *The Journal of clinical investigation*, 124(8), pp. 3419–3430.

Marondedze, C. *et al.* (2019) 'Changes in the Arabidopsis RNA-binding proteome reveal novel stress response mechanisms', *BMC Plant Biology*. doi: 10.1186/s12870-019-1750-x.

Masuda, T. *et al.* (2020) 'Mass Spectrometry-Compatible Subcellular Fractionation for Proteomics', *Journal of proteome research*, 19(1), pp. 75–84.

Mayr, C. (2017) 'Regulation by 3′-Untranslated Regions', *Annual Review of Genetics*, pp. 171–194. doi: 10.1146/annurev-genet-120116-024704.

Meisenheimer, K. M. and Koch, T. H. (1997) 'Photocross-linking of nucleic acids to associated proteins', *Critical reviews in biochemistry and molecular biology*, 32(2), pp. 101–

140.

Methawasin, M. *et al.* (2014) 'Experimentally Increasing Titin Compliance in a Novel Mouse Model Attenuates the Frank-Starling Mechanism but has a Beneficial Effect on Diastole', *Biophysical Journal*, p. 646a. doi: 10.1016/j.bpj.2013.11.3574.

Milek, M. *et al.* (2017) 'DDX54 regulates transcriptome dynamics during DNA damage response', *Genome research*, 27(8), pp. 1344–1359.

Mistry, J. *et al.* (2021) 'Pfam: The protein families database in 2021', *Nucleic acids research*, 49(D1), pp. D412–D419.

Mitchell, S. F. and Parker, R. (2014) 'Principles and properties of eukaryotic mRNPs', *Molecular cell*, 54(4), pp. 547–558.

Moore, M. J. (2005) 'From birth to death: the complex lives of eukaryotic mRNAs', *Science*, 309(5740), pp. 1514–1518.

Mukherjee, N. *et al.* (2019) 'Deciphering human ribonucleoprotein regulatory networks', *Nucleic acids research*, 47(2), pp. 570–581.

Mullari, M. *et al.* (2017) 'Specifying RNA-Binding Regions in Proteins by Peptide Cross-Linking and Affinity Purification', *Journal of proteome research*, 16(8), pp. 2762–2772.

Müller-McNicoll, M. and Neugebauer, K. M. (2013) 'How cells get the message: dynamic assembly and function of mRNA–protein complexes', *Nature Reviews Genetics*, pp. 275–287. doi: 10.1038/nrg3434.

Murayama, R. *et al.* (2018) 'Phosphorylation of the RSRSP stretch is critical for splicing regulation by RNA-Binding Motif Protein 20 (RBM20) through nuclear localization', *Scientific reports*, 8(1), p. 8970.

Muto, A. *et al.* (2018) 'The mRNA-binding protein Serbp1 as an auxiliary protein associated with mammalian cytoplasmic ribosomes', *Cell biochemistry and function*, 36(6), pp. 312–322.

Needham, E. J. *et al.* (2019) 'Illuminating the dark phosphoproteome', *Science signaling*, 12(565). doi: 10.1126/scisignal.aau8645.

Niedzwiecka, A. *et al.* (2002) 'Biophysical Studies of eIF4E Cap-binding Protein: Recognition of mRNA 5′ Cap Structure and Synthetic Fragments of eIF4G and 4E-BP1 Proteins', *Journal of Molecular Biology*, pp. 615–635. doi: 10.1016/s0022-2836(02)00328-5.

Nosella, M. L. and Forman-Kay, J. D. (2021) 'Phosphorylation-dependent regulation of messenger RNA transcription, processing and translation within biomolecular condensates', *Current opinion in cell biology*, 69, pp. 30–40.

Ochoa, D. *et al.* (2020) 'The functional landscape of the human phosphoproteome', *Nature biotechnology*, 38(3), pp. 365–373.

Okada-Katsuhata, Y. *et al.* (2012) 'N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD', *Nucleic acids research*, 40(3), pp. 1251–1266.

Okholm, T. L. H. *et al.* (2020) 'Transcriptome-wide profiles of circular RNA and RNA-binding protein interactions reveal effects on circular RNA biogenesis and cancer pathway expression', *Genome medicine*, 12(1), p. 112.

Olsen, J. V. *et al.* (2006) ‘Global, in vivo, and site-specific phosphorylation dynamics in signaling networks’, *Cell*, 127(3), pp. 635–648.

Park, C. Y. *et al.* (2021) ‘Genome-wide landscape of RNA-binding protein target site dysregulation reveals a major impact on psychiatric disorder risk’, *Nature Genetics*, pp. 166–173. doi: 10.1038/s41588-020-00761-3.

Park, J.-E. *et al.* (2016) ‘Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle’, *Molecular cell*, 62(3), pp. 462–471.

Perez-Perri, J. I. *et al.* (2018) ‘Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture’, *Nature communications*, 9(1), p. 4408.

Philippe, L. *et al.* (2018) ‘La-related protein 1 (LARP1) repression of TOP mRNA translation is mediated through its cap-binding domain and controlled by an adjacent regulatory region’, *Nucleic Acids Research*, pp. 1457–1469. doi: 10.1093/nar/gkx1237.

Plass, M., Rasmussen, S. H. and Krogh, A. (2017) ‘Highly accessible AU-rich regions in 3’ untranslated regions are hotspots for binding of regulatory factors’, *PLOS Computational Biology*, p. e1005460. doi: 10.1371/journal.pcbi.1005460.

Kraushar, M. et al (2021) ‘Protein Synthesis in the Developing Neocortex at Near-Atomic Resolution Reveals Ebp1-Mediated Neuronal Proteostasis at the 60S Tunnel Exit’, *Molecular cell*, 81(2), pp. 304–322.e16.

Queiroz, R. M. L. *et al.* (2019) ‘Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS)’, *Nature Biotechnology*, pp. 169–178. doi: 10.1038/s41587-018-0001-2.

Rai, A. K. *et al.* (2018) ‘Kinase-controlled phase transition of membraneless organelles in mitosis’, *Nature*, 559(7713), pp. 211–216.

Ramanathan, M., Porter, D. F. and Khavari, P. A. (2019) ‘Methods to study RNA–protein interactions’, *Nature Methods*, pp. 225–234. doi: 10.1038/s41592-019-0330-1.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) ‘Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips’, *Nature protocols*, 2(8), pp. 1896–1906.

Richter, F. M. *et al.* (2009) ‘Enrichment of protein-RNA crosslinks from crude UV-irradiated mixtures for MS analysis by on-line chromatography using titanium dioxide columns’, *Biopolymers*, 91(4), pp. 297–309.

Robin, X. *et al.* (2011) ‘pROC: an open-source package for R and S to analyze and compare ROC curves’, *BMC Bioinformatics*. doi: 10.1186/1471-2105-12-77.

Sabatini, D. M. (2017) ‘Twenty-five years of mTOR: Uncovering the link from nutrients to growth’, *Proceedings of the National Academy of Sciences of the United States of America*, 114(45), pp. 11818–11825.

Schneider, J. W. *et al.* (2020) ‘Dysregulated ribonucleoprotein granules promote cardiomyopathy in RBM20 gene-edited pigs’, *Nature medicine*, 26(11), pp. 1788–1800.

Schönemann, L. *et al.* (2014) ‘Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33’, *Genes & development*, 28(21), pp. 2381–2393.

Schueler, M. *et al.* (2014) 'Differential protein occupancy profiling of the mRNA transcriptome', *Genome biology*, 15(1), p. R15.

Schwanhäusser, B. *et al.* (2009) 'Global analysis of cellular protein translation by pulsed SILAC', *Proteomics*, 9(1), pp. 205–209.

Schwanhäusser, B. *et al.* (2011) 'Global quantification of mammalian gene expression control', *Nature*, pp. 337–342. doi: 10.1038/nature10098.

Sears, R. *et al.* (2000) 'Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability', *Genes & development*, 14(19), pp. 2501–2514.

Sharma, D. *et al.* (2021) 'The kinetic landscape of an RNA-binding protein in cells', *Nature*, 591(7848), pp. 152–156.

Sharma, K. *et al.* (2014) 'Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling', *Cell reports*, 8(5), pp. 1583–1594.

Silverman, I. M. *et al.* (2014) 'RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome', *Genome biology*, 15(1), p. R3.

Singh, G. *et al.* (2015) 'The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion', *Annual review of biochemistry*, 84, pp. 325–354.

Smith, E. M. *et al.* (2021) 'The mTOR regulated RNA-binding protein LARP1 requires PABPC1 for guided mRNA interaction', *Nucleic acids research*, 49(1), pp. 458–478.

Smith, I. R. *et al.* (2021) 'Identification of phosphosites that alter protein thermal stability', *Nature methods*, 18(7), pp. 760–762.

Smith, T. *et al.* (2020) 'Organic phase separation opens up new opportunities to interrogate the RNA-binding proteome', *Current opinion in chemical biology*, 54, pp. 70–75.

Sternburg, E. L. and Karginov, F. V. (2020) 'Global Approaches in Studying RNA-Binding Protein Interaction Networks', *Trends in biochemical sciences*, 45(7), pp. 593–603.

Sundararaman, B. *et al.* (2016) 'Resources for the Comprehensive Discovery of Functional RNA Elements', *Molecular cell*, 61(6), pp. 903–913.

Sun, M. *et al.* (2020) 'RBM20 phosphorylation on serine/arginine domain is crucial to regulate pre-mRNA splicing and protein shuttling in the heart', *bioRxiv*. doi: 10.1101/2020.09.15.297002.

Sysoev, V. O. *et al.* (2016) 'Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila', *Nature communications*, 7, p. 12128.

Thapar, R. (2015) 'Structural Basis for Regulation of RNA-Binding Proteins by Phosphorylation', *ACS Chemical Biology*, pp. 652–666. doi: 10.1021/cb500860x.

Thul, P. J. *et al.* (2017) 'A subcellular map of the human proteome', *Science*, 356(6340). doi: 10.1126/science.aal3321.

Trendel, J. *et al.* (2019) 'The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest', *Cell*, 176(1-2), pp. 391–403.e19.

Ule, J. *et al.* (2006) 'An RNA map predicting Nova-dependent splicing regulation', *Nature*, 444(7119), pp. 580–586.

Urdaneta, E. C. *et al.* (2019) 'Purification of cross-linked RNA-protein complexes by phenol-toluol extraction', *Nature communications*, 10(1), p. 990.

Uren, P. J. *et al.* (2012) 'Site identification in high-throughput RNA-protein interaction data', *Bioinformatics* , 28(23), pp. 3013–3020.

Van Nostrand, E. L. *et al.* (2020) 'A large-scale binding and functional map of human RNA-binding proteins', *Nature*, 583(7818), pp. 711–719.

Vieira-Vieira, C. H. and Selbach, M. (2021) 'Opportunities and Challenges in Global Quantification of RNA-Protein Interaction via UV Cross-Linking', *Frontiers in Molecular Biosciences*. doi: 10.3389/fmolb.2021.669939.

Watanabe, T., Kimura, A. and Kuroyanagi, H. (2018) 'Alternative Splicing Regulator RBM20 and Cardiomyopathy', *Frontiers in molecular biosciences*, 5, p. 105.

Watson, J., Schwartz, J.-M. and Francavilla, C. (2021) 'Using Multilayer Heterogeneous Networks to Infer Functions of Phosphorylated Sites', *Journal of proteome research*, 20(7), pp. 3532–3548.

Wells, J. N. *et al.* (2020) 'Structure and function of yeast Lso2 and human CCDC124 bound to hibernating ribosomes', *PLoS biology*, 18(7), p. e3000780.

Wheeler, E. C., Van Nostrand, E. L. and Yeo, G. W. (2018) 'Advances and challenges in the detection of transcriptome-wide protein-RNA interactions', *Wiley Interdisciplinary Reviews: RNA*, p. e1436. doi: 10.1002/wrna.1436.

Wippich, F. *et al.* (2013) 'Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling', *Cell*, 152(4), pp. 791–805.

Wiśniewski, J. R. *et al.* (2014) 'A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards', *Molecular & cellular proteomics: MCP*, 13(12), pp. 3497–3506.

Xu, C., Kim, N.-G. and Gumbiner, B. M. (2009) 'Regulation of protein stability by GSK3 mediated phosphorylation', *Cell Cycle*, pp. 4032–4039. doi: 10.4161/cc.8.24.10111.

Xu, Y. *et al.* (2019) 'Post-translational modification control of RNA-binding protein hnRNPK function', *Open biology*, 9(3), p. 180239.

Yang, X. *et al.* (2013) 'The dataset for protein-RNA binding affinity', *Protein science: a publication of the Protein Society*, 22(12), pp. 1808–1811.

Youn, J.-Y. *et al.* (2018) 'High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies', *Molecular cell*, 69(3), pp. 517–532.e11.

Zarnack, K. *et al.* (2020) 'Dynamic mRNP Remodeling in Response to Internal and External Stimuli', *Biomolecules*, 10(9). doi: 10.3390/biom10091310.

Zhang, X. *et al.* (2018) 'Nucleolin phosphorylation regulates PARN deadenylase activity during cellular stress response', *RNA biology*, 15(2), pp. 251–260.

Zhou, Y. *et al.* (2019) 'Metascape provides a biologist-oriented resource for the analysis of systems-level datasets', *Nature communications*, 10(1), p. 1523.

Zhu, C. *et al.* (2015) 'RBM20 is an essential factor for thyroid hormone-regulated titin isoform

transition', *Journal of molecular cell biology*, 7(1), pp. 88–90.

Zhu, C. *et al.* (2017) 'Insulin regulates titin pre-mRNA splicing through the PI3K-Akt-mTOR kinase axis in a RBM20-dependent manner', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, pp. 2363–2371. doi: 10.1016/j.bbadis.2017.06.023.

Zhu, Y. *et al.* (2019) 'POSTAR2: deciphering the post-transcriptional regulatory logics', *Nucleic Acids Research*, pp. D203–D211. doi: 10.1093/nar/gky830.