Immuno-Competitive Capture Mass Spectrometry, a novel unbiased approach to study endogenous protein-protein interactions

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Hélène Meistermann

aus Colmar (Frankreich)

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 26.10.2020
Stellvertretender Dekan: Prof. Dr. József Fortágh
1. Berichterstatter: Professor Dr. Ulrich Rothbauer
2. Berichterstatter: Professor Dr. Stefan Stevanović

TABLE OF CONTENTS

ABBREVIATIONS

AP	Affinity purification
AP-MS	Affinity purification coupled to mass spectrometry
ASO	Antisense oligonucleotide
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
coIP	Co-immunoprecipitation
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulfoxide
Da	Dalton
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGTA	Ethylene glycol tetraacetic acid
EIC	Extracted ion chromatogram
F2H	Fluorescent two-hybrid
FBS	Fetal bovine serum
FGF	Fibroblast growth factor
FRET	Fluorescence resonance energy transfer
GalNac	N-acetyl galactosamine
GFP	Green fluorescent protein
GST	Gluthatione S transferase
HCV	Hepatitis C virus
HRP	Horseradish peroxidase
HTS	High-throughput screening
IAA	Iodoacetamide
ICC-MS	Immuno-competitive capture mass spectrometry
Ig	Immunoglobulin
IP	Immunoprecipitation
LC	Liquid chromatography

LNA	Locked nucleic acid
MoA	Mode of action
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
N	Number of technical replicates
NMR	Nuclear magnetic resonance
OCC-MS	Oligo-competitive capture mass spectrometry
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PPI	Protein-protein interaction
PTB	Phosphotyrosine-binding
PVDF	Polyvinylidene fluoride
RNA	Ribonucleic acid
Rpm	Rounds per minutes
SDS	Sodium dodecylsulfate
SH	Src homology
SPR	Surface plasmon resonance
TBS	Tris buffered saline
v/v	Volume per volume
WT	Wild type
Y2H	Yeast two-hybrid

ZUSAMMENFASSUNG

Protein-Protein Interaktionen (PPI) kontrollieren die Mehrzahl biologischer Funktionen und sind in normalen als auch in krankhaften Abläufen die Haupttreiber zellulärer Prozesse. Ein solches Maß an Kontrolle wird nur durch einen hohen Grad an Komplexität ermöglicht: z. B durch eine große Zahl von PPIs, deren Umkehrbarkeit als auch eine Vielfalt an physikalischen und strukturellen Eigenschaften. Darüber hinaus können Bindungs-Affinitäten einen mikro-molaren bis hohen pico-molaren Bereich umfassen, und manche Proteine agieren als Knotenpunkte („hubs"), indem sie mehrere Partner haben. Diese hochentwickelte Organisation und Regulierung der PPI erklärt, weshalb Ihre Erforschung so herausfordernd ist. Kein einzelner Ansatz kann das gesamte Bild erfassen, und es gibt einen dringenden Bedarf an innovativen Plattformen, um die PPI zu studieren und zu analysieren.

Im Rahmen dieser Doktorarbeit wurde eine neue Plattform, die „Immuno-Competitive Capture Mass Spectrometry" (ICC-MS) konzipiert, um auf unvoreingenommene Weise intrazelluläre PPI zu finden. ICC-MS wurde entwickelt, um im Vergleich zur klassischen Affinitäts-Aufreinigung mit Massenspektrometrie eine höhere Spezifität zu erreichen. Dazu wurde vor der Immunopräzipitation ein weiterer Schritt, eine Verdrängung zwischen dem freien und dem bindenden Antikörper, eingeführt. Dieser auf Antikörper basierter, auf Isotopen verzichtende quantitative Ansatz wurde kombiniert mit einer rigorosen statistischen Analyse, um das zelluläre Geflecht der Wechselwirkung von Proteinen besonderen Interesses zu extrahieren und zu beleuchten, während unspezifisch bindende Protein herausgefiltert werden.

ICC-MS wurde zunächst angewendet, um das Interaktom des Hepatitis C viralen, nicht-strukturellen Protein 5A in humanen Hepatom-Zellen zu beleuchten. Dabei wurde die LATS Kinase als möglicherweise wichtiger Regulator der viralen Infektion aufgedeckt. Die Untersuchung der Wechselwirkungs-Partner Glypican-2 und HtrA1 hat die Fähigkeit der ICC-MS bestätigt, eine begrenzte Anzahl von sehr zuverlässigen Interaktionspartnern zu liefern, die vielversprechende Kandidaten für eine funktionale Validierung sind.

Interessanterweise kann ICC-MS auch angepasst werden, um die Interaktionen zwischen Proteinen und Oligonukleotiden (Oligo-Competitive Capture Mass Spectrometry, oder OCC-MS) zu untersuchen. Während es zu einem besseren Verständnis der Wirkungsweise eines SMN2 Spleiss-Modifikators beitrug, konnte die Methode jedoch nicht die Rolle von PPI in der Toxizität des Antisense-Oligonukleotides aufklären.

Insgesamt betrachtet ist dieser neue Ansatz geeignet, um den Umfang und die Richtigkeit von gegenwärtigen PPI Datenbanken im Sinne einer zutreffenden Darstellung des biologischen Interaktoms zu verbessern.

ABSTRACT

Protein-protein interactions (PPIs) are controlling the majority of biological functions and are the main driver of cellular processes observed in normal as well as pathological conditions. Such a level of controlling is only possible via a high degree of complexity; i.e. a massive number of protein-protein interactions (in the range of couple hundreds of thousands), a variety of physical and structural properties and their reversibility. Moreover, binding affinities can span from micro-molar to high pico-molar level and some proteins are acting as "hubs" by having multiple partners. This sophisticated organization and regulation of PPIs explains why their study is so challenging. No single approaches can capture the full picture and there is an urgent need for innovative platforms to study and analyze PPIs.

In this thesis, a novel platform named Immuno-Competitive Capture Mass Spectrometry (ICC-MS) was developed to screen in an unbiased fashion intracellular PPIs. ICC-MS was designed to reach higher specificity compared to classical affinity purification mass spectrometry by introducing a competition step between free and capturing antibody prior to immunoprecipitation. This antibody-based label-free quantitative approach was then combined with a rigorous statistical analysis to extract the cellular interactome of proteins of interest while filtering out non-specifically binding proteins.

ICC-MS was first applied to elucidate hepatitis C viral non-structural protein 5A interactome in human hepatoma cells revealing LATS kinases as potential important regulators of viral infection. The study of Glypican-2 and HtrA1 interacting partners further confirmed the ability of ICC-MS to deliver a limited number of highly confident interacting proteins being promising candidates for functional validation.

Interestingly, ICC-MS can also be adapted to study interactions formed between proteins and oligonucleotides (Oligo-Competitive Capture Mass Spectrometry or OCC-MS). While it contributed to a better understanding of the mode of action of an SMN2 splicing modifier, the approach could not elucidate the role of protein interactions in antisense oligonucleotides toxicity.

Taken together, this innovative approach is suitable to improve the comprehensiveness and accuracy of current protein-protein interactions databases in term of true biological interactome representation.

1. INTRODUCTION

1.1. Protein-protein interactions (PPIs)

1.1.1 Definition and main characteristics

A protein-protein interaction (PPI) is defined as a physical contact between two or more proteins that is specific and not happening by chance or accident [1]. PPIs are controlling the majority of cellular processes and signaling within and outside cells (i.e. metabolism, transport, signal transduction, cell-cycle-control, etc.). PPIs drive biological functions both in normal as well as in pathological conditions such as cancer, viral/bacterial infections, auto-immune disorders or neurodegenerative diseases [2].

Such molecular docking is usually emerging for a specific purpose and can be distinguished from generic interactions such as those that proteins experienced during their synthesis, folding or degradation (i.e. with ribosomal proteins, chaperones or proteins from degradation machinery). Those generic interactions are often stable and formed by so-called protein subunits (e.g. the spliceosome, the ribosome or nuclear pores). Nevertheless, the majority of protein-protein physical interactions are neither static nor permanent and are not necessarily present in all cells [3].

Attempts to estimate the size of the human protein interactome are reporting between 130'000 binary interactions based on empirical data [4] to a network size of 650'000 using statistical models [5]. Whatever the exact number is, current databases are far from being complete. Some proteins are part of so-called "hubs" in which interaction with multiple partners is occurring (usually above four). The study of all biological interactions in a given cell is called the interactome.

Despite such a complexity, a more comprehensive analysis of PPIs remains crucial for a better understanding of functional relevance, cellular processes and pathways as well as improve understanding of the pathology of diseases.

1.1.2 Structural types and binding affinities

PPIs vary by their nature, their function, their structure (diversity in the size and shapes of the interfaces), their reversibility, their dynamics of assembly and their binding affinities (from micro-molar to high pico-molar levels) [6, 7].

The lower level of complexity are protein oligomers, which result from proteins consisting of more than one polypeptide chain. Such oligomers can be arranged by multiple copies of the same polypeptide (homo-oligomers) or different polypeptides (hetero-oligomers). An example of high level of complexity would be the mega-complexes involved in RNA splicing (spliceosomes), transcription or translation (ribosomes) and which consist of couple hundreds of proteins interconnected.

One major distinction among PPI is based on whether interactions are permanent or transient as well as the range of their associated affinity (from micro-molar to high pico-molar levels) and temporal stability (Figure 1.1).
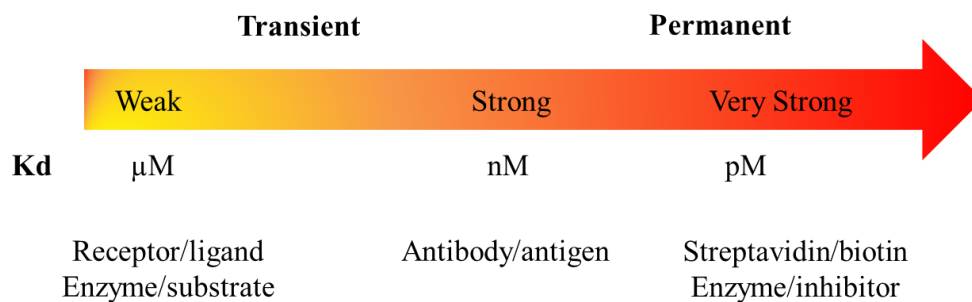
Figure 1.1: Protein-protein interactions vary in their affinity. Protein-protein interactions can be classified between permanent (with the typical example of the streptavidin-biotin biological interaction, the strongest non-covalent interaction known with a dissociation constant, Kd, in the femto-molar range) and transient interactions. Transient complexes can be either weak (Kd in the micro-molar range, with the example of receptor ligand interactions) or strong (Kd in the nano-molar range) which are associated with a greater temporal stability. The Kd is inversely correlated to the binding affinity.

Covalent binding is the highest affinity association between two proteins (via disulfide bonds or sharing of electrons) and is rarely the case of PPIs with the exception of SUMO proteins or ubiquitin. Transient and reversible PPIs are dynamic and cover most interactions involved in biochemical cascade [8]. Those interactions are mediated via hydrogen bonds, hydrophobic interactions, ionic interactions or van der Waals interactions, which are weak but can be distributed across multiple sites between the interacting proteins leading to a cumulative affinity.

PPIs can be mediated via PPI domains (peptide recognition motifs) but also outside specific amino-acid sequence or even in unfolded regions. Multiple PPI domains exist which are

evolutionary conserved among specific protein families with the most well-defined being the Src homology (SH1, SH2, SH3), Phosphotyrosine-binding (PTB), Pleckstrin homology (PH) and PDZ domains involved in cellular signal transduction [9].

1.1.3 Cellular regulation

PPIs are regulated to occur at a given time point and in a given location [10]. This spatial and temporal events are controlled at multiple levels via local protein concentration, protein affinity, presence of other proteins, co-localization of partners and physicochemical environment (pH, temperature, ions concentration). Allosteric regulation via small ligand binding (outside of the active site, discussed in paragraph 1.2.2.) or post-translational modification (PTM, such as phosphorylation, acetylation or hydroxylation) [11] will either change the amino-acid chemical properties involved in binding or affect protein folding. In addition to allosteric regulation, phosphorylation can create docking sites to interact with other proteins (e.g. SH2 and PTB recognizing phosphorylated tyrosine).

As cellular microenvironment has such an impact on protein interactions, it is critical to study PPIs as much as possible in *in vivo*-like conditions to ensure for their biological relevance.

1.2 Disease directly associated with PPIs

1.2.1 Pathologies

Unwanted aberrant or dysregulated PPIs are the cause of many and diverse pathologies [12]. Biochemical dysfunction is often caused by mutations in a binding domain or leading to protein allosteric changes. In the case of bacterial or viral infections, host-pathogen interactions are also playing a key role. Table 1.1. is listing a few pathologies having a direct link to PPIs.

Abnormal PPIs	Example of pathology	Molecular observations
Protein-DNA disruptions	Cancer	Mutation on p53's DNA binding domain preventing transcriptional activation of anti-cancer mechanisms
Protein misfolding	Von Hippel-Lindau syndrome	Y98H mutation in VHL binding site preventing interaction with HIF
Host-pathogen interactions	HPV-induced carcinogenesis	E6 and E7 viral proteins bypassing immune system via interaction with host regulators
New undesired protein interactions	Huntington's disease	Polyglutamine sequence elongation leading to huntingtin aggregation

Table 1.1: PPIs in human diseases. Aberrant PPIs are causing diverse pathologies.

1.2.2 PPIs as therapeutic targets

Due to their central role in biological processes and their links to pathologies, targeting the interfaces between interacting proteins both extracellularly and intracellularly represent promising therapeutic approaches [13, 14]. In the past, PPIs were defined as extremely challenging to target and intractable. Nevertheless, advancing our understanding of PPIs in the past decades enabled the development of potent PPI inhibitors targeting more than 40 PPIs with some of them having reached clinical trials (e.g. the anticancer drug idasanutlin occupying p53 binding pocket of MDM2). Notably, the contact surfaces are much larger (around 1'500 up to 3'000 Å) compared to those usually found between proteins and small-molecules (around 300 up to 1'000 Å). The nature of the surface is also drastically different. In protein-protein interactions contact surfaces are generally flat in contrast to deep grooves or pockets often bound by small molecules on proteins. Mutational studies revealed that not all residues from the contact surfaces are critical but that binding is occurring via so called "hot spots" which tend to be clustered at the center of the contact surface covering an area of the size of a small molecule.

Identification of small molecule has been strongly advanced by methods such as fragment-based lead discovery using biophysical methods (crystallography, surface plasmon resonance, nuclear magnetic resonance). Their binding affinities can be similar to that of the native protein partner. High-affinity small molecules can interact with residues that are not engaged with native protein partners or even deep pockets on the protein contact surface.

Although contact surfaces in PPIs often involve non-contiguous amino-acid residues, some inhibitors have been developed targeting primary peptide epitopes (short and continuous

peptides). Such inhibitors have been discovered either via fragment-based biochemical or cell-based screenings. The native ligand is usually close to the size of a small-molecule inhibitor. One example being tirofiban, mimicking a linear tripeptide Arg-Gly-Asp, the epitope of fibrinogen binding integrin IIbIIIa and thereby preventing the blood from clotting.

PPI inhibitors have been also developed to target secondary structure epitopes (extended binding sites such alpha helix or beta sheets) as well as tertiary structural epitopes (globular discontinuous binding sites). In those cases, key peptide residues are not contiguous in the primary sequence. A famous example targeting secondary structure epitopes are the Nutlins identified by HTS and which are binding MDM2 in the p53-binding pocket activating the p53 pathway in cancer cells [15].

PPIs inhibitors screens can deliver orthosteric inhibitors (binding the active site) but can also led by chance to the identification of allosteric mechanism of inhibition either inhibiting binding via altered conformation or even promoting aberrant interactions.

PPIs inhibitors are interesting molecules to treat viral infection as viruses hijack host proteins to facilitate their entry and replication. An example of a drug on the market is Maraviroc, an anti-HIV drug used as part of combination therapy and which inhibit virus entry by interacting with CCR5 found on the surface of cells of the immune system.

In this context, a better understanding of PPIs would definitely help to design PPIs inhibitors and would ultimately increase their prevalence into HTS libraries, which are currently biased as derived mostly from historical medicinal chemistry efforts in pharmaceutical companies.

1.3 Mapping the protein-protein interaction network
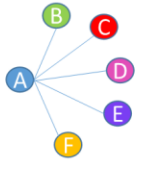
A cell can be considered as a web of complex interactions between macromolecules. Those complex systems are usually simplified by nodes (proteins) and edges (physical interactions). Edges usually do not have a direction as it is not always known which protein is functionally influencing the other [16] (the same for the distinction of activation versus repression). Interactome networks can be generated via different strategies. One strategy consists on compiling and curating existing and publicly available data. The accuracy of such maps is dependent on the methods used, which should be ideally standardized and unbiased. When combining multiple small-scale experiments, it is not possible to control for accuracy,

sensitivity and reproducibility. Another option is the use of unbiased high-throughput experimental strategies, which largely contributed in the last decade to the improvement of databases coverage [17]. On top of those empirical methods, computational techniques can be employed to predict interactions. Those predictions are usually based on factors such as amino-acid composition, structural information, phylogeny, gene conservation, domain co-occurrence, sequence similarities, post-translation modifications and sub-cellular location [30, 31]. Such computational methods differ from physical or biochemical approaches by using rather indirect information. To generate accurate map of PPIs, both completeness (number of pairs tested) and precision (proportion of true interactors) have to be taken into consideration.

1.4 Experimental strategies to profile PPIs

1.4.1 Co-complex and binary methods

In the past, biophysical methods were applied to study individual PPIs, mainly based on structural information (e.g. X-ray or NMR) [18]. More recent approaches to study PPIs are divided between those measuring direct physical interactions between protein pairs (binary methods) and those measuring physical interactions among group of proteins (co-complex methods) [1, 3]. Each method has its own strengths and weaknesses and usually a combination of approaches is needed to elucidate PPIs with sufficient confidence. Table 1.2 is summarizing the most popular approaches with their respective assets as well as limitations.

	Platform	Resulting protein pairs	Advantages	Limitations
CO-COMPLEX	AP-MS (tagged protein)		Specific antibody Whole proteome profiling	Tag leading to artificial interactions; close to endogenous level; specificity
	AP-MS (endogenous protein)		No label/tag Whole proteome profiling Endogenous target	Specificity Rely on antibody
	Co-fractionation-MS		No label/tag Whole proteome profiling	Resolution Mass limitation
BINARY	Y2H		High-throughput format	Specificity Yeast background, limited to nucleus
	SPR		Quantitative (Kd estimation) Sensitivity	Prior knowledge required No cellular context
	Protein microarrays		High-throughput format Sensitivity	High costs Labor intensive
	FRET		Living and fixed cells	Prior knowledge required Technical complexity

Table 1.2: List of most popular approaches to screen PPIs. Approaches can be divided between binary and co-complex approaches and each has its own advantages and limitations. AP-MS: Affinity purification coupled to mass spectrometry, Y2H: Yeast two-hybrid, SPR: Surface plasmon resonance and FRET: Fluorescence resonance energy transfer.

Among system-wide mapping methodologies, the most popular are the yeast two-hybrid system (Y2H) and the affinity purification coupled to mass spectrometry (AP-MS).

1.4.2 Yeast two-hybrid (Y2H): a system-wide binary method

The Y2H technique was first described in 1989 [19] and is testing the interaction between two proteins (bait and prey) via fusion to a transcription-binding domain within a yeast system. Upon interaction, the complex is activated and reporter gene are transcribed leading to detection of its product. Its main advantage is its adaptation to high-throughput screening (HTS). Y2H was the first approach able to interrogate the protein interactome on a genome scale [4, 20, 21]. On the other side, by forcing two proteins to localize to the nucleus, misfolding and aggregation of membrane proteins can occur which can explain lower detection rate as well as a bias toward non-specific interactions. In addition, mammalian proteins expressed in yeast are not necessarily harboring the PTMs usually relevant for their function and which may guide specific interactions.

1.4.3 Affinity purification coupled to mass spectrometry (AP-MS): a system-wide co-complex method

Co-complex methods are capturing both direct and indirect associations within a complex. Subsequently, additional technologies are needed to distinguish direct versus indirect interactors. The most popular technology is the combination of affinity purification and mass spectrometry (AP-MS, Figure 1.2.) where prey proteins are co-captured with an affinity captured bait protein. AP-MS appeared more than a decade ago and is combining the sensitivity of immunocapture together with the power of MS readout [22-24]. The bait protein is either the endogenous protein when a specific and selective antibody is available or a tagged version (e.g. with GST, FLAG, HA, 6xHis or Myc) expressed in cells [25, 26] ideally close to *in vivo* concentration. The first option (called label-free) might be favored whenever possible as classical tags can lead to protein misfolding and may expose hydrophobic surfaces that result in artificial interactions (false positive) [27]. However, it is crucial to select an antibody with a good efficiency of affinity depletion and high specificity. Even after rigorous antibody selection (e.g. with the help of public database mining [28]), one major drawback of classical AP-MS approach remains the lack of specificity. Indeed, the main challenge with AP-MS is to discern bona fide interactors from highly abundant cellular proteins or proteins binding nonspecifically to the affinity matrix. Strategies to overcome this limitation will be discussed in section 1.6.
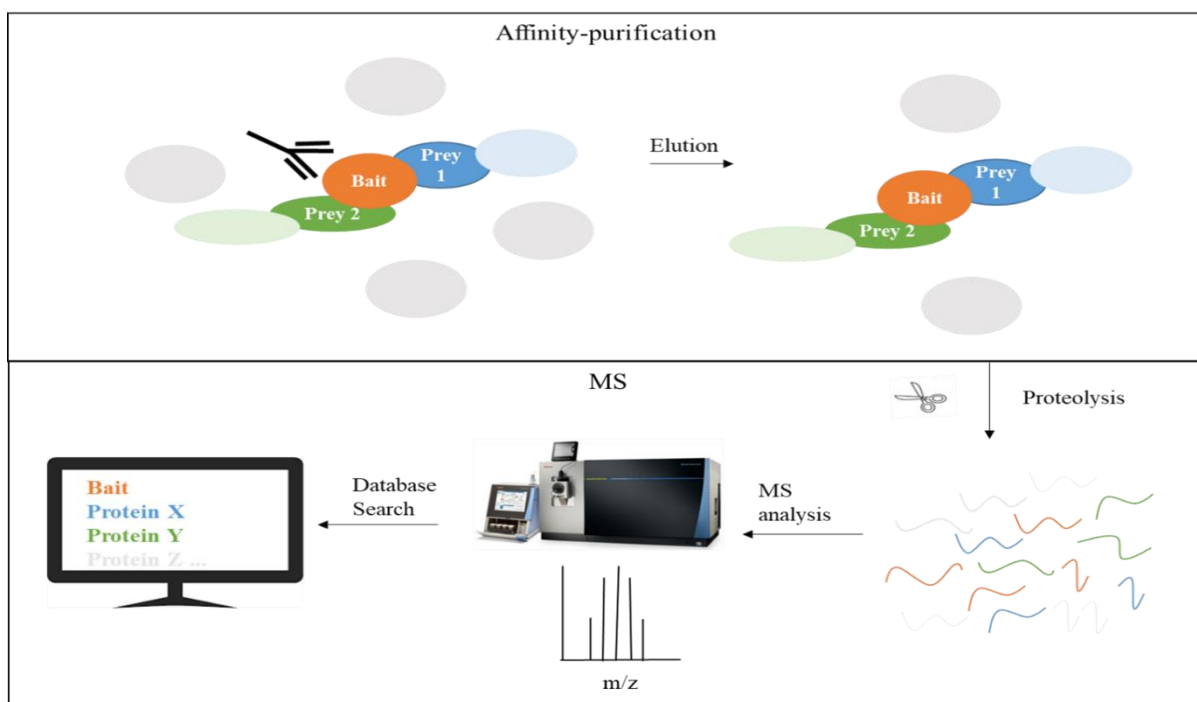
Figure 1.2: Overview of affinity purification coupled to mass spectrometry (AP-MS) workflow. A protein of interest (or the bait, in orange) is affinity purified from a cell lysate together with its direct interacting partners (or preys, in dark blue and dark green), indirect partners (in light blue and light green) as well as contaminants (in grey, typically highly abundant proteins or sticky by nature). Proteins are then digested into peptides (usually with trypsin), which are analyzed by mass spectrometry (MS). Not represented are separation of peptides by reversed-phase liquid chromatography and the two MS events. In a first scan, mass/charge ratio (m/z) of intact peptides is measured. A second scan is then applied on the most abundant peptides, which are fragmented into smaller fragment ions (e.g. by collision-induced dissociation or CID) generating a tandem MS (MS/MS) spectrum. Database searching is then matching spectra with *in-silico* predicted peptides (mass & sequence) generating a list of proteins present in the sample (including the bait, co-complexes and contaminants).

Above mentioned methodologies can be done in a low-throughput fashion (interactome of a single protein of interest) or in a high-throughput format to map global interactomes. Quite recently, four human interactome maps have been published which led to the identification of almost 93'000 unique proteins interactions and associations (indirect interactions detected by co-complex method) (reviewed in [29]). Their overlap is low which is rather explained by the fact that each map has been generated using different methodologies/protocols than by potential concerns in data quality. Only a combination of approaches can capture the different features of the human interactome.

1.5 Protein-protein interactions databases

PPIs databases are combining the knowledge arising from empirical as well as computational methods together with text mining [30]. An example of a popular database storing interaction data is BioGRID (https://thebiogrid.org/) which is currently covering more than 30'000 unique human PPIs. BioGRID together with DIP (https://dip.doe-mbi.ucla.edu) or IntAct (https://www.ebi.ac.uk/intact/) are primary databases collecting information about experimentally proven PPIs via small or large-scale experimental methods. As each database is lacking a proportion of total PPIs, it is crucial to integrate information to increase coverage [3]. Meta-databases are integrating information from different primary databases for example APID (http://apid.dep.usal.es) and PINA (https://omictools.com/pina-tool). Confidence of interactions will be based on the number of experimental proofs reported. STRING (http://string-db.org) is another example of such web-based interface combining genomic context, experimental evidences, co-expression and text mining. Depending on the level of

confidence score set, it can lead to many false positives. It is then possible to combine PPIs information with additional parameters such as gene expression, genetic interaction, colocalization or Reactome/KEGG signaling/metabolic pathways. Combining such heterogeneous data is challenging but provide additional confidence that an interaction might be leading to a biological effect. One of the most widely used open source bioinformatics tool modeling networks is Cytoscape (www.cytoscape.org).

At the end, the extent to which databases are reflecting the biologically relevant interactome remains unknown. Current knowledge of the interactome is still incomplete and noisy. This can be explained by the limitations of the methods to detect truly physiological interactions associated with the identification of many false positives and false negatives. Current databases are biased towards highly expressed proteins and well-studied pathways. To increase reliability, non-reported true negative proteins should be incorporated (i.e. when an interaction has been studied and result was negative). There is currently no effort to refine the network by eliminating false positive. Despite extensive efforts, the human interactome is still a work in progress.

1.6. Strategies to increase AP-MS specificity and improve databases' accuracy

As previously mentioned, AP-MS is a powerful and widely used approach to profile in an unbiased manner interacting partners of endogenous proteins and by such is largely contributing to interactome databases. AP-MS is applied for the analysis of protein complexes and interactions networks either of small-scale (targeted interactome of one protein of interest), medium-scale (interactome of protein members of a family) or large-scale (systematic tagging of open reading frames, ORFs [31, 32]). In any cases, filtering out background contaminants co-purified with bait proteins and subsequently identified by MS is challenging but crucial in order to improve accuracy of those databases. The majority of contaminants are proteins binding non-specifically to the affinity matrices (e.g. sepharose, agarose or magnetic beads). There is also a clear relationship between the abundance of a protein and its detection as contaminants. At the end, the percentage of true interaction partners usually represent less than 10% of all identified proteins [33].

1.6.1. Computational approaches

Various scoring methods have been developed to rank interaction partner specificity [34]. Earliest methods for filtering AP-MS datasets were based on binary comparison (presence or absence of a protein). Quantitative information was then incorporated such as spectral counts or peak intensity in label-free quantification. Including multiple replicates in an experimental design is allowing more rigorous statistics such as t-test where p-values indicate whether an observed fold change in abundance is statistically significant based on the variance across all replicates. Combining then p-values with the magnitude of fold change (for example in a Volcano plot) finally improve the distinction between false and true interactors. The latter is likely to have a high fold change and high significance (low p-value).

It is also possible to screen a dataset against a set of negative controls formerly generated by the proteomics research community (published and unpublished) for example using the CRAPome web-accessible resource (https://www.crapome.org/) [35]. It implies that raw mass spectrometry files are processed via a uniform data analysis pipeline. To score interactions, both the abundance levels of co-precipitating proteins and their occurrence in negative controls are taken into account. The type of background identified is clearly linked to the experimental conditions, that is why it is recommended whenever possible to keep some user controls in addition to the controls deposited in CRAPome. For large-scale studies in absence of negative controls, alternative scoring models exist which are based on the analysis of a prey compared to its purification with different baits (SAINT [36], ComPASS [37], MiST [26]). A true interactor in these models are proteins having a high intensity (abundance) over replicated experiments (reproducibility) but only with a few baits (specificity).

With any of the above-mentioned tools, proteins classified as "non-specific" should be tagged rather than excluded. Indeed, a highly abundant protein (such as keratin, cytoskeletal proteins, histones, ribonucleoproteins or heat-shock proteins) or a promiscuous protein can be a true interactor to a specific protein in particular conditions. Computational efforts to refine data analysis need to be combined with experimental strategies that improve the selectivity at the IP step.

1.6.2. Experimental strategies

There is a great need for novel experimental strategies aiming at reducing the rate of false positives in AP-MS experiments. It is possible to increase the stringency of purification methods (e.g. by increasing the number of washes or salts concentration) but although this would result in a reduction in the level of nonspecific binding, it would also potentially remove specific binders of low affinity or low abundance. A pre-clearing step on the matrix might also lead to a loss of information.

The best strategy to filter out contaminants to date is to combine the use of negative controls with quantitative proteomics [33, 38]. A control experiment is performed in parallel of AP of the protein of interest and abundance ratios are then quantitatively compared [38]. The three most common labeling techniques used in AP-MS experiments are ICAT (chemical labeling at level of protein), iTRAQ (chemical labeling at level of peptide) and SILAC (metabolic labeling at level of organism/cell) [39].

A mock purification using the support resin only or using a control antibody even of the same isotype are considered as poor controls. An example of a suitable negative control is a cell line that does not express the bait proteins (e.g. by knock-down or parental cell line versus the one expressing tagged protein). As minor variations in sample or sample preparation can greatly influence the capture of proteins, the closer the control is to the actual experiment, the better. Silencing approaches can be quite efficient when incorporated in the workflow to identify target related binders but is quite labor intensive [40]. Replicates are also needed to increase confidence of AP-MS experiments, at least triplicates to allow for statistics. Nevertheless, classical quantitative side-by-side comparison of a sample with its control still suffers from the fact that the control sample is not identical to the probed one and both samples can lead to the association of different nonspecific binders.

In this thesis an innovative way to filter out non-specific binders is proposed leading to an approach combining specificity, simplicity and proteome wide accessibility.

1.7. Aims and objectives

Protein-protein interactions (PPIs) are controlling the majority of biological processes. PPIs are diverse and organized in complex networks, which explains why their study is so challenging. None approaches can capture the full complexity and there is a big need for innovative platforms to ultimately improve reliability of current protein-protein databases.

The most popular platform is the combination of affinity purification and mass spectrometry (AP-MS) although with such an approach it is challenging to distinguish true interactors from non-specific binders.

The aim of this work was to overcome the limitation of classical AP-MS approaches in term of specificity by decreasing the number of false-positives. The Immuno-Competitive Capture Mass Spectrometry platform (ICC-MS) is a three steps approach combining a competitive binding (pre-competition step using the free form of the same antibody as the capture antibody) prior to subsequent affinity purification and mass spectrometry analysis. A statistical analysis is then applied on MS signals to extract the cellular interactome of proteins of interest while filtering out background proteins.

Within this dissertation, the first task was to run a proof-of-concept in a viral-host model (HCV NS5A in Con1-Huh-7). Next, the capability of ICC-MS to deliver a limited number of interacting candidates with highest potential for biological relevance was further assessed. For this purpose, Gpc2 and Htra1 cellular interactome were executed respectively in a human neural stem cell model and in a retinal pigmented epithelial model.

The last task was to adapt the protocol to the screening of oligonucleotide-protein interactions. Oligo-Competitive Capture Mass Spectrometry (OCC-MS) was evaluated in the context of RNA-protein and single-stranded antisense oligonucleotide-protein interactions.

2. MATERIAL & METHODS

2.1. Material

2.1.1 Chemicals and solutions

Chemical / solution	Manufacturer
Accutase	Sigma-Aldrich
Acetic acid	Sigma-Aldrich
Acetonitrile	Biosolve
Ammonium Bicarbonate	Fluka
ATP	SignalChem
BCA Protein Assay kit	Thermo Scientific
Bradford Assay Reagent	Thermo Scientific
$CaCl_2$	Merck
Colloidal Blue Staining Kit	Invitrogen
Dithiothreitol	Roche Applied Sciences
Dulbecco's Modified Eagle Medium (D-MEM), high glucose	Gibco
D-MEM, high glucose, GlutaMAX, Pyruvate	Gibco
Ethanol	Merck
Ethanolamine	Aldrich Chemistry
FBS	Sigma
FBS heat inactivated	Sigma
Formic acid	Fischer Chemicals
Geneticin	Gibco
Hepes	Gibco
Iodoacetamide	Sigma-Aldrich
KCl	Sigma-Aldrich
Kinase activity buffer	SignalChem
Lumi-Light Western Blotting Substrate	Roche Applied Sciences
MagicMark XP Western Protein Standard	Thermo Fischer Scientific
Mark12 unstained standard	Invitrogen
Methanol	Sigma
$MgCl_2$	Merck
Milk powder	Fluka
NaCl	Merck
Nonidet P-40	Calbiochem

Nupage Sample Reducing agent	Life Technologies
Phosphate buffered saline	Gibco
Penicillin-Streptomycin	Gibco
Phosphatase Inhibitor Cocktail PhosSTOP	Roche Applied Sciences
Pro-Q Diamond Phosphoprotein Gel Stain	Thermo Fischer Scientific
Protease inhibitor Cocktail Complete EDTA free	Roche Applied Sciences
Protector RNase inhibitor	Roche Applied Sciences
SeeBlue Plus 2 pre-stained protein standard	Invitrogen
Sodium azide	Sigma
SuperSignal West Femto	Pierce
Sypro Ruby Protein Gel Stain	Thermo Fischer Scientific
Tris base	Applichem
Tris-Glycine SDS Running Buffer	Invitrogen
TrypLE Express (cell collection)	Gibco
Trypsin (cell collection)	Gibco
Trypsin (protein digestion)	Promega
Tween 20	Sigma-Aldrich

2.1.2 Devices

Devices	Manufacturer
Automated cell counter	Bio-Rad
Blotter	Bio-Rad
Cell culture microscope Primovert	Zeiss
Cell incubator	Thermo Scientific
Concentrator	Eppendorf
Criterion tank	Bio-Rad
Laminar flow	Skan, Telstar
LTQ-Orbitrap Velos	Thermo Scientific
Luminescent image analyzer LAS4000	Fujifilm
Magnetic stirrer	IKA
Micromix5	DPC
Microplate reader	Molecular Devices
MilliQ ultrapure water system	Millipore
Mutlichannel pipette	Rainin
Multi-rotator	Grant Bio
MixMate	Eppendorf
Nanodrop spectrophotometer	Thermo Scientific

Nanoflow Easy-nLC	Thermo Scientific
Orbital shaker	IKA
pH meter	Mettler Toledo
Pipetboy	Integra
Pipettes (10, 20, 100, 200, 1000, 2500 µl)	Eppendorf
Power supply PowerPac	Bio-Rad
Thermomixer	Eppendorf
Thermo Sealer	Thermo Scientific
Q125 Sonicator	Qsonica
Tribrid Orbitrap Fusion	Thermo Scientific
Tubes and plates centrifuge	Eppendorf
Vortex	Scientific Industries
Weighing scale	Mettler Toledo
XCell Sure Lock	Invitrogen

2.1.3 Consumables

Consumables	Manufacturer
Acclaim PepMap	Thermo Scientific
Affi-Gel 10 beads	Bio-Rad
Aluminum foil lids	Beckman Coulter
AQUA C18 Trap	Phenomenex
BC2-Nanotrap	(Rothbauer, 2016)
Cell Culture Flasks T-25, T-75, T-125	Corning
Cell scraper	Sarstedt
Combitips (0.5, 2.5, 5, 10 mL)	Eppendorf
Easy spray column PepMap C18	Thermo Scientific
Falcon pipettes (2, 5, 10, 25 mL)	Corning
Falcon tube (15, 50mL)	Corning
GFP-Trap	ChromoTek
LoBind tubes (0.5, 1, 1.5, 2, 5 mL)	Eppendorf
Novex Tris-Glycine SDS sample buffer	Life Technologies
Novex 4-20% Tris-Glycine	Invitrogen
Petri dish	Corning
pH indicator strips	Macherey-Nagel
Pipette tips (10, 20, 200, 1000, 2500 µL)	Rainin, Eppendorf
PVDF membrane	Millipore
Reprosil-Pur C18-AQ	Dr. Maisch

REMP Sealing and piercing fil	Brooks Life Sciences
SteriCup	Merck
Surgical disposable scalpel	Braun
Syringe	Hamilton
Tip-Tub reservoir	Eppendorf
TopSeal	Perkin Elmer
Zeba Spin desalting column	Pierce
4-20% Criterion™TGX™ Precast gel	Bio-Rad
96 well flat-bottom	Thermo Scientific
96 well filter plate	Millipore
96 well PCR plate	Thermo Scientific
96 well perforated plate	Proxeon
96 well V-bottom	Greiner Bio-one

2.1.4 Antibodies

Following primary and secondary antibodies were used in this study

2.1.4.1 Primary antibodies

Antibody (species)	Manufacturer
Anti-Aconitase (rabbit, pAb)	Abcam
Anti-C1QTNF5 (mouse, mAb)	R&D Systems
Anti-C1QTNF5 (pig, pAb)	LSBio
Anti-C1QTNF5 (rabbit, pAb)	Sigma-Aldrich
Anti-FGF2 (rabbit, pAb)	Abcam
Anti-hnRNP A1 (ROA1) (rabbit, pAb)	Abcam
Anti-HTRA1/PRSS11 (mouse, mAb)	R&D Systems
Anti-HTRA1 (in house)	(Vierkotten, 2011)
Anti-HuR (mouse, mAb)	Thermo Scientific
Anti-H2B (rabbit, pAb)	Merck
Anti-LaminB (rabbit, mAb)	Abcam
Anti-LATS1 (rabbit, pAb)	Abcam
Anti-LATS1 C66B5 (rabbit, mAb)	Cell Signaling Technology
Anti-LATS2 (rabbit, pAb)	Abcam
Anti-LATS2 D83D6 (rabbit, mAb)	Cell Signaling Technology
Anti-NS5A 7D4 (mouse, mAb)	Santa Cruz Biotechnology
Anti-NS5A H26 (mouse, mAb)	Abcam
Anti-NS5A H110B (mouse, mAb)	Thermo Scientific

Anti-Tubulin beta KMX-1	Sigma-Aldrich
Anti-Tubulin beta III	Abcam
Anti-14-3-3 alpha beta (rabbit, pAb)	Abcam

2.1.4.2 Secondary antibodies

Antibody (species)	Manufacturer
Anti-mouse (goat) IgG	Jackson Laboratories
Anti-rabbit (goat) IgG	Jackson Laboratories

2.1.4.3 Recombinant proteins

Recombinant protein	Manufacturer
Human LATS2 kinase	SignalChem
NS5AΔ32 BK strain	Roche Nutley
Human Glypican 2	R&D Systems
FGF2	PeproTech

2.1.5 Oligonucleotides

2.1.5.1 Antisense oligonucleotides (ASOs)

Antisense oligonucleotide	Target	Sequence 5' to 3'
ASO-A	*VEGF*	TmCTcctcttccttmCAT
Bio-ASO-A	*VEGF*	5'-BioHEG-TmCTcctcttccttmCAT
ASO-B	*PCSK9*	TG^{m}Ctacaaaac^{m}C^{m}CA
GalNac ASO-B	*PCSK9*	5'-GN-C6caTG^{m}Ctacaaaac^{m}C^{m}CA
Bio-ASO-B	*PCSK9*	TG^{m}Ctacaaaac^{m}C^{m}CA-BioTEG-3'
GalNac Bio-ASO-B	*PCSK9*	5'-GN-C6caTG^{m}Ctacaaaac^{m}C^{m}CA-BioTEG-3'
ASO-C	*PCSK9*	G^{m}CtgtgtgagcttGG
GalNac ASO-C	*PCSK9*	5'-GN-C6G^{m}CtgtgtgagcttGG
Bio-ASO-C	*PCSK9*	G^{m}CtgtgtgagcttGG-BioTEG-3'
GalNac Bio-ASO-C	*PCSK9*	5'-GN-C6GmCtgtgtgagcttGG-BioTEG-3'
ASO-D	*Myd88*	CACattccttgctCTG
MOD ASO-D	*Myd88*	CACattccttgctCTG
Bio-ASO-D	*Myd88*	BioTEG-CACattccttgctCTG

Table 2.1: List of ASOs with their respective targets and 5' to 3'sequence

Antisense oligonucleotide	Provider
ASO-A	Erich Koller, Roche Innovation Center Basel
Bio-ASO-A	Susanne Kammler, Roche Innovation Center Copenhagen
ASO-B	Sabine Sewing, Roche Innovation Center Basel
GalNac ASO-B	Sabine Sewing, Roche Innovation Center Basel
Bio-ASO-B	Adrian Schäublin, Roche Innovation Center Basel
GalNac Bio-ASO-B	Adrian Schäublin, Roche Innovation Center Basel
ASO-C	Sabine Sewing, Roche Innovation Center Basel
GalNac ASO-C	Sabine Sewing, Roche Innovation Center Basel
Bio-ASO-C	Adrian Schäublin, Roche Innovation Center Basel
GalNac Bio-ASO-C	Adrian Schäublin, Roche Innovation Center Basel
ASO-D	Microsynth
MOD ASO-D	Microsynth
Bio-ASO-D	Microsynth

Table 2.2: List of ASOs with their respective providers

2.1.5.2 Oligonucleotides

Oligonucleotide	mRNA	Sequence 5' to 3'
ESE2	*SMN2 exon 7*	AAAAAGAAGGAAGG
Bio-ESE2	*SMN2 exon 7*	AAAAAGAAGGAAGG-Bio
AR	*Androgen Receptor 3' UTR*	CUGGGCUUUUUUUUCUCUUUCUCUCCUUUCUUUUUCUUCUUCCCUCCCUA
Bio-AR	*Androgen Receptor 3' UTR*	CUGGGCUUUUUUUUCUCUUUCUCUCCUUUCUUUUUCUUCUUCCCUCCCUA-Bio
IRE	*UTR of various mRNAs (e.g. ferritin)*	UCCUGCUUCAACAGUGCUUGGACGGAAC
Bio-IRE	*UTR of various mRNAs (e.g. ferritin)*	UCCUGCUUCAACAGUGCUUGGACGGAAC-Bio

Table 2.3: List of oligonucleotides with their respective mRNA and 5' to 3'sequence

Oligonucleotide	Provider
ESE2	Manaswini Sivaramakrishnan, Roche Innovation Center Basel
Bio-ESE2	Manaswini Sivaramakrishnan, Roche Innovation Center Basel
AR	Pierce™ Magnetic RNA-Protein Pull-Down Kit
Bio-AR	Sonja Meier, Roche Innovation Center Basel
IRE	Sonja Meier, Roche Innovation Center Basel
Bio-IRE	Pierce™ RNA 3' End Biotinylation Kit

Table 2.4: List of oligonucleotides with their respective providers

2.2. Methods

2.2.1 Cell culture methods

2.2.1.1 Huh7 parental cell line and human hepatocyte-derived Con1-Huh7 cells (Con1-Huh7)

Huh7 cells and Huh7 cells stably expressing Con-1 genotype 1b and the renilla luciferase reporter gene (Huh7 2209–23) [41] were provided by Junjun Gao (former Hoffmann-La Roche site in Nutley) as previously described [42]. Cells were cultured/maintained in DMEM high glucose supplemented with GlutaMAX, sodium pyruvate (110 mg/l), heat inactivated FBS and antibiotics (500 μg/ml penicillin-streptomycin, 500 μg/ml geneticin) at 37°C in a humidified chamber with a 5% CO_2 atmosphere. Cells were detached using 0.05% trypsin-EDTA (5 min at 37°C under 5% CO_2) and harvested by centrifugation (3 min at 150 x g).

2.2.1.2 Differentiated human neural progenitor (d21 NSC) and human Gpc2 transfected H4 neuroglioma cells (hGPC2 H4)

Differentiated neural progenitor cell pellets and human Gpc2 transfected H4 neuroglioma cell pellets were provided by Sebastian Kim Lugert (Roche Innovation Center Basel) [43].

2.2.1.3 Human fetal retinal pigment epithelial cell (hfRPE)

Human RPE were provided by the lab of Roberto Iacone (Esther Melo, Corinne Stucki and Frédéric Delobel, Roche Innovation Center Basel) [44]. For collection from transwells, cells were washed with PBS and incubated with accutase (at 37°C and under 5% CO_2 until the cells are detached) and scraped off. Finally cells were harvested by centrifugation (3 min at 150 x g).

2.2.1.4 Human SMA type 1 fibroblasts

SMA type 1 fibroblasts were provided by the lab of Friedrich Metzger (Roche Innovation Center Basel) and maintained in DMEM (high glucose) supplemented with GlutaMAX, 110 mg/l pyruvate and 10% heat inactivated FBS at 37°C under 5% CO_2. For collection, cells were detached using trypLE Express (1 min at 37°C under 5% $C0_2$) and harvested by centrifugation (3 min at 150 x g).

2.2.1.5 Human renal proximal tubular epithelial cell (PTEC)

PTEC-TERT1 were provided by Laurence Hilfiger from the lab of Marcel Gubler (Roche Innovation Center Basel) according to the published protocol [45]. Cells were then maintained

in DMEM/F12 supplemented with 2% FBS, detached with 0.05% trypsin-EDTA (5 min at 37°C under 5% CO_2) and harvested by centrifugation (3 min at 150 x g).

2.2.1.6 Human embryonic kidney 293 cells (Hek293)

Hek293 cells were maintained in DMEM (high glucose) containing 10% FBS at 37°C under 5% CO_2. For collection, cells were detached using 0.05% trypsin-EDTA (5 min at 37°C under 5% CO_2).

2.2.2 Biochemical methods

2.2.2.1 Mammalian cell lysis

Harvested PTEC, Hek293, d21 NSC, hGpc2 H4, hfRPE cells (fresh pellets or stored at -20°C) were resuspended in a PBS buffer supplemented with 1% (v/v) Nonidet P-40 and protease inhibitor. RNase inhibitor was also added for ASO pull-down experiments. After 15 min on ice, lysates were cleared by centrifugation at 4°C (10 min at 20'000 x g).

Harvested Huh7 and Con1-Huh7 replicon cells were resuspended in a 50 mM Tris pH 7.5 lysis buffer containing, 150 mM NaCl, 0.5% (v/v) Nonidet P-40, protease inhibitor and phosphatase inhibitor. After 15 min on ice, lysates were cleared by centrifugation at 4°C (10 min at 1'500 × g).

To generate SMA type 1 human fibroblasts nuclear extracts, harvested cells were first resuspended in Hepes 10mM pH 7.8 supplemented with MgCl2 1mM, KCl 10mM, protease inhibitor and homogenized with a potter (8 strokes at 800 rpm). Lysates were centrifuged at 4°C (10 min at 500 x g). The supernatant was collected (cytosolic fraction) and pellets were resuspended in a 50 mM Hepes pH 7.3 buffer containing 150 mM NaCl, 1 mM CaCl2, 1 mM MgCl2, 0.5% (v/v) Nonidet P-40, EDTA-free protease inhibitor and RNase inhibitor. After two sonication cycles (each 10 pulses of 1 second with 1 second of pausing between each pulse) at an amplitude of 60%, samples were finally cleared by centrifugation at 4 °C (10 min at 300 x g).

Protein concentrations were estimated using the bicinchoninic acid protein assay according to manufacturer's protocol (Thermo Fischer Scientific).

2.2.2.2 Tissue extraction

Mouse livers and kidneys (courtesy of Roche comparative pharmacology department) were weighted and 8 ml per gram of tissue of PBS buffer supplemented with 1% (v/v) Nonidet P-40 and protease inhibitor was added. After 10 min incubation on ice, tissues were extracted mechanically using a potter (8 strokes at 800 rpm) followed by two sonication cycles (each 10 pulses of 1 second with 1 second of pausing between each pulse) at an amplitude of 60%.

2.2.2.3 SDS-PAGE and Western blot

Proteins were separated by size on 4-20% Tris-Glycine denaturing polyacrylamide gels. Samples containing 1 x Laemmli sample buffer (0.1% 2-Mercaptoethanol, 0.0005% Bromophenol Blue, 10% Glycerol, 2% SDS in ddH2O) were denatured by adding reducing agent (DTT, 50mM final) followed by a boiling step at 85°C for 5 min. Separation was done either by running the gel fully (with Invitrogen precast gels), or only half its length (with Bio-Rad precast gels) according to Table 2.5. A short 1 cm migration (with Bio-Rad precast gels) was performed to eliminate SDS only. Gels were fixed for 1 hour in a fixation solution (40% ethanol, 10% acetic acid).

Migration	Voltage
Full	60 V for 20 min and 2 hrs at 125 V
Half	100 V for 10 min and 30 min at 150 V
1 cm	100 V for 10 min and 5 min at 150 V

Table 2.5: Voltage and time of migration

Separated proteins were either stained with Coomassie blue (Colloidal Coomassie blue in a methanolic solution supplemented with ammonium sulfate and phosphoric acid) or transferred to a PVDF membrane by Western blotting.

For Coomassie staining, gels were stained either overnight or at least for 3 hours under agitation at room temperature. Gels were finally rinsed in water.

Proteins were transferred from gels to 0.45 µm PVDF membranes at 100 V for 1 hour in a transfer buffer (Tris 25 mM, glycine 192 mM, methanol 5%). Membranes were blocked for 1 hour at room temperature in Tris buffered saline (TBS) buffer (Tris 10mM, NaCl 150mM) supplemented with 0.1% Tween 20 (TBST) and 5% non-fat dry Milk (MTBST). Incubation with primary antibodies (see section 2.1.4.1) was performed in MTBST overnight at 4°C. The next day, membranes were washed three times with TBST followed by incubation with

secondary antibodies (see section 2.1.4.2) diluted in MTBST for 1 hour at room temperature. After three additional washing steps, chemiluminescent substrate was added and chemiluminescent signals were detected on a Fujifilm LAS4000 Imager.

2.2.2.4 Preparation of immunoaffinity columns

NS5A, Htra1 and Gpc2 antibodies were desalted on Zeba columns following Pierce's instructions and diluted in PBS to reach a concentration of 1 mg/ml. Antibodies were then coupled to 1 ml of Bio-Rad Affi-Gel 10 agarose beads (pre-washed in ice cold water and equilibrated in PBS) for 3 hours at 4 °C with gentle rotation. Coupling efficiency was determined with a Bradford assay comparing pre-coupling antibody versus post-coupling supernatant. Unreacted binding sites were blocked by the addition of 0.2 M ethanolamine in PBS for 3 hours at 4°C with gentle rotation. Beads were finally washed four times with cold PBS, re-suspended in 1 ml PBS (50% slurry) and stored at 4 °C.

2.2.2.5 Immunoprecipitation and immuno-competitive capture

Affi-gel beads with immobilized NS5A, Gpc2 or Htra1 (40 µl of suspension corresponding to 20 µl of beads) were washed, equilibrated in cold cell lysis buffer and centrifuged 30 sec at 3'000 g. After discarding supernatant, beads were incubated 1-3 hours with cell lysates at 4 °C with gentle rotation, washed four times with lysis buffer and eluted with non-reduced SDS sample buffer 10 min at 65°C. Beads were centrifuged 1 min at 3'000 g and reducing agent (DTT, 50 mM final) was added to the supernatant followed by 5 min heating at 85°C. For competition experiments, cell lysates were pre-incubated with increasing amounts of free antibodies for 1-3 hours at 4°C before being loaded onto Affi-gel immobilized antibodies according to Table 2.6.

Study	Number of concentration	Concentrations (µg/ml)	Number of replicates
NS5A	12	0, 0.6, 1, 1.7, 2.9, 5, 8.5, 14.5, 24.6, 41.9, 71.2, 121	1
Gpc2	5	0, 1, 2.5, 5, 10	3
Htra1	5	0, 1, 2.5, 5, 10	3

Table 2.6: Competitive-binding conditions. Concentrations of free antibody and number of replicates.

2.2.2.6 Htra1 and C1QTNF5 recombinant expression and co-immunoprecipitation

Recombinant expression of Htra1 and C1QTNF5 in Hek293 cells and co-immunoprecipitation experiments were performed at the Natural and Medical Sciences Institute (NMI) in Tübingen by Philipp Kaiser. Briefly, Hek293 cells were transiently transfected with HtrA1-pTag (52.7

kDa) and Tubulin-EGFP (77.6 kDa) prior to lysis and immunoprecipitation with BC2-Trap (immunoprecipitation of HtrA1-pTag). Immunoblots were generated using either gene- or tag-specific antibodies.

2.2.2.7 ASO pull-downs and competitive assays

100 pmol of biotinylated IRE, AR, ESE2, ASO-A, ASO-B, ASO-C and ASO-D were coupled to 50 ul of nucleic-acid compatible streptavidin magnetic beads (supplied at 10 mg/ml) according to the manufacturer's instructions. Briefly, magnetic beads were washed in 50 µl Tris HCl 20 mM pH 7.5, equilibrated in 50 µl of Tris HCl 20 mM pH 7.5, 1M NaCl, 1mM EDTA and incubated with biotinylated oligos for 30 min at room temperature under agitation. Beads with immobilized oligos were equilibrated in lysis buffer containing RNase inhibitor and incubated with cell extracts at 4°C for 1 hour under agitation. Flow-through samples were collected and beads were washed three times with cell lysis buffer. Elution was performed by adding SDS sample buffer containing DTT 5 min at 85°C. For competition experiments, cell lysates were pre-incubated for 1 hour with increasing amounts of free oligos in triplicates prior to be loaded on streptavidin immobilized oligos.

2.2.2.8 In-gel digestion and peptide extraction

Gel bands of maximum 1 cm were cut with a scalpel. Each band was further cut in squares of ≈ 1 mm x 1 mm and loaded in wells from perforated or filter based 96 well plates. For protein in-gel digestion, an adapted protocol of Shevchenko and colleagues [46] was used. Proteins were reduced with 50 mM DTT in 100 mM ammonium bicarbonate for 1 hour at room temperature and alkylated with 55 mM IAA in 100 mM ammonium bicarbonate for 30 min at room temperature in the dark. 50 ng of trypsin in 10% acetonitrile/50 mM ammonium bicarbonate was added per well. Digestion was performed overnight at room temperature.

Peptides were extracted twice with 1:2 (v/v) acetonitrile/25 mM ammonium bicarbonate, and twice with 1:2 (v/v) acetonitrile/5% formic acid, each steps for 15 min at room temperature. Peptides were dried down using a speedvac, and stored at −20°C before analysis.

2.2.2.9 In vitro NS5A phosphorylation

Recombinant NS5A Δ32 (BK strain, residues 33–447) was expressed in Escherichia coli and isolated from the soluble fraction as reported (30). The purified protein (2 µg) was premixed with 500 ng active human LATS2 (residues 480–1088) in kinase activity buffer (12.5 mM β-glycerol phosphate, 25 mM MgCl2, 5 mM EGTA, 2 mM EDTA, 0.25 mM DTT in 25 mM MOPS pH 7.2). The reaction was initiated by the addition of 1 mM ATP (15 µl reaction

volume), kept at 30 °C for 0, 20 and 120 min, and stopped by the addition of SDS sample buffer. Proteins were separated by SDS-PAGE and stained with Pro-Q Diamond and SYPRO Ruby. The same experiment was repeated with 400 ng of purified protein and 400 ng LATS2 for 0, 20, 40, 60, and 80 min and processed for LC-MS analysis as described in section 2.2.2.

2.2.3 Molecular biological methods

2.2.3.1 Transient Replicon Assay

At formerly Roche Nutley (Sophie Le Pogam and Sailaja Sankabathula), transient wild type replicon Con1 (expressing firefly luciferase reporter gene) (16) was used to introduce the Ser71Ala mutation in the NS5A region using Quick Change site-directed mutagenesis kit following manufacturer's instructions (Stratagene, La Jolla, CA). The introduced mutation was confirmed by DNA sequencing. Lunet Huh7 cells ($4 \times 10E6$) were transfected with 10 µg of *in vitro* transcribed HCV transient subgenomic replicon RNAs (16) (WT or NS5A mutant). The normalized replication efficiency of the NS5A Ser71Ala mutant was determined as the firefly luciferase signal at 96 hours post-transfection divided by the signal at 4 hours. The replication capacity of the NS5A Ser71Ala mutant replicons was expressed as its normalized replication efficiency compared with that of the WT replicon set at a value of 100%. Average and standard deviation were calculated from four independent experiments.

2.2.3.2 siRNA Knockdown Experiments

All siRNA experiments were performed at Roche Nutley (Sophie Le Pogam and Sailaja Sankabathula) and all siRNAs used in the study were obtained from Integrated DNA Technologies, Inc., Coralville, IA. Transfection of siRNA in the HCV replicon cells was performed using Lipofectamine RNAiMAX (Invitrogen) following the manufacturer's protocol. 5'000 replicon cells, seeded in a 96 well plate 24 hours pre-siRNA transfection, were transfected in duplicate with 1 pmol of siRNA. The level of knockdown for each siRNA used in the study was evaluated by real-time PCR (17). Briefly, total cellular RNA was extracted with RNeasy (Qiagen, Venlo, Netherlands) according to manufacturer's instructions. Reverse transcription was carried out using the Taqman RT reagents (Applied Biosystems, Foster City, CA). To quantitate the level of HCV replicon, cDNA was amplified using Taqman Universal PCR mix (PE Biosystems, Norwalk, CT) with a set of primers and probe complementary to a region of the NPTII gene (fluorogenic probe labeled with FAM (6-carboxyfluorescein), 5′-TCC TGC CGA GAA AGT ATC CAT CAT GGC T-3′; forward primer, 5′-GCT GCT ATT

GGG CGA AGT G-3′; reverse primer, GCC GCC GCA TTG CA; all obtained from Integrated DNA Technologies, Inc.). A second set of primers and probe complementary to the β-actin gene (Applied Biosystems) was used to quantitate the β-actin mRNA level as the endogenous control. For the quantitation of PI4KA, LATS1 and LATS2, sets of primers and probes (Applied Biosystems) were used according to the manufacturer's instructions. Level of knockdown for each target gene was normalized to the one obtained using negative control siRNAs. Replicon renilla luciferase signal was read 72 hours post-siRNA transfection using the Renilla Luciferase Assay System (Promega, Fitchburg, WI, USA) following the manufacturer's protocol. The signal for each gene-specific siRNA experiment was normalized to the one obtained using negative control siRNAs. Average and standard deviation were calculated from 8 to 12 independent experiments.

2.2.3.3 Antisense oligonucleotides synthesis

ASO-A was synthesized and biotinylated at Roche Innovation Center Copenhagen. ASO-B and ASO-C were synthesized and biotinylated (3' biotin TEG, LGC, Biosearch Technologies) by Adrian Schäublin at Roche Innovation Center Basel using a MerMade 12 synthesizer from BioAutomation. ASO-D, MOD ASO-D and biotinylated ASO-D were obtained from Microsynth.

2.2.4 LC-MS/MS analysis and data processing

2.2.4.1 LC-MS/MS Analyses (NS5A and Gpc2)

Dried samples were reconstituted in 2% acetonitrile/5% formic acid and analyzed with a nanoflow ultrahigh pressure liquid chromatography Easy-nLC system (Proxeon, Odense, Denmark) connected to Orbitrap LTQ Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Peptides were concentrated on an AQUA C18 trap (100 μm × 10 mm, Phenomenex, Torrance, CA) before being separated on a ReproSil-Pur C18-AQ (75 μm × 200 mm, 3 μm particle size, 120 Å, Dr. Maisch GmbH, Ammerbuch-Entringen Germany) analytical column using a 50 min gradient of 0–35% acetonitrile (with 0.6% acetic acid) at 250 nl/min. Acquisitions were cycled between full scan at 60,000 resolution (at m/z 400) in the Orbitrap and 10 data-dependent collision-induced dissociation scans in the ion trap. Full-scan MS was recalibrated on the fly using the polycyclodimethylsiloxane at m/z 445.120084 as a lock mass [47]. Ions were selected only with assigned charge states >1+ and then excluded for 30 s.

2.2.4.2 LC-MS/MS Analyses (Htra1,SMN, ASO pull-downs)

Samples were reconstituted in 2% acetonitrile/5% formic acid and analyzed with a nanoflow ultrahigh pressure liquid chromatography Easy-nLC system (Proxeon, Odense, Denmark) connected to Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Samples were concentrated on an Acclaim PepMap C18 trapping column (100 μm × 20 mm, 5 μm particle size) and peptides were separated on an Acclaim PepMap C18 EASY-spray column (75 μm × 500 mm, 2 μm particle size) using the following gradient at 300 nl per min: 7–50% B in 45 min, 50–80% B in 2 min, 80% B for 13 min, corresponding to a total time of 60 min (buffer A: 0.1% formic acid; buffer B: 0.1% formic acid/acetonitrile). Data were on-the-fly recalibrated using ambient air hexacyclodimethylsiloxane at m/z 445.1200238. The ten most intense precursor ions, with charge states between 2 and 6, a minimum intensity of 5E3, were monoisotopically selected for higher-energy collisional dissociation, using a quadrupole isolation of m/z 0.7, automatic gain control target of 1E4, maximum IT of 35 ms, collision energy of 30%, and ion trap readout with rapid scan rate. Only a single charge state per precursor was selected for MS2. Interrogated precursor ions were dynamically excluded for 75 s using a ±10 ppm mass tolerance.

2.2.4.3 Data processing (NS5A and Gpc2)

MS raw files were converted into .dta files using Extract-MSn (version 1.0.0.8) and data searched using Sequest (version 27.0, revision 12, both Thermo Fisher Scientific). A database consisting of the human part of the SwissProt (June 2009, 34,275 entries, including splice variants) was used and was augmented with the five viral nonstructural proteins from the Con1 replicon NS3, NS4A, NS4B, NS5A, and NS5B for NS5A interactome. Searching parameters were trypsin (full) as an enzyme, one missed-cleavage and a mass tolerance of 10 ppm, and 1.0 Da for precursor and fragment ions, respectively. Oxidized methionines (+15.9949 Da), phosphorylated serine, threonine, and tyrosines (+79.9663 Da) were set as differential, while carbamidomethylated cysteines (+57.0215 Da) were set as static modifications. The spectral false discovery rate (specFDR) was restricted to 2.5% or 1.0% by performing a target-decoy search using a concatenated decoy database [48]. Processing and analysis, as described below, was performed with the 2.5% specFDR dataset, and the final protein-binding partners were verified with the 1.0% specFDR search. Data from NS5A *in vitro* phosphorylation experiment were processed using Proteome Discoverer (version 1.2) with the same parameters mentioned above and selected extracted ion chromatograms (EICs) generated with XCalibur's Qual Browser (Thermo Fisher Scientific).

2.2.4.4 Data processing (Htra1,SMN, ASO pull-downs)

MS raw files were imported using Progenesis QI 2.1 (Nonlinear Dynamics; Newcastle, UK) and ions m/z were aligned to compensate for drifts in retention time between runs (maximum charge state set at +5). Peptides were identified by searching with Mascot Server 2.5.1 (Matrix Science, London, UK) together with the UniProt/SwissProt human or mouse protein database using trypsin/P as an enzyme (cleavage at carboxyl side of lysine or arginine also if a proline follows), a maximum of two missed cleavage sites, 10 ppm, and 0.5 Da as the precursor, and fragment ion tolerances, respectively. Carbamidomethylated cysteines (+57.02146 Da) were set as static while oxidized methionines (+15.99492 Da) were set as dynamic modifications. The specFDR was restricted to 1.0% by performing a target-decoy search using a concatenated decoy database.

2.2.4.5 Statistical analysis

Peptide extracted ion counts (EICs) were estimated as measure of peptide quantitation. Data normalization was performed in Progenesis by applying a scalar multiple to each feature abundance measurement with the assumption that most peptide ions do not change in abundance (similar abundance distributions globally).

Normalized log2 scaled EIC was summed to calculate the relative protein abundance. In order to control quality, data were subjected to principal component analysis in Progenesis. In order to detect dose–response relationships, a linear model was then fit using a set of contrasts [49]. The contrasts were used to compare the protein abundance values above and below each concentration point. Then the maximum of the contrasts moderated t-statistics [50] was obtained for each protein. Multiple testing adjusted significance (p-values) was derived by permutation testing. Proteins with lowest adjusted p-values and highest signal reduction were considered as specific binders. Data computations and visualizations were performed with R (http://www.r-project.org/).

3. RESULTS

The aim of this work was to develop a platform to screen endogenous protein-protein interactions (PPIs) in an unbiased fashion. Such an *in vivo* like global proteomics profiling can be achieved by combining affinity purification with mass spectrometry (AP-MS). However, this often results in an extensive list of potential interactors including many non-specific binders. To improve the specificity of the established AP-MS technique, a new approach was assessed in which protein extracts are pre-incubated with increasing concentration of free antibody prior to immunoprecipitation (IP) of the target protein with the same antibody immobilized on agarose beads (Figure 3.1). This proteome-wide platform was named Immuno-Competitive Capture Mass Spectrometry (ICC-MS) and its rationale is based on the hypothesis that the capture of the target protein (the bait) as well as its interacting partners (the preys) should be gradually reduced upon increasing concentrations of the free form of the antibody, offering an innovative way to filter out non-specific binders.

Technically, cellular lysates are incubated with competitive free antibody prior to IP. Proteins bound to immobilized antibodies and/or the matrix are eluted, digested by trypsin and further processed for analysis by nanoflow liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS). After protein identification, MS data are normalized and relative protein abundances are translated from peptide extracted ion chromatograms (EICs). For downstream analysis, the relative abundance is plotted versus the increasing amounts of free target-specific antibody to obtain a concentration-dependent displacement profile. As previously postulated, the displacement of the target protein should be similar to the one from its interacting partners, which are extracted from the entire proteome by applying a robust statistical approach analyzing dose dependent relationships.

ICC-MS specificity as well as its ability to deliver a short list of potential interactors having a high probability to be functionally relevant, have been evaluated in three experimental settings.

In this thesis ICC-MS was first applied to study the interactome of the viral non-structural protein 5A (NS5A), a key regulatory protein of hepatitis C virus (HCV), in a cellular HCV replication system (Results 3.1.1). The approach was further assessed to identify interacting partners of the cell surface proteoglycan protein Glypican-2 (Gpc2) in differentiated neural progenitor cells (Results 3.1.2). Finally, the interactome of HtrA1, a protease described to be

associated with Age-related macular degeneration (AMD), was studied in retinal pigmented epithelial cells (Results 3.1.3). In the three experimental conditions, a comprehensive evaluation was performed to check whether ICC-MS applied in relevant cellular systems can be used to reliably identify interacting partners leading to a better understanding of the biological function of the target proteins.

To expand the field of application of ICC-MS, the approach was in a second stage adapted to study interactions between proteins and oligonucleotides. Thus, the immobilized capture antibody was replaced by a specific oligonucleotide sequence and protein captures were competed by increasing concentrations of free forms of the same oligonucleotide sequence. After a feasibility assessment on well-described RNA-protein complexes (Results section 3.2.1 and 3.2.2), the approach was applied to analyze the particular mode of action of a splicing modifier (Results section 3.2.3) and to explore potential mechanisms behind antisense oligonucleotides (ASOs) toxicity (Results section 3.3).
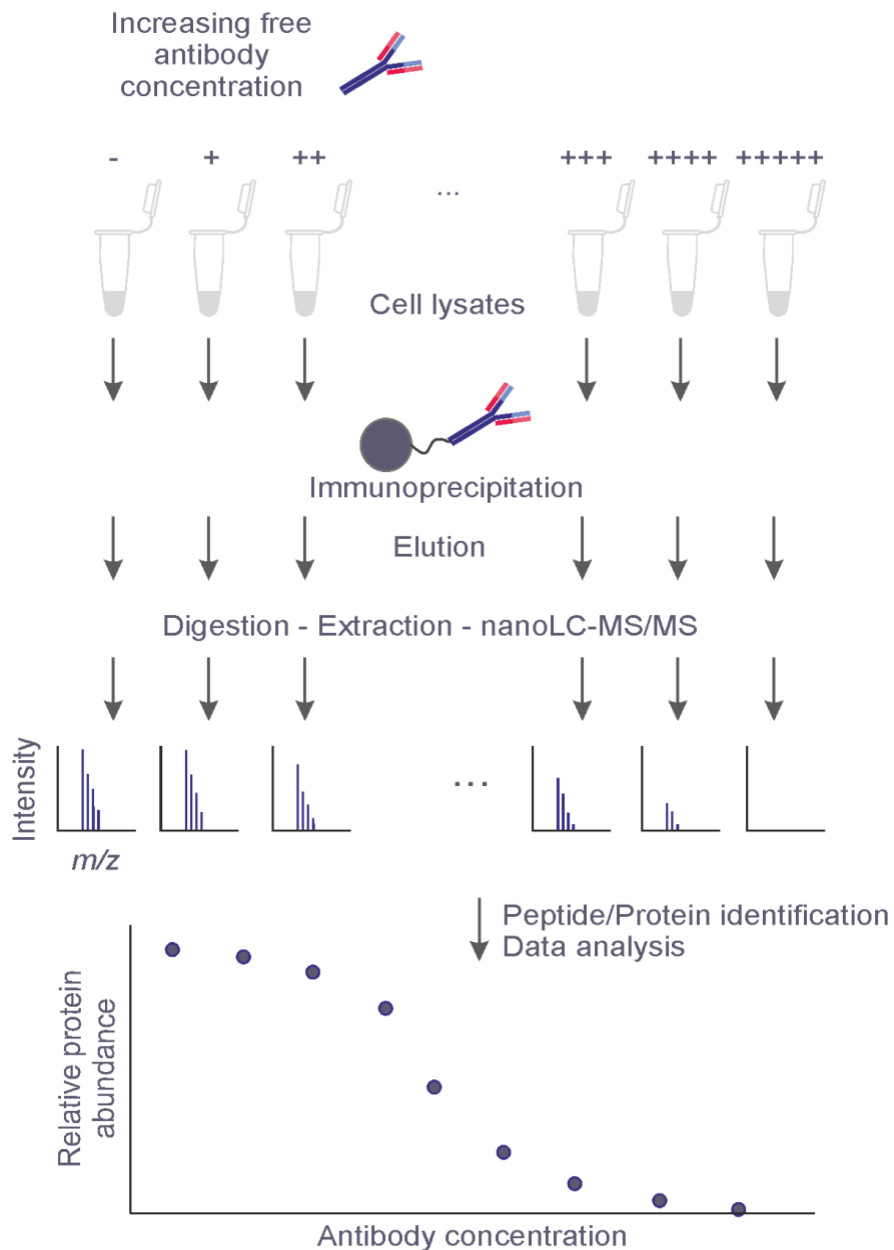
Figure 3.1: Schematic outline of the ICC-MS experimental procedure to identify protein interactors. Protein extracts are first pre-incubated with increasing concentrations of the free form of a protein-specific antibody. Extracts are then individually immunoprecipitated using the same antibody immobilized on agarose beads. The unbound fractions are discarded and beads are washed several times. Bound proteins are eluted from the beads and digested with trypsin. Extracted peptides are analyzed by LC-MS/MS on an Orbitrap mass spectrometer. Peptide extracted ion chromatograms (EICs) are summarized into relative abundance for each corresponding protein quantitation group. Concentration-dependent signal decrease of a specific interactor is finally detected by applying a robust statistical analysis.

3.1. Immuno-Competitive Capture Mass Spectrometry (ICC-MS) to study proteome-wide binding to proteins of interest

3.1.1 Characterization of NS5A protein interactome in a cellular model of hepatitis C virus replication

The first application of ICC-MS was dedicated to test its ability to identify novel and biological relevant interacting partner(s) of the viral NS5A protein.

Worldwide approximately 71 million people are chronically infected with HCV as reported by the World Health Organization [51]. Chronic HCV infection can lead to fibrosis, cirrhosis, and often hepatocellular carcinoma. NS5A is a multi-faceted regulatory protein encoded by HCV which plays a key role in both genome replication and modulation of host response [52].

Direct interactions of NS5A with host proteins, such as phosphatidylinositol 4 kinase III alpha (PI4KA) and 33 kDa vesicle-associated membrane protein-associated protein A (VAPA), have been described and are required to support viral genome replication [53, 54]. Other NS5A-interacting proteins have also been identified using Y2H [55] or classical co-expression and co-IP methods [56], but mostly in a setting where viral genome replication was absent.

Exploring the complete NS5A interactome in a cellular HCV replication model using a MS-based approach should lead to a better characterization of its function.

In this context, ICC-MS was performed in Con1-Huh7, a human hepatocyte-derived cellular model containing the HCV subgenomic replicon, where active viral genome replication occurs [42]. In this experimental system, nonstructural HCV proteins NS3/NS4A, NS4B, NS5A, and NS5B are expressed at physiological levels and all steps of viral RNA replication are recapitulated. The cell model was provided by Junjun Gao (former Hoffmann-La Roche site in Nutley).

A prerequisite, as for any antibody-based approach, was to use an antibody with high selectivity and sufficient capture efficiency. The anti-NS5A H26 antibody with an epitope mapped to NS5A amino-acid sequence 228-278 has been selected for evaluation. Sequence alignment using BLAST did not identify sequence similarities that would lead to potential cross-reactivity and anti-NS5A H26 antibody was then tested to precipitate NS5A from protein extracts of Con1-Huh7 cells (Figure 3.2 A). The lysis buffer was containing a low concentration of non-denaturating detergent (0.5% NP40) to favor proteins solubilization while preserving their

native conformation and protein interactions. NS5A was captured with high efficiency (no signal detected in the unbound fraction) and high specificity (no signal detected after precipitation from the parental Huh7 cell line which does not express NS5A). To provide further support towards anti-NS5A H26 antibody specificity, an immunoblot of Con1-Huh7 lysates where anti-NS5A H26 antibody was used as primary detection antibody was performed (Figure 3.2 B). In Con1-Huh7 lysate, a clear band was detected at the molecular weight of NS5A and cross-reactivity to other proteins can be reasonably excluded.
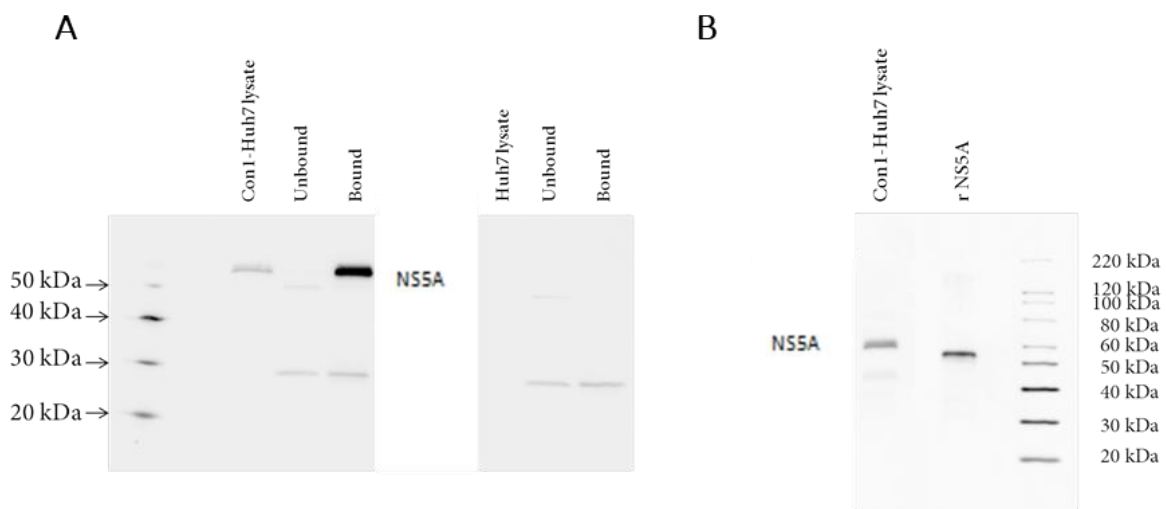
Figure 3.2: Immunoblot-based validation of NS5A-antibody H26 specificity. (A) Con1-Huh7 and Huh7 cell lysate (20 µg) were probed for NS5A by immunoblot using anti-NS5A clone 7D4. Unbound and bound fractions from IP experiments performed using H26 antibody in both Con1-Huh7 and Huh7 parental cell line extracts were probed for NS5A as well. (B) Detection of NS5A in Con1-Huh7 (10 µg) using H26 antibody. Recombinant NS5A (rNS5A) was loaded as a reference (NS5AΔ32 BK strain residues 33-447).

In a next step the feasibility to introduce a competition between the free form and the immobilized form of NS5A antibody H26 was assessed by running a pilot experiment with five different concentrations of free antibody (Figure 3.3). Data obtained for the NS5A immunoblot confirmed that the free antibody prevents binding of the target protein to the immobilized capture antibody. Already with 3 µg/ml, the signal of the immunoprecipitated NS5A protein was reduced and the largest drop was observed when incubating with 9 µg/ml. This signal was then barely detectable after a pre-incubation of the lysate with 100 µg/ml of NS5A-specific antibody.
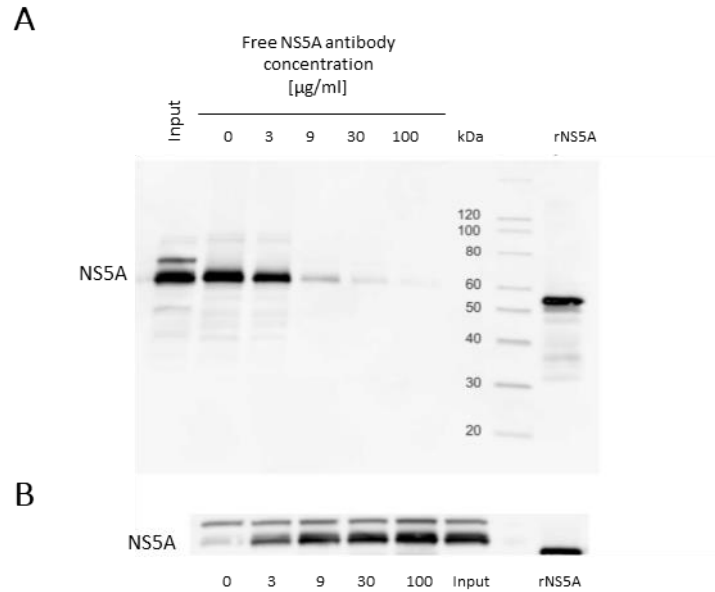
Figure 3.3: NS5A immunoblot from bound and unbound fractions from immuno-competitive precipitations in Con1-Huh7 lysates. 230 µg of cell extracts were pre-incubated with 0, 3, 9, 30 or 100 µg/ml of free anti-NS5A H26 antibody before immunoprecipitation with anti-NS5A H26 antibody immobilized on agarose beads. (A) 8.5 µl from 100 µl total elution (B) 20 µg equivalent of Flow-through, were probed for NS5A in an immunoblot. Input: Con1-Huh7 lysate, rNS5A: Recombinant NS5A (NS5AΔ32 BK strain residues 33-447).

To generate a higher-resolved displacement profile in order to increase the probability identifying a limited number of highly specific interactors, the number of concentrations of free antibody was increased and samples were analyzed by LC-MS/MS. Thus twelve concentrations of free NS5A antibody were selected. According to the results of the pilot with five concentrations, additional concentrations were selected below 3 µg/ml and between 3 µg/ml and 9 µg/ml. A concentration of 0.6 µg/ml was selected as the first one and then an even interval of 1.7 was applied (Figure 3.4). Bounds fractions were separated on SDS-PAGE (Figure 3.4 A), which was cut into four gel slices followed by in-gel digestion and further processed for protein identification by LC-MS/MS (2 injections, i.e. a total of 96 LC-MS/MS measurements). After data quality control of identified peptide peaks, extracted ion counts (EIC, log2 scaled) were normalized and summarized to relative protein abundance. A linear model was fit using a set of contrasts comparing the protein abundance values above and below each concentration point [49] to extract proteins displaced with increasing concentration of free anti-NS5A antibody. The maximum of the contrasts moderated t-statistics was obtained for

each protein [50] and permutation testing was applied to generate multiple testing adjusted significance (p-values). The MS output obtained when Jens Lamerz (Roche Innovation Center Basel) applied the above mentioned statistical analysis extracting dose-response relationships perfectly matched the corresponding immunoblot (Figure 3.4 B and C).

In addition to the correlation between the immunoblot and the LC-MS/MS output, the target protein NS5A was ranked as the best hit (adjusted p value of 6.7 x 10-6) and low technical variance for replicate analyses was observed. Altogether, it validated the use of label-free quantitative MS as a reliable readout for ICC-MS experiments.
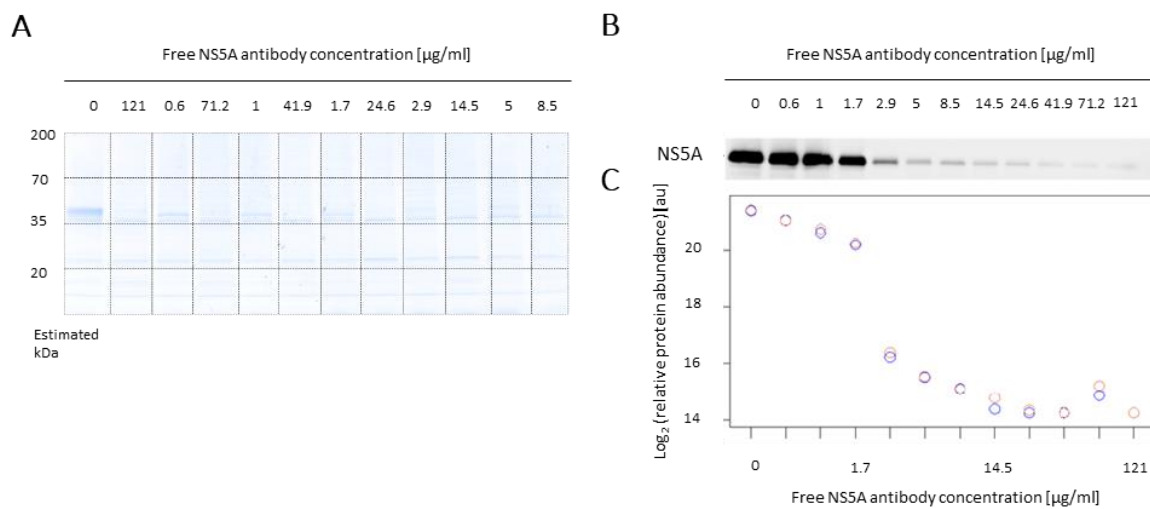
Figure 3.4: ICC-MS of NS5A in Con1-Huh7 cell extracts. NS5A was immunoprecipitated from Con1-Huh7 cell lysates (300 µg total protein) which were pre-incubated with 0, 0.6, 1, 1.7, 2.9, 5, 8.5, 14.5, 24.6, 41.9, 71.2 or 121 µg/mL of free NS5A antibody. (A) Bound fractions (25 µL from 60 µL elution volume) were separated by SDS-PAGE migrated half its length and proteins were stained with Coomassie blue. Four bands (from 20 to 150 kDa) were cut and in-gel trypsin digested for protein identification by LC-MS/MS. (B) NS5A protein in bound fractions was probed by Western Blot. (C) NS5A protein in bound fractions was identified and quantified by LC-MS/MS. Displacement curves obtained after statistical analysis performed by Jens Lamerz, Roche Innovation Center Basel (y axis: relative abundance; x axis: free NS5A antibody concentrations; replicates: red and blue).

In total 769 proteins were identified and quantified in the NS5A-specific antibody bound fraction [57]. Among those proteins, only ten were highlighted as having a significant displacement upon increasing free NS5A-specific antibody concentrations (with adjusted p

values below 5%): NP1L4, PKHG2, FBW1B, NP1L1, PI4KA, UBP19, VAPA, LATS1, LATS2 and PGAM5 (Figure 3.5 and Table 3.1).
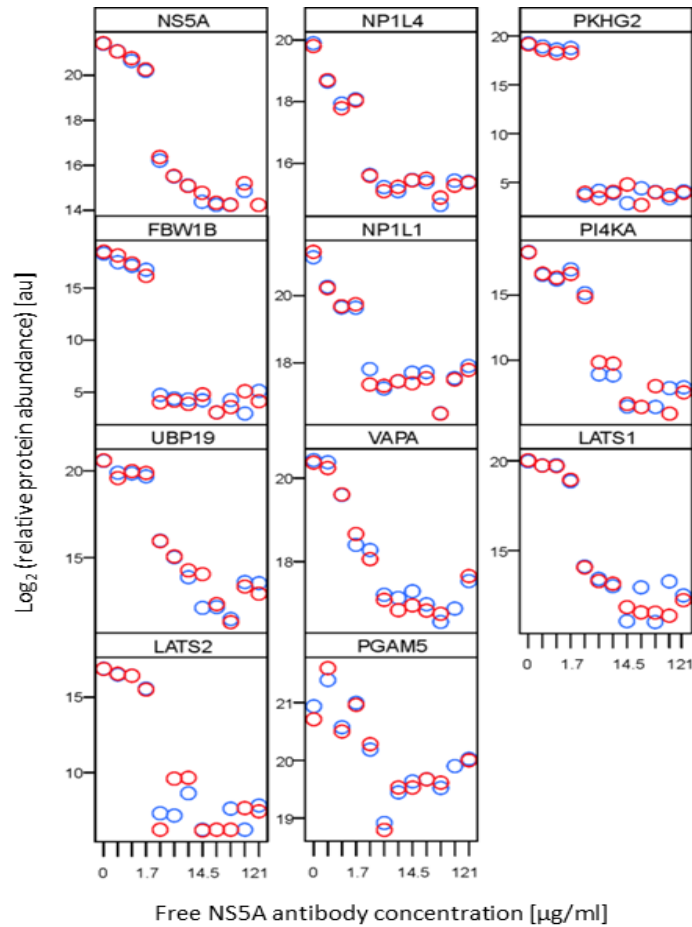
Figure 3.5: Statistically significant NS5A interactors identified by ICC-MS of NS5A in Con1-Huh7 cell extracts. Candidates are ranked by increasing adjusted p value (replicates: red and blue).

Of the obtained list of interactors, few candidates were already described in the literature as NS5A host interacting proteins. Notably, PI4KA was previously described as being required for optimal HCV replication and as a direct interactor of NS5A leading to its activation [53]. VAPA was described to play a role in the formation of the HCV replication complex and to co-localize with NS5A in membrane lipid rafts [54]. NP1L1 and UBP19 were also already reported as potential interactors [58, 59]. The identification of these previously described binding partners indicate that ICC-MS is suitable to identify specific binders while efficiently filtering out non-specific ones.

Other candidates were much less convincing such as PKHG2, FBW1B and PGAM5 whose identifications were based on one peptide only (the latter having in addition a limited signal reduction below 50%).

SwissProt entry	Gene name	pAdj	Signal reduction [%]	Total peptides	Unique peptides
NS5A_POLG_HCVCO		6.7E-06	99	41	13
NP1L4_HUMAN	NAP1L4	6.7E-06	95	9	6
PKHG2_HUMAN	PLEKHG2	6.7E-06	100	1	1
FBW1B_HUMAN	FBXW11	3.2E-04	100	1	1
NP1L1_HUMAN	NAP1L1	8.0E-04	90	10	5
PI4KA_HUMAN	PI4KA	3.1E-03	100	10	7
UBP19_HUMAN	USP19	3.4E-03	99	18	11
VAPA_HUMAN	VAPA	5.6E-03	86	7	2
LATS1_HUMAN	LATS1	1.8E-02	99	10	7
LATS2_HUMAN	LATS2	2.0E-02	100	4	4
PGAM5_HUMAN	PGAM5	3.5E-02	43	1	1

Table 3.1: NS5A interacting candidates identified after statistical analysis of MS data. pAdj: p value adjusted by the Westfall-Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free antibody; Total peptides: number of signals including different charge states; Unique peptides: number of peptides assigned to a protein quantitation group. Candidates ranked by increasing adjusted p-values.

Interestingly, two candidates were never reported in the literature; the two members of the nuclear Dbf2-related (NDR) family of AGC kinases, [60, 61], LATS1 and LATS2.

NS5A phosphorylation status was previously described to play an important role in HCV RNA replication without being completely understood [62-64]. NS5A has been indeed reported to be phosphorylated at multiple sites but the responsible upstream host kinase(s) remain largely unidentified, except for casein kinase I and II [65, 66].

LATS Ser/Thr kinases are known to phosphorylate a specific substrate consensus sequence His-X-Arg/His/Lys-X-X-Ser/Thr. This motif could be mapped in domain 1 of NS5A (His65 to Ser/Thr71) and was found to be highly conserved among all HCV viral genotypes (GT 1a, 1b, 2a, 2b, 3a, 4a, 5a and 6a; Table 3.2) with Ser/Thr at position 71 100% conserved.

	Observed residues in genotype (counts/percentage)						
Position	**GT 1a**	**GT 1b**	**GT 2a, b**	**GT 3a**	**GT 4a**	**GT 5**	**GT 6**
66	His 330/100	His 382/99.74 Arg 1/0.26	His 24/61.54 Asn 14/35.90	His 34/100	His 15/100	His 2/100	His 15/100
67	Val 330/100	Val 381/99.48 Phe 2/0.52	Val 39/100	Val 34/100	Ile 14/93.33 Val 1/6.67	Val 2/100	Val 15/100
68	Lys 324/98.18 Arg 6/1.82	Lys 381/99.48 Thr 1/0.26 Arg 1/0.26	Arg 39/100	Lys 34/100	Lys 14/93.33	Lys 2/100	Lys 15/100
69	Asn 326/98.79 Thr 3/0.91	Asn 380/99.22 Thr 3/0.78	Met 19/48.72 Leu 20/51.28	Asn 34/100	Asn 15/100	Asn 2/100	Asn 15/100
70	Gly 330/100	Gly 383/100	Gly 39/100	Gly 34/100	Gly 15/100	Gly 2/100	Gly 15/100
71	Thr 320/96.97 Ser 9/2.73	Ser 383/100	Thr 24/61.54 Ser 15/39.46	Ser 34/100	Ser 15/100	Thr 2/100	Thr 13/86.67 Ser 2/13.34

Table 3.2: His-X-Arg/His/Lys-X-X-Ser/Thr motif in NS5A across all HCV genotypes. Amino-acid alignment of 818 HCV clinical isolates (sequences collected from HCVdb at Los Alamos National Laboratory).

The displacement of the capture of LATS1 and LATS2 over increasing concentrations of NS5A antibody have been also visualized on a western blot (Figure 3.6 A) and those co-captures were confirmed to take place only in the cell line expressing NS5A (Figure 3.6 B). To confirm LATS kinases as novel NS5A-interacting partners, additional co-immunoprecipitation (coIP) experiments were performed. Thus Con1-Huh7 as well as Huh7 extracts were immunoprecipitated using LATS1 and LATS2 antibodies. LATS1 and LATS2 were detected in the bound fractions of respectively LATS1 and LATS2 IP and NS5A was co-immunoprecipitated in both experiments increasing the confidence towards LATS kinases as

being true NS5A-interactors (Figure 3.6 C). To check whether NS5A is a bona fide substrate of LATS kinases, an *in vitro* phosphorylation assay was performed. Catalytically active recombinant LATS2 (500 ng) was premixed with purified recombinant NS5A derived from *E. coli* (2 µg) and reaction was initiated by adding 1mM ATP. The reaction was stopped after 0, 20 and 120 min. Phosphoproteins were then stained on SDS-PAGE with Pro-QDiamond and revealed that LATS2 kinases was inducing phosphorylation of NS5A already after 20 minutes of incubation (Figure 3.7 B; A: total protein staining using SYPRO).
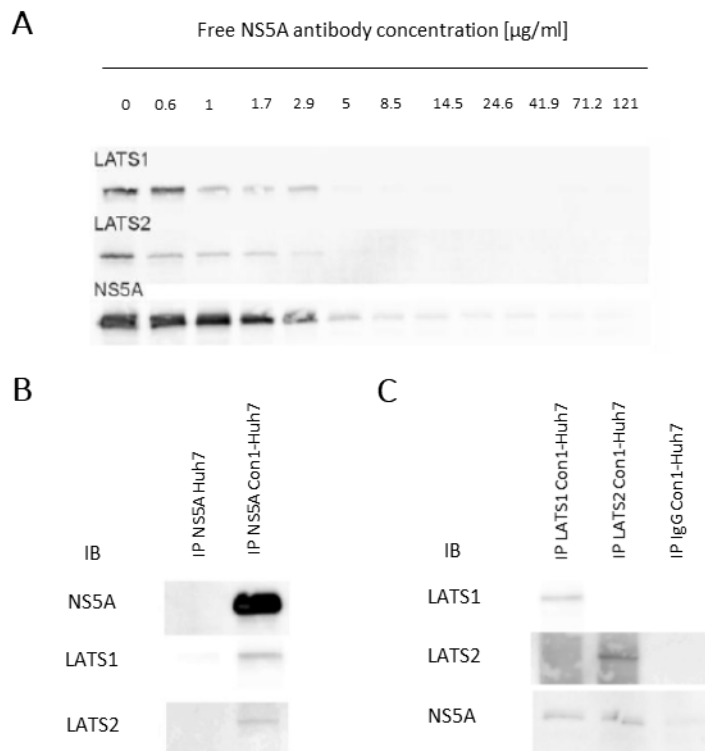
Figure 3.6: Binding confirmation of LATS1 and LATS2 to NS5A. (A) NS5A ICC-MS in Con1-Huh7. 380 µg of cell lysates were incubated with 0, 0.6, 1, 1.7, 2.9, 5, 8.5, 14.5, 24.6, 41.9, 71.2 and 121 µg/ml of free NS5A antibody prior to NS5A immunoprecipitation. 25 of 60 µl total elution were probed for LATS1 and LATS2. (B) NS5A IP in Con1-Huh7 and Huh7 parental cell line. NS5A, LATS1 and LATS2 immunoblots (IB). (C) LATS1/2 IP in Con1-Huh7 cell line (330 µg protein extracts, 5.5 of 60 µl total elution loaded) and NS5A IB (18 of 60 µl total elution loaded).
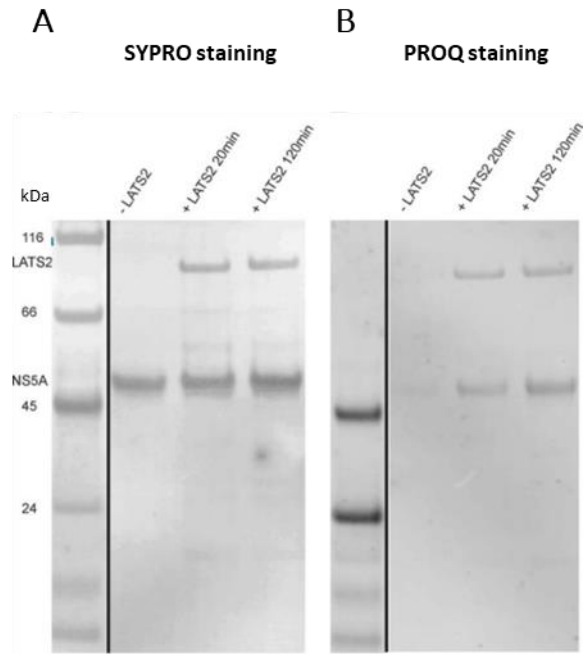
Figure 3.7: LATS2 is phosphorylating NS5A *in vitro*. NS5A purified recombinant protein (residues 33-447) was mixed with recombinant LATS2 for 20 or 120min. Reaction was initiated by the addition of 1mM ATP. A negative control without kinase was included as well. (A) Total protein staining with SYPRO. (B) Phosphoprotein staining with Pro-Q Diamond.

To identify the site of phosphorylation, the same experiment was repeated with recombinant NS5A (400 ng) incubated with LATS2 (400 ng) in presence of ATP for 0, 20, 40, 60 and 80 min. LC-MS/MS analysis of reaction samples unambiguously localized the site of NS5A phosphorylation on Ser71, i.e. on the Serine of the consensus sequence. Semi-quantitative assessment of the phosphorylation stoichiometry was performed by comparing the peak areas of the unphosphorylated and the phosphorylated peptide (Figure 3.8) and revealed that after 80 min of reaction, the phosphorylated peptide [69]NGSMRIVGPK[78] was almost nine times more abundant than its unmodified version.
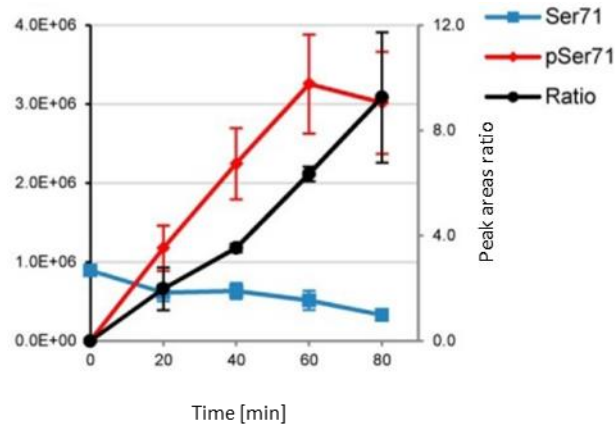
Figure 3.8: NS5A Ser71 phosphorylation by LATS2. Relative quantitation of phosphorylated and unphosphorylated NGSMRIVGPK peptide by LC-MS/MS. Ratios of EICs represented. Error bars, S.E. 2 replicates.

To further assess the biological relevance of these findings, the impact of a point mutation preventing phosphorylation at Ser71 on Con1-Huh7 replication capacity was explored. At formerly Roche Nutley, Sophie Le Pogam and Sailaja Sankabathula introduced the Ser71Ala mutation by site-directed mutagenesis in the NS5A region of the wild type (WT) transient genotype 1b Con1 replicon, which expresses the firefly luciferase reporter gene. Replication efficiency was determined by the luciferase signal at 96 hours post-transfection divided by the signal at 4 hours and compared to the WT. Infecting Huh7 cells with Ser71Ala mutant replicon resulted in a clear impairment of viral replication capacity (only 14% compared with the WT replicon replication, Figure 3.9) suggesting that phosphorylation of Ser71 is essential for optimal viral genome replication.
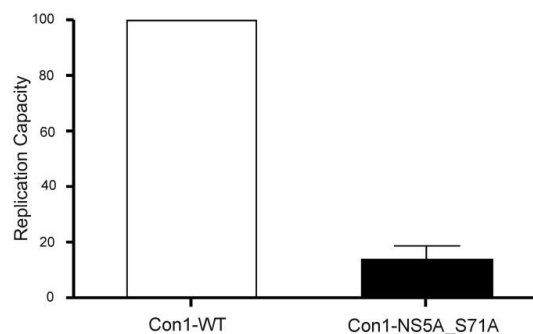
Figure 3.9: Replication capacity of Con1-NS5A-Ser71Ala. Normalized replication efficiency of mutant transient replicon compared with reference strain Con1-WT (set at 100%) (n=4).

In the same lab, the effect of siRNAs targeting LATS1 and LATS2 kinases were further evaluated by measuring the HCV replicon renilla luciferase signal (Figure 3.10). An inhibition of 63% and 38% respectively was detectable suggesting a relevant role for LATS kinases in HCV life cycle probably by modulating NS5A phosphorylation.
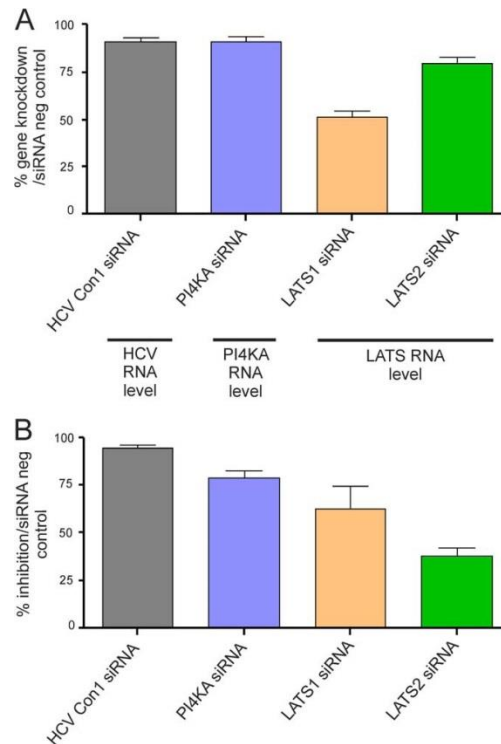

Figure 3.10: LATS1 and LATS2 silencing. (A) Gene knockdown after siRNA transfection (as positive controls, Con1 HCV replicon sequence and PI4KA gene, n=5 to 9). (B) HCV replicon renilla luciferase signal inhibition normalized to negative control siRNA (n=8 to 12).

The application of ICC-MS approach to identify novel interacting partners of the viral NS5A protein resulted in only a handful of potential interactors. The study revealed an implication of Ser71 phosphorylation in viral replication and that LATS kinases are probably important regulators of NS5A function.

3.1.2 Characterization of Glypican-2 protein interactome in a human neural stem cell model

Next, the ICC-MS approach was applied to identify interacting partners of Glypican-2 (Gpc2), a cell surface proteoglycan recently discovered as a potential new marker of neurogenesis by Ducret *et al.* using an unbiased proteomics screen in a stem cell model of human brain development [43]. Glypicans (Gpcs) have been indeed described to regulate the activity of several growth factors involved in morphogenesis [67], nevertheless a functional role of Gpcs in adult neurogenesis has never been reported.

Gpcs as part of Heparan sulfate proteoglycans (HSPGs) family are bound to the cell via a glycosylphosphatidylinositol (GPI) linkage (Figure 3.11). Six members exist in vertebrates, Gpc1-Gpc6 [68], with Gpc2 (or cerebroglycan) being exclusively expressed in central nervous system [69]. Heparan sulfate chains are highly negatively charged polysaccharides prone to interact with many proteins with a net positive charge but with little specificity. Nevertheless, extracellular matrix proteins such as chemokines, growth factors as well as their receptors were shown to have preferred HS binding structures indicating a higher specificity [70]. Heparan sulfate proteoglycans having a core protein with heparan sulfate chains and a kind of glycosaminoglycan, have been shown to be involved in signaling pathway regulation (Wnt Hedgehog, transforming growth factor-ß, fibroblast growth factor pathways) and synaptic organization [70-74] thus probably playing a crucial role in CNS development.
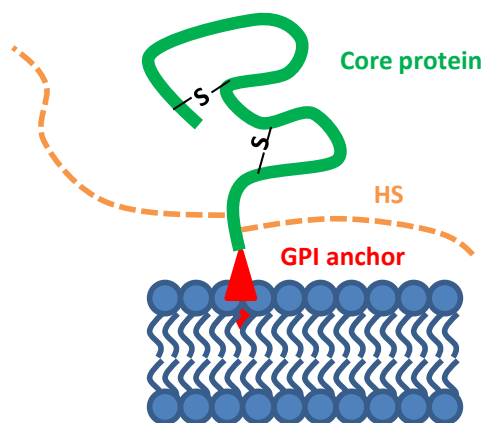

Figure 3.11: Glypican structure. N-terminal domain is globular and stabilized by disulphide bonds. The core protein is linked to the cell membrane via a glycolipid anchor (glycosylphosphatidylinositol = GPI). Heparan sulphate chains are linear polysaccharides with alternating glucosamine and uronic acid residues. HS are carried at the N-terminal and varies in their modifications state, such as deacetylation or sulfation. Upon cleavage at GPI anchor, an HS proteoglycan is shed from the cell surface.

Identifying proteins interacting with Gpc2 in a human neural stem cell model would lead to a better understanding of its potential link to adult neurogenesis. Consequently, ICC-MS was applied to filter out non-specific binders which in turn may facilitate the selection and the validation of true interactors.

Five anti-Gpc2 antibodies generated in the lab of Jean-Philippe Carralot (Roche Innovation Center Basel) were tested for Gpc2 capture in H4 neuroglioma cells transfected with His-tagged human Gpc2 (Figure 3.12, cells provided by Sebastian Lugert, Roche Innovation Center Basel). Clone 2/34 did not show a high yield of capture and with clone 2/38 additional bands were detected indicating a lower specificity towards Gpc2. Among the best performing antibodies, clone 2/43 was further selected to capture Gpc2 in a cellular model for neurogenesis. Neural progenitor cells (NPCs) were cultured by Sebastian Lugert under differentiating conditions for 21 days (d21NPCs). In the starting material (d21 NPC protein extract) and the bound fraction, the core protein was detected at a molecular weight of ~ 60kDa (Figure 3.13.) as well as a smear observed at higher molecular masses which might correspond to different sugar modification patterns. An additional band between 30 and 40 kDa was also detected which could correspond to a cleaved version of Gpc2. Indeed, proteolytic cleavages in the core or at GPI linkage are known to generate different Gpc forms [73]. In addition to the large diversity in Gpc structures, cleaved/uncleaved Gpc ratio among Gpc members and tissues are also varying [67].

A fraction of the bound fraction was separated by SDS-PAGE and further processed for LC-MS/MS protein identification. As expected, Gpc2 was identified in the different bands (Table 3.3) and enriched in band 3 to 6 with a C terminal peptide missing in the low molecular weight band accordingly to the assumption of a cleaved form.
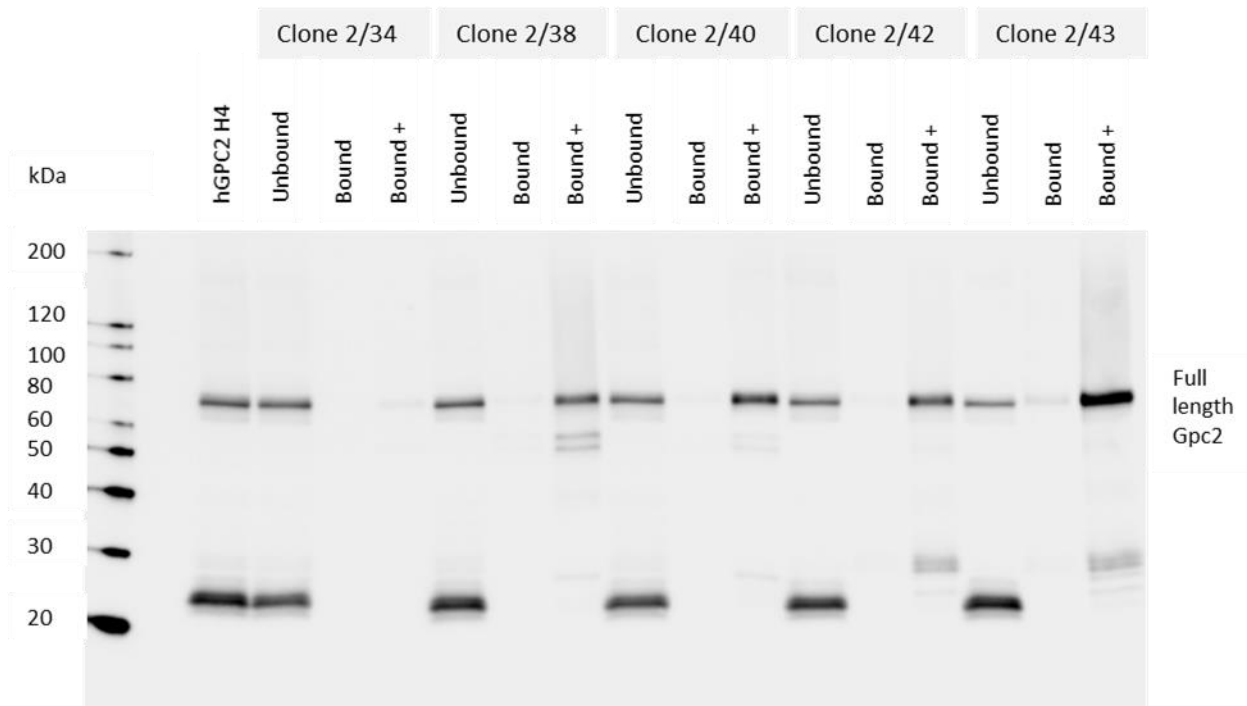
Figure 3.12: Gpc2 immuno-capture in human Gpc2-transfected H4 cell extracts. Gpc2 was immunoprecipitated from Gpc2-transfected H4 protein extracts (220 µg) using anti-Gpc2 antibody clone 2/34, 2/38, 2/40, 2/42 or 2/43 immobilized on agarose resin (1 hour at 4°C). Western blot was probed with anti-penta His (619) antibody generated by Jean-Philippe Carralot. Loading: hGPC2 H4 extract and unbound: 10µg; Bound: 2.7 µl from 60 µl elution; Bound +: 18.9 µl from 60 µl elution.
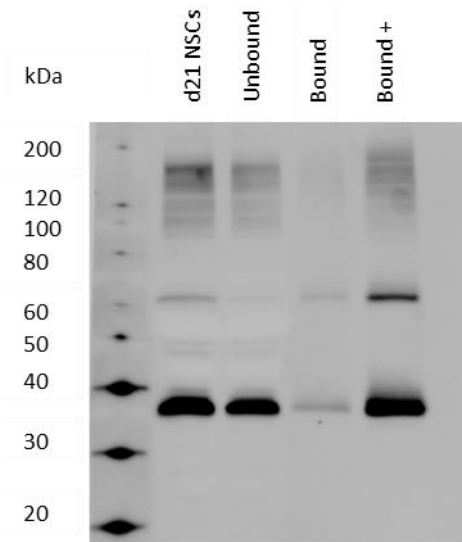

Figure 3.13. Gpc2 immuno-capture in d21 NPC. Gpc2 was immunoprecipitated from d21NPC protein extracts (730 µg) using anti-Gpc2 antibody clone 2/43 immobilized on agarose resin (1 hour at 4°C). Western blot revealed with clone 2/40. Loading: d21 NSCs and unbound: 50 µg; Bound: 4 µl from 60 µl elution; Bound +: 16 µl from 60 µl elution.

| Band	Number of unique peptides	Coverage (%)
7	0	/
6	20	40
5	20	37
4	22	46
3	15	31
2	0	/
1	0	/

Table 3.3: Gpc2 LC-MS/MS identification in d21NPC Gpc2 immunoprecipitates. Eluate from Gpc2 immunoprecipitation in d21NPCs (18 µL) was separated by SDS-PAGE and stained by Coomassie blue. After in gel-digestion with trypsin, samples were processed for LC-MS/MS identification with a nanoflow Easy-nLC system connected to an LTQ-Orbitrap Velos mass spectrometer. Gpc2 was identified with 15, 22 and 20 unique peptides respectively in SDS-PAGE gel bands 3, 4 and 5/6.

As previously mentioned, antibody specificity is a crucial prerequisite for the success of ICC-MS experiments. In order to confirm the specificity of clone 2/43, a protein extract from d21 NSC was probed using 2/43 as primary antibody (Figure 3.14). The clone 2/43 did not perform as well as 2/40 in immunoblot. No clear bands were detected but a smear was observed around 60 kDa and at higher molecular masses. Even if it is corresponding to the size range of the core protein and of the different sugar modification patterns previously described, an unambiguous Gpc2 detection by clone 2/43 cannot be concluded from this experiment. This could be due to the fact, that 2/43 does not recognize a linear epitope.
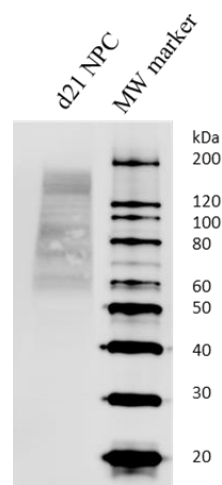
Figure 3.14: d21NPC extract probed with anti-Gpc2 clone 2/43. Anti-Gpc2 clone 2/43 was used to probe d21NPC protein extract (50µg).

Nevertheless, the superiority of clone 2/43 in terms of Gpc2-capture efficiency (Figure 3.12) was considered as the most important factor to use this antibody in an ICC-MS approach. A pilot experiment with seven different concentrations of competitor confirmed that a competition between free and immobilized anti-Gpc2 clone 2/43 was indeed happening in d21 NSC protein extracts (Figure 3.15). Notably, a concentration of 0.5 µg/ml of free Gpc2 antibody did not have an impact on Gpc2 capture. The largest drop was occurring with 5 µg/ml and a close to total displacement was achieved with 10 µg/ml.
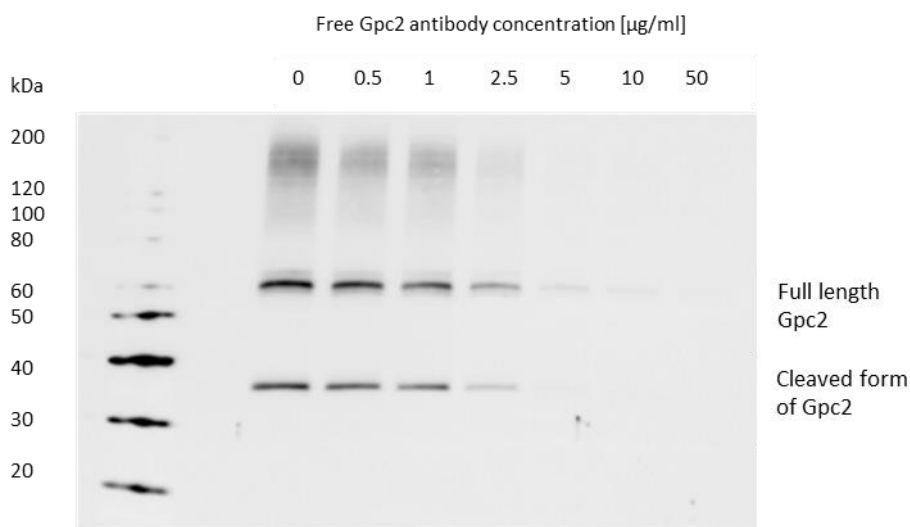
Figure 3.15: Gpc2 competitive immuno-capture in d21NPCs. Gpc2 capture from d21NPCs (570 µg protein extracts per conditions) pre-incubated with 0, 0.5, 1, 2.5, 5, 10 and 50 µg/ml of free Gpc2 2/43 antibody was probed by Western blot with anti GPC clone 2/40.

Based on these observations, the ICC-MS experimental design was slightly modified compared to 3.1.2., by decreasing the number of free antibody competitor concentrations to five and by increasing the number of replicates to three. Again the Western Blot image was matching the MS readout (Figure 3.16) and 5 different concentrations turned out to be sufficient to capture the displacement.
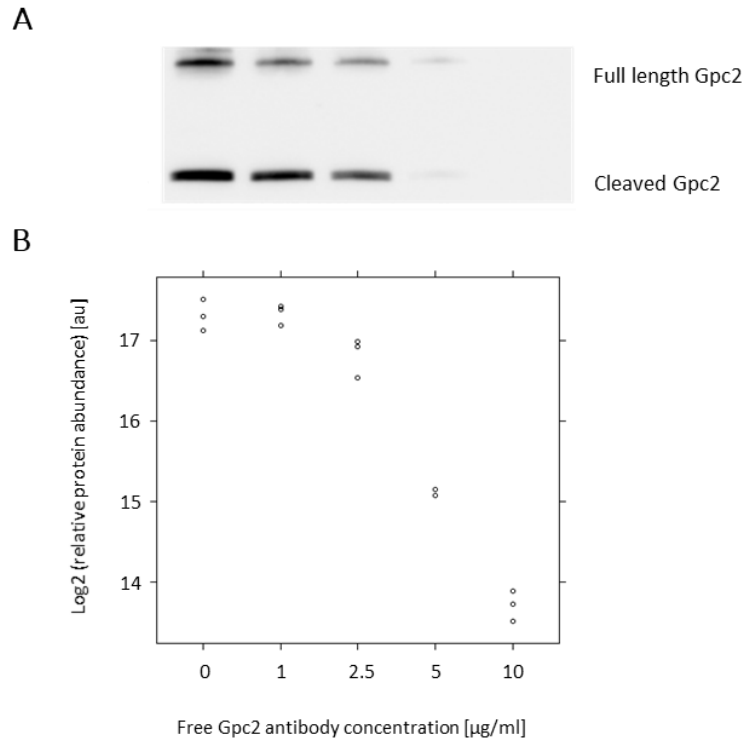
Figure 3.16: ICC-MS of Gpc2 in d21NPC extracts. Gpc2 was immunoprecipitated from d21NPC lysates (855 µg protein extracts per conditions) pre-incubated with 0, 1, 2.5, 5 and 10 µg/ml of free Gpc2 antibody. (A) Gpc2 protein in bound fractions was probed by Western Blot. (B) Gpc2 protein in bound fractions was identified and quantified by LC-MS/MS (y axis: relative abundance; x axis: free Gpc2 antibody concentrations; each black dot is a replicates; displacement curve from band corresponding to core protein).

In total 512 proteins were identified and quantified in the Gpc2-specific antibody bound fraction. MS signals statistical analysis revealed fifteen proteins significantly displaced in addition to Gpc2 (pAdj below 1% and signal reduction above 50%, Table 3.4).

SwissProt entry	pAdj	Signal reduction (%)	Peptides
GPC2_HUMAN	0.000681533	92	19
ECHB_HUMAN	0.000582401	90	22
ECHA_HUMAN	0.000686401	90	46
FGF2_HUMAN	0.000898094	89	1
GSLG1_HUMAN	0.000970668	79	5
SQSTM_HUMAN	0.000593186	79	2
ACOT8_HUMAN	0.000585952	77	3
HIC2_HUMAN	0.000706589	74	1
LPPRC_HUMAN	0.000990681	74	5
SGPL1_HUMAN	0.000589547	68	4
TF3C1_HUMAN	0.001001001	60	1
EIF3A_HUMAN	0.000980572	54	13
EIF3D_HUAMN	0.000696349	53	7
SRP14_HUMAN	0.000906567	53	2
EIF3C_HUMAN	0.000691339	52	7
EIF3L_HUMAN	0.000657592	51	5

Table 3.4: Gpc2 interacting candidates identified after statistical analysis of MS data. pAdj: p value adjusted by the Westfall-Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free antibody; peptides: number of peptides assigned to a protein quantitation group. Candidates were ranked by decreasing signal reduction.

Identified Gpc2 interacting candidates are connected to different biological processes: lipid metabolism (ECHB, ECHA, SGPL1), translation (EIF3A, EIF3D, EIF3C, EIF3L), transcription (TF3C1, LPPRC, HIC2) and differentiation (FGF2, SQSTM). FGF2 was the only heparan binding growth factor from the list. Its displacement - although highly significant - was based on one peptide only, probably reflecting either the fact that Gpc2 is acting as a low-affinity binding protein towards extracellular signaling molecules and/or that FGF2 is a small and a low abundant protein in NPCs extracts. Interestingly, a receptor in the FGF signaling pathway, GSLG1, was also identified (Table 3.4 and Figure 3.17 A).

To confirm a direct binding of FGF2 to Gpc2, a reverse pull-down using FGF2 as bait was performed. Differentiated NPCs were immunoprecipitated using a FGF2 antibody and the bound fractions were probed for Gpc2 by immunoblot. Gpc2 was identified in the bound fraction of FGF2 IP (Figure 3.17) providing further support towards a true interaction between FGF2 and Gpc2.
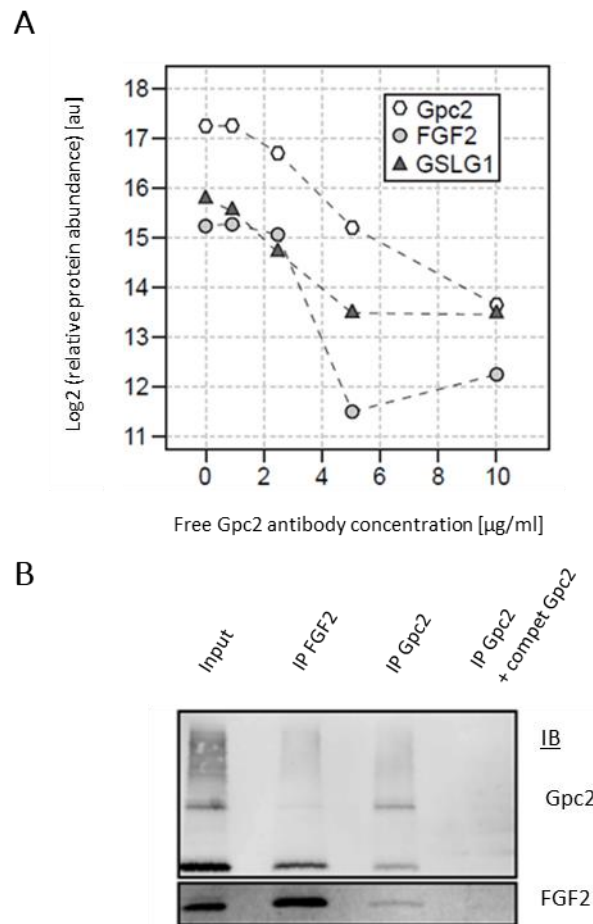
Figure 3.17: Glypican-2 is specifically interacting with FGF2 and GSLG1 in d21NPC extracts.
(A) FGF2 and GSLG1 showed a statistically significant concentration-dependent decrease in MS signal.
(B) Gpc2 and FGF2 were co-captured in 21 days differentiated NSCs. Captures were displaced by pre-incubation with free Gpc2 antibody (IP Gpc2 + compet Gpc2, 10 µg/ml).

Notably, FGF2 has been previously described to play an important role in adult NPC proliferation [75-78]. It is also known from the literature that HSPGs are having either stimulatory or inhibitory effects on FGF2 depending on cell type. For example, an increased proliferative effect towards FGF was observed in keratinocytes [79] and an anti-proliferative effect has been described in morphogenesis [67]. It was also mentioned that an opposite activity can be observed depending on the type of growth factor. Thus FGF2 was an attractive candidate for additional follow up experiments.

To assess the effect of Gpc2 on FGF2, hNPCs were co-cultured in Lugert's lab in the presence of FGF2 in combination with increasing amounts of recombinant Gpc2 (Figure 3.18). The pro-

proliferative effect of FGF2 turned out to be inhibited by the addition of Gpc2 in a concentration-dependent manner. Gpc2 could be potentially sequestering FGF2 to modulate the transition from proliferation to differentiation as described for Gpc1 in skeletal muscle where it prevent FGF2 interaction with transducing receptors [80].
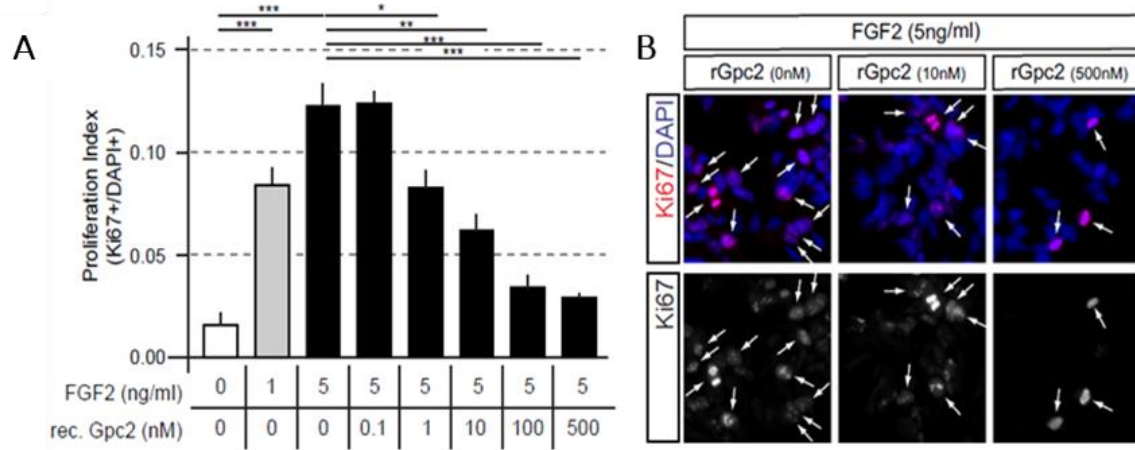
Figure 3.18: GPC2 inhibits the proliferative effect of FGF2 which in return leads to differentiation Endogenous application of GPC2 inhibits FGF2 induced neural stem cell mitogenic activity.

Gpc2 was found to be abundant in human cerebrospinal fluid (CSF) and markedly decreased in a longitudinal cohort of aging CSF [43]. Collectively, these results are strongly suggesting that Gpc2 could act as a niche factor involved in regulation of adult neurogenesis offering the potential to assess in a non-invasive manner the degree of neurogenesis in human physiology and disease.

More than five hundreds proteins were initially identified in Gpc2 immunoprecipitates from dNPCs. By applying ICC-MS approach, only fifteen proteins were extracted from the background of non-specific binders. With such a limited number of candidates, FGF2 was rapidly identified as an attractive target for further follow-up experiments. This is another illustration of how ICC-MS led to successful identification of biologically relevant protein-protein interactions.

3.1.3 Characterization of the HtrA1 protein interactome in a polarized retinal pigment epithelial (RPE) disease model

HtrA1, part of the high-temperature requirement A (HtrA) family of serine proteases, has been reported to play an important role among other pathologies in the pathogenesis of Age-related macular degeneration (AMD), the primary cause of irreversible blindness in aging population [81]. HtrA1 is composed of four domains (Figure 3.19) including a highly conserved trypsin-like protease domain. HtrA1 is referred as a secreted protease having a signal peptide for its secretion but an intracellular localization has been reported as well, both in the cytoplasm and the nucleus [82].

HtrA1 overexpression has been shown to lead to changes in the Bruch's membrane (BM) in a transgenic mouse model by cleavage of extracellular matrix components (ECM) [83] and furthermore, elevated levels of HtrA1 have been reported in the retinal pigment epithelial (RPE) cells of patients genotyped for the AMD risk variant [81]. Still the correlation between HtrA1 levels and how RPE functions are affected is poorly understood and it is still unknown whether HtrA1 has an intracellular activity related to AMD.
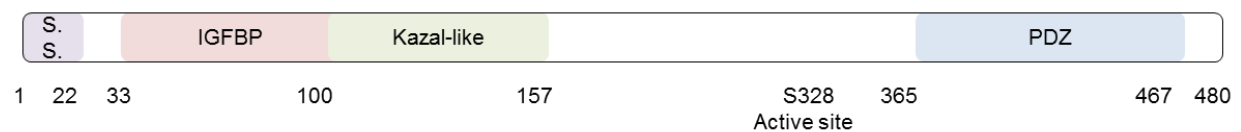
Figure 3.19: HtrA1 structural domains. HtrA1 protein has four distinct protein domains. From N- to C-terminus: IGFBP (insulin-like growth factor binding protein domain) - a kazal domain – active domain (trypsin-type serine protease) – PDZ (regulatory domain). SS: signal sequence (for secretion).

More clarity about the intracellular role of HtrA1 and its association to AMD could be brought by the identification of HtrA1-interacting partners in a relevant cellular model. Thus ICC-MS was applied in a human fetal retinal pigment epithelial (hfRPE) cell culture model, in which cells are fully differentiated into a polarized functional monolayer and in which HtrA1 protein is overexpressed to a pathological level [44, 84]. Such a pertinent cellular model resembling the RPE layer in the retina tissue and mimicking the increase of HtrA1 transcriptional levels associated with AMD predisposition and pathogenesis was successfully implemented in the lab of Roberto Iacone (Roche Innovation Center Basel) by Esther Melo, Frédéric Delobel and Corinne Stucki [44]. hfRPE cells were seeded and cultivated in transwell plates coated with laminin (Figure 3.20). After attachment and proliferation (one week), cells were transferred to a RPE maturation medium for three weeks. Differentiation into a polarized

functional monolayer was confirmed by RPE markers gene expression profiling [44]. After RPE monolayer formation, cells were stably transfected with a recombinant adenovirus containing wild type human HtrA1 mRNA (RPE HtrA1 WT) or an enzymatically inactive variant having a conversion by site-directed mutagenesis of the active domain Ser328 to Alanine (RPE HtrA1 S328A). Based on previous data generated by Esther Melo [85], the level of HtrA1 overexpression was assumed to be similar using HtrA1 WT or S328A constructs and in the range of tenfold higher compared to endogenous HtrA1 expression in untransfected RPE.
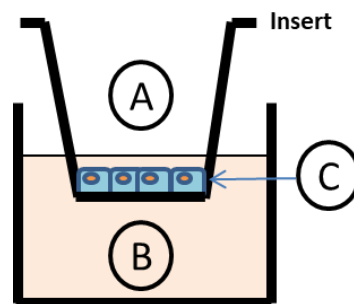

Figure 3.20: Human fetal retinal pigment epithelial (hfRPE) cell culture model. hfRPE cells were seeded in transwells. (A) Apical medium (mimicking retina secretion) (B) Basolateral medium (mimicking choroid secretion) (C) Intracellular (Fetal RPE cells on laminin-coated membrane).

Similarly to the previous applications, first the specificity of the commercially available anti-HtrA1 monoclonal antibody (MAB2916) against HtrA1 was evaluated. Proteins were extracted from differentiated hfRPE untransfected or overexpressing HtrA1 with a detergent-based buffer in non-denaturant non-ionic mild conditions (1% NP40). HtrA1 was specifically immunoprecipitated by the selected anti-HtrA1 antibody in cells overexpressing HtrA1 (Figure 3.21) whereas endogenous HtrA1 could not be detected due to low HtrA1 expression levels in the RPE cells. Notably, the capture antibody has also been used for immunoblot detection. In RPE lysates it detected in addition to HtrA1 (~ 50 kDa), a few high molecular bands. Those bands could represent HtrA1 aggregates as the protein has been described to form multimers [85]. Altogether, the performance of the antibody was considered to be sufficient for ICC-MS to further identify intracellular HtrA1 interacting partners.
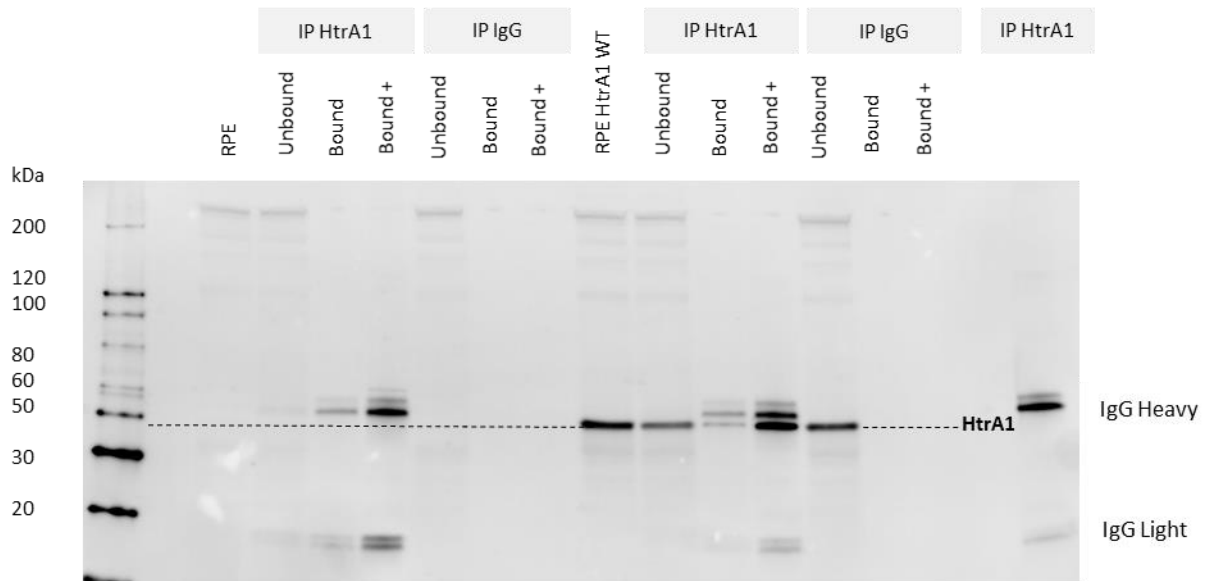
Figure 3.21: HtrA1 immuno-capture in differentiated hfRPE. HtrA1 was immunoprecipitated from protein extracts of hfRPE differentiated for 21 days (900 µg), untransfected (RPE) or overexpressing HtrA1 (RPE HtrA1 WT) using anti-HtrA1 antibody (R&D MAB2916). Western blot was detected with the same antibody. Loading: soluble fraction of total protein extract from RPE and RPE HtrA1 WT: 50 µg; Unbound: 50µg; Bound: 10 µl from 60 µl elution; Bound +: 15 µl from 60 µl elution. Immunoprecipitation with human IgG as a control as well as HtrA1 immunoprecipitation in absence of protein extracts (No RPE).

ICC-MS (five concentrations of free anti-HtrA1 antibody in triplicates) was then performed both in 500 µg of soluble protein extracts from differentiated hfRPE either transfected with HtrA1 WT or the enzymatically inactive variant HtrA1 S328A. A similar capture and displacement upon increasing concentrations of HtrA1 antibody were observed for both transfections (Figure 3.22 A and B).
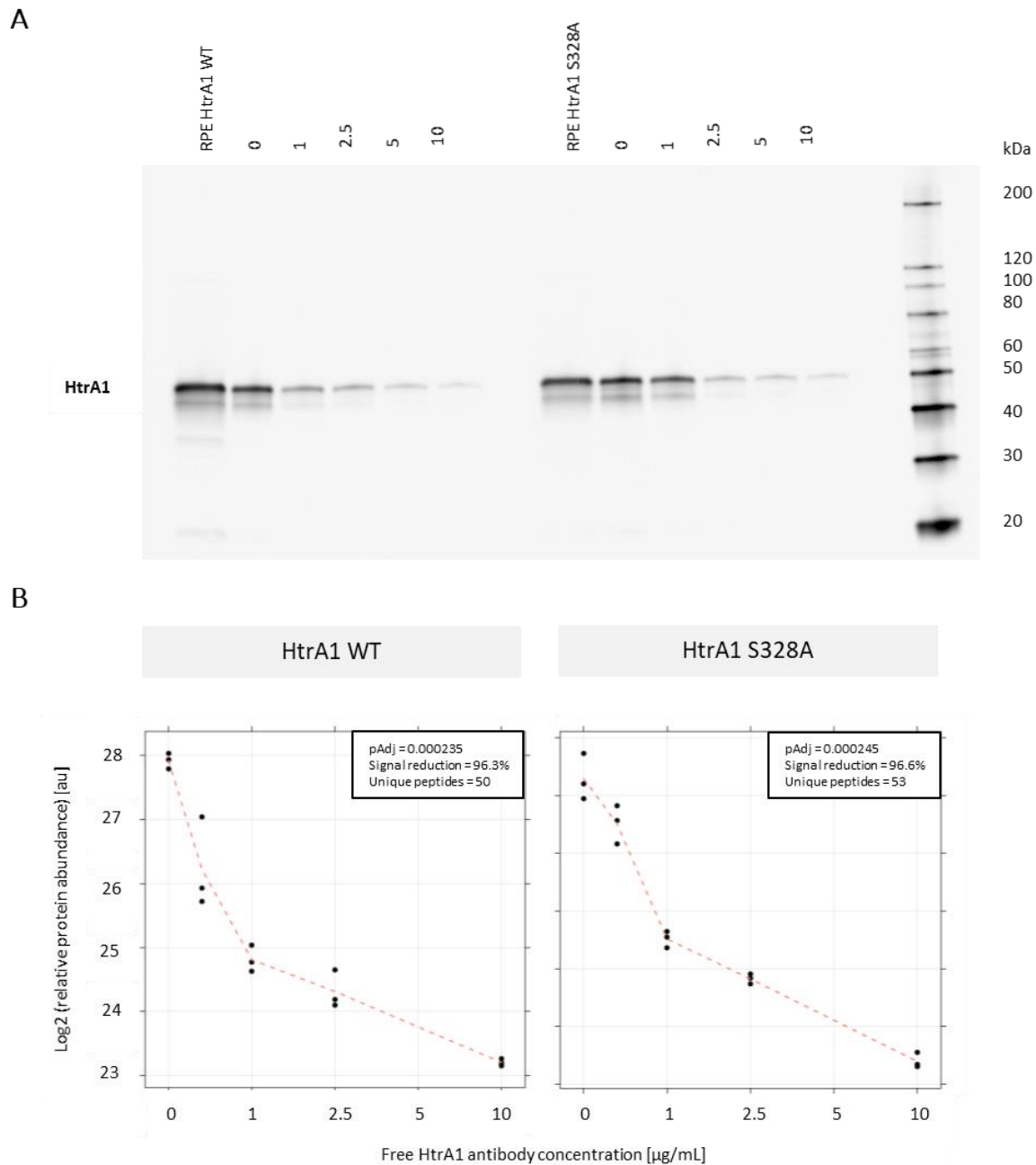
Figure 3.22: ICC-MS of HtrA1 in hfRPE extracts. HtrA1 was immunoprecipitated from 21 days differentiated hfRPE (500 µg protein extracts per conditions; both proteolytic active HtrA1 protein and enzymatically inactive S328A variant transfections) pre-incubated with 0, 1, 2.5, 5 and 10 µg/ml of free HtrA1 antibody. HtrA1 protein in bound fractions was (A) probed by Western Blot with an anti-HtrA1 antibody [83] (B) identified and quantified by LC-MS/MS (y axis: relative abundance; x axis: free HtrA1 antibody concentrations; each black dot is a replicate).

In total, 868 proteins were identified and quantified respectively in hfRPE overexpressing HtrA1 WT and among them nineteen proteins were identified as HtrA1-interacting proteins (adjusted p-value below 1% and at least two unique peptides; Table 3.5).

| SwissProt entry	pAdj	Signal reduction (%)	Unique Peptides
HTRA1_HUMAN	0.0002	96	37
C1QTNF5_HUMAN	0.0002	96	3
TBB4B/1/8/8L_HUMAN	0.0002	85	2
TBB3_HUMAN	0.0002	94	2
ECHB_HUMAN	0.0002	75	4
NPM_HUMAN	0.0004	98	5
RS27/L_HUMAN	0.0005	85	4
PGAM5_HUMAN	0.0005	72	4
TBA1A/3C/3E/4A_HUMAN	0.0005	87	4
COL12_HUMAN	0.0011	65	4
DJC10_HUMAN	0.0039	92	3
TBA4A_HUMAN	0.0055	97	2
GCST_HUAMN	0.0063	66	2
DNJA2_HUMAN	0.0064	86	3
TBB5_HUMAN	0.0071	90	6
TBB4A_HUMAN	0.0071	90	12
TBB6_HUMAN	0.0071	94	4
TBB4B/8/8L_HUMAN	0.0076	90	2
TBB3/1_HUMAN	0.0081	91	12
TBA1C/4B/AL3_HUMAN	0.0087	89	4

Table 3.5: HtrA1 co-immunoprecipitated proteins after statistical analysis of MS data (ranked by pAdj then signal reduction). Besides proteins from the tubulin family, 9 proteins (C1QTNF5, ECHB, NPM, RS27L, PGAM5, COL12, DJC10, GCST and DNJA2) were identified as HtrA1-interacting proteins with adjusted p-values below 1% and at least 2 unique peptides. pAdj: p value to test displacement adjusted by the Westfall-Young approach; Signal reduction: signal difference between the highest and the lowest concentration of competitor antibody; Unique peptides: number of peptides of a respective protein quantitation group.

Interestingly from the list, ten proteins were part of the tubulin family. Comparable interactions, predominantly with ß-isoforms, were observed when performing an ICC-MS in HtrA1 S328A hfRPE indicating tubulin binding irrespectively of the protease activity (Figure 3.23). Similar observations describing HtrA1 as a microtubule associated protein (MAP) were already published [86, 87].
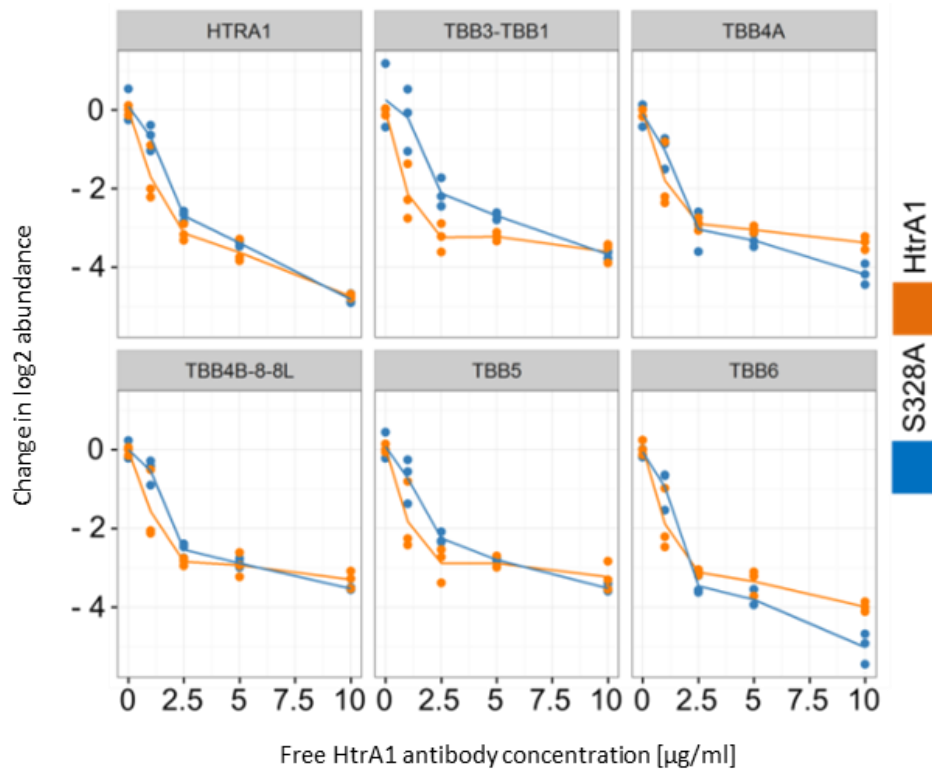
Figure 3.23: ß-tubulins identification as specific HtrA1 interacting partners by ICC-MS. HtrA1 was immunoprecipitated from 21 days differentiated hfRPE (500 µg protein extracts per conditions; both proteolytic active HtrA1 protein and enzymatically inactive S328A variant transfections) pre-incubated with 0, 1, 2.5, 5 and 10 µg/ml of free HtrA1 antibody. HtrA1 protein in bound fractions was (A) probed by Western Blot with an in-house antibody (B) identified and quantified by LC-MS/MS (y axis: change in ß-tubulins log2 abundance; x axis: free HtrA1 antibody concentrations; each dot is a replicate).

However, to confirm this interaction further validation experiments via co-IP in 21 days differentiated hfRPE overexpressing HtrA1 WT or HtrA1 S328A were performed (Figure 3.24 A) as well as in cotransfected Hek293 cells (Figure 3.24 B, data generated at NMI by Dr Philipp Kaiser). Both approaches unambiguously confirmed tubulins as being specific HtrA1 interacting partners.
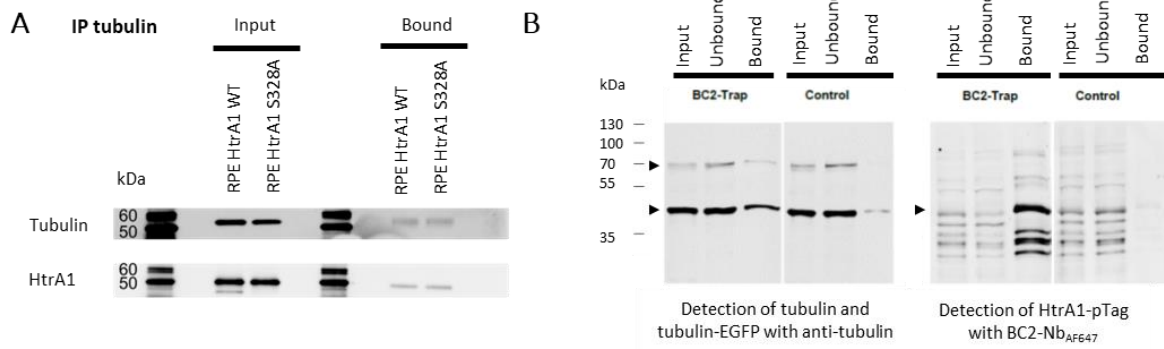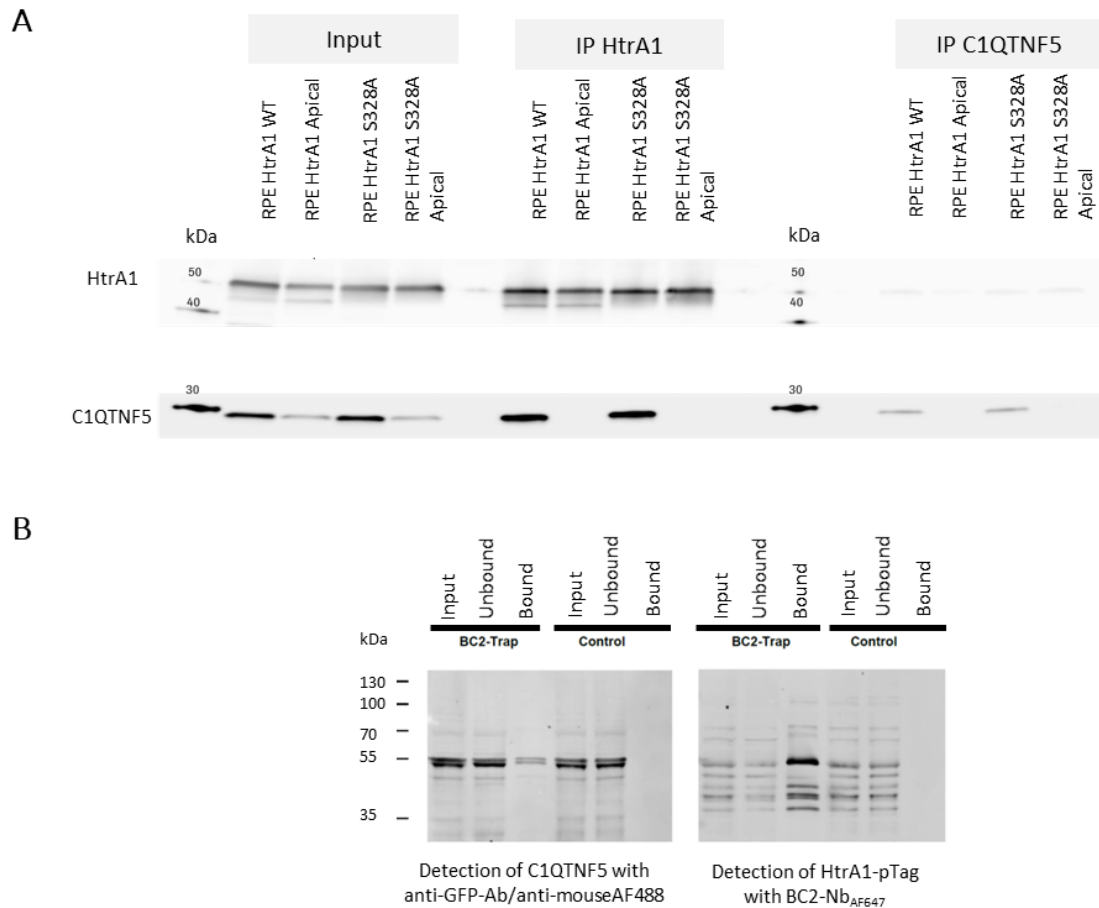
Figure 3.24: HtrA1 interacts with tubulin (A) HtrA1 was co-captured in 21 days differentiated hfRPE overexpressing HtrA1 WT (RPE HtrA1 WT) and HtrA1 S328A (RPE HtrA1 S328A) immunoprecipitated with a ß-tubulin antibody (MAB3408). Inputs (10 µg cell lysates) and bound fraction (1/4 elution volume) revealed by either anti-ß-tubulin (Ab52623) or anti-HtrA1 (in-house) antibodies. (B) Hek293 cells were transiently transfected with HtrA1-pTag (52.7 kDa) and tubulin-EGFP (77.6 kDa) prior to lysis and immunoprecipitation with BC2-Trap which specifically binds to the pTag. Immunoblots were probed with anti-ß-tubulin antibody (MAB3408) or anti-BC2 nanobody conjugated to Alexa647. Experiments are performed in biological replicates. Shown are representative results of one experiment. Control for unspecific binding to the nanobody scaffold and/or bead matrix: beads coupled with a BSA-specific nanobody (data generated at NMI by Dr Philipp Kaiser, material from ChromoTek). Additional negative controls indicated that interactions are not mediated by the interaction with beads loaded with EGFP or EGFP-pTag (data not shown).

In addition to HtrA1 direct interaction with tubulins, subsequent degradation could then be demonstrated [44] in alignment with published literature [87] and supporting the hypothesis that HtrA1 is having an intracellular protease activity.

In the list of HtrA1 interacting partners, candidates related to the described HtrA1 low temperature chaperone activity [88, 89] were also identified (i.e. DJC10 and DNAJ2) as well as interactors that were never reported in the literature such as Complement C1q tumor necrosis factor-related protein 5 (C1QTNF5). This latter is definitively an interesting candidate for further validation experiments as proteins of the complement system are at the basis of the development of AMD [90]. C1QTNF5 has been described as a component of basement membranes which may play a role in cell adhesion and interestingly mutations in this gene have been associated with late-onset retinal degeneration [91]. Intracellular binding confidence was further confirmed by coIP experiments (Figure 3.25 A and B). Although HtrA1-C1QTNF5 interaction was also detected in the apical media from differentiated RPE using ICC-MS (data

not shown), this was not observed by immunoblot probably due to lower protein concentrations in the apical medium.

Figure 3.25: C1QTNF5 interacts with HtrA1. (A) HtrA1 was co-captured in 21 days differentiated hfRPE overexpressing HtrA1 (RPE HtrA1 WT) and HtrA1 S328A (RPE HtrA1 S328A) immunoprecipitated with C1QTNF5 antibody (LS-C345051). IP were also performed in apical medium mimicking retina secretion (compartment A Figure 3.20). Inputs (10 µg cell lysates or 15 µl apical medium) and bound fractions (1/4 elution volume) revealed by either anti-C1QTNF5 (Sigma HPA038604) or anti-HtrA1 (in-house) antibodies. (B) Soluble protein extracts of Hek293 cells transiently expressing HtrA1-pTag (52.7 kDa) and C1QTNF5-EGFP (77.6 kDa) were incubated with the BC2-Trap [92] to precipitate HtrA1-pTag. Immunoblots were probed with tag-specific antibodies. Control for unspecific binding to the nanobody scaffold and/or bead matrix: beads coupled with a BSA-specific nanobody (data generated at NMI by Dr Philipp Kaiser, material from ChromoTek). Additional negative controls indicated that interactions are not mediated by the interaction with beads loaded with EGFP or EGFP-pTag (data not shown).

In summary the results demonstrated that the ICC-MS approach was a relevant method to identify novel and biologically relevant Htra1-interacting partners.

Taken together, the results obtained from all the above-mentioned applications support the validity of ICC-MS as a technically straightforward and widely applicable new approach for studying in an unbiased fashion biologically relevant protein-protein interactions (Figure 3.26). Notably, the combination of immuno-competition and quantitative mass spectrometry increases specificity compared to more classical AP-MS approaches by efficiently filtering out background proteins. Without the need for prior knowledge, a shorter list of candidates is generated having a greater chance to be functionally validated.
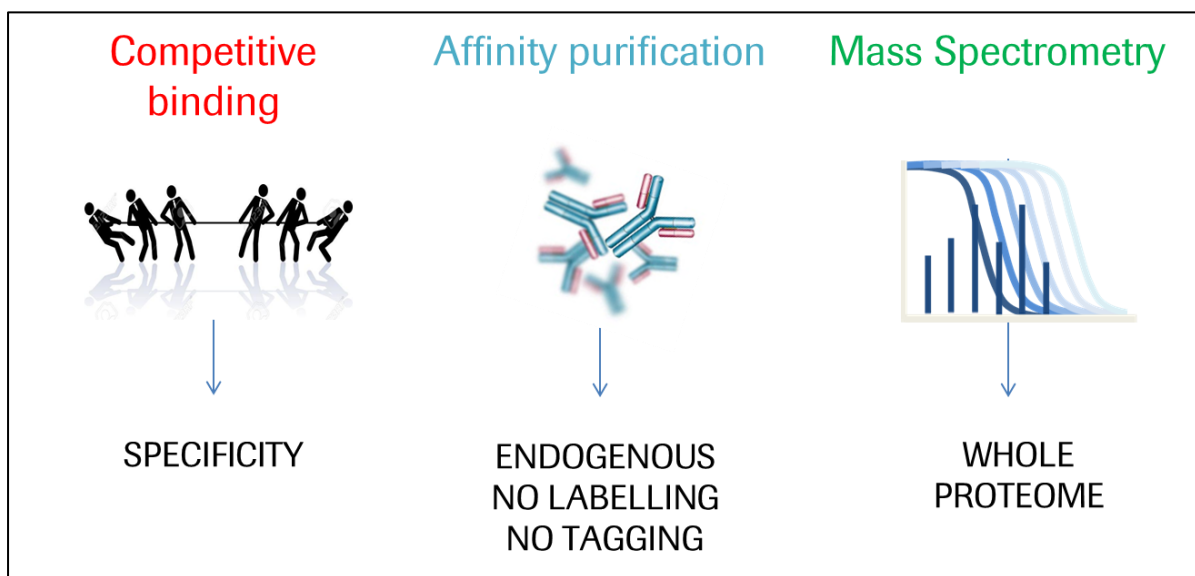
Figure 3.26: Schematic overview of the three main components of the ICC-MS platform. ICC-MS is a versatile approach delivering limited number of candidates with a high potential for functional relevance. The competitive binding step is allowing the ranking of binders based on their specificity. The endogenous profiling in *in-vivo* like conditions is increasing the chance of identifying biologically relevant interactions. Finally, there is no need for prior knowledge and there is no restriction to previously published interactions by the combination with a MS readout.

The molecular apparatus of a cell being not only controlled by protein-protein interactions but also largely by protein-nucleic acid interactions [93], the next step was to adapt ICC-MS to the study of proteome-wide binding to RNA.

3.2. ICC-MS-derived platform as an RNA-centric approach to study proteome-wide binding

The molecular basis of protein-RNA interactions is not completely understood but the role of RNA-binding proteins (RBPs) in cells is central from RNA fate to its synthesis and decay [94, 95]. Innovative technologies are heavily needed to complement our knowledge in that field. An approach derived from ICC-MS could fill the gap by offering an increased specificity and accuracy compared to more classical protein centric (RIP-chip) and RNA-centric (RNA pull-downs) approaches.

The ICC-MS platform has been developed and optimized to be used with an antibody immunoprecipitating a protein of interest together with its interacting partners. In the first chapter, the high specificity of ICC-MS could be demonstrated in three different cellular models. In this second chapter, it was tested whether a similar platform can be applied as a RNA-centric approach to screen in an unbiased fashion for protein-RNA interactions (Figure 3.27). The biggest challenge of such an Oligo-Competitive Capture Mass Spectrometry (OCC-MS) approach is to deal with an even higher amount of background proteins due to the fact that RNA molecules are large in size and negatively charged (phosphate groups). Thus conditions ensuring capture to an extent that allows detection of interacting partners while efficiently filtering out non-specific binders needed to be established and validated. For a proof-of-principle study two well characterized RNA-protein complexes were chosen: the Iron Regulatory Element-Aconitase complex and the Androgen Receptor-Hu-antigen R complex.

After validating the technical feasibility, the approach was applied to dissect the mode of action of survival of motor neuron (SMN) splicing modifiers.
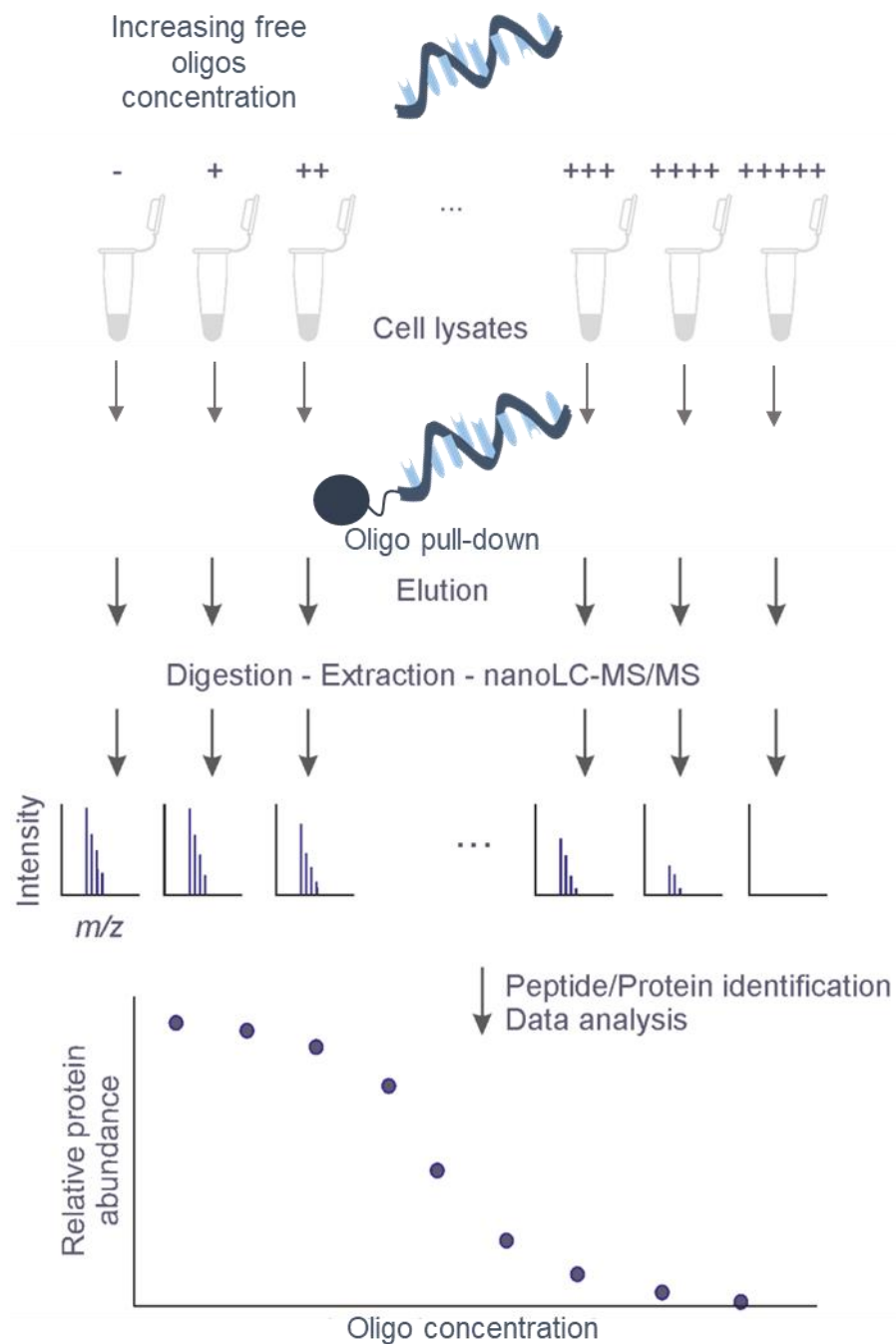
Figure 3.27: Schematic outline of the OCC-MS experimental procedure to identify protein-RNA interactions. Protein extracts are first pre-incubated with increasing concentrations of the free form of specific oligonucleotide sequence. Extracts are then individually loaded on agarose beads with the same oligonucleotide sequence immobilized. The unbound fractions are discarded and beads are washed several times. Bound proteins are eluted from the beads and digested with trypsin. Extracted peptides are analyzed by LC-MS/MS on an Orbitrap mass spectrometer. Relative protein abundances are translated from peptide extracted ion chromatograms (EICs) and finally subjected to a statistical analysis to derive concentration-dependent signal decrease of the specific interactors.

3.2.1 Technical feasibility study using the Iron Regulatory Element-Aconitase complex

The protein Aconitase is a bifunctional protein catalyzing the isocitrate/citrate interconversion but acting also as an RBP binding to Iron Regulatory Element (IRE) that are part of a family of non-coding mRNA sequences regulating iron homeostasis [96, 97].

To perform affinity purifications, 3' biotinylated IRE sequences were selected and used as baits on streptavidin magnetic beads in human fibroblasts.

The ultimate goal of such a RNA centric strategy is to better understand the role of specific RBPs in mRNA transcription and processing rather in the nucleus than in mRNA fate in the cytoplasm (i.e., subcellular localization, metabolism or translation). Thus, it was important to evaluate from the beginning whether nuclear fractions of cells could serve as starting material. Thus, a procedure allowing a proper separation of nuclear from cytosolic fractions while ensuring proteins solubilization, native conformation and protein interactions was required.

Human fibroblasts subcellular fractionation was performed by re-suspending cell pellets in a hypotonic HEPES buffer. After mechanical disruption and centrifugation, nuclear pellets were re-suspended in 0.5% NP40 lysis buffer containing protease and RNase inhibitors. To confirm the efficiency of fractionation, aliquots of cytosolic and nuclear fractions were loaded on SDS-PAGE and transferred for immunoblotting against nuclear and cytoplasmic control proteins (Figure 3.28). The nuclear proteins Lamin B and Histone H2B were exclusively detected in the nuclear fractions whereas the cytosolic 14-3-3 αß protein was detected in the cytosolic fraction. Thus, this protocol is efficiently extracting nuclei while being compatible with the approach having a low concentration of non-denaturating detergent.

The amount of biotinylated IRE RNA immobilized on streptavidin magnetic beads was the next parameter to be explored.

50 µl of nucleic-acid compatible streptavidin magnetic beads (supplied at 10 mg/ml) were mixed with 0, 50, 100, 250 or 500 pmol of biotinylated IRE for 30 minutes and incubated then with 150 µg of human fibroblast nuclear extracts for 1 hour. After a few washing steps, IRE binders were eluted, separated by SDS-PAGE and transferred for Aconitase immunoblotting (Figure 3.29).

With an amount of 50 pmol of biotinylated IRE RNA immobilized on 50 µl of streptavidin magnetic beads, Aconitase was almost completely captured from the human fibroblast nuclear extracts. Only a weak signal was remaining in the unbound fraction. Based on these findings the amount of 100 pmol was selected to increase the chance in future experiments to identify lower abundant proteins as well. Higher amounts such as 250 and 500 pmol turned out to be slightly suboptimal probably due to steric hindrance.
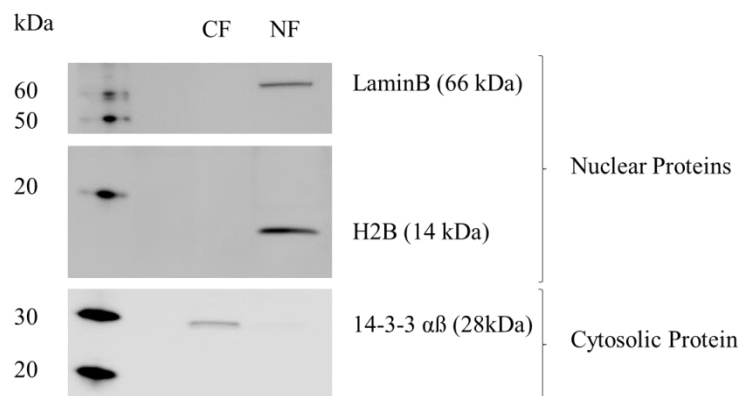

Figure 3.28: Nuclear extracts quality check from human fibroblasts. 10 µg of human fibroblasts nuclear (NF) and cytosolic extracts (CF) were separated by SDS-PAGE, transferred on PVDF membrane and probed for Histone H2B (nuclear protein), Lamin B (nuclear protein) and 14-3-3 αß (cytosolic protein).
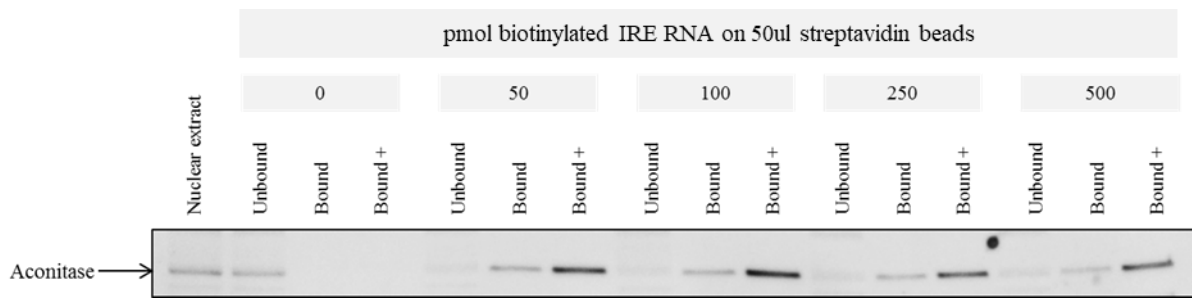
Figure 3.29: Capture of Aconitase by increasing amounts of IRE RNA immobilized on beads. IRE pull-downs were performed in human fibroblast nuclear extracts (150 µg of total protein) with 0, 50, 100, 250 or 500 pmol of biotinylated IRE (5'- UCCUGCUUCAACAGUGCUUGGACGGAAC -3') immobilized on streptavidin magnetic beads. Nuclear extracts (10 µg), unbound (10 µg) and bound fractions (Bound: 20 µl of elution, Bound +: 30 µl of elution; from a total elution volume of 100 µl) from IRE pull-down experiments were probed for Aconitase by immunoblot using anti-Aconitase EPR7225.

A prerequisite to benefit from the theoretical specificity of the OCC-MS approach was to demonstrate the feasibility to combine an efficient capture of RBP with a competitive binding between immobilized and free RNAs. This was assessed by pre-incubating human fibroblast nuclear extracts (150 µg of total protein) with 0, 50, 100, 500 and 1000 pmol of free IRE RNA prior to IRE pull-down (Figure 3.30). A reduced signal of the captured Aconitase protein was detected upon applying 50 pmol of free IRE competitor and the largest drop was observed by adding 500 pmol. Finally pre-incubation of the cellular extract with 1000 pmol led to complete loss of Aconitase signal in the bound fraction. Thus, it is technical feasible to introduce a binding competition and ultimately to benefit from this strategy to identify specific binders.
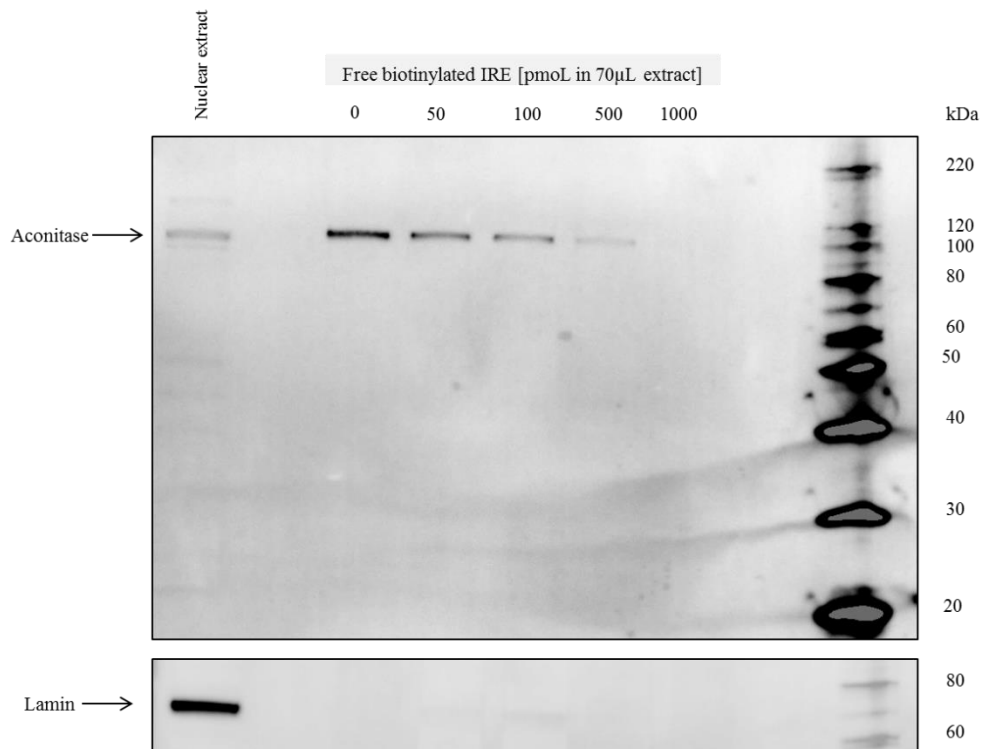
Figure 3.30: Aconitase immunoblot from IRE RNA-competitive pull-downs in cellular nuclear extracts. 150 µg of human fibroblast nuclear extracts were pre-incubated with 0, 50, 100, 500 and 1000 pmol of free IRE RNA sequences before affinity purification using biotinylated IRE immobilized on streptavidin magnetic beads (100 pmol on 50 ul of beads). Nuclear extract (10 µg) and bound fractions (30 µl from 100 µl total elution volume) were probed for Aconitase immunoblot using a secondary antibody coupled to HRP and chemiluminescence as a mode of detection. Lamin added as a non-RBPs control.

3.2.2 Technical feasibility study using Androgen Receptor-Hu-antigen R complex

To further confirm, the technical feasibility of an OCC-MS approach to profile RNA-proteins interactions, a second model was explored. The Hu-antigen R protein (HuR), a well described RBP part of the Elav/Hu family which has been shown to bind avidly to UC-rich region within the 3'-Untranslated Region (UTR) of the Androgen Receptor (AR) messenger RNA (mRNA) [98]. Thus 3' biotinylated AR RNA were selected for immobilization on streptavidin magnetic beads (100 pmol on 50 µl of beads).

HuR was efficiently affinity purified from human fibroblast nuclear extracts (140 µg) although some signal could still be detected in the unbound fraction (Figure 3.29). Subsequently, nuclear extracts were pre-incubated with 0, 100, 200 and 300 pmol of free AR RNA prior to be individually loaded on immobilized biotinylated AR RNA. Notably, the amount of 300 pmol

of free AR RNA was not enough for complete loss of the HuR in the bound fraction signal. However, a concentration-dependent signal decrease was observed, successfully assigning HuR as an AR RNA binder (Figure 3.31).

At this stage, it was important to verify that a signal decrease upon increasing concentrations of free AR RNA was not observed for a non-specific binder. The heterogeneous nuclear ribonucleoprotein A1 (ROA1), a well-characterized RBP involved in many aspects of RNA processing [99] and being highly and ubiquitously expressed, was selected for these analysis. ROA1 was detected in the AR RNA bound fraction in absence of competing free AR RNA and could have been identified as a potential AR binder in the context of a binary comparison. However, incubation with increasing concentrations of free AR RNA showed no impact on its capture, which is a characteristic behavior of a non-specific binder in such an experiment.

It is important in this second chapter to clearly differentiate "non-specific binders" from proteins binding to RNA substrates at "non-specific sites". The focus of this thesis is on "non-specific binders", i.e. highly abundant or 'sticky' proteins interacting with a matrix or a chemical backbone without any relation to biological activity. The latter also often named as "background" does not include the numerous proteins which are having important biological roles by interacting with RNAs irrespectively of specific sequences or structure signature.
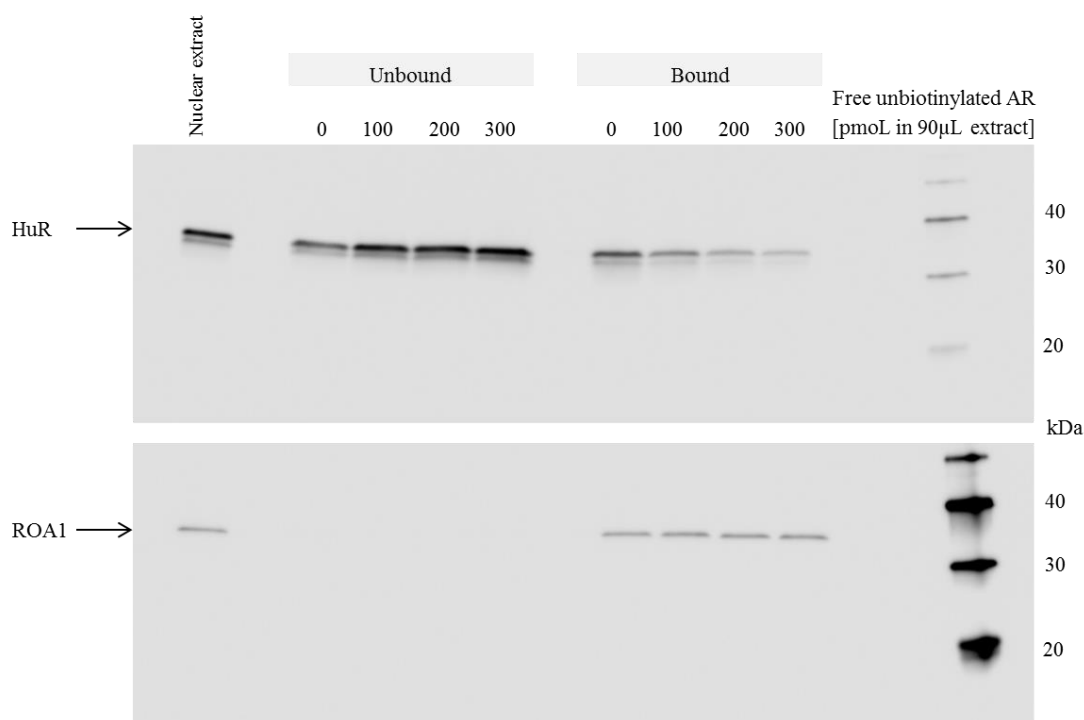
Figure 3.31: AR RNA-competitive pull-downs in cellular nuclear extracts. 140 µg of human fibroblast nuclear extracts were pre-incubated with 0, 100, 200 and 300 pmol of free AR RNA (5´-CUGGGCUUUUUUUUCUCUUUCUCUCCUUUCUUUUUCUUCUUCCCUCCCUA -3´) prior to affinity purification using biotinylated AR RNA immobilized on streptavidin magnetic beads. Nuclear extract (10 µg) Unbound (20 µg) Bound fractions (20 µl from 100 µl total elution volume) probed for HuR and ROA1.

Altogether, these results confirmed the technical feasibility of OCC-MS to profile RNA-protein interactions. The next step being to test the combination to MS as a readout which is offering then the possibility to screen in an unbiased fashion the whole proteome.

3.2.3 Application to decipher survival of motor neuron (SMN) splicing modifiers mode of action

In eukaryotic organisms, RNA splicing is a complex post-transcriptional RNA processing mechanism based on removal of introns from precursor messenger mRNA (pre-mRNA) followed by exon ligation ultimately leading to mature mRNA [100]. A single gene can then code for different protein isoforms via pre-mRNA alternative splicing [101]. This step is regulated by the spliceosome, a large ribonucleoprotein complex [102], composed of multiple proteins (trans-acting factors) recognizing exonic or intronic sequence-specific sites (cis-acting elements) which are acting as splicing enhancers or silencers.

Mis-splicing events due to hereditary or somatic mutations can have dramatic pathological consequences [103]. One example is spinal muscular atrophy (SMA), the leading genetic cause of infant mortality (1 in 10'000 births) associated with progressive degeneration of spinal motoneurons [104, 105] due to loss of survival motor neuron (SMN) full-length protein. SMN protein is encoded by *SMN1* and *SMN2* genes having nine exons. *SMN2* exon 7 is often skipped due to a C-to-T transition at position 6 which was suggested to either disrupt an Exonic Splicing Enhancer (ESE) motif or to create an exonic splicing silencer (ESS) [106, 107]. This is leading to mRNA translation into a truncated and unstable protein. In SMA patients, *SMN1* is mutated or deleted and the level of *SMN2* mis-splicing is directly related to the severity of the disease (Figure 3.32). SMA type 1, accounting for almost 60% of all cases, is the most severe form of the disease with life expectancies of less than eighteen months. Therapeutic strategies targeting *SMN2* alternative splicing are highly attractive although targeting splicing machinery can be associated with weak specificity.
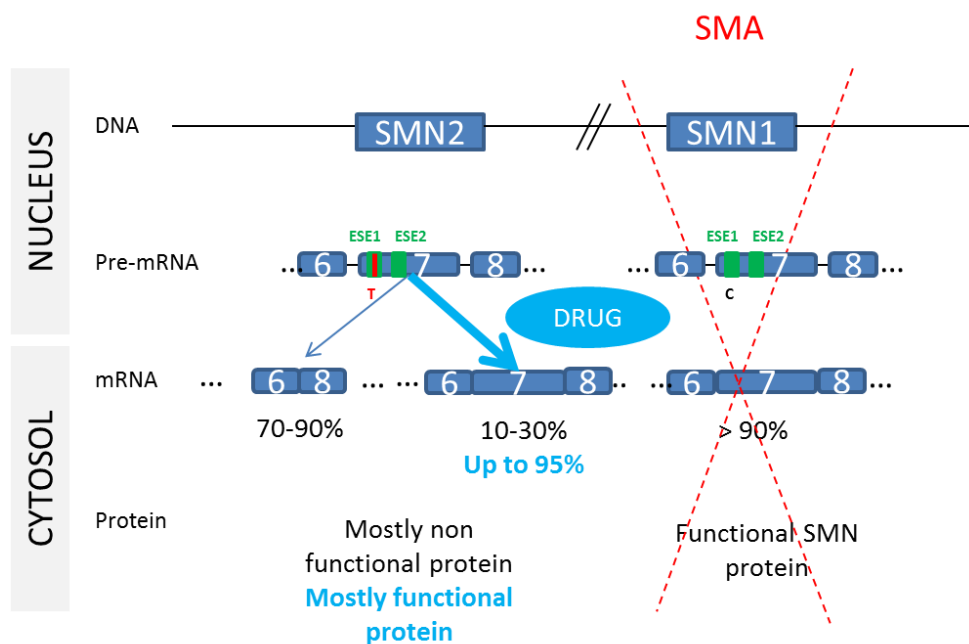
Figure 3.32: Representation of *SMN1* and *SMN2* genes and protein product in SMA patients. In SMA, *SMN1* is mutated or deleted. *SMN2* a nearly identical copy of *SMN1* is differing by a C-to-T transition at position 6 of exon 7 (probably in an ESE region) leading to frequent exon skipping. Mis-spliced mRNAs are translated into truncated and unstable proteins. The lower the level of functional SMN protein is, the more critical the symptoms of the disease are. In light blue, how *SMN2* splicing modifiers would act to ensure proper production of SMN functional protein.

In this context, Roche Innovation Center Basel developed the SMN-C class of molecules which are promising *SMN2* splicing modifiers [108] showing high selectivity towards *SMN2* via a unique mode of action [109]. SMN-C compounds were shown to interact with two distinct sites, the exonic splicing enhancer 2 (ESE2) region of exon 7 and the 5'splice site (ss). Both sites were revealed to be important for the specific activity of SMN-C in promoting exon 7 inclusion in *SMN2* transcripts.

To study in further depth SMN-C mode of action, an OCC-MS approach was applied to identify trans-acting splicing factors associated with the ESE2 motif in *SMN2* exon7 (Figure 3.33 A) followed by an exploration how SMN-C compounds influence their binding properties (Figure 3.33 B).
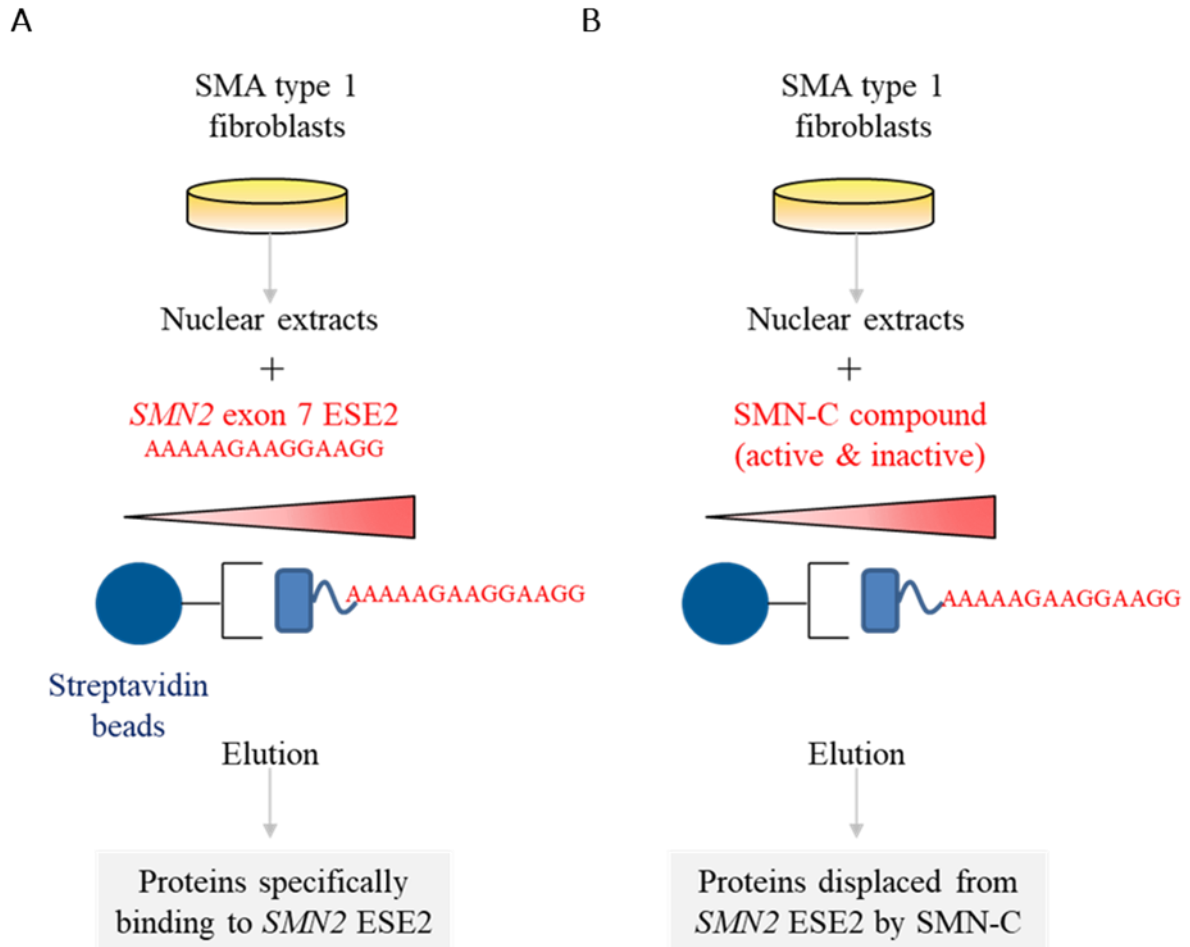
Figure 3.33: Schematic outline of OCC-MS approach applied for SMN-C mode of action deciphering. Nuclear extracts from SMA type 1 patient fibroblasts were first pre-incubated with increasing concentrations of free *SMN2* ESE 2 (A) or free SMN-C compound (B). Extracts were then individually affinity purified on biotinylated *SMN2* ESE2 immobilized on streptavidin magnetic beads. After washing the beads several times, bound proteins were eluted and further processed for protein identification by LC-MS/MS as described in section 3.1. Similar statistical analysis to the one described in section 3.1 was applied to derive (A) concentration-dependent signal decrease of *SMN2* ESE2 specific interactors or (B) *SMN2* ESE2 interactors displaced by SMN-C compound. *SMN2* ESE2 sequence: 5' AAA AAG AAG GAA GG 3'.

3.2.2.1. Identification of trans-acting splicing factors associated with ESE2 motif in SMN2 exon7

To identify *SMN2* ESE2-interacting proteins, *SMN2* ESE2 pull-downs (100 pmol biotinylated *SMN2* ESE2 immobilized on 50 µl streptavidin magnetic beads) were performed in 170 µg of SMA type 1 fibroblast nuclear extracts previously incubated with increasing quantities of free *SMN2* ESE2 (0, 50, 100, 500 and 1000 pmol in 70 µl total volume) in triplicates. 3' end biotinylated *SMN2* ESE2 sequences were provided by Manaswini Sivaramakrishnan and Sonja Meier from the lab of Friedrich Metzger at Roche Innovation Center Basel. The fifteen bound fractions (five concentrations of free *SMN2* ESE2 in triplicates) were randomized, separated on SDS-PAGE and cut into five gel slices (Figure 3.34). The resulting 75 samples were in-gel digested and further processed for protein identification by LC-MS/MS. In total, 1413 proteins were identified and quantified in the bound fraction. A dose-response statistical analysis was then applied on the MS output by Balazs Banfai, Gonzalo Duran-Pacheco and Jens Lamerz [109] to check if specific binders can be efficiently extracted from such a high number of background proteins. Among the 1413 proteins identified and quantified, 42 proteins only were significantly displaced upon increasing free *SMN2* ESE2 concentrations (Adj-p value < 5%, identified with at least two unique peptides and signal reduction > 50%) (Table 3.6).
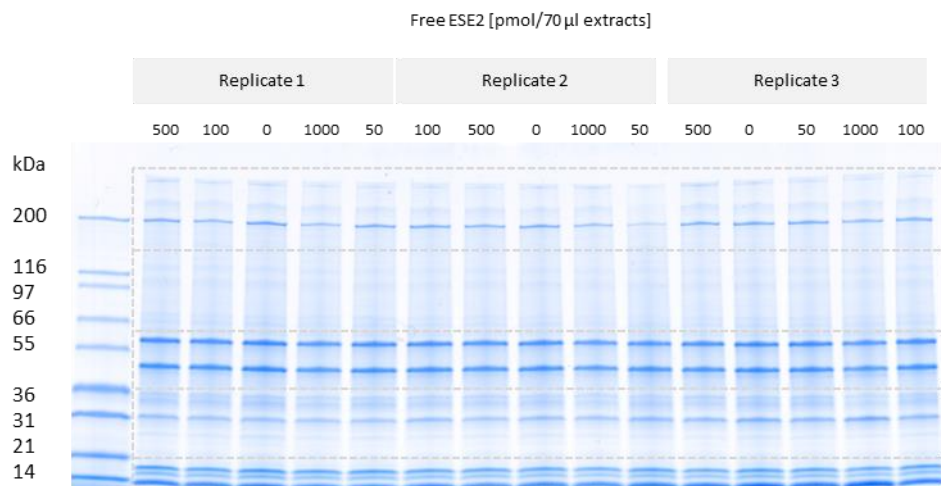
Figure 3.34: SDS-PAGE separation from ESE2 bound fractions in SMA type 1 patient fibroblasts nuclear extracts pre-incubated with free ESE2. ESE2-interacting proteins were extracted (100 pmol of biotinylated ESE2 on 50 µl of magnetic beads) from SMA type 1 patient fibroblasts nuclear extracts (170 µg total protein) pre-incubated with 0, 50, 100, 500 and 1000 pmol of free *SMN2* ESE2 sequences. Bound fractions (10 µl from 100 µl elution volume) were separated by SDS-PAGE migrated half its length and proteins were stained with Coomassie blue. Five bands (from 10 to > 200 kDa) were cut and in-gel trypsin digested for protein identification by LC-MS/MS.

SwissProt entry	pAdj	Signal reduction	Unique peptides
SRSF5_HUMAN	0.000177692	92%	3
TREX1_HUMAN	0.000189166	94%	4
SRSF1_HUMAN	0.000190289	86%	15
NUD16_HUMAN	0.000190667	84%	8
RU1C_HUMAN	0.000191046	83%	3
SNRPA_HUMAN	0.000191426	85%	15
RSMB_HUMAN	0.000192192	77%	8
TRA2B_HUMAN	0.000193742	75%	7
GRSF1_HUMAN	0.000218898	97%	18
NUCL_HUMAN	0.000219397	98%	38
ACOT9_HUMAN	0.0002199	85%	5
PABP2_HUMAN;EPAB2_HUMAN	0.000220404	97%	6
HNRH2_HUMAN	0.000220911	82%	18
HNRPF_HUMAN	0.000221931	87%	23
STRAP_HUMAN	0.000222961	96%	14
DAZP1_HUMAN	0.000223479	93%	5
HNRH1_HUMAN	0.000224524	72%	23
CSDE1_HUMAN	0.000225049	94%	3
LA_HUMAN	0.000225578	95%	5
PLOD1_HUMAN	0.000226108	64%	3
CNBP_HUMAN	0.000338367	98%	12
SSBP_HUMAN	0.000340766	70%	6
SRSF2_HUMAN;SRSF8_HUMAN	0.000341979	75%	4
SMD2_HUMAN	0.000345669	80%	9
SRSF1_HUMAN;SRSF9_HUMAN	0.000346917	75%	3
SMD3_HUMAN	0.000348174	74%	10
RBM3_HUMAN	0.00034944	68%	7
ROA1_HUMAN;RA1L2_HUMAN	0.000384531	65%	31
SRSF6_HUMAN	0.000407221	84%	3
TRA2A_HUMAN	0.000408039	79%	2
SRSF4_HUMAN	0.00040886	93%	2
PURA_HUMAN	0.00043592	55%	11
PURB_HUMAN;PURG_HUMAN	0.000437938	73%	14
CIRBP_HUMAN	0.000871242	67%	7
RL19_HUMAN	0.001205751	57%	13
RU17_HUMAN	0.00148403	85%	12
RUXE_HUMAN	0.00177071	57%	5
RUXF_HUMAN	0.003313015	61%	4
SMD1_HUMAN	0.005371784	59%	5
RTCA_HUMAN	0.013982808	99%	4
PCBP2_HUMAN;PCBP3_HUMAN	0.014945568	71%	2
ZCCHV_HUMAN	0.015148165	69%	3

Table 3.6: ESE2-interacting candidates identified after statistical analysis of MS data. pAdj: p value adjusted by the Westfall-Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free antibody; Unique peptides: number of peptides assigned to a protein quantitation group. Ranked by increasing adjusted p-values, then decreasing signal reduction and number of peptides.

Notably, most of the identified proteins are previously reported as RBPs. To check whether some of them are interacting with each other, the 42 ESE2-interacting protein candidates were uploaded in the Search Tool for the Retrieval of Interacting proteins database (SRINGdb). Actually, it appeared that many of them were previously reported as interacting with each other (Figure 3.35) which indicate a probability for some of them to be captured as part of a complex rather than as direct binders.

In the list, many serine-rich (SR) proteins were identified which are indeed known to bind and promote exon splicing in regions with ESEs. Many heterogeneous ribonucleoprotein particles (hnRNPs) were also found in the list of ESE2-interacting proteins although those proteins are more often associated with blocking of exon splicing in regions with ESSs via RNA-binding domains and splicing inhibitory domains (e.g. glycine-rich motifs) [110].
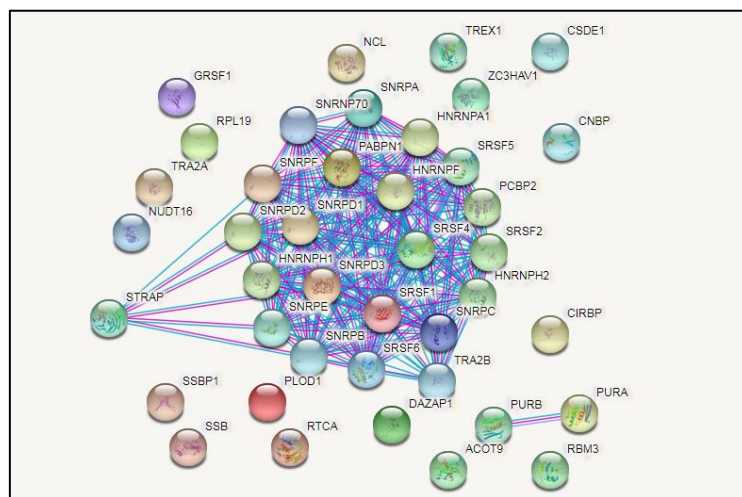
Figure 3.35: PPI network of ESE2-interacting candidates. The 42 protein candidates identified as ESE2-interacting proteins were loaded in the STRING DB (https://string-db.org/) using the highest confidence (0.900) in term of interaction score from curated databases (blue lines) as well as experimentally determined (pink lines).

The validity of the approach is confirmed by the presence of Tra2-ß1 in the list of ESE2-interacting protein candidates (pAdj 0.000193742, Figure 3.36). Indeed, Tra2-ß1 is a well described ESE2 trans-acting factor acting as an important actor of exon7 inclusion. The RNA recognition motif (RRM) of Tra2-ß1 was shown to recognize specifically a 5' AGAAA 3' motif within ESE2 [111] and this interaction is proposed in current models to lead to the recruitment of two other splicing factors SRSF9 and RBMX [112, 113]. Those proteins, known

to play a major role in *SMN2* exon7 inclusion, in contrary to Tra2- ß1 were not identified as ESE2-interacting partners (Figure 3.36). It is known that RBMX is not requiring a specific motif to bind SMN exon 7 RNA [113] and that SRSF9 is a weak affinity RNA binder [112]. Those proteins are also described to interact directly with Tra2-ß1. The fact that none were identified as co-eluting with Tra2-ß1 could indicate either a weak/transient affinity or that the complex formed by ESE2, Tra2-ß1, SRSF9 and RBMX might be active only during pre-mRNA processing.
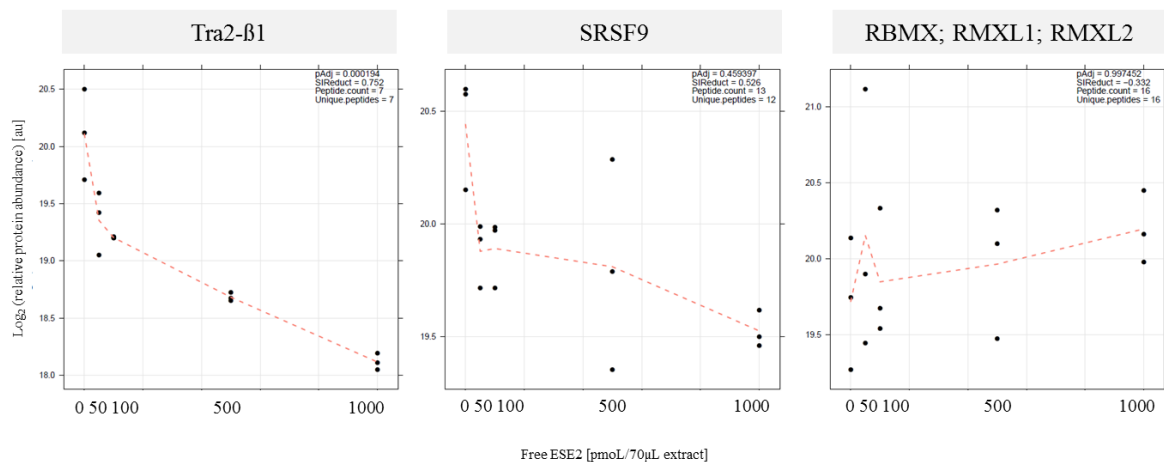
Figure 3.36: Tra2-ß1 identified as ESE2-interacting candidate. Displacement profiles of Tra2-ß1, SRSF9 and RBMX obtained from ESE2 RNA-competitive pull-downs in human SMA type 1 nuclear extracts. 170 µg of human fibroblast nuclear extracts were pre-incubated with 0, 50, 100, 500 and 1000 pmol of free ESE2 prior to pull-down with biotinylated ESE2 immobilized on streptavidin magnetic beads. Proteins from bound fractions were identified by LC-MS/MS and MS signals are subjected to a statistical analysis to derive concentration-dependent signal decrease of specific interactors.

It should be noted that a protein group formed by SRSF1 and SRSF9 was found in the list of potential ESE2-interacting candidates (Table 3.6) but the displacement observed was probably driven by SRSF1 peptides as observed in Figure 3.37.
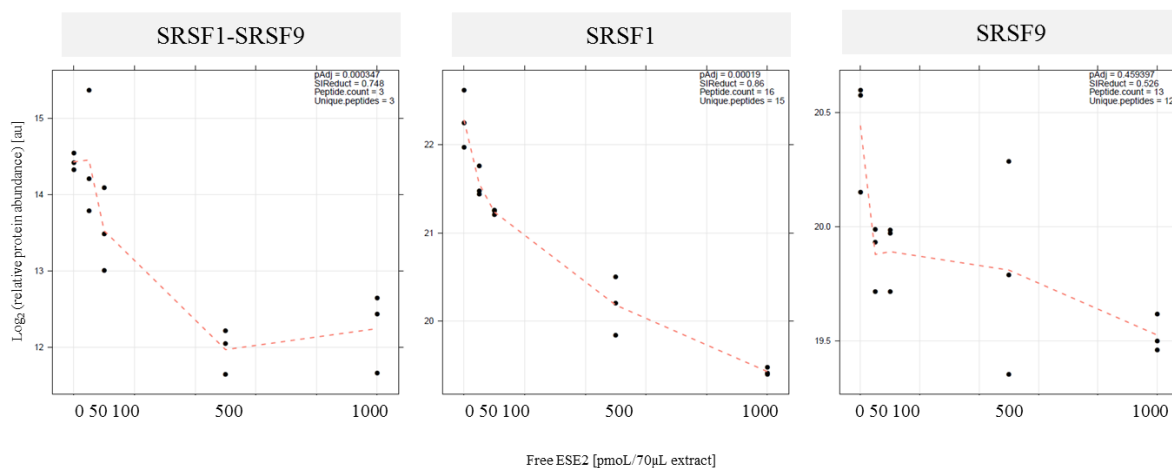
Figure 3.37: SRSF1 but not SRSF9 is identified as ESE2-interacting candidate. Displacement profiles of SRSF1, SRSF9 and the protein group SRSF1-SRSF9 obtained from ESE2 RNA-competitive pull-downs in human SMA type 1 nuclear extracts. 170 µg of human fibroblast nuclear extracts were pre-incubated with 0, 50, 100, 500 and 1000 pmol of free ESE2 prior to pull-down with biotinylated ESE2 immobilized on streptavidin magnetic beads. Proteins from bound fractions were identified by LC-MS/MS and MS signals were subjected to a statistical analysis to derive concentration-dependent signal decrease of specific interactors.

Altogether, these findings suggest a high relevance of combining competitive binding to RNA pull-downs in order to identify specific RBPs. However, additional factors than SR and SR-like proteins such as hnRNP proteins were identified and are probably involved in exon inclusion through direct or indirect interactions.

3.2.2.2. SMN-C compound effect on trans-acting splicing factors associated with ESE2

The OCC-MS approach revealed a protein complex associated with ESE2 motif in *SMN2* exon7. The impact of SMN-C splicing modifier on that complex was then explored (Figure 3.33 B). Studying the effect of SMN-C splicing modifier directly on ESE2 trans-acting splicing factors would indeed offer an opportunity to get a deeper understanding of its mode of action.

Two SMN-C compound derivatives were synthesized by Hasane Ratni (Roche Innovation Center Basel) to assess their effect on ESE2 interacting partners (Figure 3.38). Both compounds were tested by Hasane Ratni for their ability to modify *SMN2* splicing in a cellular assay using type 1 patient fibroblast. One compound was having an EC50 (concentration of drug that gives half-maximal response) of 43 nM in modifying *SMN2* splicing (Figure 3.38, "Active SMN-

C"). The additional carbonyl introduced at one extremity of the other compound led to a drop in its potency (EC50 > 1 µM, Figure 3.38, "Inactive SMN-C") [109].
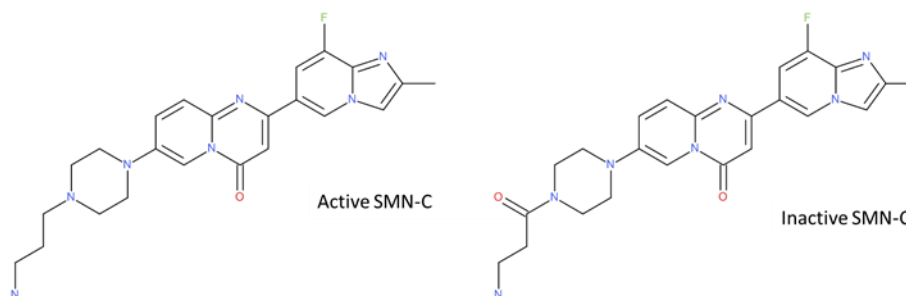
Figure 3.38: Chemical structures of SMN-C active and inactive derivatives. Active SMN-C and inactive SMN-C were synthesized and tested for their potency in modifying *SMN2* splicing in a cellular assay using type 1 patient fibroblasts by Hasane Ratni. The carbonyl group introduced at one extremity of the SMN-C inactive analog molecule is resulting in EC50 shift from 43 nM to above 1 µM (data not shown).

To assess the effect of both compounds to ESE2 interactome, affinity purifications were performed in human SMA type 1 fibroblasts nuclear (150 µg total proteins). Prior to the capture with 100 pmol of biotinylated ESE2 immobilized on streptavidin magnetic beads, nuclear extracts were pre-incubated with increasing concentrations (0, 0.5, 1, 5 and 10 µM in triplicates) of either active or inactive SMN-C derivatives. Thirty bound fractions (five concentrations of free SMNC active or inactive in triplicates) were randomized, separated on SDS-PAGE and cut into five gel slices (Figure 3.39).
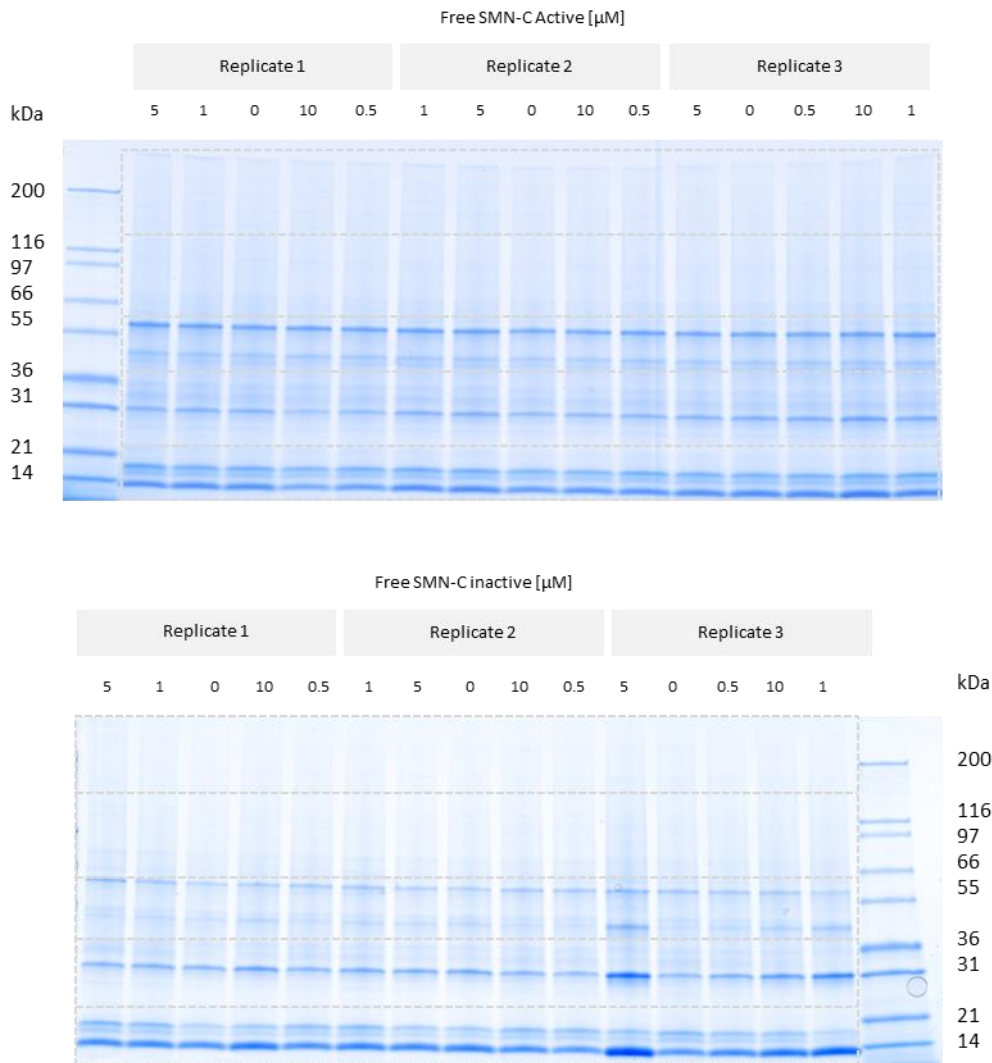
Figure 3.39: SDS-PAGE separation from ESE2 bound fractions in SMA type 1 patient fibroblasts nuclear extracts after competition with free SMN-C compounds. ESE2-interacting proteins were extracted (100 pmol of biotinylated ESE2 on 50 µl of magnetic beads) from SMA type 1 patient fibroblasts nuclear extracts (150 µg total protein) pre-incubated with 0, 0.5, 1, 5 and 10 µM of free SMN-C active or inactive compounds in triplicates. Bound fractions (10 µl from 100 µl elution) were separated by SDS-PAGE migrated half its length and proteins were stained with colloidal blue. Five bands (from 10 to > 200 kDa) were cut and in-gel trypsin digested for further protein identification by LC-MS/MS.

The resulting 150 samples were in-gel digested and further processed for protein identification by LC-MS/MS. In total, 850 proteins were identified and quantified in ESE2 RNA pull-down both in presence of active SMN-C or of its inactive analog. Notably, none of those proteins were found to be significantly displaced by increasing concentrations of SMN-C inactive analog (Adj-p value < 5% and identified with at least two unique peptides) after extracting

dose-response relationships. On the other hand, increasing concentrations of active SMN-C led to the specific displacement of ten proteins from ESE2 sequence (Adj-p value < 5% and unique peptides > 1; Table 3.7). Based on STRINGdb (Figure 3.40), four of these proteins are described to be part of a complex: U1 snRNP proteins A and 70, FUS and more interestingly *SMN2* exon 7 inclusion promoter RBMX.

| SwissProt entry	Gene name	pAdj	Signal reduction	Unique peptides
NUCL_HUMAN	NCL	0.000196516	36%	45
FUS_HUMAN	FUS, TLS	0.000206215	43%	3
SNRPA_HUMAN	SNRPA	0.000206658	38%	5
RBMX; RBMXL1-2-3_HUMAN	HNRPG	0.000207551	29%	15
RU17_HUMAN	SNRNP70	0.00033	63%	3
STRAP_HUMAN	STRAP	0.000384531	53%	11
CSDE1_HUMAN	CSDE1	0.0003861	49%	11
DAZP1_HUMAN	DAZAP1	0.000387683	49%	2
RBM45_HUMAN	RBM45	0.000388479	52%	3
EWS_HUMAN	EWSR1	0.000394144	45%	2

Table 3.7: SMN-C interacting candidates identified after ESE2 pull-down. pAdj: p value adjusted by the Westfall-Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free antibody; Unique peptides: number of peptides assigned to a protein quantitation group. Ranked by increasing adjusted p-values, then decreasing signal reduction and number of peptides.
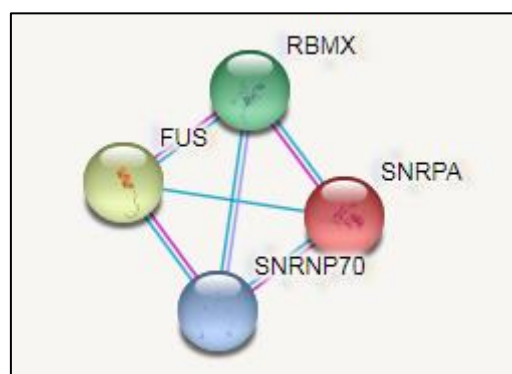

Figure 3.40: PPI network of SMN-C-interacting proteins on ESE2. The ten protein candidates identified as SMN-C-interacting proteins after affinity enrichment with the *SMN2* ESE2 RNA sequence as bait were loaded in the STRING DB using the highest confidence (0.900) in term of interaction score from curated databases (blue lines) as well as experimentally determined (pink lines).

The identification of RBMX, and not Tra2-ß1 or SRSF9, as specific interactor of SMN-C compound raised particular interest (Figure 3.41 and 3.42). This suggests a direct interaction of SMN-C at the RNA-interaction site of RBMX. However, it should be noted that although highly statistically relevant, the level of displacement upon increasing concentration of SMN-C was only of 29% (Figure 3.41). The observed signal reductions were actually below 60 % for all candidates (Table 3.7) indicating that the amount of free compound added prior to the ESE2 affinity purification was probably a limiting factor. Nevertheless direct binding of SMN-C active compound to RBMX (and not to Tra2-ß1) and its subsequent displacement from ESE2 could be confirmed via orthogonal methods, i.e. Surface Plasmon Resonance (SPR)-binding analyses and Nuclear Magnetic Resonance (NMR) chemical-shifts [109].
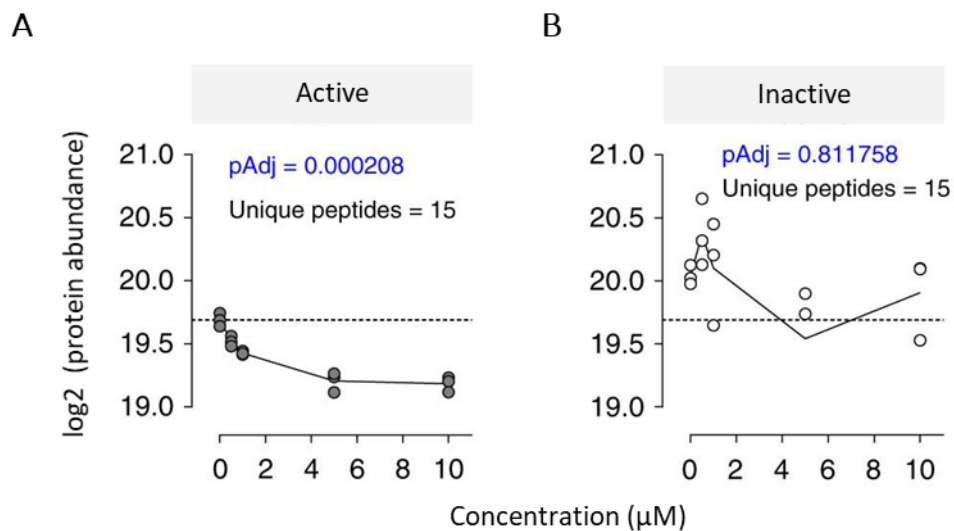
Figure 3.41: Displacement of RBMX from SMN2 exon 7 ESE2 by SMN-C splicing modifier. Protein abundance (log2) of RBMX after affinity enrichment with the SMN2 ESE2 RNA sequence as bait, in the presence of increasing concentrations of (A) Active SMN-C or (B) Inactive SMN-C ligand.
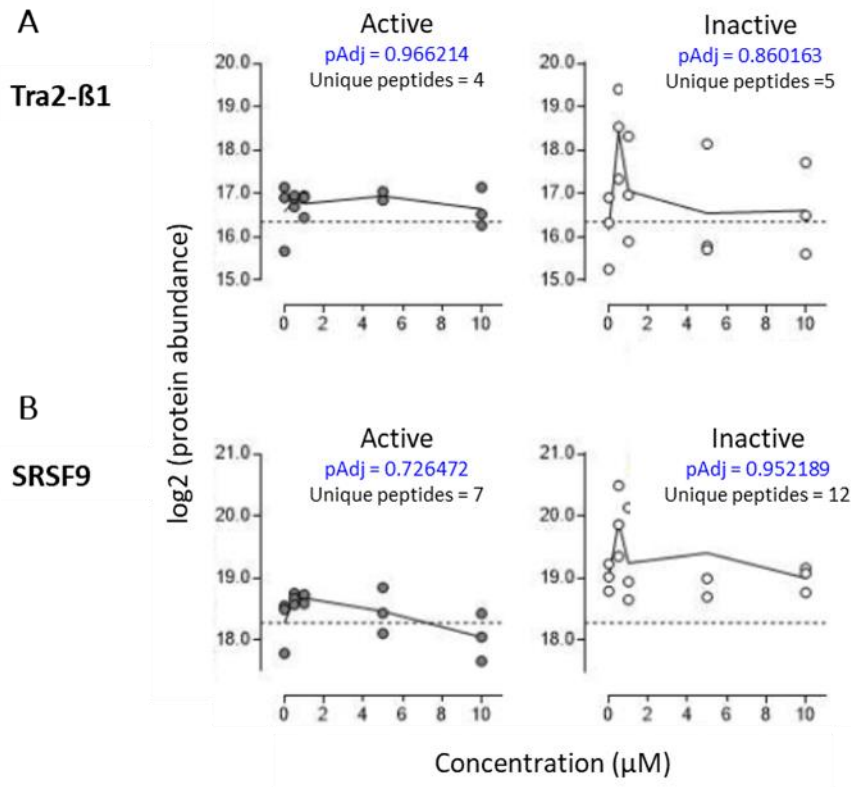
Figure 3.42: Affinity enrichment of Tra2-ß1 and SRSF9. Protein abundance (log2) of Tra2-ß1 (A) and SRSF9 (B) after affinity enrichment with the *SMN2* ESE2 RNA sequence as bait, in the presence of increasing concentrations of (A) Active SMN-C or (B) Inactive SMN-C ligand.

These results indicate that the exquisite specificity of SMN-C as positive regulator of *SMN2* exon 7 inclusion might be explained by its direct interaction to multiple trans-acting splicing factors. SMN-C might stabilize RBMX and U1 snRNP interaction to ESE2 or on contrary reorganize their position or even lead to recruitment from other proteins.

To summarize the results of this section, OCC-MS successfully dissected ESE2 interactome and contributed to a better understanding of *SMN2* splicing modifier mode of action. It is validating OCC-MS as a relevant approach to screen RNA-protein interactions. The specificity given by the competitive capture is allowing to deal with the high background of non-specific binders. Moreover, the conditions in which the capture is performed is allowing the capture of protein networks.

3.3. ICC-MS-derived platform to study protein interactions to single-stranded antisense oligonucleotides (ASOs)

Antisense oligonucleotides (ASOs) are synthetic polymers, usually between 15 to 20 nucleotides and their 3'-5' sequence is complementary to the sense sequence of a specific RNA transcript according to Watson-Crick base pairing.

Drug discovery approaches based on RNA targeting using ASOs leading to an altered synthesis of a particular protein, are highly promising. Antisense drugs bind (hybridize) to a target RNA forming a duplex further preventing proper production of disease-causing proteins via different approaches. The most widely used method is the activation of the endonuclease Ribonuclease H (RNase H) [114], but additional mechanisms having an impact on RNA splicing and translation machineries can be considered as well [115, 116].

The potential of antisense drugs has been illustrated in 2016 by the approval of Nusinersen which has been developed to treat SMA pathology [117].

To reach market approval, this approach has been evolving since the last three decades. Multiple ASO modifications led to improvement in nuclease resistance, increased half-life, higher target affinity, better distribution and cellular uptake [118]. However, there are still limitations due to severe side effects as toxic findings especially in the liver and the kidney.

Interactions of ASOs with proteins have been shown to influence their potency and cellular localization [119]. However, the overall impact of protein interactions on the efficacy, pharmacokinetics and safety of therapeutic ASOs is still poorly understood.

Profiling ASOs-protein interactions might lead to a better understanding of the role of ASO binding proteins and ultimately might help in designing safer ASOs. As for RNAs, ASOs are by nature highly charged and large, therefore a strategy to filter out non-specific binders could represent an interesting approach to identify relevant interacting proteins. Mapping ASOs protein interactome is surely even more challenging as many of them are having a phosphorothioate (PS) backbone (Figure 3.43) which is increasing their half-lives and cellular uptake [120] but which binds even more avidly to proteins than a phosphodiester. In addition, some ASOs are locked nucleic acid (LNA, Figure 3.43) with constrained ethyls increasing target affinity but which are highly hydrophobic 2' modification binding also many more

proteins. Some ASOs are conjugated to N-acetylgalactosamine (GalNAc, Figure 3.43) moieties which is allowing a liver-targeted uptake via the binding to the asialoglycoprotein receptor (ASGPR) expressed specifically and at a high copy number on the surface of hepatocytes (e.g. for diseases such as hepatitis B) [121-123].
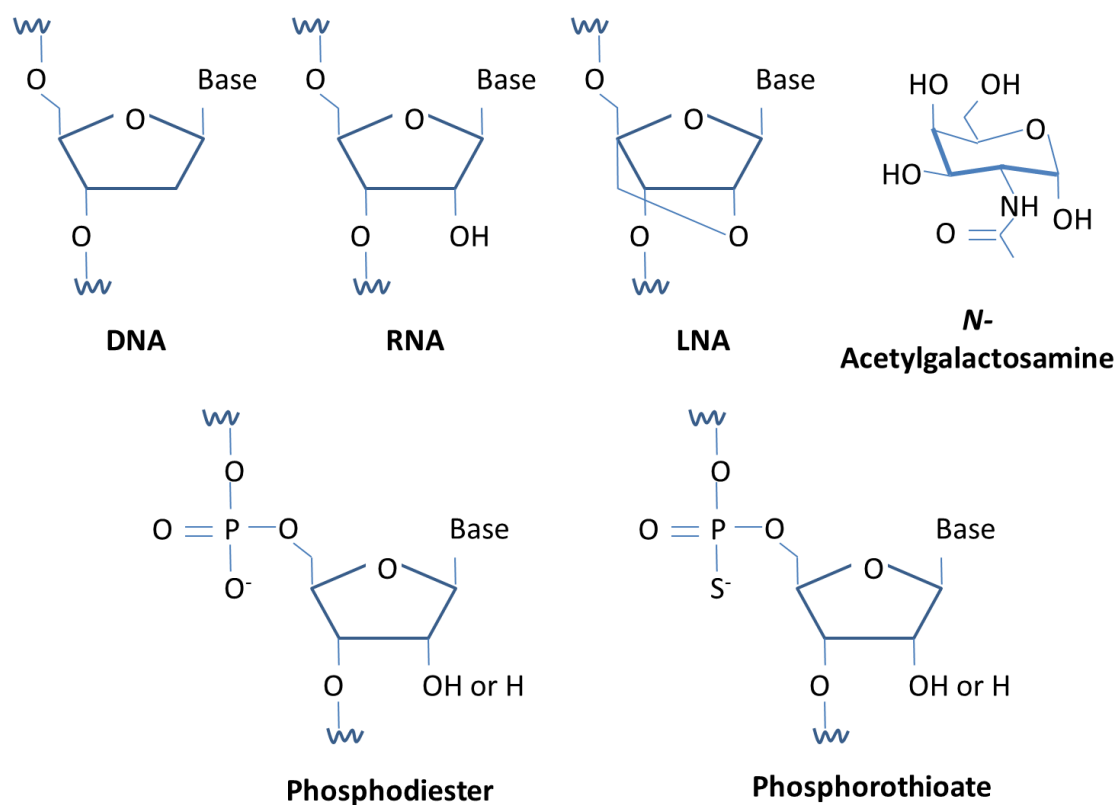
Figure 3.43: Common ASOs chemical modifications. ASOs are single-stranded sequences of DNA with a phosphate backbone and sugar rings. In ASOs backbone, non-bridging phosphodiester oxygen is often replaced by a sulfur (phosphorothioate) to increase stability toward nucleases. Locked nucleic acids (LNAs) are having a 2', 4'-methylene linkage which is locking the ribose in a favorable conformation for interacting with RNA target. Coupling ASOs to N-acetylgalactosamine (GalNAc, usually a tricluster), an amino sugar derivative of galactose, is often used when a mediated delivery to hepatocytes which abundantly express its receptor (asialoglycoprotein receptor, ASGR) is required.

The overall technical feasibility was first assessed in mouse tissue extracts (liver and kidney). Then the relation between protein binding profiles and ASO safety assessment was explored *in vitro* in cellular models designed to mimic *in vivo* nephrotoxicity (human renal proximal tubule epithelial cells, PTECs) and hepatotoxicity (mouse 3T3 fibroblast cells).

3.3.1 Phosphorothioated ASO interactome in mouse liver and kidney tissue extracts

To evaluate whether an OCC-MS approach is suitable to decipher the interactome of large and negatively charged ASOs, an assessment in tissue extracts from organs that are typical targets of toxicity (i.e. liver and kidney) have been performed.

The tested ASO (ASO-A) consisted of a LNA gapmer (chemically modified wings and a DNA gap in-between), arbitrarily chosen from a library available at Roche Innovation Center Basel (Table 3.8). Three nucleotides on each end were LNA modified and the backbone was fully phosphorothioated. A 5' biotinylated version was provided by Erich Koller (Roche Innovation Center Basel) and Susanne Kammler (Roche Innovation Center Copenhagen) for immobilization on streptavidin magnetic beads. A hexa-ethylene glycol (HEG) spacer arm was incorporated to minimize steric hindrance between the biotin moiety and the oligonucleotide upon attachment to magnetic beads.

ASO	Target	Conjugation	3' Biotinylation	Sequence (5' to 3')
ASO-A	VEGF	none (naked)	No	TmCTcctcttccttmCAT
		none (naked)	Yes	5'-BioHEG-TmCTcctcttccttmCAT

Table 3.8: ASO-A: sequences and backbone modifications. Single strand oligonucleotide sequence is indicated in blue font. Uppercase: Wing segments of nucleotides with LNA sugar modifications. ^mC: 5-methylcytidine. Lower case: Gap segment of 2'-deoxynucleotides. Internucleotide linkages are phosphorothioated. BioTEG: biotin extended with a 15 atom triethylene glycol spacer arm.

Livers from C57 BlbcJ mice were mechanically homogenized in a PBS buffer 1% NP40 with protease inhibitor. Samples were sonicated and centrifuged at 16'000 g to remove insoluble material. After adding RNase inhibitor, 180 µg of protein extract was incubated with different quantities of ASO-A (0, 50, 100, 500 and 1000 pmol in 100 µl in triplicates) prior to capture with immobilized biotinylated ASO-A. Bound proteins were eluted with SDS sample buffer and loaded for a short SDS-PAGE to eliminate SDS that is LC-MS incompatible and would have denatured the trypsin enzyme used for proteins cleavage. Samples were cut as one band, in gel-digested and processed for LC-MS/MS analysis.

Due to the inherent nature of LNAs, it was expected to identify a high number of proteins in the ASO-A bound fraction and indeed 1792 proteins were identified and quantified (1325 with at least two peptides). A Principal Component Analysis (PCA) was performed on the normalized protein level data to identify the linear combination of observed variables having maximum variance. This revealed that increasing the concentration of free ASO-A was the main contributor to the observed variability (Figure 3.44 A). Additionally, dose-response relationships were extracted and visualized on a scatter plot of signal reduction versus pAdj values (Figure 3.44 B). On such a plot, interesting candidates having a high chance to be specific binders of ASO-A are located in the upper right corner (low pAdj value and high signal reduction). In line with the previous PCA observations, many proteins were covering both criteria, i.e. their capture on immobilized ASO-A was significantly affected by free ASO-A in a dose-dependent manner. Also it was not possible to set a strict and clear cut-off pAdj value to discriminate specific versus non-specific binders. After visual inspection, it appeared that roughly 10% of the proteins identified (> 170 proteins) were displaying a convincing displacement profile. Those proteins can be either direct interactors of the immobilized ASO-A or indirect (part of a complex of proteins). Despite the high number of potential ASO-A binder candidates, the approach still efficiently filtered out a large proportion of background proteins which were not affected by the presence of the competitor.
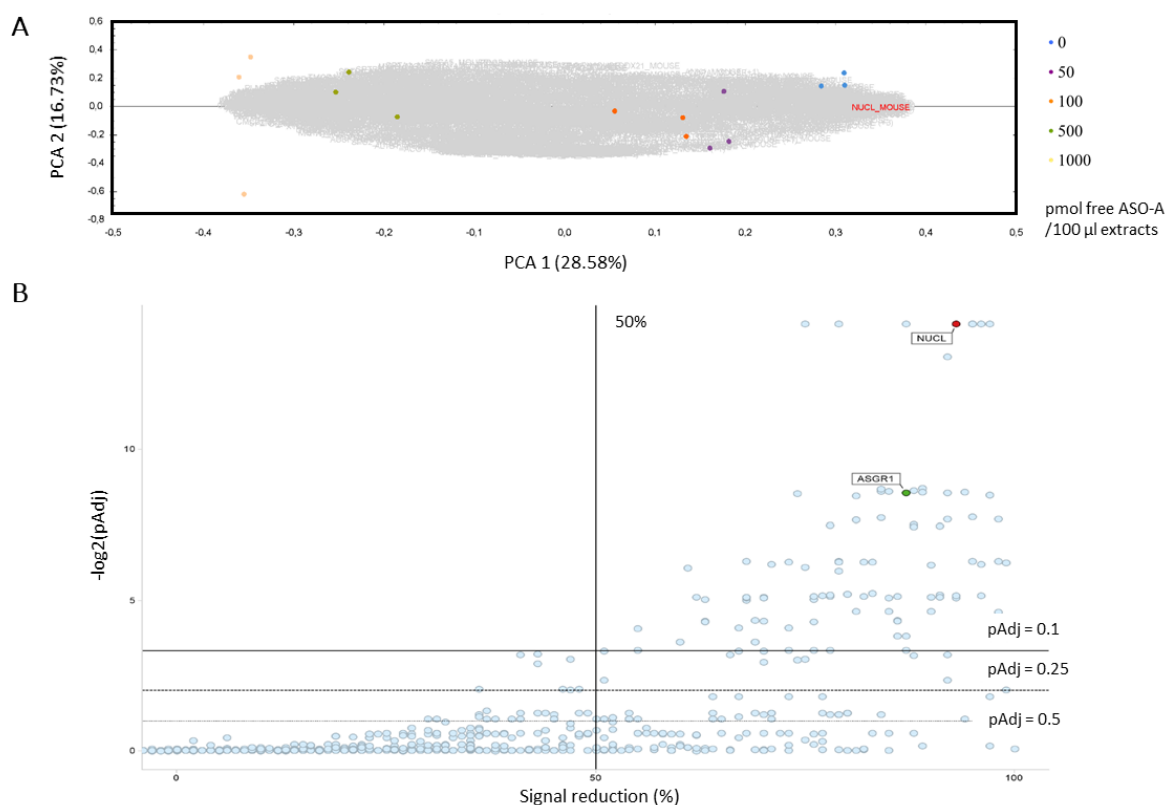
Figure 3.44: PCA and scatter plot (signal reduction versus pAdj values) after applying OCC-MS to ASO-A in mouse liver extracts. 1792 protein groups were identified & quantified after applying an OCC-MS approach to ASO-A. A linear model was fit using non-monotonic contrasts (Hirotsu contrasts). P values from the tests were adjusted for multiple testing by permutations. (A) PCA on the normalized protein level data. Blue dots: 0 pmol; purple dots: 50 pmol, orange dots: 100 pmol, green dots: 500 pmol, yellow dots: 1000 pmol, of ASO-A in 100 µl (in triplicates). (B) Scatter plot displaying the signal reduction (in %) versus the adjusted p value (-log 2 transformed). pAdj: p value adjusted by the Westfall-Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free ASO-A. In red, the protein with the lowest pAdj value (nucleolin, NUCL_MOUSE, 5.4e-05) and in green asialoglycoprotein receptor 1 (ASGR1_MOUSE). Horizontal lines at pAdj value 0.1, 0.25 and 0.5 and vertical line at 50% signal reduction.

The protein with the lowest pAdj value (5.4e-05) was the abundant nucleolar protein nucleolin (NUCL, signal reduction of 93%) known to have 4 consensus RNA-binding domains called RNA recognition motifs (RRM) [124]. Nucleolin was previously described as a binder of PS-ASOs and nucleolin-specific aptamer-mediated delivery of ASOs has been evaluated to increase the activity of splicing-modulation ASOs in different cancer cells [119, 125]. ASGR1 has also been identified as a specific binder (with a pAdj value of 0.002648 and a signal reduction of 87%, Figure 3.45) which was surprising as ASO-A was not GalNac modified.

Nevertheless, this finding was in agreement with recent published data indicating that even unconjugated phosphorothioate containing ASOs may interact with ASGR1 which may play a small but significant role in their uptake into hepatocytes and their activity (Tanowitz, 2017).
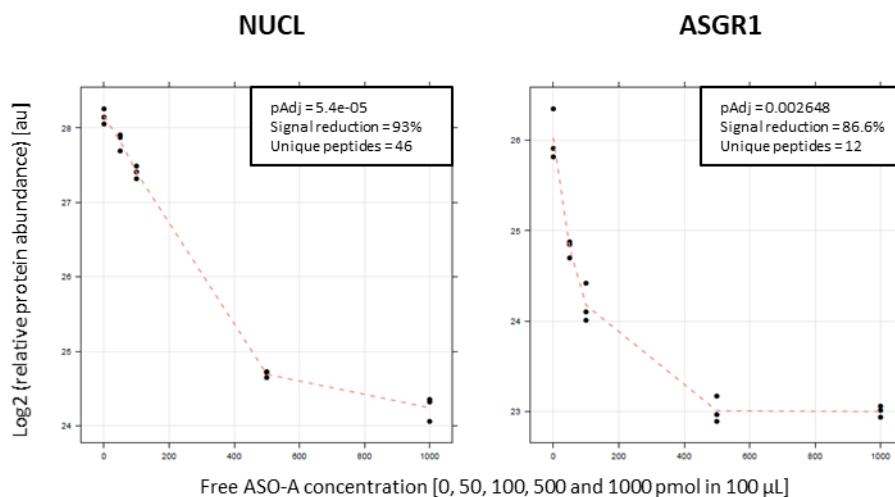

Figure 3.45: Displacement of NUCL and ASGR1 from ASO-A by ASO-A. Protein abundance (log2) of NUCL and ASGR1 after affinity enrichment in mouse liver extracts (180 µg) with ASO-A as bait, in the presence of increasing concentrations of free ASO-A (0, 50, 100, 500 and 1000 pmol of free ASO-A in 100 µl in triplicates). Proteins from bound fractions were identified by LC-MS/MS and MS signals were subjected to a statistical analysis to derive concentration-dependent signal decrease of specific interactors.

Next was investigated whether decreasing the amount of tissue extracts (from 180 µg to 100 µg) and shifting the concentrations of free ASO-A to a higher range (0, 100, 500, 1000 and 2000 pmol in 100 µl) improve the discrimination of specific from non-specific binders. In addition, kidney extracts were processed in parallel to evaluate the overlap of the findings when profiling different tissues. As for the first trial in liver, a high number of proteins were identified and quantified in the ASO-A bound fractions (2005 and 2020 proteins respectively in the liver and in the kidney extracts). The same observations were made after PCA and after plotting the signal reduction versus pAdj values than in the first experiment (Figure 3.46) with a clear grouping of the samples based on concentrations of competitor (being the principle axe of abundance variation) and an ambiguous cut-off between specific and non-specific binders. Thus decreasing the amount of starting material did not have an effect on the number of background proteins identified and likewise, a similar percentage (around 10%) of the proteins after visual inspection were having a displacement profile. Nucleolin was again among the 10 first candidates in both extracts (Figure 3.47). ASGR1 specific binding to ASO-A observed in

the first trial in liver was confirmed in the second liver experiment but of course not in the kidney as ASGR1 is exclusively expressed in liver hepatocytes.
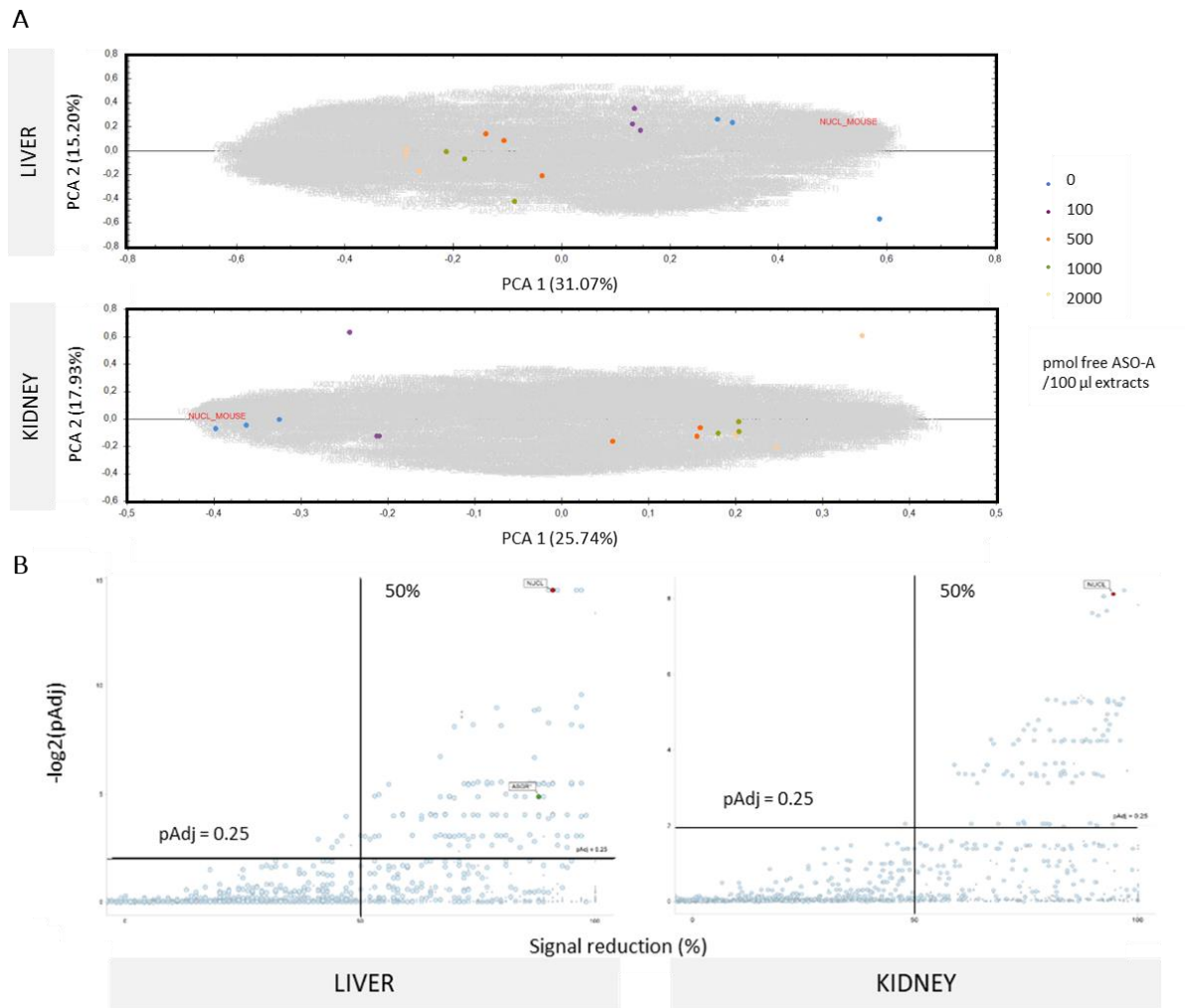

Figure 3.46: PCA and scatter plots (signal reduction versus pAdj values) after applying OCC-MS to ASO-A in mouse liver and kidney extracts. 100 µg of mouse liver or kidney protein extracts was incubated with different quantities of ASO-A (0, 100, 500, 1000 and 2000 pmol in 100 µl in triplicates) prior to captures with immobilized biotinylated ASO-A. Bound proteins were eluted with SDS sample buffer and loaded for a short SDS-PAGE to eliminate SDS. Samples were cut as one band, in gel-digested and processed for LC-MS/MS analysis. 2005 protein groups were identified & quantified in liver extracts and 2020 in kidney extracts. A linear model was fit using non-monotonic contrasts (Hirotsu contrasts). P values from the tests were adjusted for multiple testing by permutations. (A) PCA on the normalized protein level data. Blue: 0 pmol, purple: 100 pmol, orange: 500 pmol, green: 1000 pmol and yellow: 2000 pmol of ASO-A in 100 µl (in triplicates). (B) Scatter plot displaying the signal reduction versus the adjusted p value. pAdj: p value adjusted by the Westfall-Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free antibody. Proteins identified with only one unique peptide depicted with a smaller circle.
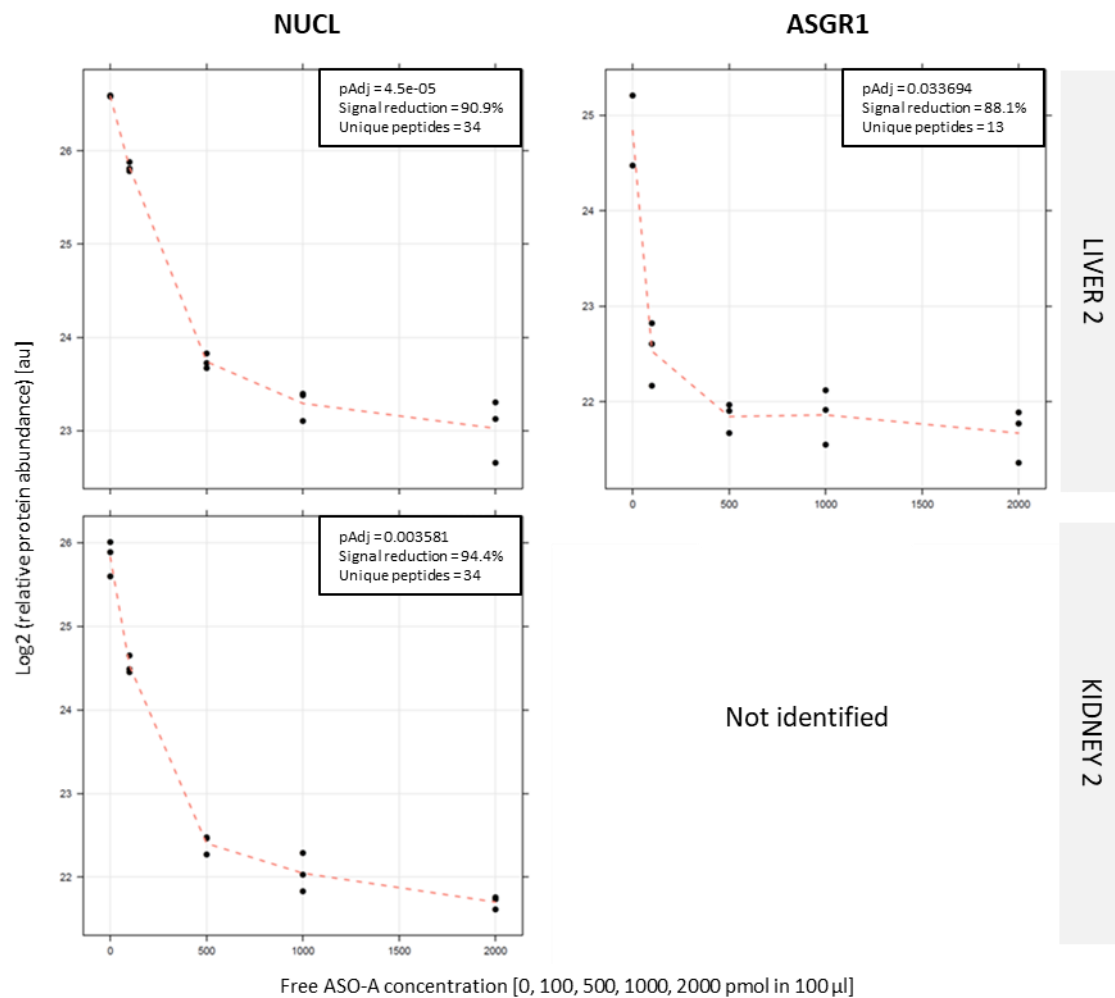
Figure 3.47: Displacement of NUCL and ASGR1 from ASO-A by ASO-A in liver and kidney extracts. Protein abundance (log2) of NUCL and ASGR1 after affinity enrichment in mouse liver and kidney extracts (100 µg) with ASO-A as bait, in the presence of increasing concentrations of free ASO-A (0, 100, 500, 1000 and 2000 pmol of free ASO-A in 100 µl in triplicates). Proteins from bound fractions were identified by LC-MS/MS and MS signals were subjected to a statistical analysis to derive concentration-dependent signal decrease of specific interactors.

To check whether the same proteins were identified as potential ASO-A binders, the overlap among the first 200 proteins ranked by increasing pAdj values was explored (Figure 3.48).
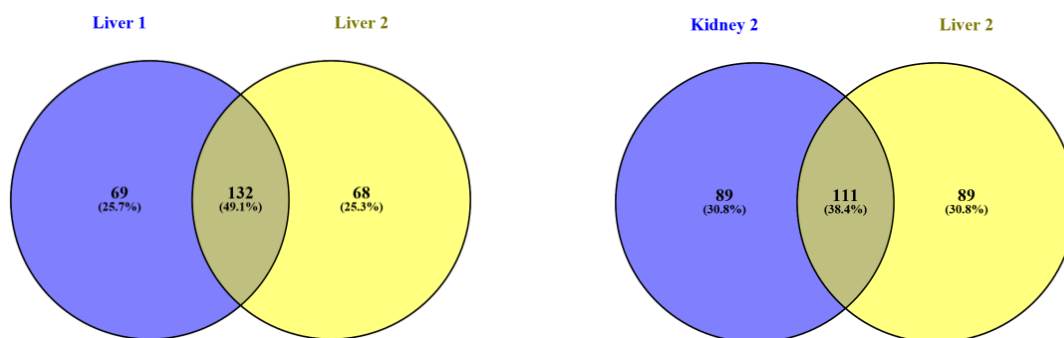
Figure 3.48: Comparison of the 200 first hits extracted from OCC-MS applied to ASO-A. The list of the 200 first hits when ranking with increasing pAdj values in each experiment was compared using the tool Venny 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/).

The overlap between the two liver profiling was around 50% (same extracts but different experimental conditions) and close to 40% between liver and kidney (different extracts but same experimental conditions). The percentage of similarity was not higher when restricted to the first 50 candidates (data not shown). The limited reproducibility observed between the two liver experiments was highlighting the importance of the starting conditions (amount of protein extracts, level of expression of each individual proteins and range of competitor concentrations) and a need for multiple biological replicates to increase the chance to have a manageable number of potential candidates (i.e. to identify the "core" proteome). In such an untargeted approach compared to the applications previously described in chapter 1 and 2, there were no positive controls (i.e. a known interactor) to optimize the experimental conditions. Despite this constraint, the approach was further evaluated to compare ASOs with distinct *in vitro* toxicity prediction in cellular models.

3.3.2 Application to GalNac and naked ASOs with distinct *in vitro* nephrotoxicity prediction

In vitro screening strategies are critical in drug development to predict risk of toxicity and for selection of the most appropriate therapeutic candidate. One of the major safety liability of nucleic acid therapeutics is related to ASOs accumulation in kidney tubules and there is a big need for *in vitro* models that would mimic *in vivo* nephrotoxicity. In 2017 Annie Moisan *et al.* published an *in vitro* screening strategy using renal proximal tubular epithelial cells (PTEC TERT1) to investigate potential risk of nephrotoxicity [126]. In such a screen, ASOs inducing a decrease in intracellular ATP, an accumulation of epidermal growth factor (EGF) in PTEC supernatants (inhibition of uptake) and morphological changes are highlighted as having a risk of safety liability. Interestingly, the team recently reported that GalNac conjugation (ASOs

linked at the 5' terminus to three GalNac molecules via a triantennary linker, triGalNac) was having as an effect to mask *in vivo* nephrotoxicity potential in their assay [45]. Sabine Sewing *et al.* showed that this effect was not due to a poor uptake and lower intracellular content nor explained by reduced target gene knockdown. In addition, it has been shown that free triGalNac by itself was not masking ASOs cytotoxic effects [45].

Thus comparing the protein interactome of GalNac and naked versions might help to better understand the protective effect of GalNac conjugation in renal tubular cells *in vitro*.

ASOs having known *in vivo* kidney liabilities but with an inert or attenuated effect of their GalNac versions in PTECs were selected, i.e. anti-*PCSK9* ASO-B and ASO-C, respectively moderately and highly nephrotoxic in rat (Table 3.9) [45]. Biotinylated versions were then generated by Adrian Schäublin (Roche Innovation Center Basel) at the 3' end since 5' labeling would not have been possible for GalNac sequences.

ASO	Target	Conjugation	3' Biotinylation	Sequence (5' to 3')
B	*PCSK9*	none (naked)	No	TG^{m}Ctacaaaac^{m}C^{m}CA
		GalNac	No	5'-GN-C6caTG^{m}Ctacaaaac^{m}C^{m}CA
		none (naked)	Yes	TG^{m}Ctacaaaac^{m}C^{m}CA-BioTEG-3'
		GalNac	Yes	5'-GN-C6caTG^{m}Ctacaaaac^{m}C^{m}CA-BioTEG-3'
C	*PCSK9*	none (naked)	No	G^{m}CtgtgtgagcttGG
		GalNac	No	5'-GN-C6G^{m}CtgtgtgagcttGG
		none (naked)	Yes	G^{m}CtgtgtgagcttGG-BioTEG-3'
		GalNac	Yes	5'-GN-C6G^{m}CtgtgtgagcttGG-BioTEG-3'

Table 3.9: List of *PCSK9* ASOs with corresponding sequences and backbone modifications. Single strand oligonucleotide sequence is indicated in blue font. Uppercase: Wing segments of nucleotides with locked nucleic acid (LNA) sugar modifications. ^{m}C: 5-methylcytidine. Lower case: Gap segment of 2'-deoxynucleotides. Internucleotide linkages are phosphorothioated. 5'-GN: trivalent N-acetylgalactosamine (GalNAc) cluster. C6/C6ca: triantennary linker. BioTEG: biotin extended with a 15 atom triethylene glycol spacer arm.

For the first experiment, 2E+07 PTEC cells were resupended in 1% NP40 lysis buffer containing protease and RNase inhibitors. After centrifugation to remove insoluble material, supernatants (90 µg total protein each) were spiked with increasing concentrations (0, 100, 500, 1000, 2000 pmol in 130 µl total volume) of either naked or GalNac ASO B in triplicates.

Extracts were then loaded on their respective biotinylated ASO (50 µl of magnetic beads coated with 100 pmol of naked or GalNac biotinylated ASO).

From the pilot experiment in tissue extracts, a high background and variability was expected with no clear cut-off to separate specific from non-specific binders. Thus the whole experiment was repeated twice to increase confidence in the findings, resulting in a total of 60 individual bound fractions (five concentrations in triplicates, GalNac or Naked, in two biological replicates). Fractions were loaded for a short SDS-PAGE to eliminate SDS, were cut as one band, in gel-digested and processed for LC-MS/MS analysis.

Around 1'600 proteins in total have been identified in each bound fractions. At the level of protein identification, a similarity around 70% was observed between the replicates as well as between naked and GalNac conditions indicating overall a comparable background of proteins (data not shown). PCA plots as for the pilot in liver extracts were clearly indicating a strong effect on the dataset of increasing concentrations of free ASOs (Figure 3.49 A). Likewise, using pAdj values together with the signal reduction to distinguish specific versus non-specific binders did not allow identifying a clear cut-off (Figure 3.49 B).
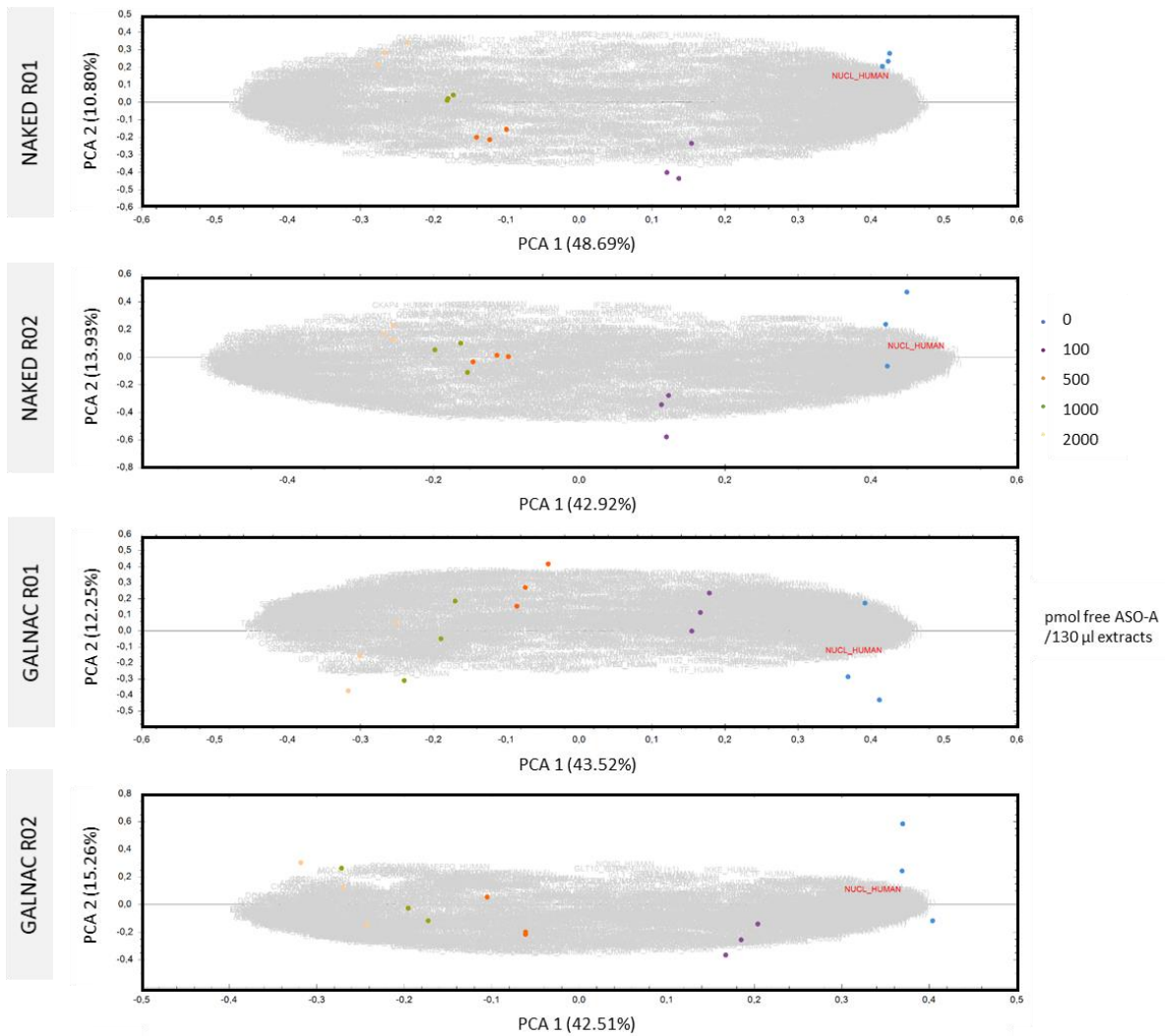
Figure 3.49: PCA and scatter plots (signal reduction versus pAdj values) after applying OCC-MS to naked and GalNac ASO-B in PTECs extracts. 90 µg of PTEC TERT1 protein extract was incubated with different quantities of naked or GalNac ASO B (0, 100, 500, 1000, 2000 pmol in 130 µl in triplicates with two biological replicates) prior to captures with immobilized biotinylated ASO B, respectively naked or GalNac. Bound proteins were eluted with SDS sample buffer and loaded for a short SDS-PAGE to eliminate SDS. Samples were cut as one band, in gel-digested and processed for LC-MS/MS analysis. Around 1'600 protein groups were identified & quantified in each experiments. A linear model was fit using non-monotonic contrasts (Hirotsu contrasts). P values from the tests were adjusted for multiple testing by permutations. (A) PCA on the normalized protein level data; Blue: 0, purple: 100, orange: 500, green; 1000 and yellow: 2000 pmol in 130 µl in triplicates.
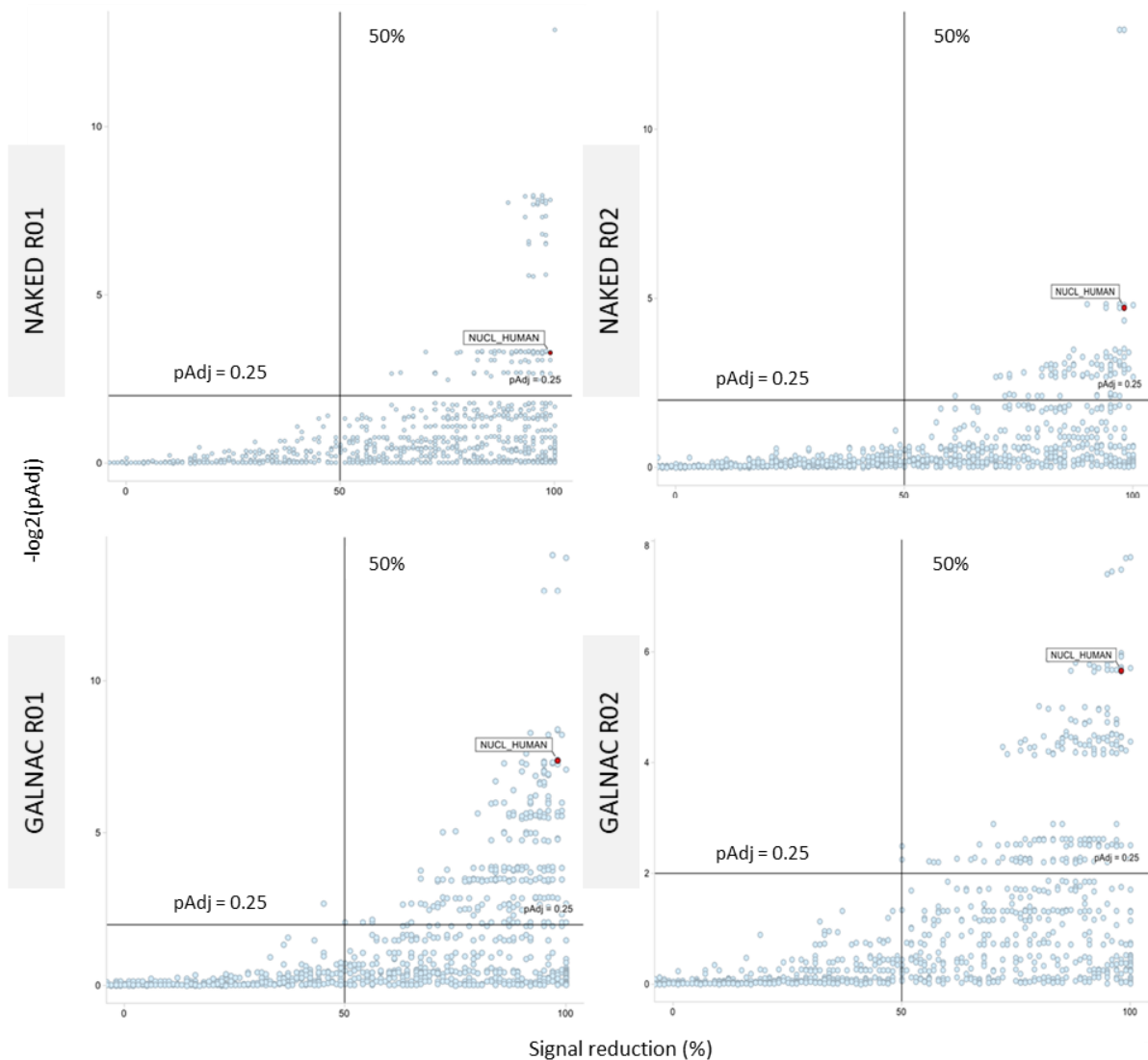
(B) Scatter plot displaying the signal reduction versus the adjusted p value. pAdj: p value adjusted by the Westfall-Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free antibody.Young approach; Signal reduction: Signal difference between the highest and lowest concentration of free antibody.

After visual inspection, again at least 10% of the identified proteins were having a displacement profile upon increasing concentration of their free ASO either naked or GalNac modified.

Plotting the adjusted p values between replicates indicated a limited reproducibility (Figure 3.50). A few proteins (such as nucleolin) were identified repeatedly as specific binders but most of the proteins were highlighted as binders in only one replicate.
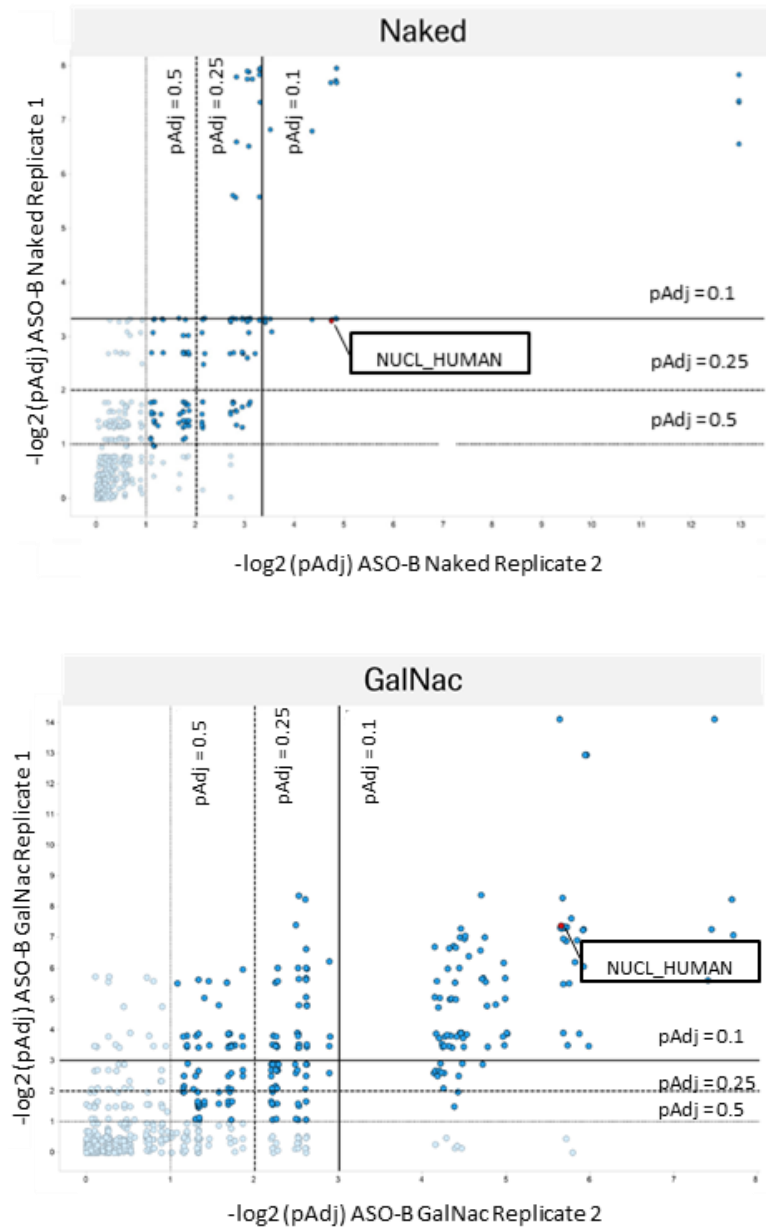
Figure 3.50: Comparison of pAdj values distribution between replicates after OCC-MS applied to ASO-B in PTECs extracts (GalNac and naked). Scatter plots displaying the adjusted p value (pAdj) among replicates (GalNac and naked). pAdj: p value adjusted by the Westfall-Young approach. Lines arbitrarily set at 3.32 (-log$_2$ pAdj 0.1), 2 (-log$_2$ pAdj 0.25) and 1 (-log$_2$ pAdj 0.5). Proteins identified with at least two peptides in one replicate. Red dot: nucleolin, NUCL_HUMAN.

Keeping as a postulate that the proteins with the lower p-values were having the higher chance to be direct binders, a comparison of GalNac and naked datasets using this parameter was performed (Figure 3.51).
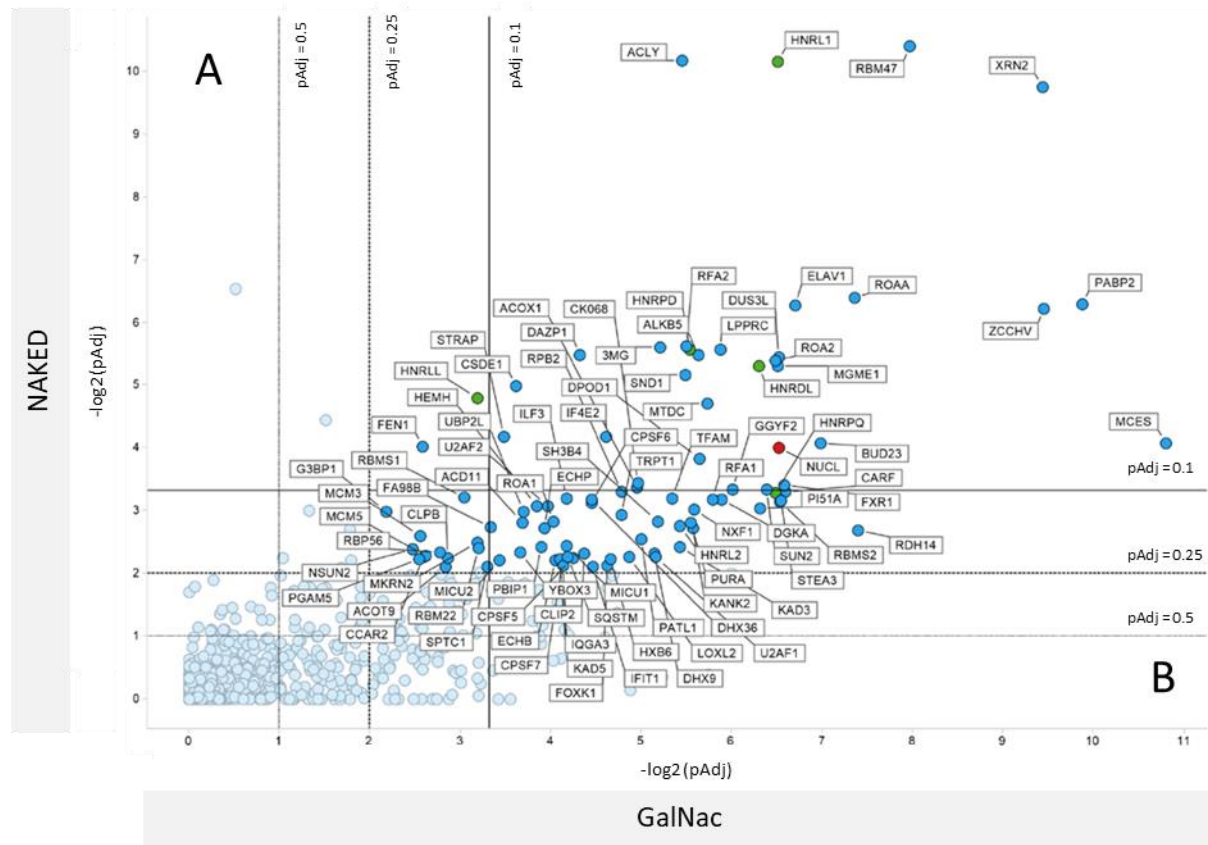
Figure 3.51: Comparison of pAdj values distribution between GalNac ASO-B and naked ASO-B after OCC-MS in PTECs extracts. Scatter plot displaying the adjusted p value (pAdj) GalNac versus naked (average of the replicates). pAdj: p value adjusted by the Westfall-Young approach. Lines arbitrarily set at 3.32 (-log2 pAdj 0.1), 2 (-log2 pAdj 0.25) and 1 (-log2 pAdj 0.5). Proteins identified with at least two peptides in one replicate. Red dot: nucleolin, NUCL_HUMAN. Green dots: hNRPs. Quadrant A: lower pAdj values in naked only; Quadrant B: lower pAdj values in GalNac only.

Many proteins are binding to ASO-B irrespectively of having or not a GalNac moiety attached. As an example and similarly to the pilot in liver extracts, the protein nucleolin was found as a specific binder in both datasets. The majority of the common proteins identified, were part of biological processes such as transcription regulation, mRNA processing and DNA damage. A couple of heterogeneous nuclear ribonucleoproteins (hNRPs, in green in Figure 3.51) known to interact with each other were identified indicating potentially the capture of a complex of proteins. Only a few proteins were showing a more specific binding to the naked (Quadrant A

Figure 3.51, Table 3.10) although many proteins were showing a preference towards the GalNac version (Quadrant B Figure 3.51, Table 3.10). This observation was not surprising as GalNac molecules are bigger and thus more prone to interact with proteins and could explain the masking of *in vitro* nephrotoxic prediction in PTECs.

Protein name	ASO-B GalNac	ASO-B Naked	Protein name	ASO-B GalNac	ASO-B Naked
PUS7	0.025993756	0.767446321	HABP4	0.396731898	0.861334662
FMNL3	0.030789717	0.791636935	TIAR	0.424610522	0.863232092
RUSD2	0.038175699	0.986863407	RO60	0.428867055	0.846032941
DUS1L	0.094450158	1	SP1	0.429094213	0.96663205
HNF1B	0.106641084	1	PAPS2	0.434796437	0.997307229
ERI1	0.11197637	0.770312739	P5CR3	0.435441583	0.982479351
FMNL2	0.121231469	0.960448636	TRIPC	0.43613775	0.980597667
RNH1	0.133946117	0.753093267	RUSD3	0.436848484	0.949582006
BCKD	0.143279302	0.840220835	SRSF3	0.440895299	0.775948391
CHST3	0.148970855	1	MGMT	0.452513995	0.789580287
FARP2	0.150863945	1	RRP44	0.45290776	0.971164346
AKP8L	0.156750204	1	TRM61	0.453054514	0.884105035
PTH	0.201197007	0.754144751	NEUA	0.454683539	0.998037844
FABD	0.235264991	0.75336949	DEFM	0.455002796	0.96008043
VANG1	0.23822981	0.796426056	QOR	0.465968042	0.964974736
BORG5	0.243228886	0.88452288	KCY	0.473562586	0.866010465
CMC1	0.260715007	0.857214176	ABCD1	0.476506682	0.983687816
E41L2	0.275067274	0.928365882	RL7	0.725880259	0.374086823
IQGA1	0.308216134	0.860036938	KCC2D	0.732048198	0.408763376
AKAP8	0.31829988	0.808031669	RFIP5	0.761760333	0.579127138
UHRF2	0.318648969	0.929145181	RBM7	0.762117851	0.4633264
MTCH1	0.324125559	0.853319294	RLA0	0.772976908	0.468950034
TSC2	0.333735549	0.952646557	GMPPA	0.778023036	0.384222943
RPC1	0.334173034	1	RACK1	0.78027272	0.515431957
L2GL1	0.334715079	0.8872133	FUBP3	0.816188193	0.513867589
HS2ST	0.35653925	0.809942136	CCDC8	0.834221443	0.310668174
KPCI	0.366360463	1	PDIP3	0.834318334	0.56032769
SYTL2	0.366461432	1	CE170	0.848888035	0.549834185
AGK	0.366746201	1	PCBP3	0.863604082	0.51547017
RPOM	0.374920893	0.90019794	LYRIC	0.937093996	0.487020203
RT02	0.377010255	0.78801492	GTF2I	0.974074923	0.572198296
JUND	0.382865472	1	BCD1	1	0.559013806
TRM1L	0.384869455	0.823504497		pAdj naked < pAdj GalNac	

Table 3.10: List of proteins having a distinct binding profile to ASO-B GalNac versus ASO-B naked. Ranked by increasing ASO-B GalNac pAdj value. Lower pAdj appear in green and higher in red. pAdj: p value adjusted by the Westfall-Young approach (average of the replicates).

Next OCC-MS was applied to ASO-C in PTECs extracts and proteins from Table 3.10 were mapped on the obtained scatter plot of signal reduction versus pAdj values (Figure 3.52 A and B).
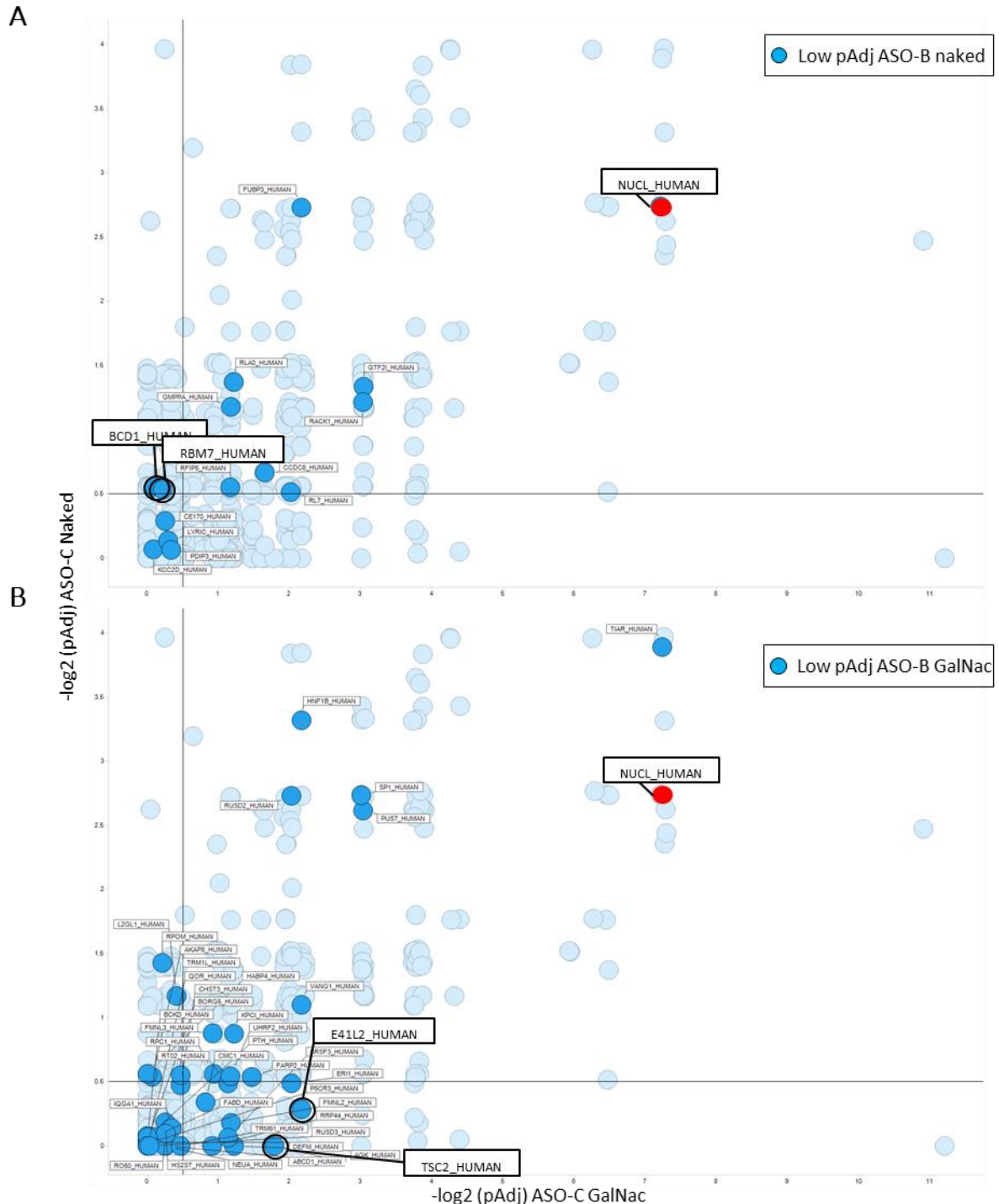

Figure 3.52: Comparison of pAdj values distribution between GalNac ASO-C and naked ASO-C after OCC-MS in PTECs extracts. Scatter plots displaying the adjusted p value (pAdj) of GalNac and naked ASO-C. pAdj: p value adjusted by the Westfall-Young approach. Lines arbitrarily set at –log 2

of 0.5 (pAdj ~ 70%). Proteins identified with at least two peptides in one replicate. (A) Proteins highlighted in dark blue: proteins previously identified with a lower pAdj for naked ASO-B versus GalNac ASO-B. (B) Proteins highlighted in dark blue: proteins previously identified with a lower pAdj for GalNac ASO-B versus naked ASO-B. Nucleolin (NUCL_HUMAN) in red as a reference. Proteins associated to lower pAdj values for both ASO-B and ASO-C either naked (A) or GalNac (B) circled in black.

Again a higher number of binders were observed for the GalNac version. ASO-B "naked specific" proteins were either similarly captured by ASO-C naked and GalNac or not captured at all. Only Box C/D snoRNA protein 1 (BCD1) and RNA-binding protein 7 (RBM7) were having a slightly better displacement profile for naked ASO-C compared to GalNac ASO-C (Figure 3.53 A).

ASO-B "GalNac specific" proteins could be found in all three quadrants, the two most confident candidates being band 4.1-like protein 2 (E41L2) and Tuberin (TSC2) (Figure 3.53 B).

B

E41L2   TSC2

Free ASO concentration [0, 100, 500, 1000, 2000 pmol in 130 µl]
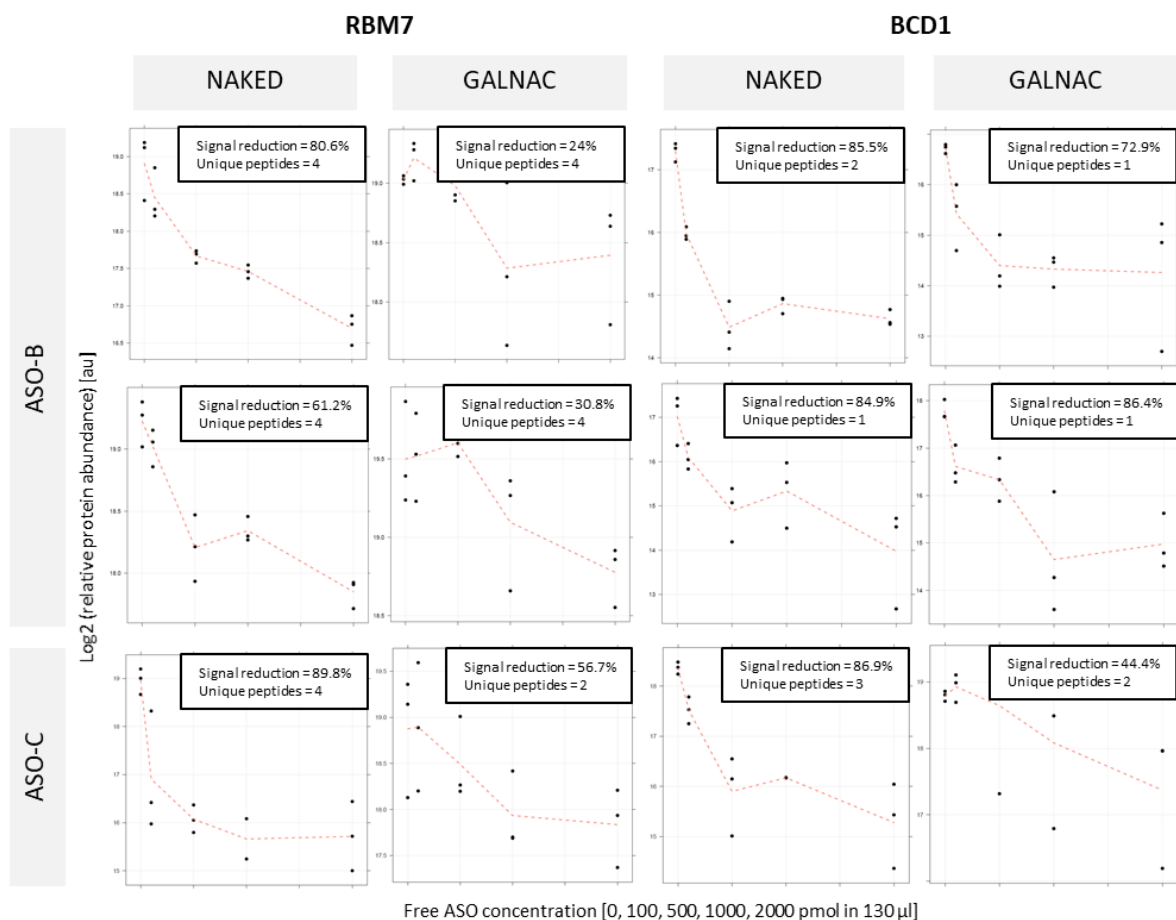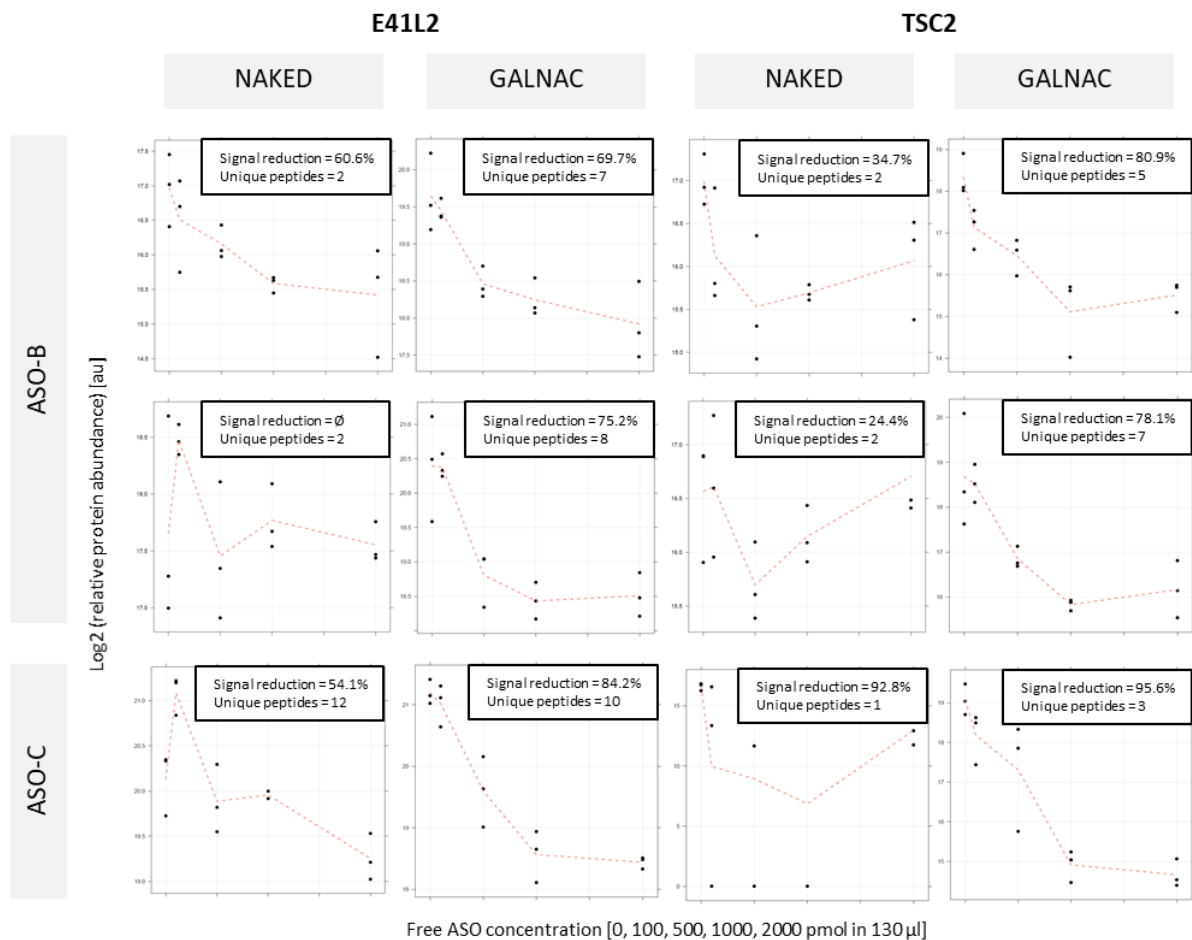
Figure 3.53: Displacement profiles of RBM7, BCD1, E41L2 and TSC2. Protein abundance (log2) of (A) RBM7, BCD1 (B) E41L2 and TSC2 after affinity enrichment in PTECs (90 µg) with ASO-B or ASO-C as bait, in the presence of increasing concentrations of free ASO-B or ASO-C (0, 100, 500, 1000, 2000 pmol in 130 µl in triplicates). Proteins from bound fractions were identified by LC-MS/MS and MS signals were subjected to a statistical analysis to derive concentration-dependent signal decrease of specific interactors.

Taking into account the variability observed in the different trials, none were considered confident enough for further follow-up. The masking effect of GalNac ASO seemed to be rather driven by the extent of binding (number of binders) than by particular proteins. Consequently, naked versions of ASO should always be used for *in vitro* toxicity prediction.

In a final attempt to decipher sequence-independent binding of ASOs to cellular proteins which might interfere with their biological function and thus may have safety implications, a comparison was made between two additional ASOs having again distinct *in vitro* profiles but differing by a much minor modification than GalNac.

3.3.3 Application to ASOs with distinct *in vitro* hepatoxicity profiles

LNA modifications can improve the potency of ASOs but are often associated with higher risk of hepatotoxicity in animals [127]. Potential hepatic liabilities can be predicted *in vitro* in order to de-risk and prioritize ASO candidates. Andreas Dieckmann *et al.* published recently a new cellular model using transfected mouse 3T3 fibroblasts [128]. By delivering ASOs via lipotransfection, a clear association between *in vivo* hepatotoxicity in mouse and apoptosis in cells was shown.

In addition Andreas Dieckmann, Yvonne Burki and Grzegorz Sarek (Roche Innovation Center Basel) observed that the insertion of a minor chemical modification into ASO-D hepatotoxic oligonucleotide (Table 3.11, and referred as LNA41 in [128]), strongly suppresses toxicity *in vitro* (Figure 3.54).

ASO	Target	Sequence (5' to 3')
ASO-D	*Myd88*	CACattccttgctCTG
Modified ASO-D (MOD ASO-D)	*Myd88*	CACattccttgctCTG
5' Biotinylated –ASO-D (BIO ASO-D)	*Myd88*	BioTEG-CACattccttgctCTG

Table 3.11: ASO-D sequences and backbone modifications. Single strand oligonucleotide sequence is indicated in blue font. Uppercase: Wing segments of nucleotides with locked nucleic acid (LNA) sugar modifications. Lower case: Gap segment of 2'-deoxynucleotides. Internucleotide linkages are all phosphorothioated. BioTEG: biotin extended with a 15 atom triethylene glycol spacer arm. MOD : modification leading to reduced caspase activity in mouse 3T3 fibroblast cellular assay.
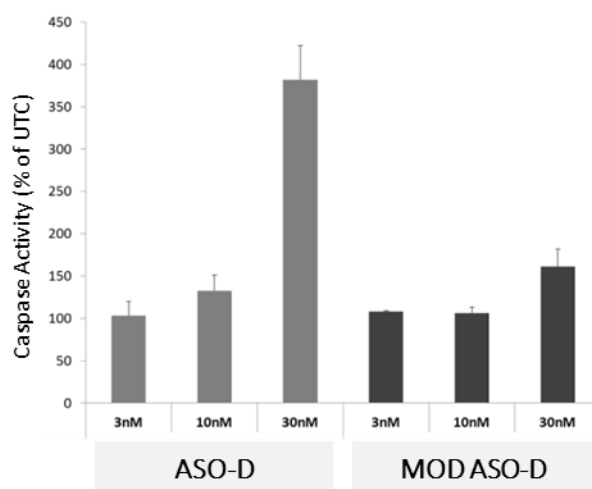
Figure 3.54: Modification of toxic ASO-D strongly suppresses *in vitro* toxicity. Mouse 3T3 fibroblast cells were transfected using Lipofectamine 2000 with ASO-D and MOD ASO-D, respectively. Caspase 3/7 activity was measured 24 hours later. Triplicate transfections were performed; the results are shown as a percentage change relative to untreated cells (UTC). Data sere means ± SD. Data generated in the lab of Andreas Dieckmann (Roche Innovation Center Basel).

Our previous attempt to compare protein binding profiles of ASOs with distinct *in vitro* characteristics was particularly complex due to the size of the GalNac modification and the use of different hooks for capture (3' biotinylated GalNac versus 3' biotinylated naked).

Here the same biotinylated ASO-D was used for capture and competed by either ASO-D or MOD ASO-D. The chemical difference between ASO-D and MOD ASO-D is minor and not modifying the oligonucleotide backbone.

The experiment was performed in Hek293 cell extracts. Briefly, 100 µg total protein were spiked with increasing concentrations (0, 5, 25, 100 and 500 pmol in 120 µl total volume) of either naked ASO-D or MOD ASO-D prior to being loaded on biotinylated ASO-D (50 µl of magnetic beads coated with 100 pmol of biotinylated ASO-D).

Fractions were loaded for a short SDS-PAGE to eliminate SDS, were cut as one band, in gel-digested and processed for LC-MS/MS analysis.

Around 1'300 proteins in total have been identified in each bound fractions (1'000 identified with at least two unique peptides). The background of proteins was rather similar with 80% overlap (at least two unique peptides, data not shown). Scatter plots comparing pAdj values revealed common binders such as nucleolin (Figure 3.55 A) and clearly indicated that many more proteins were preferentially binding to ASO-D (Figure 3.55 B) compared to MOD ASO-

D. The chemical modification led to a "cleaner" protein binding profile which could explain the strong suppression of *in vitro* toxicity. This observation is in accordance with recent published data [129] reporting that those protein interactions might be associated with LNA mislocalization and cellular stress.

Among the proteins preferentially binding to ASO-D versus MOD ASO-D (i.e. for which the binding was potentially lost upon modification of ASO-D), multiple members from one of the most abundant and best characterized DNA binding domain, the Zinc-finger protein family, were identified (Figure 3.55 B and 3.56). Although Zinc finger proteins are usually described as DNA-binding transcription factors, roles of RNA-binding proteins regulating RNA metabolism are reported [130, 131]. The suppression of toxicity observed *in vitro* with MOD ASO-D would need first to be confirmed *in vivo* before investigating further a potential implication of Zinc-finger proteins.
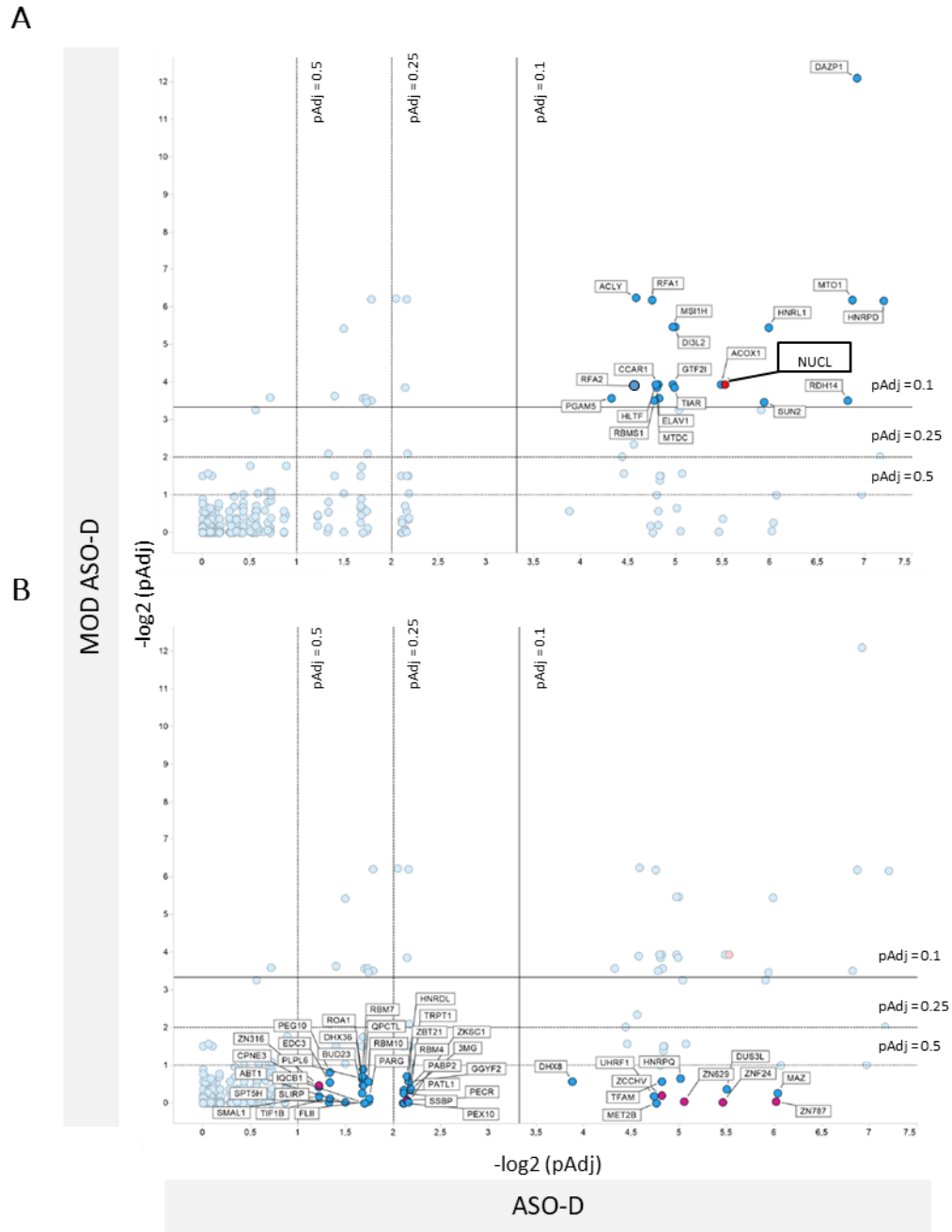
Figure 3.55: Comparison of pAdj values distribution between ASO-D and MOD ASO-D after applying OCC-MS in Hek293 extracts. Scatter plot displaying the adjusted p value (pAdj) when competing with ASO-D versus MOD ASO-D. pAdj: p value adjusted by the Westfall-Young approach. Lines arbitrarily set at at 3.32 (-log2 pAdj 0.1), 2 (-log2 pAdj 0.25) and 1 (-log2 pAdj 0.5). Proteins identified with at least two unique peptides. Proteins highlighted in dark blue (A) proteins specifically binding to both ASO-D and MOD ASO-D (pAdj below 10%) (B) proteins with a lower pAdj for ASO-D versus MOD ASO-D. Nucleolin (NUCL) highlighted as a reference. In pink: Zinc-finger protein family members.
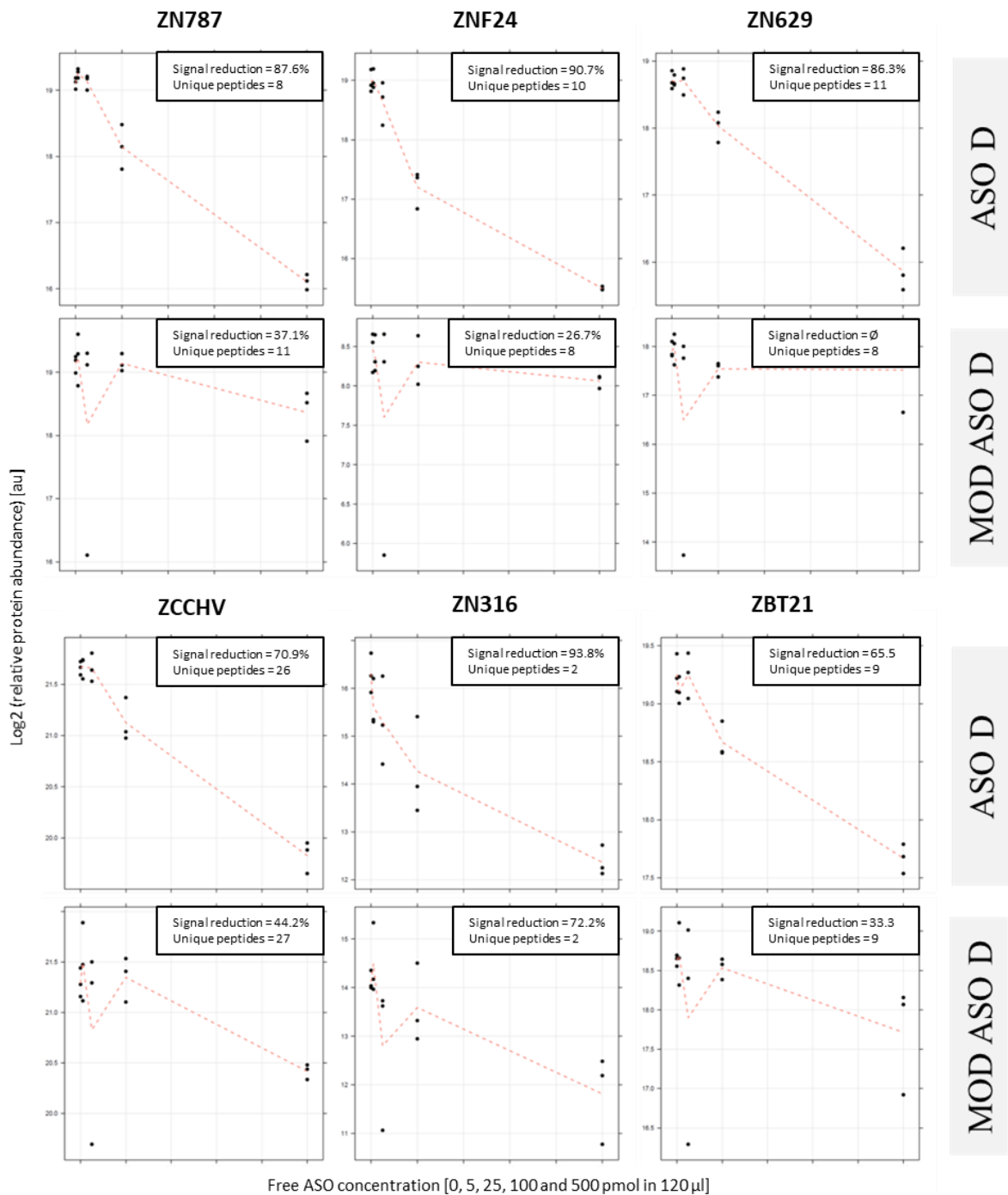
Figure 3.56: Displacement profiles of Zinc-finger proteins from ASO-D in Hek293. Protein abundance (log2) of ZN767, ZNF24, ZN629, ZCCHV, ZN316 and ZBT21 after affinity enrichment in Hek293 (100 µg) with ASO-D as bait, in the presence of increasing concentrations of free ASO-D or MOD ASO-D (0, 5, 25, 100, 500 pmol in 120 µl in triplicates). Proteins from bound fractions were identified by LC-MS/MS and MS signals were subjected to a statistical analysis to derive concentration-dependent signal decrease of specific interactors.

To summarize, it is not possible to conclude from these results whether ASO safety profiles could be partly driven by interaction with specific proteins or by their propensity to bind many proteins or a combination of both. Nonetheless, the results show that minor engineering of ASOs can lead to a dramatic impact on protein binding properties offering options for more favorable safety profiles (i.e. improvement of ASO therapeutic index).

4. DISCUSSION

Understanding the biological function of a protein relies crucially on the identification of its interactome in a cellular context. Such sophisticated networks are usually complex to dissect and innovative strategies are required to improve the coverage and accuracy of current interaction databases. In this thesis, an immuno-competitive capture mass spectrometry (ICC-MS) approach was implemented and tested in different biological backgrounds. By combining competitive capture with MS, ICC-MS is filtering out background proteins to generate highly specific protein-protein interaction maps. In each application a limited number of potential interactors haven been extracted from the whole proteome and efficiently validated functionally. In the first application, LATS kinases were shown to interact with HCV NS5A protein and to play a role in viral replication efficiency. Then, FGF2 was revealed as a Gpc2 interacting candidate having probably an impact on the cell transition from a proliferative to a differentiated state. Finally, HtrA1 was identified as a microtubule associated protein in alignment with previous observations. The approach has been adapted to capture RNA-protein interactions (oligo-competitive capture mass spectrometry or OCC-MS) and contributed to the deciphering of SMN splicing modifier mechanism of action by analyzing the effect of the presence of the compound on RNA-protein interactions. An attempt to extend OCC-MS to extract ASO-protein interactions highlighted some challenges associated to the nature of those chemically modified nucleic acid therapeutics. In the following section the observed results are critically debated in the context of current available techniques to screen PPIs.

4.1 Affinity purification method selection

To study endogenous PPIs, many strategies exist (see section 1.4) having all their strengths and limitations. A growing interest towards MS-based approaches emerged recently to study pathogens' interactome [132-135]. Among them, AP-MS is one of the most popular screening approach in which the target protein together with its interaction partners are isolated by affinity binding to an antibody immobilized on a solid support.

In this thesis, an antibody based co-complex approach has been developed in which proteins of interest were captured in their endogenous form. One obvious advantage towards a capture through an affinity epitope tag is to ensure isolation of protein complexes close to their physiological state and without the need for cloning or tagging. A prerequisite then is to have access to an antibody with high affinity (high yield of capture) and specificity (no cross-

reactivity towards other proteins). Thus to ensure a successful ICC-MS, an extensive validation of the capture antibody is required (e.g. validation of NS5A antibody via sequence alignment of the epitope, blotting with the capture antibody and IP in the parental cell line which does not express the target protein). Although any types of engineering on an endogenous protein is associated with a risk to impact protein function, expression, folding or localization (especially large protein tags such as GST or GFP), promising alternative such as nanobody based affinity systems emerged [92]. In any cases, the outcome of an antibody based profiling exercise will always be dependent on the particular epitope of the antibody selected and information will be lost at the antibody-binding site. Therefore, using multiple antibodies directed against different epitopes on the target protein will likely increase the coverage of the full interactome.

Low-affinity interactions (such as receptor/ligand) or transient interactions (such as enzyme/substrate) might not be detected with ICC-MS, that's why a combination with approaches specifically designed to overcome this limitation would increase mapping coverage, e.g. proximity-based labeling methods using either engineered ascorbate peroxidase (APEX) or mutated *E- coli* biotin ligase BirA (BioID) [136], chemical-crosslinking [137] or optimized cell surface receptor screens [138].

In this thesis, protein-metabolite interactions have not been investigated although there role in controlling cellular processes is critical. A recently developed chemoproteomics approach named Limited Proteolysis Mass Spectrometry (LiP MS) allow for a systematic and unbiased analysis without need for chemical labelling [139].

4.2 Cellular model selection and protein extraction conditions

A protein often has multiple interactors associated with different functions in response to specific stimulus or cellular state. To ensure that the primary scientific question is addressed properly, profiling a protein in an appropriate biological system is essential. Cellular models were critically selected for their relevance in each ICC-MS applications reported in this thesis (Con1-Huh7 cells, d21 NSC, hfRPE, SMA type 1 fibroblasts and PTECs). It is also crucial that the bait protein is reasonably expressed in the cellular model selected which may require to optimize the amount of cells utilized as well as the number of cell passages.

Every steps of the ICC-MS protocol will be influencing the generated list of interacting partners (protein amount, salts concentration, detergents, number of washes, incubation time, etc.). Cell

lysis should be performed in conditions ensuring protein solubilization but without disruption of protein complexes. In this thesis a low concentration of non-denaturing detergent has been shown to be appropriate (0.5-1% NP40).

The majority of the experiments described in this thesis were performed in whole cell lysate. Whole extracts have the advantage to potentially yield more target protein and protein extraction protocol is faster and milder to preserve native interactions. Many proteins have multiple subcellular localizations which means that a whole extract will be then composed of a mix of proteins from different subcellular compartments associated to distinct protein complexes. For certain applications it can make sense to include a subcellular fractionation step depending on the scientific question, e.g. in nuclear extracts to study SMN splicing modifier mode of action (MoA). Profiling of subcellular compartments could be a strategy to extract functionally relevant ASO-binders as discussed in 4.5.

## 4.3 Competitive binding strategy and statistical model

ICC-MS markedly increases the specificity of classical IP by introducing a pre-competition step between free and capturing antibody followed by statistical analysis of the quantified MS signals. The ability of a free form of an antibody to prevent target capture by an immobilized version of the same antibody was demonstrated first using an antibody against HCV NS5A protein. The extent of displacement being observed in a dose-dependent manner (Figure 3.4), it allows the use of MS-based label-free quantitation (i.e. no need for stable isotopes or isotopic tags) to compare the relative abundance of a protein across multiple LC-MS/MS experiments. The twelve-concentration antibody competition profile as well as the five-concentration tested for Gpc2 and Htra1 provided sufficient power to the statistical analysis. The use of a gradual increase of free antibody concentrations is providing additional confidence in the obtained results compared to more classical binary comparisons.

Experimental parameters such as the number of antibody concentrations and replicates, are adjustable and may need to be fine-tuned for each particular experiment. For example in this thesis, decreasing the number of concentrations from twelve to five has been compensated by increasing the number of replicates from one to three.

Reducing even further the number of concentration was tested on the NS5A dataset. An analysis using only three concentrations (0, 1.7 and 24 μg/ml) resulted in a different list with

twenty-five significant hits (Figure 4.1, pAdj ≤ 0.05, as applied in the original analysis). Among the twenty-five, nine proteins (NS5A, NP1L4, PKHG2, FBW1B, NP1L1, PI4KA, UBP19, LATS1 and LATS2) overlapped with the twelve concentrations analysis. UBP19 is found with four protein quantitation groups, thus three being redundant. A visual inspection of the displacement profiles of the thirteen additional proteins clearly flags eleven of them (OST48, NCRP1, ELAV1, EFTU, PLAK, RS5, XRN2, FOXK1, CSN2, FOXK2, and DSC1) as false positives, if considering the three- as instead of the twelve-concentrations analyses. Two proteins are not found at all (bottom row, VAPA and PGAM5) whereas two other proteins (ANKY2 and SMYD3) could be considered as potential candidates (p-values of 7% and 40% in the twelve-concentrations experiment, respectively).

In summary, the statistical analysis applied would exceed the limits of its validity by reducing the sample size from twelve to three concentrations and the effort to generate additional data points is largely compensated by a lower investment to validate fewer hits with higher confidence.
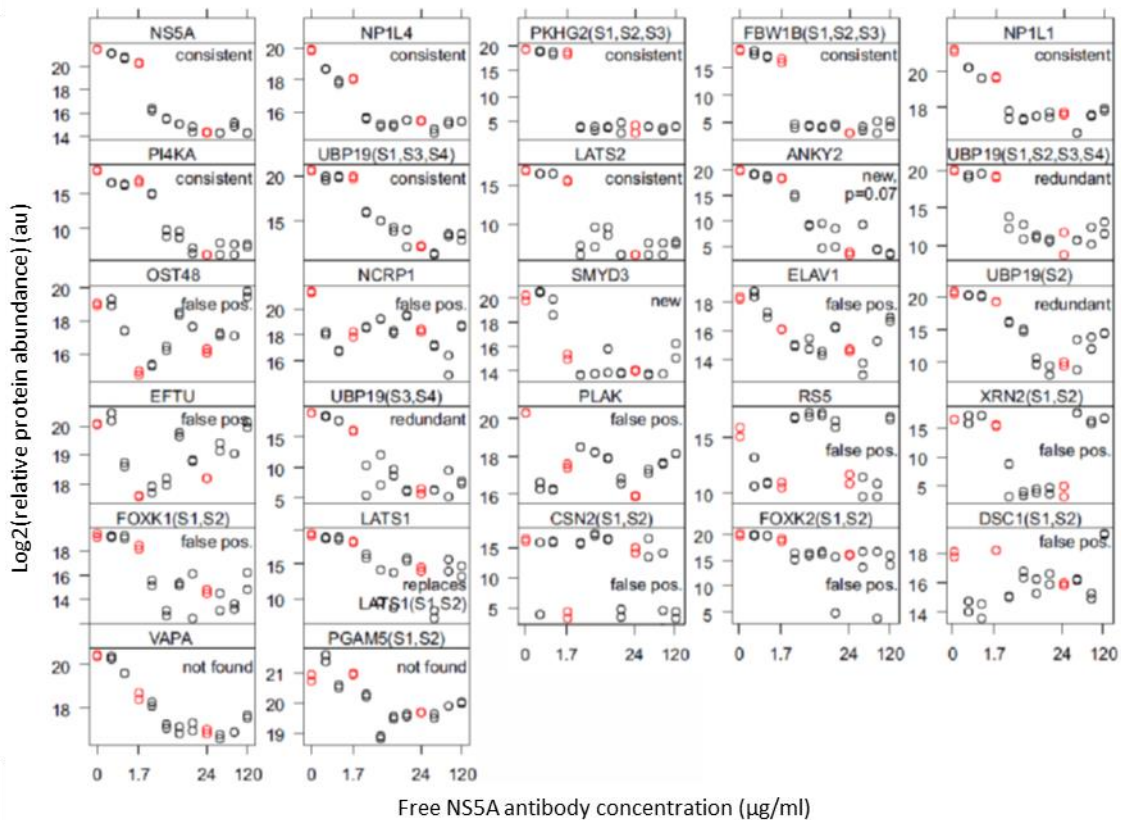
Figure 4.1: Statistical power of ICC-MS model. Relative abundance plotted against all 12 concentrations, with three concentrations (0, 1.7 and 24 µg/ml) plotted in red.

The ultimate goal of ICC-MS is to provide a manageably short list of proteins, which have a high chance of being successfully validated in orthogonal follow-up experiments. P-values are indicative and allow ranking based on confidence of interaction. Nevertheless, it cannot be claimed that proteins below a given statistical threshold are absolute binders of a specific protein, and directly above the threshold are non-binding. The more rigorous the threshold is, the shorter the list, and subsequently the higher the confidence. At the same time, the chance of missing some true binders is also increased.

4.4 Further validation of interacting candidates

ICC-MS is a powerful approach to identify physical interactions in a protein complex. Once the list of interacting candidates is generated, some open questions remains:

1. Is the identified interaction confirmed by another approach?
2. Is the interaction identified in a test tube also occurring in a compartmentalized living cell?
3. Is the identified interaction ultimately leading to a biological function?

To address the first question, combination with orthogonal methods is the most efficient strategy to improve confidence in identified PPIs. In this thesis, coIP experiments have been performed and validated ICC-MS findings (LATS/NS5A, Gpc2/FGF2, Htra1/tubulin, Htra1/C1TNF5). Although only a handful of candidates are identified when applying ICC-MS, text mining as well as biological expertise will always be taken into account to decide on which proteins to follow up. With ICC-MS data only, it is not possible to distinguish direct from indirect interactions. SPR is a method of choice to confirm direct binding and to derive affinity values as well as kinetics [140] but it requires the production of a purified recombinant protein retaining the structural integrity of the endogenous form.

The second question can be addressed by fluorescence microscopy using genetically expressed fluorescent proteins such as fluorescent two-hybrid cellular assay (F2H) [141]. This technique relies on co-localization studies of fluorescently-tagged proteins in BHK cells comprising a lac operator stably integrated in the nucleus. An attempt have been performed to further confirm HtrA1 binding to tubulin and C1QTNF5 in collaboration with Philip Kaiser from NMI (data not shown) but the experiment failed probably due to the high aggregation rate of HtrA1. At

the end only a minor amount of HtrA1 protein was most likely accessible for binding to C1QTNF5 or tubulin respectively.

Affinity, kinetic and spatial co-localization are giving some indications about the nature of the interaction at the cellular level but at the end are not a guarantee of biological functional relevance. Assays applied for functional validation are numerous and will greatly depend on the type of protein identified (e.g. *in vitro* phosphorylation for LATS kinases, replication capacity of Con1-NS5A-Ser71Ala, impact of LATS silencing on viral replication, proliferative effect of increasing concentration of Gpc2).

It is highly challenging to prove or disprove the functional relevance of an interaction and at the end any interactions can be ultimately considered as an evolutionary reservoir for future functional interactions. In addition, although many strategies are available to increase the confidence towards true interactors, the issue of false negative or missed interactions remain, a subset of the interactome only often being probed.

4.5 Applicability to the profiling of ASO-protein interactions

The combination of competitive capture with MS is a valuable tool to screen not only protein-protein interactions (ICC-MS) but also RNA-protein interactions (OCC-MS). Nevertheless, the specific study of ASO-protein interactions turned out to be more challenging as ASO are by nature binding to many proteins/complex of proteins by offering an extended surface with multiple points of contact (not like an hydrophobic pocket as for small molecule-protein interactions). Many binders are competing or cooperating for a specific site or backbone which will be influenced by multiple factors such as the concentration of the protein and of the ASO, the presence of recognition motifs, specific chemical modifications, etc.

OCC-MS alone will not be sufficient to confidently link an interaction with a specific intracellular protein to an observed *in vivo* toxicity profiles. Nevertheless it revealed that a greater number of binders is potentially translating in stronger induced *in vitro* hepatotoxicity (see section 3.3.3.). Previously published data using protein microarrays is aligned with such hypothesis [142]. Efforts are currently ongoing to further support this observation by profiling ASOs with distinct *in vivo* toxicity profiles directly in tissue extracts (liver and kidney).

Despite the inherent complexity of ASO-protein interactions, OCC-MS has been capable of revealing a cellular "core proteome", i.e. a few interacting partners consistently identified

irrespectively of the cell type or of the sequence of the oligonucleotide. Indeed, comparing ASO-B/C/D cellular pull-downs, 34 proteins have been identified as ASO-interacting partners across all datasets (Figure 4.2 A) with an handful already published in another RNA centric study [119] (Figure 4.2 B, in red).
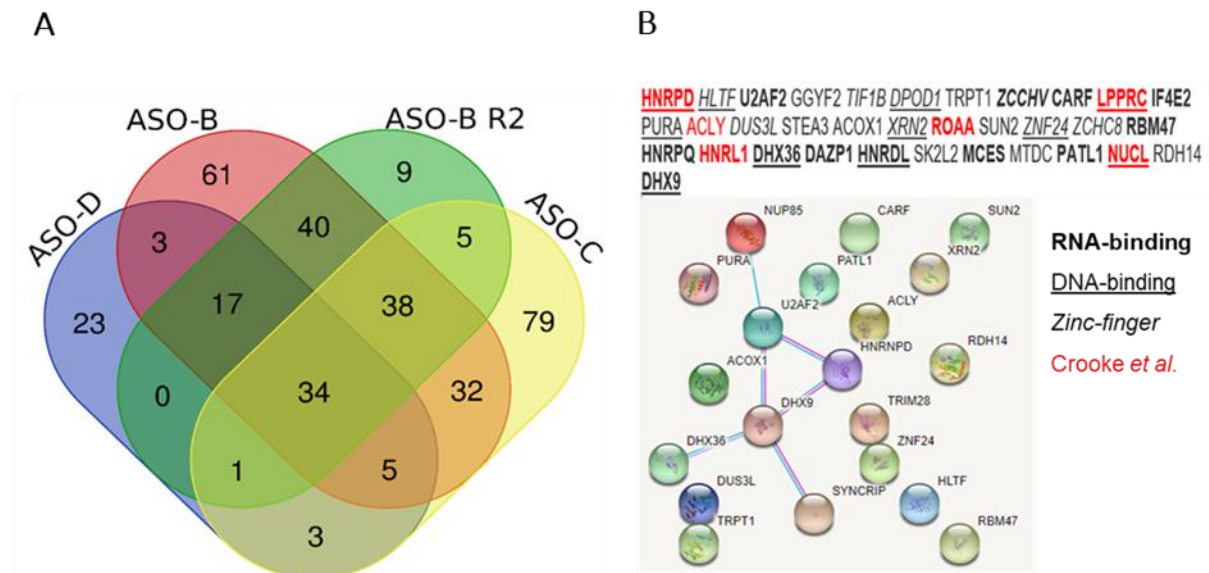
Figure 4.2: Cellular core binders. Statistical relevant interacting partners consistently identified in pull-downs from cellular extracts (identified with at least 2 unique peptides, pAdj < 0.5, signal reduction > 50% and with highest abundance when no competition). (A) Venn diagram (http://bioinformatics.psb.ugent.be/) comparing ASO-D in Hek293, ASO-B (including R2: replicate 2) and ASO-C in PTEC TERT1. (B) The 34 protein candidates were loaded in STRING DB (https://string-db.org/) using the highest confidence (0.900) in term of interaction score from curated databases (blue lines) as well as experimentally determined (pink lines). In bold, known RNA-binding proteins; underlined, known DNA-binding proteins; in italic, zinc-finger proteins; in red, proteins also reported in [119].

The further confirmation of functionally relevant interactions is even more challenging given the high number of potential candidates identified. Follow-up experiments to distinguish direct versus indirect binders or to verify biological relevance for so many candidates is not feasible. OCC-MS statistical model can probably be optimized to deal better with the high number of potential binders. Indeed, even at high pAdj values (up to ~50%) some candidates were still displaying a reasonable displacement with increasing concentrations of free ASO indicating a failure in the power of our test for statistical significance. This can be explained by the fact that the model is assuming that the majority of the proteins are not affected by the increase in free competitor, which is not true for ASO pull-downs. The model has been designed to reduce as

much as possible the identification of false positives but not false negatives. In any cases, although a definite nominal value cannot be set, pAdj values can still be used for ranking (from the most significant to the least).

It should be noticed that our approach might be too artificial when dealing with ASO as the vast majority of ASOs remains trapped in endocytic compartment and travel via particles [119]. Hence many of the identified ASO-binding proteins would even not be encountering the ASO in a system where compartmentalization is preserved. Techniques like UV crosslinking immunoprecipitation (CLIP) or RNA-protein interaction detection (RaPID) using proximity-dependent protein labeling in intact live cells can help addressing this limitation [143-145].

As observed in section 3.3.3, a minor chemical modification can have a dramatic impact on ASO protein binding profile. Modification having no impact on RNA affinity and antisense activity while decreasing the number of interacting partners are representing promising strategy to design safer ASO. OCC-MS could support the selection of the most appropriate modification.

Therefore OCC-MS could be used as a qualitative way to profile ASOs, ranking them based on their protein binding affinity. OCC-MS could guide chemistry to mitigate cellular toxicity and ultimately *in vivo* hepatotoxicity or nephrotoxicity. Overall, OCC-MS could represent an additional approach contributing to the improvement of ASO therapeutic index.

4.6 The relevance of ICC-MS and OCC-MS approaches in therapeutic research

ICC-MS and OCC-MS have been shown to have the potential to contribute to:

- The identification of novel therapeutic targets, e.g. Further investigations are of course required to clearly define the role of LATS kinases but LATS1 and LATS2 could be considered as interesting new host targets for the development of future classes of HCV therapeutic agents.
- The discovery of potential novel biomarkers, e.g. FGF2 together with Glypican-2 as CSF biomarker for neurogenesis.
- The relevance of targets for a specific pathology, e.g. Htra1 for AMD.
- MoA understanding, e.g. SMN splicing modifiers.

In addition determination of sites of PPIs by mutagenesis, protein purification, peptide mapping/arrays or X-ray crystallography can provide extremely useful information for designing specific small molecules. Those approaches are usually challenging as well as time consuming, as some proteins are difficult to express or crystallize.

Developing a new drug is associated with extreme efforts and late stage failure of a drug candidate is a dramatic loss primarily for the patients and also in terms of investments. Investing efforts in target biology and disease mechanisms understanding during the drug development process is part of the strategies developed by pharmaceutical companies to select the most efficacious and safest molecules for clinical testing [146]. Indeed a better understanding of target biology can not only bring confidence towards its relevance to treat a specific pathology but can support early safety de-risking as well as guide the selection of appropriate pre-clinical models. In case of failure, a deep understanding of target biology can even lead to drug repositioning.

4.7 Outlook

In this thesis, a novel unbiased platform to screen protein-protein interactions in protein extracts with high specificity has been developed. The number of candidates emerging being limited, this allows for interaction confirmation via orthogonal methods and further functional validation. The effort of combining AP-MS with competitive binding is largely compensated by a lower investment to validate fewer hits with higher confidence.

The workflow is simple but would greatly benefit both in term of time and reproducibility from an automation option. A 96 well plate format using AssayMAP Bravo platform from Agilent is currently being assessed using streptavidin cartridges to capture biotinylated antibodies. In addition to the benefit of automation, the nature of streptavidin/biotin interaction will greatly reduce any issue associated with antibody leakage. The samples processing for further MS analysis will be also associated with faster protocols based on on-beads digestion altogether allowing for an improvement in overall throughput.

Based on the proof-of-principle studies described in this thesis and the future improvements made to the protocol, ICC-MS should remain an attractive approach to better understand targets biology and support their relevance as therapeutic targets.

5. REFERENCES

1.   Berggard, T., Linse, S., & James, P., *Methods for the detection and analysis of protein-protein interactions.* Proteomics, 2007. **7**: p. 2833-2842.
2.   Braun, P., & Gingras, A.C., *History of protein-protein interactions: from egg-white to complex networks.* Proteomics, 2012: p. 1478-1498.
3.   De Las Rivas, J., & Fontanillo, C., *Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks.* PLOS Computational Biology, 2010. **6**.
4.   Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., Yildirim, M.A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J.M., Cevik, S., Simon, C., de Smet, A.S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R.R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M.E., Roth, F.P., Hill, D.E., Tavernier, J., Wanker, E.E., Barabási, A.L., & Vidal M., *An empirical framework for binary interactome mapping.* Nature Methods, 2009. **6**: p. 83-90.
5.   Stumpf, M.P.H., Thorne, T., de Silva, E., Stewart, R., Jun An, H., Lappe, M., & Wiuf C., *Estimating the size of the human interactome.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**: p. 6959-6964.
6.   Nooren, I.M.A., & Thornton, J.M., *Diversity of protein-protein interactions.* The EMBO Journal, 2003. **22**: p. 3486-3492.
7.   Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., & Orengo, C., *Transient Protein-Protein Interactions: Structural, Functional, and Network Properties.* Structure, Cell Press, 2010. **18**: p. 1233-1243.
8.   Westermarck, J., Ivaska, J., & Corthals, G.L., *Identification of protein interactions involved in cellular signaling.* Molecular and Cellular Proteomics, 2013. **12**: p. 1752-1763.
9.   Lakey, J.H., & Gokce, I., *Protein-Protein Interactions.* Encyclopedia of Life Sciences, 2001.
10.   Beckett, D., *Multilevel regulation of protein–protein interactions in biological circuitry.* Physical Biology, 2005. **2**: p. 67-73.
11.   Deribe, Y.L., Pawson, T., & Dikic, I., *Post-translational modifications in signal integration.* Nature Structural & Molecular Biology, 2010. **17**: p. 666-672.
12.   Gonzalez, M.W., & Kann, M.G., *Chapter 4: Protein Interactions and Disease.* PLOS Computational Biology, 2012. **8**.
13.   Arkin, M.R., Tang, Y. & Wells, J.A., *Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality.* Chemistry & Biology, 2014. **21**: p. 1102-1114.
14.   Wells, J.A.M., C.L., *Reaching for high-hanging fruit in drug discovery at protein-protein interfaces.* Nature, 2007. **450**: p. 1001-1009.
15.   Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., & Liu, E.A., *In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2.* Science, 2004. **303**: p. 844-848.
16.   Vidal, M., Cusick, M.E., & Barabási, A.L., *Interactome Networks and Human Disease.* Cell, 2011. **144**: p. 986-998.
17.   Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., Kamburov, A., Ghiassian, S.D., Yang, X., Ghamsari, L., Balcha, D., Begg, B.E., Braun, P., Brehme, M., & Vidal, M., *A Proteome-Scale Map of the Human Interactome Network.* Cell, 2014. **159**: p. 1212-1226.
18.   O'Connell, M.R., Gamsjaeger, R., & Mackay, J.P., *The structural analysis of protein–protein interactions by NMR spectroscopy.* Proteomics, 2009. **9**: p. 5224-5232.
19.   Fields, S., & Song, O.K. , *A novel genetic system to detect protein-protein interactions* Nature, 1989. **340**: p. 245-246.

20. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., & Sakaki, Y., *A comprehensive two-hybrid analysis to explore the yeast protein interactome.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **10**: p. 4569-4574.

21. Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., & Rothberg, J.M., *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.* Nature, 2000. **403**: p. 623-627.

22. Dunham, W.H., Mullin, M., & Gingras, A.C., *Affinity-purification coupled to mass spectrometry: basic principles and strategies.* Proteomics, 2012. **12**: p. 1576-1590.

23. Gingras, A.C., Gstaiger, M., Raught, B., & Aebersold, R., *Analysis of protein complexes using mass spectrometry.* Nature Reviews, 2007. **8**: p. 645-654.

24. Miteva, Y., Budayeva, H.G., & Cristea, I.M., *Proteomics.based methods for discovery, quantification, and validation of protein-protein interactions.* Analytical Chemistry, 2013. **85**: p. 749-768.

25. Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y.V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J.P., Duewel, H.S., Stewart, I.I., Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S.L., Moran, M.F., Morin, G.B., Topaloglou, T., & Figeys, D., *Large-scale mapping of human protein–protein interactions by mass spectrometry.* Molecular Systems Biology, 2007. **3**: p. 89.

26. Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., Hernandez, H., Jang, G.M., Roth, S.L., Akiva, E., Marlett, J., Stephens, M., D'Orso, I., Fernandes, J., Fahey, M., Mahon, C., O'Donoghue, A.J., Todorovic, A., Morris, J.H., Maltby, D.A., Alber, T., Cagney, G., Bushman, F.D., Young, J.A., Chanda, S.K., Sundquist, W.I., Kortemme, T., Hernandez, R.D., Craik, C.S., Burlingame, A., Sali, A., Frankel, A.D., & Krogan, N.J., *Global landscape of HIV–human protein complexes.* Nature, 2011. **481**: p. 365-370.

27. Mackay, J., Sunde, M., Lowry, J.A., Crossley, M. & Matthews, J.M., *Protein interactions: is seeing believing?* Trends in Biochemical Sciences, 2007. **32**: p. 530-531.

28. Björling, E., & Uhlén, M., *Antibodypedia, a portal for sharing antibody and antigen validation data.* Molecular and Cellular Proteomics, 2008. **7**: p. 2028-2037.

29. Luck, K., Sheynkman, G.M., Zhang, I., & Vidal, M., *Proteome-scale human interactomics.* Trends in Biochemical Sciences, 2017. **42**: p. 342-354.

30. Lehne, B., & Schlitt, T., *Protein-protein interaction databases: Keeping up with growing interactomes.* Human Genomics, 2009. **3**: p. 291-297.

31. Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B., & Superti-Furga, G., *Proteome survey reveals modularity of the yeast cell machinery.* Nature, 2006. **440**: p. 631-636.

32. Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrín-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., &

Greenblatt J.F., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.* Nature, 2006. **440**: p. 637-643.

33. Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urci,a R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H., & Lamond, A., *Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes.* Journal of Cellular Biochemistry, 2008. **183**: p. 223-239.

34. Nesvizhskii, A.I., *Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments.* Proteomics, 2012. **10**: p. 1639-1655.

35. Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, JP., St-Denis, N.A.,Li, T., Miteva, Y.V., Hauri, S., Sardiu, M.E., Low, T.Y., Halim, V.A., Bagshaw, R.D., Hubner, N.C., al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W.H., Goudreault, M., Lin, Z.Y., Gonzalez Badillo, B., Pawson, T., Durocher, D., Coulombe, B., Aebersold, R., Superti-Furga, G., Colinge, J., Heck, A.J.R., Choi, H., Gstaiger, M., Mohammed,S., Cristea, I.M., Bennett, K.L., Washburn, M.P., Raught, B., Ewing, R.M. Gingras, AC., & Nesvizhskii, A.I., *The CRAPome: a contaminant repository for affinity purification–mass spectrometry data.* Nature Methods, 2013. **10**: p. 730-736.

36. Hyungwon Choi, B.L., Zhen-Yuan Lin, Ashton Breitkreutz, Dattatreya Mellacheruvu, Damian Fermin, Zhaohui S Qin, Mike Tyers, Anne-Claude Gingras & Alexey I Nesvizhskii, *SAINT: probabilistic scoring of affinity purification–mass spectrometry data.* Nature Methods, 2011. **8**: p. 70-73.

37. Sowa, M.E., Bennett, E.J., Gygi, S.P., & Harper, J.W., *Defining the Human Deubiquitinating Enzyme Interaction Landscape.* Cell, 2009. **138**: p. 389-403.

38. Paul, F.E., Hosp, F., & Selbach, M., *Analyzing protein–protein interactions by quantitative mass spectrometry.* Methods, 2011. **54**: p. 387-395.

39. Trinkle-Mulcahy, L., *Resolving protein interactions and complexes by affinity purification followed by label-based quantitative mass spectrometry.* Proteomics, 2012. **12**: p. 1623-1638.

40. Selbach, M., & Mann, M. , *Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK).* Nature Methods, 2006. **3**: p. 981-983.

41. Klumpp, K., Lévêque, V., Le Pogam, S., Ma, H., Jiang, W.R., Kang, H., Granycome, C., Singer, M., Laxton, C., Qi Hang, J., Sarma, K., Smith, D.B., Heindl, D., Hobbs, C.J., Merrett, J.H., Symons, J., Cammack, N., Martin, J.A., Devos, R., & Nájera, I., *The Novel Nucleoside Analog R1479 (4'-Azidocytidine) Is a Potent Inhibitor of NS5B-dependent RNA Synthesis and Hepatitis C Virus Replication in Cell Culture.* The Journal od Biological Chemistry, 2006. **281**(3793-3799).

42. Lohmann, V., Hoffmann, S., Herian, U., Penin, F., & Bartenschlager, R. , *Viral and cellular determinants of hepatitis C virus RNA replication in cell culture.* Journal of Virology, 2003. **77**: p. 3007-3019.

43. Lugert, S.e.a., *Glypican-2 levels in cerebrospinal fluid predict the status of adult hippocampal neurogenesis.* Nature Scientific Reports, 2017. **7**.

44. Melo, E., Oertle, P., Trepp, C., Meistermann, H., Burgoyne, T., Sborgi, L., Cabrera, A.C., Chen, C., Hoflack, J.C., Kam-Thong, T., Schmucki, R., Badi, L., Flint, N., Ghiani, Z.E., Delobel, F., Stucki, C., Gromo, G., Einhaus, A., Hornsperger, B., Golling, S., Siebourg-Polster, J., Gerber, F., Bohrmann, B., Futter, C., Dunkley, T., Hiller, S., Schilling, O., Enzmann, V., Fauser, S., Plodinec, M., and Iacone, R., *HtrA1 Mediated Intracellular Effects on Tubulin Using a Polarized RPE Disease Model.* EBioMedicine, 2018. **27**: p. 258-274.

45. Sewing, S., Gubler, M., Gérard, R., Avignon, B., Mueller, Y., Braendli-Baiocco, A., Odin, M., & Moisan, A., *GalNAc Conjugation Attenuates the Cytotoxicity of Antisense Oligonucleotide Drugs in Renal Tubular Cells.* Molecular Therapy, Nucleic Acids, 2019. **14**: p. 67-79.

46.   Shevchenko, A., Wilm, A., Vorm, O., and Mann, M., *Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels.* Analytical Chemistry, 1996. **68**: p. 850-858.

47.   Olsen, J.V., de Godoy, L.M., Li, G, Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., & Mann, M., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap.* Molecular and cellular Proteomics, 2005. **4**: p. 2010-2021.

48.   Elias, J.E., & Gygi, S.P., *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nature Methods, 2001. **4**: p. 207-214.

49.   Augustin, A., Lamerz, J., Meistermann, H., Golling, S., Scheiblich, S., Hermann, J.C., Duchateau-Nguyen, G., Tzouros, M., Avila, D.W., Langen, H., Essioux, L. and Klughammer, B., *Quantitative Chemical Proteomics Profiling Differentiates Erlotinib from Gefitinib in EGFR Wild-Type Non–Small Cell Lung Carcinoma Cell Lines.* Molecular Cancer Therapeutics, 2013.

50.   Smyth, G., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Statistical Application in Genetics and Molecular Biology, 2004.

51.   WHO, *WHO Global Hepatitis Report, 2017.* 2017.

52.   He Y, S.K., & Tan SL, *HCV NS5A: A Multifunctional Regulator of Cellular Pathways and Virus Replication in Hepatitis C viruses.* Genomes and Molecular Biology, 2006(ed Tan SL (Horizon Bioscience, Norfolk): p. 267-292.

53.   Reiss S, e.a., *Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment.* Cell Host Microbe, 2011. **9**(1): p. 32-45.

54.   Gao L, A.H., He JW and Lai MM, *Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft.* Journal of Virology, 2004. **78**: p. 3480-3488.

55.   Ahn, J., et al., *Systematic identification of hepatocellular proteins interacting with NS5A of the hepatitis C virus.* Journal of Biochemistry and Molecular Biology, 2004. **37**(6): p. 741-748.

56.   Macdonald A, a.H.M., *Hepatitis C virus NS5A: tales of a promiscuous protein.* Journal of General Virology, 2004: p. 2485-2502.

57.   Meistermann, H., & al, *A Novel Immuno-Competitive Capture Mass Spectrometry Strategy for Protein-Protein Interaction Profiling Reveals That LATS Kinases Regulate HCV Replication Through NS5A Phosphorylation.* Molecular and Cellular Proteomics, 2014.

58.   Superti-Furga, G., et al, *Viral immune modulators perturb the human molecular network by common and unique strategies.* Nature 2012. **487**: p. 486-490.

59.   Goonawardane, N., Gebhardt, A., Bartlett, C., Pichlmair, A., and Harris, M., *Phosphorylation of serine 225 in hepatitis C virus NS5A regulates protein-protein interactions.* Journal of Virology, 2017.

60.   Pearce, L.R., Komander, D., and Alessi, D.R., *The nuts and bolts of AGC protein kinases.* Nature Reviews, 2010. **11**: p. 9-22.

61.   Visser, S., & Yang, X., *LATS tumor suppressor: A new governor of cellular homeostasis.* Cell Cycle, 2010. **9**(19): p. 3892-3903.

62.   LeMay, K.L., Treadaway, J., Angulo, I., and Tellinghuisen, T.L., *A Hepatitis C Virus NS5A Phosphorylation Site That Regulates RNA Replication.* Journal of Virology, 2013. **87**(2): p. 1255-1260.

63.   Evans, M.J., Rice, C.M. and Goff, S.P., *Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(35): p. 13038-13043.

64.   Huang, Y., Staschke, K., De Francesco, R. and Tan, S.L., *Phosphorylation of hepatitis C virus NS5A nonstructural protein: A new paradigm for phosphorylation-dependent viral RNA replication?* Virology, 2007. **364**(1): p. 1-9.

65.   Suzuki, T.e.a., *Involvement of Hepatitis C Virus NS5A Hyperphosphorylation Mediated by Casein Kinase I-alpha in Infectious Virus Production.* Journal of Virology, 2014. **88**(13): p. 7541-7555.

66.   Kim, J., Lee, D., and Choe, J., *Hepatitis C Virus NS5A Protein Is Phosphorylated by Casein Kinase II.* Biochemical and Biophysical Research Communications, 1999. **257**: p. 777-781.

67.   Song, H.H., and Filmus, J., *The role of glypicans in mammalian development.* Biochimica et Biophysica Acta, 2002. **1573**: p. 241-246.

68.   De Cat, B., and David, G., *Developmental roles of the glypicans.* Seminars in Cell & Developmental Biology, 2001. **12**: p. 117-125.

69.   Ivinsa, J.K., Litwack, D.E., Kumbasara, A., Stipp, C.S., Landera, A.D., *Cerebroglycan, a Developmentally Regulated Cell-Surface Heparan Sulfate Proteoglycan, Is Expressed on Developing Axons and Growth Cones.* Developmental Biology, 1997. **184**: p. 320-332.

70.   Kirn-Safran, C., Farach-Carson, M.C., and Carson, D.D., *Multifunctionality of extracellular and cell surface heparan sulfate proteoglycans.* Cellular and Molecular Life Sciences, 2009. **66**: p. 3421-3434.

71.   Tanaka, S.S., Kojima, Y., Yamaguchi, Y.L., Nishinakamura, R., and Tam, P.P.L., *Impact of WNT signaling on tissue lineage differentiation in the early mouse embryo.* Development, Growth & Differentiation, 2011. **53**: p. 846-856.

72.   de Wit, J., O'Sullivan, M.L., Savas, J.N., Condomitti, G., Caccese, M.C., Vennekens, K.M., Yates, J.R. 3rd and Ghosh A. , *Unbiased discovery of glypican as a receptor for LRRTM4 in regulating excitatory synapse development.* Neuron, 2013. **79**(4): p. 696-711.

73.   Fico, A., Maina, F., and Dono, R., *Fine-tuning of cell signaling by glypicans.* Cellular and Molecular Life Sciences, 2011. **68**: p. 923-929.

74.   Lin, X., *Functions of heparan sulfate proteoglycans in cell signalling during development.* Development, 2004. **131**: p. 6009-6021.

75.   Kerever, A., Schnack, J., Vellinga, D., Ichikawa, N., Moon, C., Arikawa-Hirasawa, E., Efird, J.T., Mercier, F., *Novel extracellular matrix structures in the neural stem cell niche capture the neurogenic factor fibroblast growth factor 2 from the extracellular milieu.* Stem Cells, 2007. **25**(9): p. 2146-2157.

76.   Häcker, U., Nybakken, K., Perrimon, N., *Heparan sulphate proteoglycans: the sweet side of development.* Nature Reviews. Molecular Cell Biology, 2005. **6**(7): p. 530-541.

77.   Coles, C.H., Shen, Y., Tenney, A.P., Siebold, C., Sutton, G.C., Lu, W., Gallagher, J.T., Jones, E.Y., Flanagan, J.G., Aricescu, A.R., *Proteoglycan-Specific Molecular Switch for RPTPs Clustering and Neuronal Extension.* Science, 2011. **332**: p. 484-488.

78.   Herndon, M.E., Stipp, C.S., Lander, A.D., *Interactions of neural glycosaminoglycans and proteoglycans with protein ligands: assessment of selectivity, heterogeneity and the participation of core proteins in binding.* Glycobiology, 1999. **9**(2): p. 143-155.

79.   Bonneh-Barkay, D., Shlissel, M., Berman, B., Shaoul, E., Admon, A., Vlodavsky, I., Carey, D.J., Asundi, V.K., Reich-Slotky, R., and Ron, D., *Identification of glypican as a dual modulator of the biological activity of fibroblast growth factors.* The Journal od Biological Chemistry, 1997. **272**: p. 12415-12421.

80.   Gutiérrez, J., and Brandan, E., *A Novel Mechanism of Sequestering Fibroblast Growth Factor 2 by Glypican in Lipid Rafts, Allowing Skeletal Muscle Differentiation.* Molecular and Cellular Biology, 2010. **30**: p. 1634-1649.

81.   Yang, Z., Camp, NJ., Sun, H., Tong, Z., Gibbs, D., Cameron, DJ., Chen, H., Zhao, Y., Pearson, E., Li, X., Chien, J., Dewan, A., Harmon, J., Bernstein, PS., Shridhar, V., Zabriskie, NA., Hoh, J., Howes, K. and Zhang, K., *A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration.* Science, 2006. **314**(5801): p. 992-993.

82.   Clawson, G.A., Bui,V., Xin, P., Wang, N., and Pan, W., *Intracellular Localization of the Tumor SuppressorHtrA1/Prss11 and its Association With HPV16 E6and E7 Proteins.* Journal of Cellular Biochemistry, 2008. **105**: p. 81-88.

83. Vierkotten, S., Muether, P., and Fauser, S. , *Overexpression of HTRA1 Leads to Ultrastructural Changes in the Elastic Layer of Bruch's Membrane via Cleavage of Extracellular Matrix Components.* PLoS ONE, 2011. **6**(8): p. 1-12.

84. Sonoda, S., Spee, C., Barron, E., Ryan, S.J., Kannan, R., & Hinton, D.R., *A protocol for the culture and differentiation of highly polarized human retinal pigment epithelial cells.* Nature Protocols, 2009. **4**: p. 662-673.

85. Cabrera, A.C., Melo, E., Roth, D., Topp, A., Delobel, F., Stucki, C., Chen, C., Jakob, P., Banfai, B., Dunkley, T., Schilling, O., Huber, S., Iacone, R., and Petrone, P., *HtrA1 activation is driven by an allosteric mechanism of intermonomer communication.* Nature Scientific Reports, 2017. **7**.

86. Chien, J., Ota, T., Aletti, G., Shridhar, R., Boccellino, M., Quagliuolo, L., Baldi, A., and Shridhar, V., *Serine Protease HtrA1 Associates with Microtubules and Inhibits Cell Migration.* Molecular and Cellular Biology, 2009. **29**: p. 4177-4187.

87. Chien, J., He, X., & Shridhar, V., *Identification of tubulins as substrates of serine protease HtrA1 by mixture-based oriented peptide library screening.* Journal of Cellular Biochemistry, 2009. **107**: p. 253-263.

88. Clausen, T., Kaiser, M., Huber, R., and Ehrmann, M., *HTRA proteases: regulated proteolysis in protein quality control.* Nature Reviews, 2011. **12**: p. 152-162.

89. Spiess, C., Beil, A., and Ehrmann, M., *A Temperature-Dependent Switch from Chaperone to Protease in a Widely Conserved Heat Shock Protein.* Cell, 1999. **97**: p. 339-347.

90. Bradley, D.T., Zipfel, P.F., & Hughes, A.E., *Complement in age-related macular degeneration: a focus on function.* Eye, 2011. **25**: p. 683-693.

91. Stanton, C., Borooah, S., Drake, C., Marsh, JA., Campbell, S., Lennon, A., Soares, DC., Vallabh, NA., Sahni, J.,Cideciyan, AV., Dhillon, B., Vitart, V., Jacobson, SG., Wright, AF., and Hayward, C., *Novel pathogenic mutations in C1QTNF5 support a dominant negative disease mechanism in late-onset retinal degeneration.* Nature Scientific Reports, 2017. **7**: p. 12147.

92. Braun, M.B., Traenkle, B., Koch, P.A., Emele, F., Weiss, F., Poetz, O., Stehle, T., & Rothbauer, U., *Peptides in headlock – a novel high-affinity and versatile peptidebinding nanobody for proteomics and microscopy.* Nature Scientific Reports, 2016. **6**: p. 1-10.

93. Castello, A., Fischer, B. Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., Krijgsveld, J., & Hentze, M.W. and E.M.B. Laboratory, *Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins.* Cell, 2012. **149**: p. 1393-1406.

94. Glisovic, T., Bachorik, JL., Yong, J., and Dreyfuss, G., *RNA-binding proteins and post-transcriptional gene regulation.* FEBS Lett., 2008. **582**(14): p. 1977-1986.

95. Re, A., Joshi, T., Kulberkyte, E., Morris, Q., and Workman, C.T., *RNA-protein interactions: an overview.* Methods in Molecular Biology, 2014. **1097**: p. 491-521.

96. Theil, E.C., *Iron regulatory elements (IREs): a family of mRNA non-coding sequences.* Biochemical Journal, 1994. **304**: p. 1-11.

97. Pantopoulos, K., *Iron metabolism and the IRE/IRP regulatory system: an update.* Annals of the New York Academy of Sciences, 2004. **1012**: p. 1-13.

98. Yeap, B.B., Voon, DC., Vivian, JP., McCulloch, R.K., Thomson, A.M., Giles, K.M., Czyzyk-Krzeska, M.F., Furneaux, H., Wilce, M.C.J., Wilce, J.A., and Leedman, P.J., *Novel Binding of HuR and Poly(C)-binding Protein to a Conserved UC-rich Motif within the 3'-Untranslated Region of the Androgen Receptor Messenger RNA.* The Journal od Biological Chemistry, 2002. **277**: p. 27183-27192.

99. Burd, C.G.a.D.G., *RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing.* EMBO Journal, 1994. **13**: p. 1197-1204.

100. Shi, Y., *Mechanistic insights into precursor messenger RNA splicing by the spliceosome.* Nature Reviews, 2017. **18**: p. 655-670.

101.   Smith, C.W.J.V., J., *Alternative pre-mRNA splicing: the logic of combinatorial control.* Trends in Biochemical Sciences, 2000. **25**: p. 381-388.

102.   Matera, A.G., & Wang, Z., *A day in the life of the spliceosome.* Nature Reviews, 2014. **15**: p. 108-121.

103.   Scotti, M.M., & Swanson, M.S., *RNA mis-splicing in disease.* Nature Reviews Genetics, 2016. **17**: p. 19-32.

104.   Verhaart, I.E.C., Robertson, A., Wilson, I.J., Aartsma-Rus, A., Cameron, S., Jones, C.C., Cook, S.F., & Lochmüller, H., *Prevalence, incidence and carrier frequency of 5q–linked spinal muscular atrophy – a literature review.* Orphanet Journal of Rare Diseases, 2017. **12**: p. 1-15.

105.   Nlend, R., Meyer, K., & Schümperli, D., *Repair of pre-mRNA splicing: Prospects for a therapy for spinal muscular atrophy.* RNA Biology, 2010. **7**: p. 430-440.

106.   Lorson, C.L., Hahnen, E., Androphy, E.J., & Wirth, B., *A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**: p. 6307-6311.

107.   Kashima, T., & Manley, J.L., *A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy* Nature Genetics, 2003. **34**: p. 460-463.

108.   Naryshkin, N.A., Weetall, M., Dakka, A., Narasimhan, J., Zhao, X., Feng, Z., Ling, K.K., Karp, G.M., Qi, H., Woll, M.G., Chen, G., Zhang, N., Gabbeta, V., Vazirani, P., Bhattacharyya, A., Furia, B., Risher, N., Sheedy, J., Kong, R., Ma, J., Turpoff, A., Lee, C.S., Zhang, X., Moon, Y.C., Trifillis, P., Welch, E.M., Colacino, J.M., Babiak, J., Almstead, N.G., Peltz, S.W., Eng, L.A., Chen, K.S., Mull, J.L., Lynes, M.S., Rubin, L.L., Fontoura, P., Santarelli, L., Haehnke, D., McCarthy, K.D., Schmucki, R., Ebeling, M., Sivaramakrishnan, M., Ko, C.P., Paushkin, S.V., Ratni, H., Gerlach, I., Ghosh, A., Metzger, F., *Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy.* Science, 2014. **345**: p. 688-693.

109.   Sivaramakrishnan, M., McCarthy, K.D., Campagne, S., Huber, S., Meier, S., Augustin, A., Heckel, T., Meistermann, H., Hug, M.N., Birrer, P., Moursy, A., Khawaja, S., Schmucki, R., Berntenis, N., Giroud, N., Golling, S., Tzouros, M., Banfai, B., Duran-Pacheco, G., Lamerz, J., Liu, Y.H., Luebbers, T., Ratni, H., Ebeling, M., Cléry, A., Paushkin, S., Krainer, A.R., Allain,F.H.T., & Metzger, F., *Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers.* Nature Communications, 2017. **8**: p. 1476.

110.   Wang, Z., and Burge, C.B., *Splicing regulation: From a parts list of regulatory elements to an integrated splicing code.* RNA, 2008. **14**: p. 802-813.

111.   Cléry, A., Jayne, S., Benderska, N., Dominguez, C., Stamm, and S., Allain, FH., *Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1.* Nature Structural and Molecular Biology, 2011. **18**: p. 443-450.

112.   Young, P.J., DiDonato, C.J., Hu, D., Kothary, R., Androphy, E.J., and Lorson, CL., *SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1.* Human Molecular Genetics, 2002. **11**: p. 577-587.

113.   Hofmann, Y., and Wirth, B., *hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1.* Human Molecular Genetics, 2002. **11**: p. 2037-2049.

114.   Wu H., L.W.F., & Crooke S.T., *Properties of cloned and expressed human RNase H1.* The Journal od Biological Chemistry, 1999. **274**: p. 28270-28278.

115.   Havens, M.A., & Hastings, M.L., *Splice-switching antisense oligonucleotides as therapeutic drugs.* Nucleic Acids Research, 2016. **44**: p. 6549-6563.

116.   Bennett, C.F., & Swayze, E.E., *RNA Targeting Therapeutics: Molecular Mechanisms of Antisense Oligonucleotides as a Therapeutic Platform.* The Annual Review of Pharmacology and Toxicology, 2010. **50**: p. 259-293.

117.   Messina, S., *New Directions for SMA Therapy.* Journal of Clinical Medicine, 2018. **7**: p. 1-8.

118.	Crooke, S.T., Witztum, J.L., Bennett, C.F., & Baker, B.F., *RNA-Targeted Therapeutics.* Cell Metabolism, 2018. **27**: p. 714-739.
119.	Crooke, T.L., Wang, S., Vickers, T.A., Shen, W., & Liang, X., *Cellular uptake and trafficking of antisense oligonucleotides.* Nature Biotechnology, 2017. **35**: p. 230-237.
120.	Schoch, K.M.M., T.M., *Antisense Oligonucleotides: Translation from Mouse Models to Human Neurodegenerative Diseases.* Neuron, 2017. **94**: p. 1056-1070.
121.	Huang, Y., *Preclinical and Clinical Advances of GalNAc-Decorated Nucleic Acid Therapeutics.* Molecular Therapy, Nucleic Acids, 2017. **6**: p. 116-132.
122.	Sewing, S., Gubler, M., Gérard, R., Avignon, B., Mueller, Y., Braendli-Baiocco, A., Odin, M., & Moisan, A., *GalNAc Conjugation Attenuates the Cytotoxicity of Antisense Oligonucleotide Drugs in Renal Tubular Cells.* Molecular Therapy, Nucleic Acids, 2018. **14**: p. 67-79.
123.	Javanbakht, H., Mueller, H., Walther, J., Zhou, X., Lopez, A., Pattupara, T., Blaising, J., Pedersen, L., Albæk, N., Jackerott, M., Shi, T., Ploix, C., Driessen, W., Persson, R., Ravn, J., Young, J., & Ottosen S., *Liver-Targeted Anti-HBV Single-Stranded Oligonucleotides with Locked Nucleic Acid Potently Reduce HBV Gene Expression In Vivo.* Molecular Therapy, Nucleic Acids, 2018. **11**: p. 441-454.
124.	Tajrishi, M.M., Tuteja, R., & Tuteja, N., *Nucleolin; The most abundant multifunctional phosphoprotein of nucleolus.* Communicative & Integrative Biology, 2011. **4**: p. 267-275.
125.	Kotula, J.W., Pratico, E.D., Ming, X., Nakagawa, O., Juliano, R.L., & Sullenger, B.A., *Aptamer-Mediated Delivery of Splice-Switching Oligonucleotides to the Nuclei of Cancer Cells.* Nucleic Acid Therapeutics, 2012. **22**: p. 187-195.
126.	Moisan, A., Gubler, M., Zhang, J.D., Tessier, Y., Erichsen, K.D., Sewing, S., Gérard, R., Avignon, B., Huber, S., Benmansour, F., Chen, X., Villaseñor, R., Braendli-Baiocco, A., Festag, M., Maunz, A., Singer, T., Schuler, F., & Roth, A.B., *Inhibition of EGF Uptake by Nephrotoxic Antisense Drugs In Vitro and Implications for Preclinical Safety Profiling.* Molecular Therapy, Nucleic Acids, 2017. **6**: p. 89-105.
127.	Swayze, E.E., Siwkowski, A.M., Wancewicz, E.V., Migawa, M.T., Wyrzykiewicz, T.K., Hung, G., Monia, B.P., & Bennett C.F., *Antisense oligonucleotides containing locked nucleic acid improve potency but cause significant hepatotoxicity in animals.* Nucleic Acids Research, 2007. **35**: p. 687-700.
128.	Dieckmann, A., Hagedorn, P.H., Burki, Y., Brügmann, C., Berrera, M., Ebeling, M., Singer, T., & Schuler, F., *A Sensitive In Vitro Approach to Assess the Hybridization-Dependent Toxic Potential of High Affinity Gapmer Oligonucleotides.* Molecular Therapy, Nucleic Acids, 2018. **10**: p. 45-54.
129.	Shen, W., De Hoyos, C.L., Migawa, M.T., Vickers, T.A., Sun, H., Low, A., Bell, T.A., Rahdar, M., Mukhopadhyay, S., Hart, C.E., Bell, M., Riney, S., Murray, S.F., Greenlee, S., Crooke, R.M., Liang, X.H., Seth, P.P., & Crooke, S.T., *Chemical modification of PS-ASO therapeutics reduces cellular protein-binding and improves the therapeutic index.* Nature Biotechnology, 2019. **37**: p. 640-650.
130.	Font, J., & Mackay, J.P., *Beyond DNA: Zinc Finger Domains as RNA-Binding Modules.* Methods in Molecular Biology, 2010. **649**: p. 479-491.
131.	Fu, M., & Blackshear, P.J. , *RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins.* Nature Reviews Immunology, 2017. **17**: p. 130-143.
132.	Cristea, I., Graham, D., *Virology meets Proteomics.* Proteomics, 2015. **15**(12): p. 1941-1942.
133.	Douam, F., and Ploss, A., *Proteomic approaches to analyzing hepatitis C virus biology.* Proteomics, 2015. **15**(12): p. 2051-2065.
134.	Jean Beltran, P.M., Federspiel, J.D., Sheng, X., and Cristea, I.M., *Proteomics and integrative omic approaches for understanding host-pathogen interactions and infetious diseases.* Molecular Systems Biology, 2017. **13**.

135. Lum, K.K.a.C., I.M., *Proteomic approaches to uncovering virus–host protein interactions during the progression of viral infection.* Expert Review of Proteomics, 2016. **13**(3): p. 325-340.

136. Trinkle-Mulcahy, L., *Recent advances in proximity-based labeling methods for interactome mapping.* F1000Research, 2019. **8**: p. 1-12.

137. Clinton, Y.L., H., *Cross-Linking Mass Spectrometry (XL-MS): an Emerging Technology for Interactomics and Structural Biology.* Analytical Chemistry, 2018. **90**: p. 144-165.

138. Martinez-Martin, N., Marcandalli, J., Huang, C.S., Arthur, C.P., Perotti, M., Foglierini, M., Ho, H., Dosey, A.M., Shriver, S., Payandeh, J., Leitner, A., Lanzavecchia, A., Perez, L., & Ciferri C., *An Unbiased Screen for Human Cytomegalovirus Identifies Neuropilin-2 as a Central Viral Receptor.* Cell, 2018. **174**: p. 1158-1171.

139. Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., & Picotti, P., *A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication.* Cell, 2018. **172**: p. 358-372.

140. Douzi, B., *Protein-Protein Interactions: Surface Plasmon Resonance.* Methods in Molecular Biology, 2017. **1615**: p. 257-275.

141. Yurlova, L., Derks, M., Buchfellner, A., Hickson, I., Janssen, M., Morrison, D., Stansfield, I., Brown, C.J., Ghadessy, F.J., Lane, D.P., Rothbauer, U., Zolghadr, K., & Krausz, E., *The fluorescent two-hybrid assay to screen for protein-protein interaction inhibitors in live cells: targeting the interaction of p53 with Mdm2 and Mdm4.* Journal of Biomolecular Screening, 2014. **19**: p. 516-525.

142. Kakiuchi-Kiyota, S., Whiteley, L.O., Ryan, A.M., & Mathialagan, N., *Development of a Method for Profiling Protein Interactions with LNA-Modified Antisense Oligonucleotides Using Protein Microarrays.* Nucleic Acid Therapeutics, 2016. **26**.

143. Ramanathan, M., Majzoub, K., Rao, D.S., Neela, P.H., Zarnegar, B.J., Mondal, S., Roth, J.G., Gai, H., Kovalski, J.R., Siprashvili, Z., Palmer, T.D., Carette, J.E., & Khavari, P.A., *RNA–protein interaction detection in living cells.* Nature Methods, 2018. **15**: p. 207-212.

144. Stork, C., & Zheng, S., *Genome-Wide Profiling of RNA–Protein Interactions Using CLIP-Seq.* Methods in Molecular Biology, 2016. **1421**: p. 131-151.

145. Lin, C., & Miles, W.O., *Beyond CLIP: advances and opportunities to measure RBP–RNA and RNA–RNA interactions.* Nucleic Acid Research, 2019. **47**: p. 5490-5501.

146. Mohsa, R.C., & Greigb, N.H., *Drug discovery and development: Role of basic biological research.* Alzheimers Dementia, 2017. **3**: p. 651-657.

6. ANNEX

6.1 Publications

Primary literature

Parts of this thesis are published in:

Meistermann H., Gao J., Golling S., Lamerz J., Le Pogam S., Tzouros M., Sankabathula S., Gruenbaum L., Nájera I., Langen H., Klumpp K., Augustin A. A novel immuno-competitive capture mass spectrometry strategy for protein-protein interaction profiling reveals that LATS kinases regulate HCV replication through NS5A phosphorylation. Mol Cell Proteomics. 2014 Nov;13. https://www.mcponline.org/content/13/11/3040/tab-e-letters

Lugert S, Kremer T, Jagasia R, Herrmann A, Aigner S, Giachino C, Mendez-David I, Gardier AM, Carralot JP, **Meistermann H**, Augustin A, Saxe MD, Lamerz J, Duran-Pacheco G, Ducret A, Taylor V, David DJ, Czech C. Glypican-2 levels in cerebrospinal fluid predict the status of adult hippocampal neurogenesis. Sci Rep. 2017 Apr;7. https://www.nature.com/articles/srep46543

Melo E, Oertle P, Trepp C, **Meistermann H**, Burgoyne T, Sborgi L, Cabrera AC, Chen C, Hoflack JC, Kam-Thong T, Schmucki R, Badi L, Flint N, Ghiani ZE, Delobel F, Stucki C, Gromo G, Einhaus A, Hornsperger B, Golling S, Siebourg-Polster J, Gerber F, Bohrmann B, Futter C, Dunkley T, Hiller S, Schilling O, Enzmann V, Fauser S, Plodinec M, Iacone R. HtrA1 Mediated Intracellular Effects on Tubulin Using a Polarized RPE Disease Model. EBioMedicine, 2018 Jan; 27. https://www.ebiomedicine.com/article/S2352-3964(17)30492-9/abstract

Sivaramakrishnan M, McCarthy KD, Campagne S, Huber S, Meier S, Augustin A, Heckel T, **Meistermann H**, Hug MN, Birrer P, Moursy A, Khawaja S, Schmucki R, Berntenis N, Giroud N, Golling S, Tzouros M, Banfai B, Duran-Pacheco G, Lamerz J, Liu YH, Luebbers T, Ratni H, Ebeling M, Cléry A, Paushkin S, Krainer AR, Allain FHT, Metzger F. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. Nature Communications, 2017 Nov; 8. https://www.nature.com/articles/s41467-017-01559-4

Conferences:

Meistermann H., Golling S., Banfai B., Koller E., Sewing S., Kustermann S., Moisan A., Dunkley T. Development of a method to profile specific interactions between proteins and antisense oligonucleotides (ASOs). 66th American Society for Mass Spectrometry annual conference, 2018, San Diego, USA (Poster).

6.2 Contributions

ICC-MS has been designed and optimized at Roche Pharma Research & Early Development, Roche Innovation Center Basel in collaboration with Dr. Angélique Augustin and Sabrina Golling.

ICC-MS statistical model has been designed at Roche Pharma Research & Early Development Roche Innovation Center Basel in collaboration with Dr. Jens Lamerz. Statistical analysis were performed in collaboration with Dr. Jens Lamerz (NS5A, SMN, Gpc2), Dr. Balazs Banfai (SMN, OCC-MS), Dr. Gonzalo Duran-Pacheco (SMN, Gpc2) and Dr. Juliane Siebourg-Polster (Htra1) at Roche Pharma Research & Early Development Roche Innovation Center Basel.

- <u>Biological applications of ICC-MS & OCC-MS</u>

- Characterization of NS5A protein interactome in a cellular model of hepatitis C virus replication (Biology leader Dr. Jujun Gao).
- Characterization of Glypican-2 protein interactome in a human neural stem cell model (Biology leader Dr. Sebastian Lugert).
- Characterization of the HtrA1 protein interactome in a polarized retinal pigmented epithelial (RPE) disease model (Biology leader Dr. Esther Melo).
- Application to decipher survival of motor neuron (SMN) splicing modifiers mode of action (Biology leaders Dr. Friedrich Metzger and Dr. Manaswini Sivaramakrishnan).
- Application to GalNac and naked ASOs with distinct *in vitro* nephrotoxicity prediction (Biology leaders Dr. Sabine Sewing and Dr. Annie Moisan).
- Application to ASOs with distinct *in vitro* hepatoxicity profiles (Biology leader Dr. Andreas Dieckmann).

6.3 Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die zur Promotion eingereichte Arbeit mit dem Titel: «A novel unbiased approach to study endogenous protein-protein interactions» selbständing verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich erkläre, dass die Richtlinien zur Sicherung guter wissenschaftlicher Praxis der Universität Tübingen (Beschluss des Senats vom 25.5.2000) beachtet wurden. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir is bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe von bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen den, 22.05.2020 _____

Hélène Meistermann

6.4 Acknowledgements