Functional characterization of a novel p97:ASPL protein complex

Dissertation to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

Submitted to the Department of Biology, Chemistry and Pharmacy

Freie Universität Berlin

by

Simona Kostova

January 2020

This work was conducted from January 2016 until December 2019

under the supervision of Prof. Dr. Erich E. Wanker

at the Max Delbrück Center for Molecular Medicine

in the Helmholtz-Association.

Hereby I declare that I have authored this thesis independently

and used explicitly the declared sources.

1st Reviewer: Prof. Dr. Erich E. Wanker

2nd Reviewer: Prof. Dr. Stephan Sigrist

Thesis defense on: 03.07.2020

To A. S.

To the person who always encouraged me to never stop learning…

as one should only regret the lack of knowledge.

Summary

Investigating protein-protein interactions (PPIs) in the human cell has proven to be a long-lasting endeavor that required the elaborate work of a few generations of scientists. Nowadays, interactome mapping approaches gain more and more interest as they not only deliver insights into the molecular processes which shape cell function and fate, but also represent a so far unexploited source of potential new therapeutic targets. The protein standing in the focus of my work is p97 and despite the great amount of literature describing its functional diversity, still not much is known how mutations in the p97 gene lead to the manifestation of a multisystem disorder called IBMPFD. Nevertheless, some publications reported perturbations in the p97 interaction network to be caused by disease-associated single amino acid exchanges in the protein. This suggests that there is a connection between disease onset or progression and the regulation of the p97 interactome.

In order to test this hypothesis, I decided to follow a systematic approach and study in detail second order interaction partners of p97. My aim was to gather information about the mechanisms under their regulation and their disease relevance. One particularly good target for an investigation of this kind is the stable p97:ASPL complex. ASPL in human cells is predominantly present in a p97 bound state and finding ASPL interaction partners increases the chance of detecting complex binders. A SILAC analysis of immunoprecipitated endogenous ASPL and the following validation of its putative interaction partners by LuTHy revealed first indications of p97:ASPL being involved in processes related to pre-mRNA splicing. Three of the validated interaction partners of the complex are involved in the initiation of splicing and spliceosome assembly by recognizing the 3' splice site in intron sequences. These are the structural homologs PUF60, RBM39 and U2AF2. The interaction between U2AF2 and p97:ASPL is specific and the splicing factor appears to bind to a surface which is only present in the complex. This I was able to show by applying ASPL mutants in the interaction tests. They bind to p97 with a reduced affinity, but most importantly do not form an oligomeric structure with p97 as the wild-type ASPL molecules do. In a domain mapping analysis, I was able to define the domains of U2AF2 responsible for its association with the complex. These are the N-terminal RS-ULM region, which is essential

for binding to p97, and the C-terminal UHM domain which is sufficient for the interaction with ASPL. These experiments suggest a bipartite association mode between p97:ASPL and U2AF2. Additionally, I was able to demonstrate their endogenous interaction by a proximity ligation assay (PLA). By a thorough RNA-Seq analysis performed in HEK293 cells I showed that the depletion of ASPL affects the splicing pattern of functionally diverse transcripts in human cells. This effect is dependent on intrinsic sequence features such as the strength of the 3' splice site. Cassette exons under the regulation of weak 3' splice sites tend to get skipped more frequently upon ASPL knock-out. I was able to confirm this observation with a specific U2AF2 minigene reporter applied in HeLa cells.

Finally, I analyzed the effects of single amino acid exchanges in 26 mutated p97 variants regarding their binding to ASPL and U2AF2. The majority of the tested point mutations affected strongly the interaction between p97 and ASPL and a few of them modified the interaction signal detected between p97 and U2AF2. However, it remains unclear if these mutations lead to fluctuations in alternative splicing in IBMPFD patient derived cells similar to changes observed in ASPL-KO cells.

In summary, I can claim that p97:ASPL complexes in human cells have a general role in early splicing events. They are required for maintaining cassette exons with weak 3' splice sites in matured transcripts. The interaction of p97:ASPL with U2AF2 and other related splicing factors, such as PUF60 and RBM39, is critical for this process.

Zusammenfassung

Die Untersuchung von Protein-Protein-Wechselwirkungen in der menschlichen Zelle hat sich als langwieriges Unterfangen erwiesen, das die aufwendige Arbeit einiger Generationen von Wissenschaftlern erforderte. Heutzutage gewinnen Ansätze zum Studieren des Interaktoms zunehmend an Interesse, da sie nicht nur Einblicke in die molekularen Prozesse liefern, die die Funktion und das Schicksal der Zellen beeinflussen, sondern auch eine bisher ungenutzte Quelle zur Identifizierung potenzieller neuer therapeutischer Targets darstellen. Das Protein, das im Mittelpunkt meiner Arbeit steht, ist p97. Bis heute ist trotz einer umfassenden Menge an Literatur, die seine funktionelle Vielfalt beschreibt, noch nicht viel darüber bekannt, wie Mutationen im Gen, das p97 kodiert, zur Manifestation einer Multisystemstörung mit dem Namen IBMPFD führen. Dennoch berichten einige Veröffentlichungen, dass Änderungen im p97-Interaktionsnetzwerk durch den Austausch einzelner krankheitsassoziierten Aminosäuren im Protein verursacht werden. Dies deutet darauf hin, dass ein enger Zusammenhang zwischen dem Ausbruch oder dem Krankheitsverlauf und der Störung des p97-Interaktoms besteht.

Um diese Hypothese zu testen, beschloss ich, einen systematischen Ansatz zu verfolgen und die p97-Interaktionspartner zweiter Ordnung im Detail zu untersuchen. Damit wollte ich Informationen über die von p97-Interaktionspartnern regulierten zellulären Mechanismen und deren Krankheitsrelevanz sammeln. Als besonders geeignet hierfür erweist sich der stabile p97:ASPL-Komplex. ASPL ist in der menschlichen Zelle überwiegend in einem p97-gebundenen Zustand vorhanden, und das Identifizieren von ASPL-Interaktionspartnern erhöht die Wahrscheinlichkeit Proteine zu finden, die an den p97:ASPL Komplex binden. Eine SILAC-Analyse von immunpräzipitiertem endogenem ASPL und die folgende Validierung seiner mutmaßlichen Interaktionspartner durch LuTHy lieferte erste Hinweise darauf, dass p97:ASPL an Prozessen beteiligt ist, die mit dem Reifen von prä-mRNA Molekülen in Verbindung stehen. Drei der validierten Interaktionspartner des Komplexes sind an der Initiierung vom Splicing und der Assemblierung des Spliceosoms beteiligt, indem sie die 3'-Splice Stelle in Intronsequenzen erkennen. Dies sind die strukturellen homologen Proteine PUF60, RBM39 und U2AF2. Die Wechselwirkung zwischen U2AF2 und p97:ASPL ist

spezifisch und der Splicing Faktor scheint an eine Oberfläche zu binden, die nur vorhanden ist, wenn die Proteine p97 und ASPL einen Komplex eingehen. Dieses Ergebnis konnte ich durch den Einsatz von ASPL-Mutanten und deren Untersuchung in spezifischen Interaktionstests erzielen. Sie binden mit einer verringerten Affinität an p97, bilden aber vor allem keine oligomeren Komplexe mit p97, die normalerweise mit Wildtyp-ASPL-Molekülen entstehen. In einer Domain Mapping Analyse konnte ich die Domänen in U2AF2 identifizieren, die für die Assoziation mit dem Komplex verantwortlich sind. Dies sind die N-terminale RS-ULM-Region, die für die Bindung an p97 essentiell ist, und die C-terminale UHM-Domäne, die für die Wechselwirkung mit ASPL ausreicht. Zusätzlich konnte ich die endogene Wechselwirkung zwischen den Proteinen mit einem Proximity Ligation Assay (PLA) nachweisen. Durch eine gründliche RNA-Seq-Analyse, die in HEK293-Zellen durchgeführt wurde, zeigte ich, dass die unterbundene Expression von ASPL das Splicing Muster von Transkripten beeinflusste, die nicht funktionell verwandt sind. Dieser Effekt ist abhängig von intrinsischen Sequenzmerkmalen wie der Stärke der 3'-Splice Stelle. Kassetten-Exons, die unter der Regulierung schwacher 3'-Splice Stellen stehen, neigen dazu, beim ASPL-Knock-out häufiger übersprungen zu werden. Diese Beobachtung konnte ich mit einem spezifischen U2AF2-Minigen-Reporter in HeLa-Zellen bestätigen.

Schließlich analysierte ich den Effekt des Austausches einzelner Aminosäuren in 26 mutierten p97-Varianten hinsichtlich ihrer Bindung an ASPL und U2AF2. Die Mehrzahl der getesteten Punktmutationen beeinflusste die Wechselwirkung zwischen p97 und ASPL stark und einige von ihnen modifizierten das ermittelte Interaktionssignal zwischen p97 und ASPL. Es bleibt jedoch unklar, ob diese Mutationen zu ähnlichen Schwankungen im alternativen Splicing von Zellen von IBMPFD Patienten führen, wie ich es in ASPL-KO-Zellen beobachtet hatte.

Zusammenfassend kann ich sagen, dass meine Arbeiten darauf hindeuten, dass p97:ASPL-Komplexe eine wichtige Rolle beim Splicing von mRNA Transkripten in menschlichen Zellen spielen. Der Komplex ist erforderlich, um die Integrationsraten von Kassetten-Exons mit schwachen 3'-Splice Stellen aufrechtzuerhalten. Die Wechselwirkung zwischen p97:ASPL und U2AF2 und den anderen verwandten Splicing Faktoren wie PUF60 und RBM39 ist von großer Bedeutung für diesen Prozess.

Abbreviations

Abbreviations	Meaning
2xPA	double **P**rotein-**A**
AAA+ ATPase	**A**TPase **A**ssociated with diverse cellular **A**ctivities
ASPL	**A**lveolar **S**oft **P**art sarcoma **L**ocus
ATP	**A**denosine triphosphate
AU	**a**rbitrary **u**nits
BCA	**Bic**inchoninic **a**cid assay
BN PAGE	**b**lue **n**ative **p**oly**a**crylamide **g**el **e**lectrophoresis
BP	att**B** att**P**
BRET	**B**ioluminescence **R**esonance **E**nergy **T**ransfer
BSA	**B**ovine **s**erum **a**lbumin
cBRET	**c**orrected **BRET**
cLuC	**c**orrected **LuC**
CoIP	**co**-**i**mmuno**p**recipitation
CRISPR	**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats
d.m.	**d**ouble **m**utant
DAVID	**D**atabase for **A**nnotation, **Vi**sualization and **I**ntegrated **D**iscovery
DNA	**D**eoxyiribo**n**ucleic **a**cid
HEK293	**H**uman **E**mbryonic **K**idney 293 cells
HEPES	4-(2-**h**ydroxy**e**thyl)-1-**pi**perazine**e**thane**s**ulfonic acid
IBMPFD	**I**nclusion **B**ody **M**yopathy with **P**aget's disease of the bone and **F**rontotemporal **D**ementia
kb	**k**ilo**b**ases
kDa	**K**ilo**da**lton
LR	att**L** att**R**
LuC	**Lu**minescence-based **C**o-precipitation
LuTHy	**Lu**minescence-based **t**wo-**hy**brid assay
mCit	**mCit**rine
mRNA	**m**essenger **r**ibo**n**ucleic **a**cid
NL or **NanoLuc**	**Nano Luc**-Luciferase
PBS	**P**hosphate-**B**uffered **S**aline
PdB	Pladienolide B
PEI	**P**oly**e**thylen**i**mine
PLA	**P**roximity **L**igation **A**ssay
PPI	**P**rotein-**P**rotein **I**nteraction
PSI	**p**ercentage **s**pliced **i**n
RBP	**R**NA **b**inding **p**rotein
RRM	**R**NA **r**eocgnition **m**otif
RS-domain	Arg/Ser rich domain
s.m.	**s**ingle **m**utant
SDS	**S**odium **d**odecyl **s**ulfate
SDS-PAGE	**SDS p**oly**a**crylamide **g**el **e**lectrophoresis
sgRNA	single guide RNA
SILAC	**S**table **I**sotope **L**abeling with **A**mino acids in **C**ell culture
snRNP	**s**mall **n**uclear **Ri**bo**n**uclear **p**rotein
STRING	**S**earch **T**ool for the **R**etrieval of **I**nteracting **G**enes/Proteins
U2AF2	**U2** Small Nuclear RNA **A**uxiliary **F**actor **2**
UBX	**ubi**quitin-like domain
UHM	**U**2AF **h**omology **m**otif
ULM	**U**HM **l**igand **m**otif
VCP	**v**alosin-**c**ontainin **p**rotein

Table of contents

List of Figures and Tables

1. Introduction

1.1 Protein-protein interactions and their scientific relevance

Thinking of the human proteome as a rigid entity of a finite number of polypeptides is certainly convenient when the composition of a specific sample is the only focus of interest. However, in terms of understanding cellular functionality and disease pathology at the molecular level this might be too much of a simplification[1]. Events like alternative splicing and post-translational modifications are strongly increasing the proteome complexity[2,3], but protein-protein interactions (PPIs) and the networks they comprise are what manifests cellular characteristics[1,4,5].

The significance of studying PPIs is elevated not only by the central role they play in cellular physiology, but also by the fact that ~60% of disease-associated mutations lead to perturbations in the interaction pattern of the mutated protein rather than affecting its folding properties or stability[6]. What has been referred to as the "edgotype" of human cells is a comprehensive network model including macromolecules as "nodes" and their interactions as "edges". Thus, the edgotype model links the genotype with the phenotype[7] (**FIG. I1**). Proteome and interactome-wide studies of complex diseases, such as autism[8,9] or schizophrenia[10], performed in the last years, demonstrated the potential of edgotype analysis for revealing disease-related molecular pathways and common mechanisms in these genetically heterogeneous disorders.

Besides point mutations, changes in the expression levels of essential proteins can shift the equilibrium of processes under their control by prioritizing certain interactions over others. One of the best studied interaction networks with a central role in the regulation of programmed cell death comprises proteins of the BCL-2 family[11]. A very common mechanism leading to survival of cancer cells in different lymphomas is the upregulation of BCL-2 and other anti-apoptotic proteins[12]. BCL-2 interacts with the pro-apoptotic protein BAD and inhibits its role in cell death initiation. Thus, it is not surprising that a very successful therapeutic approach was established by developing a small molecule which mimics the binding surface of BAD and binds into the grooved pocket of BCL-2 with high affinity resulting in the release of BAD molecules[13]. This interaction inhibitor – named Venetoclax – was released to the market by AbbVie

in 2016 and selling numbers are increasing ever since. It is expected that the compound will turn into one of the top 15 best-selling cancer drugs by 2022 (Fierce Pharma press release from January 2017).

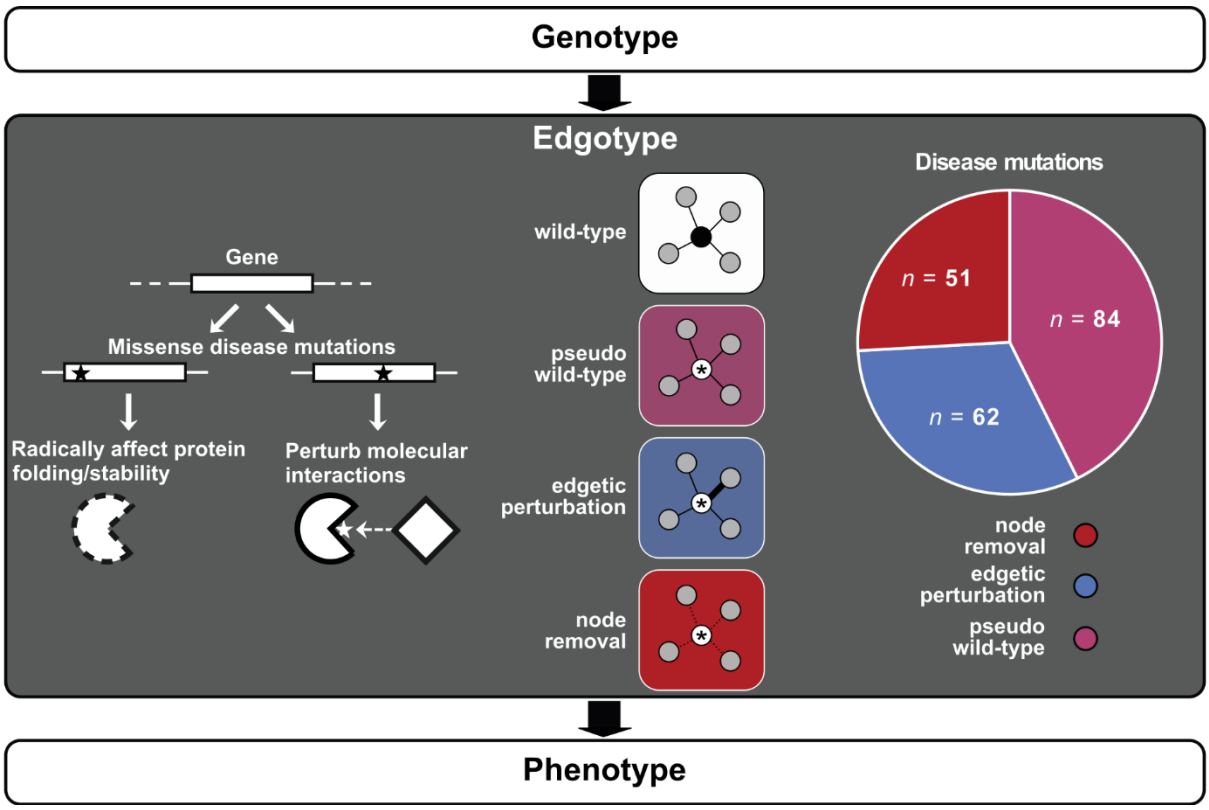


FIG. I1: Disease-associated mutations affect PPIs.

The human interactome as a link between genotype and disease phenotype as suggested by Sahni *et al.* (2015 & 2013). Taken together, ~60% of the analyzed missense disease-related mutations lead either to perturbations or complete loss of PPIs.

Having a quick glance at a few numbers is already sufficient to pinpoint the obvious demand on further interactome investigations. For instance, from the total number of predicted PPIs within the human interactome, estimated to be more than 650,000 by Stumpf *et al.*[14] in 2008 (**FIG. I2A**), less than 0.3% were identified, according to a commentary published by Amaral in PNAS in the same year[15]. Until 2014 around 41,000 PPIs have been annotated as reported by Szilagyi and Zhang[16], accounting for ~6% of the search space, meaning that in a time frame of six years, from 2008 to 2014, only 5.7% of the human interactome were unraveled. If we further follow on the progress in this field, we can see the immense increase in the number of annotated interactions

from the generation of the first human PPI network by Stelzl and colleagues[17] in 2005 until the release of the second update of the **HIPPIE** database in 2017 (**H**uman **I**ntegrated **P**rotein-**P**rotein **I**nteraction r**e**ference) (**FIG. I2B**). Stelzl *et al.* (2005)[17] reported 911 high-confidence PPIs. In comparison, in the publication by Alanis-Lobato *et al.* (2017)[18] in the **HIPPIE v2.0** database ~270,000 experimentally detected PPIs were described. Despite this increase in PPIs, it becomes clear that twelve years of research were not enough to annotate the interactions taking place in the human cell in their full scope.

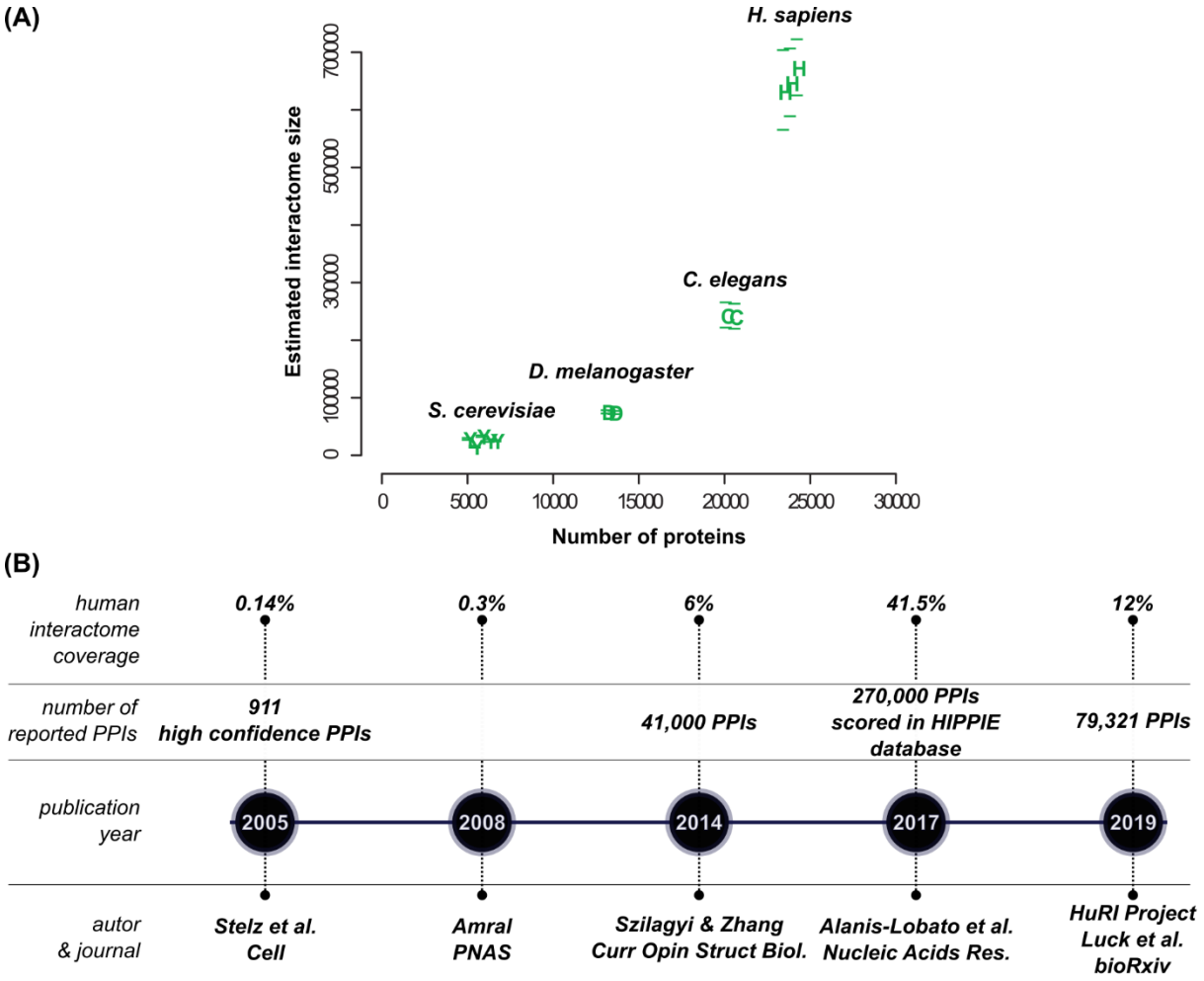


FIG. I2: Human interactome exploration from 2005 until 2019.

(A) Numbers of proteins and their interactions in different species as estimated by Stumpf *et al.* (2008). **(B)** Timeline of publications contributing to the process of deciphering the human interactome. Coverage percentage was calculated by dividing the numbers of PPIs reported in the listed papers by 650,000, the total number of estimated PPIs in the human interactome by Stumpf *et al.* (2008).

In comparison, the human genome project was officially launched in 1990 and in 2003 a draft sequence was made public, covering 99% of our genome[19]. Additionally, some interactome large-scale studies are still ongoing, e.g. **HuRI**, the **Hu**man **R**eference Protein **I**nteractome Mapping Project directed from the Center for Cancer Systems Biology (CCSB), which includes initial data published by Rual *et al.*[20] in 2005. So far 79,321 PPIs have been identified by a systematic high-throughput yeast two-hybrid approach. Collectively, these facts highlight the challenges of an endeavor of such a scale. One plausible explanation is the chemical and structural heterogeneity of polypeptides compared to other macromolecules, accompanied by multilateral regulatory mechanisms controlling the proteome[2]. However, the next hurdle on the way towards systematic mapping of the human interactome will be the assignment of specific functional characteristics to newly detected PPIs.

1.2 p97 is one of the central interaction hubs of the human cell

Proteins in the cell interact in a strictly orchestrated manner upon specific molecular signals (phosphorylation, methylation, acetylation, glycosylation, etc.) that can result in a modified binding hierarchy within a PPI network. One particularly good example of a protein which plays a central part in a large interaction network that is strictly regulated is p97, an essential hexameric AAA+ ATPase also known as VCP (valosin-containing protein) in vertebrates or Cdc48 in yeast cells[21,22]. It is highly conserved in archaea and eukaryotes[23] and represents one of the most abundantly expressed proteins in the human cell[24]. Each protomer in the symmetric structure of the p97 hexamer has the same globular domain architecture comprised of an N-terminal domain, two consecutive ATPase domains (D1 and D2) and a disordered C-terminus (**FIG. I3A**). Both ATPase domains are stacked upon each other forming the typical double-barrel ring-shaped structure of the enzyme with a central pore of ~23Å in diameter[25]. The pore serves mainly for the threading of polyubiquitinated substrates and their unfolding which makes them accessible for the proteasome or facilitates the extraction of proteins from complexes of higher molecular weight[26,27]. The N-terminus is situated outside of the ring and can be structurally divided in two distinct subdomains – Nn and Nc. They are connected by a short linker sequence and form a

hydrophobic binding pocket, which binds a great number of p97 interaction partners[28]. The ATPase activity is not only providing p97 with energy for the extraction of polyubiquitylated substrates from membranes or higher order molecular structures[29], it also induces specific intra-[30] and interdomain[31] conformational changes leading to a constant up and down movement of the N-terminus[32]. According to Schuller *et al.* (2016)[32] these "swinging motions" of the N-terminal domains relative to the ring-like structure of the p97 hexamer are translating the energy from the ATP hydrolysis cycle into the mechanic force required for its segregase activity.

What makes p97 one of the best examples of the AAA+ ATPase family are the numerous and manifold cellular functions it executes. This major keeper of protein homeostasis was very appropriately described as the "swiss army knife" of the human cell by Baek and colleagues (2013)[21] in their review article on p97. This label appears to be even more suitable when one goes through the myriad of processes p97 is involved in, starting with ER-associated degradation (ERAD) and other proteasome degradation pathways, such as ribosomal- (RAD), mitochondrial- (MAD) and chromatin-associated degradation (CAD) followed by Golgi reassembly and endosomal or membrane trafficking, as well as cell cycle progression and DNA damage response and repair mechanisms[21,33–35]. The different fates a substrate can face once processed by p97, such as extraction, segregation, degradation or recycling[27,33,36], is one of the factors explaining the diversity in its molecular activities. However, the driving force behind this functional versatility are p97 binding partners, also termed cofactors[35]. p97 interacts with more than 30 cofactors[34,35] and one of the best described p97 interaction partners are the members of the UBX protein family[37] (**FIG. I3B**). Characteristic for this group is the presence of an ubiquitin-like domain, UBX, which often mediates the interaction with p97. A few other domains, such as PUB (PNGase/UBA or UBX containing proteins), VIM (VCP-interacting motif) or the SHP-binding motif (BS1, binding segment 1), complement the UBX domain in p97 interacting proteins and also can influence the association of partner proteins with p97[31].

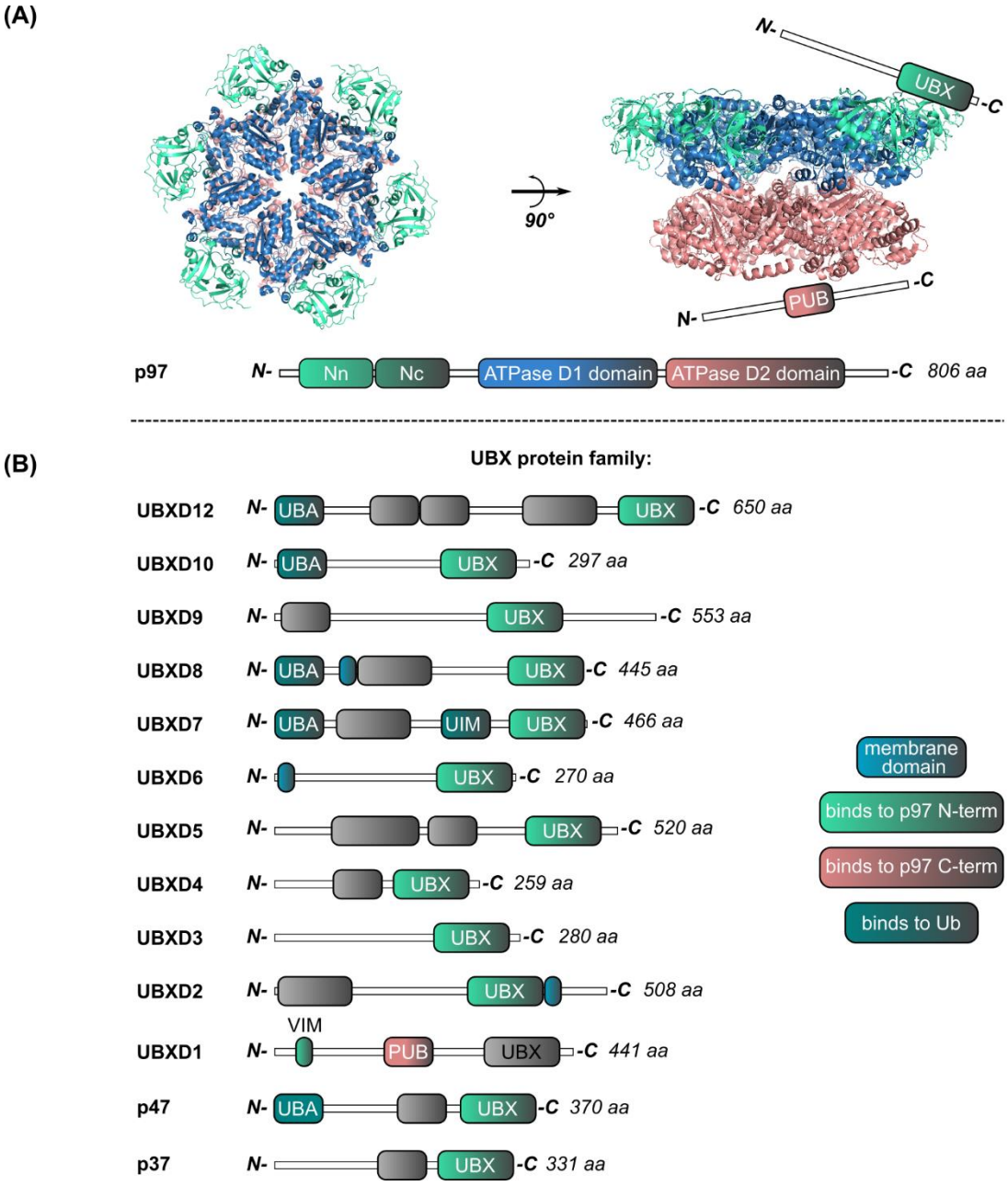


FIG. I3: p97 and its interaction partners from the UBX protein family.

(A) Ribbon representations of the structure of the p97 hexamer in top and side view. PDB structure accession number: 5FTK. In the side view the binding of p97 interaction partners containing a UBX or PUB domain is presented schematically. The domain composition of each 97 protomer includes an N-terminal domain which can be divided in two subdomains – Nn and Nc – and two ATPase domains (D1& D2). **(B)** Schematic overview of the domain structure of the 13 members of the UBX protein family expressed in human cells. The characteristic UBX domain is depicted in green. In UBXD1 this domain is not highlighted as it does not participate in the interaction with p97. For simplicity in all listed proteins the domains which are not essential for p97 or ubiquitin binding are depicted as gray unlabeled boxes. UBX: ubiquitin-like domain; UBA: ubiquitin-associated domain; VIM: VCP-interacting motif; PUB: Peptide:N-glycanase/UBA or UBX-containing domain.

1.3 ASPL: the unique p97 interaction partner

Among all p97 interaction partners there is one which stands out in a very noticeable way. This protein is **ASPL**, also known as UBXD9 or ASPSCR1; the abbreviation stands for **a**lveolar **s**oft **p**art sarcoma **l**ocus. The name originates from a genetic aberration which results in an in-frame gene fusion between the N-terminal part of ASPL and exon-3 or exon-4 of the transcription factor TFE3 after a translocation affecting chromosomes X and 17[38]. This abnormal gene product is believed to contribute to the uncontrolled cell proliferation in soft tissues, mainly in the deep soft tissues of the extremities of young adults, which often spreads to the lungs[39,40]. Despite the occurrence of other TFE3 oncogenic fusions described in patients with different types of alveolar soft part carcinoma (ASPS), ASPL-TFE3 is the most frequent[41]. Besides the observed up-regulation of the Met receptor tyrosine kinase the molecular role of ASPL-TFE3 remains unclear[42].

The wild-type full-length ASPL protein is a 553 amino acid long polypeptide, which possesses a characteristic UBX domain discriminating it from other UBX proteins. The exhaustive investigation of the structure of ASPL in complex with p97 executed by Arumughan and colleagues (2016)[43] contributed enormously to our current knowledge about this protein. In this study, it was shown that the canonical UBX amino acid sequence in ASPL is not sufficient for its binding to p97. The so called extended UBX domain (eUBX) in ASPL includes an additional β-grasp fold and two α-helices forming a lariat preceding the UBX domain which is followed by further two α-helical extensions (**FIG. I4A**). Apart from these extra features in the eUBX domain ASPL is the only p97 interaction partner that is capable of disrupting the stable hexamer. As demonstrated by Arumughan *et al.* (2016)[43], this process is accompanied by the formation of a new stable oligomeric structure, most probably a heterotetramer consisting of two ASPL and two p97 molecules (**FIG. I4B**). One of the molecular hallmarks of this p97:ASPL complex is its abolished ATPase activity caused by the drastic conformational change and spatial reorientation of the D2 domain in the heterotetramer. Thus, overexpression of ASPL in human cells affects ERAD and induces cell death[43]. Based on these structural and biochemical findings it was assumed that ASPL primary functions as an inhibitor of p97 activity. Whether the cellular role of ASPL is to prevent the excess of functioning

p97 hexamers needs to be determined and this important question will be addressed in the present study.

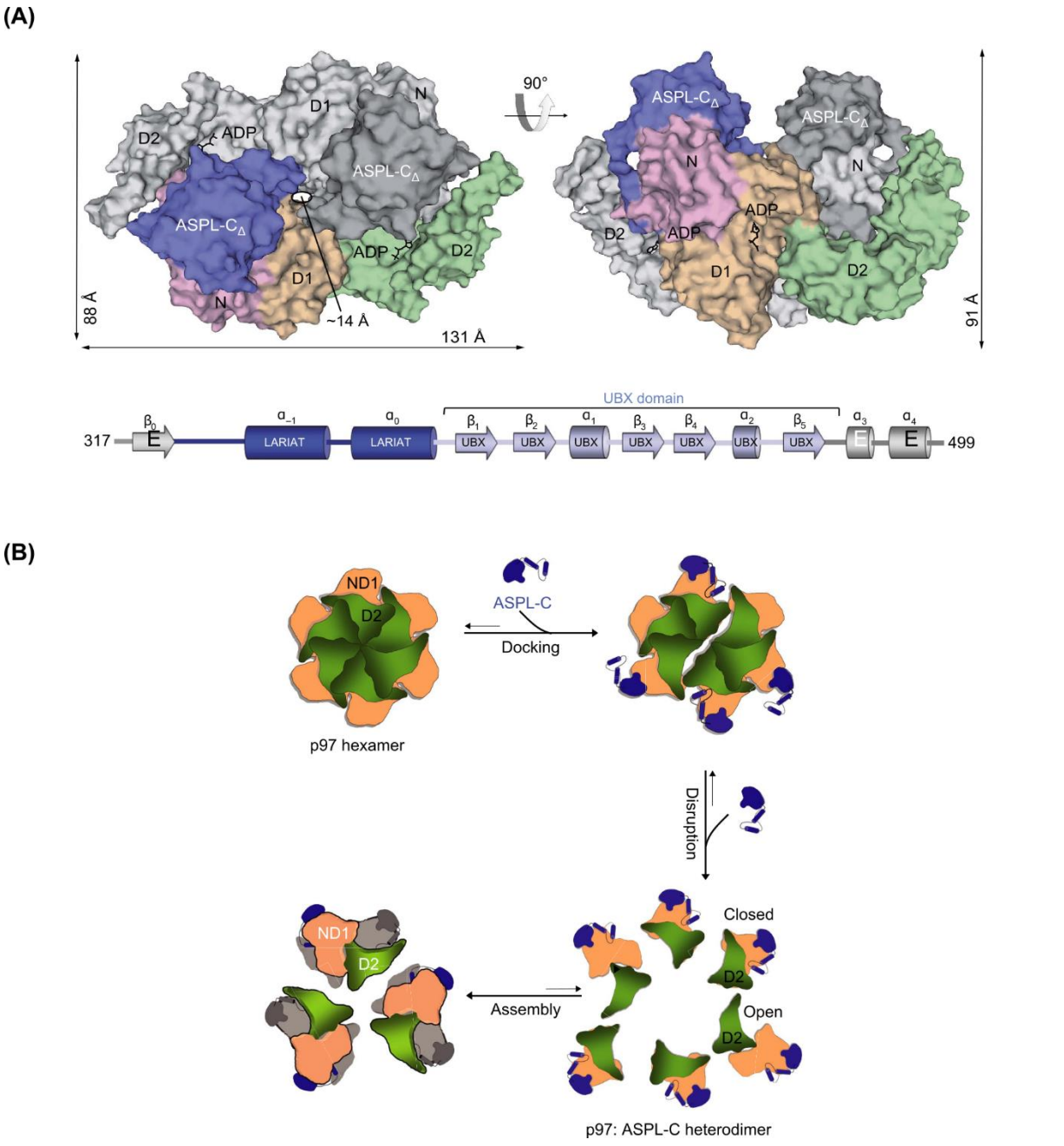


FIG. I4: Characteristics of the interaction between p97 and ASPL.

(A) p97:ASPL heterotetramer containing two ASPL and two p97 molecules shown here as surface representation of a structure obtained with the depicted C-terminal ASPL fragment (ASPL-C) and p97-ND1-D2. **(B)** Model elucidating the disruption of a p97 hexamer by ASPL and formation of the p97:ASPL heterotetramer. Images from Aramughan *et al.* (2016), Nature Communications.

1.4 p97 is a disease-relevant protein

Mutations in the gene encoding p97 are described as causative of a complex late-onset multisystem proteinopathy named IBMPFD (Inclusion Body Myopathy with Paget Disease of the Bone and Frontotemporal Dementia)[44]. Behind the pivotal role p97 plays in the cell there are multiple regulatory mechanisms enabling the fine tuning and control of its interactions. This involves binding site competition of interacting proteins, hierarchical protein binding, mutual exclusion of binding partners, bipartite binding of interacting proteins and conformational changes in p97 and interacting proteins[28]. Therefore, it is not surprising that there is a great body of literature emphasizing the disturbances point mutations cause in the p97 interaction network[40-42]. Furthermore, p97 has been shown to be related to cancer and aging[48–50], which supports the relevance of this protein as a research target.

Taking this into account, studying the interactome of p97 appears to be an important approach to gain new information not only about fundamental cellular pathways, but also to identify disease-related targets. Hence, my PhD project is focused on the detailed analysis and characterization of the so far less known p97:ASPL heterotetrameric complex[43]. This strategy aims at the identification of novel p97:ASPL interaction partners and the investigation of their potential as therapeutic targets in the long-term.

1.5 LuTHy: more than just a beautiful name

Methods and tools for PPI prediction, detection and quantification have been developing rapidly in the last few decades and new techniques arise constantly[51,52]. From yeast two-hybrid and immunoprecipitation methods, over resonance energy transfer-based assays to mass spectrometry-based technologies, each approach is characterized by its specific advantages and limitations. The importance of combining different assay versions in order to increase the overall PPI detection efficiency in a certain proteome space was recently highlighted by Choi and colleagues (2019)[53]. They were able to successfully modify the NanoBit technology and to perform 24 measurements per binary PPI which dramatically raised the total PPI detection propensity.

Along the line of these thoughts, a particularly good example of a combinatory assay with high PPI recovery rates when compared to single-readout methods is LuTHy, developed by Trepte *et al.* (2018)[54]. LuTHy is a double-readout approach which combines the in-cell detection of the PPI of interest via bioluminescence resonance energy transfer (BRET) with a luminescence-based co-precipitation assay (LuC) in cell extracts (**FIG. I5**).

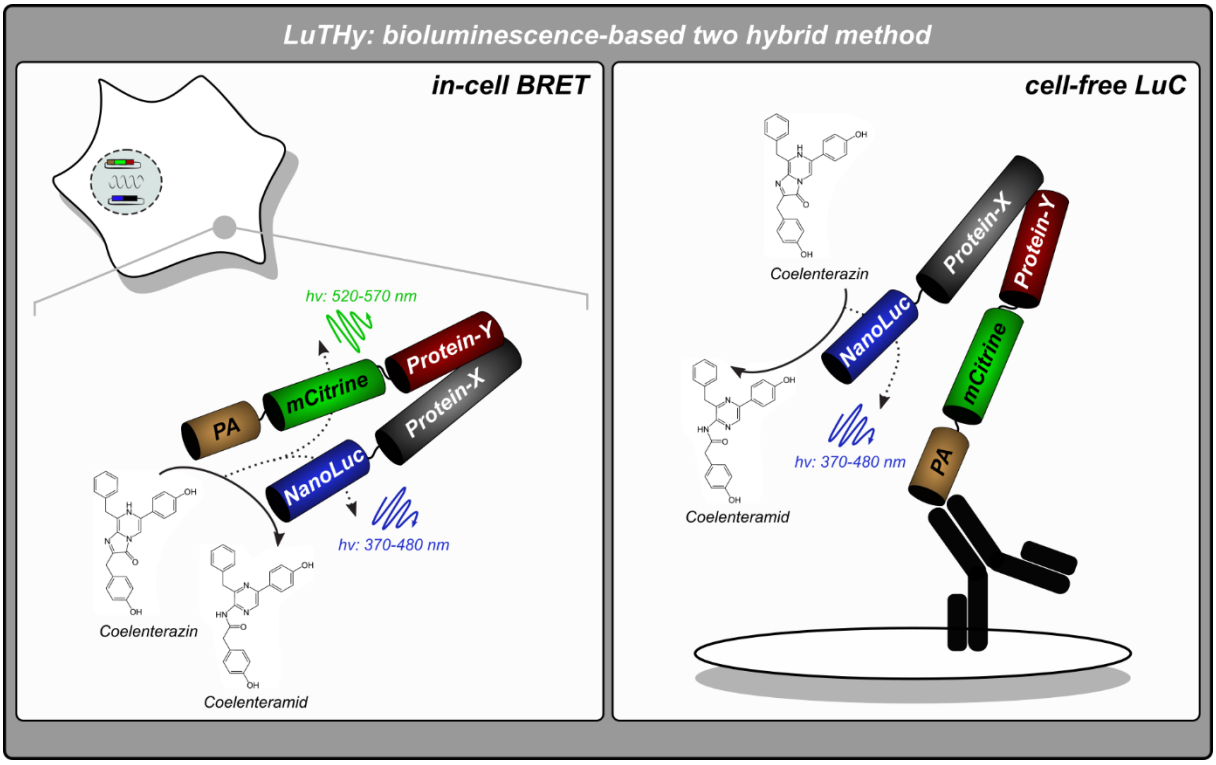


FIG. I5: LuTHy: a double-readout bioluminescence-based approach for PPI quantification.

The graphical representation of the BRET and LuC components of LuTHy was adapted from Mikko Taipale's comment (Europe PMC) on Trepte *et al.* (2018). During the in-cell BRET measurement bioluminescence energy transfer quantification provides the information about the tested binary PPI of interest. LuC is a distance independent readout since no energy transfer, but only the luminescence emitted by the co-precipitated interaction partner is detected.

In order to apply LuTHy for PPI quantification, HEK293 cells are first co-transfected with plasmids for the expression of fusion constructs comprised of the proteins of interest with the respective donor or acceptor tags. Proteins tagged as donors are fused to NanoLuc (NL) and the ones applied as acceptors are labeled with a double Protein-A-mCitrine construct (2xPA-mCit). For every tested PPI two negative controls

are measured. In this case each protein of interest is respectively co-expressed with either NanoLuc or mCitrine only constructs depending on the tag it carries. Following the addition of the specific NanoLuc substrate Coelenterazin-h to the cells it gets converted to Coelenteramid. The energy released by the chemical reaction leads to the emission of visible light with a maximum peak at 460 nm. The resulting energy can be transferred to the acceptor, mCitrine, only if the distance between the interacting proteins is below 10 nm. In this case mCitrine is excited and emits light with a maximum at 530 nm. Luminescence is then measured in the long wavelength area (LWL: 520-570 nm) and in the short wavelength area (SWL: 370-480 nm), respectively. The ratio between them (**Formula (1)**) is corrected for the donor bleed-through measured with a PA-NanoLuc control sample ($Cf_{(BRET)}$) (**Formula (2)**). The corresponding BRET ratios for both negative controls are calculated accordingly and the cBRET (corrected BRET) for each tested binary PPI results from subtracting the higher BRET of the two corresponding negative controls. The empirically estimated threshold by Trepte *et al.* (2018)[54] for cBRET values is 0.01.

$$\boldsymbol{Formula}\ (\mathbf{1}){:}\ BRET\ ratio = \frac{LWL}{SWL} - Cf$$

$$\boldsymbol{Formula}\ (\mathbf{2}){:}\ Cf_{(BRET)} = \frac{LWL_{PA-NL}}{SWL_{PA-NL}}$$

In the LuC readout the luminescence from the NanoLuc in the range between 370-480 nm is detected in lysates incubated in coated (NL_{OUT}) or non-coated microtiter plates (NL_{IN}). Only in the IgG coated plates PA-tagged-mCitrine proteins can be precipitated and the strength of the luminescent signal correlates with the amount of the interacting protein which is tagged with NanoLuc. The ratio between NL_{OUT} and NL_{IN} is normalized to the emission measured with a control sample containing PA-tagged NanoLuc ($LuC_{(PA-NL)}$) (**Formulas (3)** and **(4)**). The multiplication factor in front of NL_{IN} in **Formula 3** and **4** corresponds to the difference in the volume of the lysates applied to the non-coated plates compared to the coated ones. The resulting cLuC values are corrected by subtracting the negative control values, similar to the cBRET calculations, and classified as positive if they are above the cLuC threshold of 0.03.

$$\textbf{Formula (3)}: LuC_{PA-NL} = \frac{NL_{OUT}}{2 \times NL_{IN}}$$

$$\textbf{Formula (4)}: LuC = \frac{NL_{OUT} / (2 \times NL_{IN})}{LuC_{PA-NL}}$$

Trepte *et al.*[54] demonstrated not only the higher sensitivity of the LuTHy technology in comparison to single-readout methods, but also experimentally proved its suitability for the quantification of the effects caused by treatment with small molecule, stress induction or point mutations on PPIs. Combined with the relatively short preparation and execution times, these factors convinced me that LuTHy is the most suitable technique for the PPI investigations I was planning to conduct.

1.6 Aim and strategy

The main goal of my study was the identification and validation of proteins interacting with the p97:ASPL complex. By finding novel binding partners I aim to elucidate the functionality of the complex, gain new insights into its molecular role and contribute to the exceptional amount of knowledge already existing on p97.

In order to study the interactome of the p97:ASPL complex in depth in human cells, I decided to apply a combinatory strategy involving (1) the initial characterization of ASPL interaction partners, which (2) was followed by systematic validation of their binding to p97 in the presence or absence of additional ASPL molecules. Proteins which interact with p97 in an ASPL-dependent manner were selected for further investigations and functional studies.

2. Results

2.1 Identification and validation of p97:ASPL binding partners

2.1.1 Studying interactions within the p97 PPI network by LuTHy

Prior to analyzing the ASPL interactome in human cells, the suitability of LuTHy as a tool for the investigation and quantification of specific p97 PPIs was tested. For this purpose, full-length p97 and a set of known p97 binding partners, including ASPL and nine further members of the UBX protein (UBXP) family, were shuttled into four LuTHy plasmids (**FIG. 1A**). An exemplarily expression validation experiment by western blotting for p97 and ASPL constructs is presented in **FIG. S1**.

All eight possible donor/acceptor combinations and the respective negative controls were co-transfected in HEK293 cells. In-cell BRET measurements were performed 48 hours post transfection (**FIG. 1B**), followed by cell lysis and cell-free analysis of the interaction via the LuC readout (**FIG. 1C**). The only protein delivering positive cBRET and cLuC ratios for all tested binary combinations was ASPL. In the case of the other UBX proteins, there were a few orientations which delivered results below the respective threshold values of 0.01 for cBRET and 0.03 for cLuC. However, all tested PPIs were confirmed in both LuTHy readouts which demonstrates the applicability of the assay for quantification of p97 PPIs. The cBRET and cLuC values of all measured interactions are listed in **Table S1**.

Nevertheless, it became evident that certain interaction directions enabled a more efficient resonance energy transfer than others. If one takes a closer look at the results from the interaction between p97 and ASPL, for instance, the highest cBRET ratios were obtained with the combinations in which p97 was applied as a donor and ASPL was tagged with PA-mCitrine at the C-terminus. The values were 0.220 (± 0.005 SEM) when NanoLuc was fused to the p97 C-terminus and 0.209 (± 0.003 SEM), when it was attached to the N-terminus (**FIG. 1B**). Similarly, when p97 was tested as an acceptor, the highest values were calculated when ASPL was expressed as a fusion construct with NanoLuc at its C-terminus. In this case, fusion of PA-mCitrine to the N-terminus of p97 was more efficient for the energy transfer with a cBRET value of 0.044 (± 0.001 SEM)

over 0.024 (± 0.001 SEM) for p97-mCitrine-PA. These results show very clearly that for the BRET readout it is better to apply p97 as a donor and ASPL as an acceptor compared to the respective reciprocal combinations. This might be an indication for differences in the flexibility and relative orientation of the tags, NL and mCitrine-PA, to one another when expressed as fusion constructs. Such constraints can affect the efficiency of the energy transfer.

In the cell-free LuC readout (**FIG. 1C**) a similar tendency was observed when p97 was applied as a donor, with the highest cLuC values calculated for the combination between NL-p97 and ASPL-mCitrine-PA (0.875 ± 0.070 SEM), followed by p97-NL and ASPL-mCitrine-PA (0.726 ± 0.073 SEM). In comparison, all cLuC values obtained with ASPL as the donor and p97 as the acceptor were lower. In this case tagging p97 with PA-mCitrine at the N-terminus delivered a cLuC value of 0.520 (± 0.018 SEM) with NL-ASPL and 0.376 (± 0.046 SEM) with ASPL-NL. The reciprocal acceptor constellation with p97-mCitrine-PA was even less favorable for the LuC measurement with 0.272 (± 0.005 SEM) for NL-ASPL and 0.2117 (± 0.017 SEM) for ASPL-NL, respectively. From these results it can be concluded that ASPL applied as an acceptor enabled a more efficient co-precipitation of p97 than the reversed combination. In this case the observed difference between ASPL and p97 in their roles as acceptors in the LuC readout might result from the difference in their expression levels. In **FIG. S1** one can clearly see that ASPL fused to mCitrine is expressed better than p97 labeled with mCitrine. This logically results in a greater number of ASPL molecules being immobilized on the surface of the assay plate and a higher amount of p97 labeled with NL which can be co-precipitated. Nevertheless, in the LuC readout all tested directions delivered ratios which surpass the threshold of 0.03 by several orders of magnitude.

It is noteworthy, that the PPI orientation in which the fusion-tags are placed at the N-terminus of p97 and C-terminus of ASPL, corresponds to the interacting domains of p97 and ASPL as published by Arumughan *et al.* (2016)[43]. Additionally, according to the published crystal structure of the complex, the two p97:ASPL heterodimers are orientated in a mirror-inverted manner towards each other (**see FIG. I4A**) which could favor the energy transfer between p97 present in one heterodimer with ASPL present in the other; this cold also explain why energy transfer is observed, when both proteins are both tagged at their C-termini.

In order to get a general understanding of the best fusion-tag orientation in the analyzed set of PPIs, for each interaction the cBRET and cLuC values were normalized to the highest of the eight values acquired in both readouts, respectively (**FIG. 1D**).

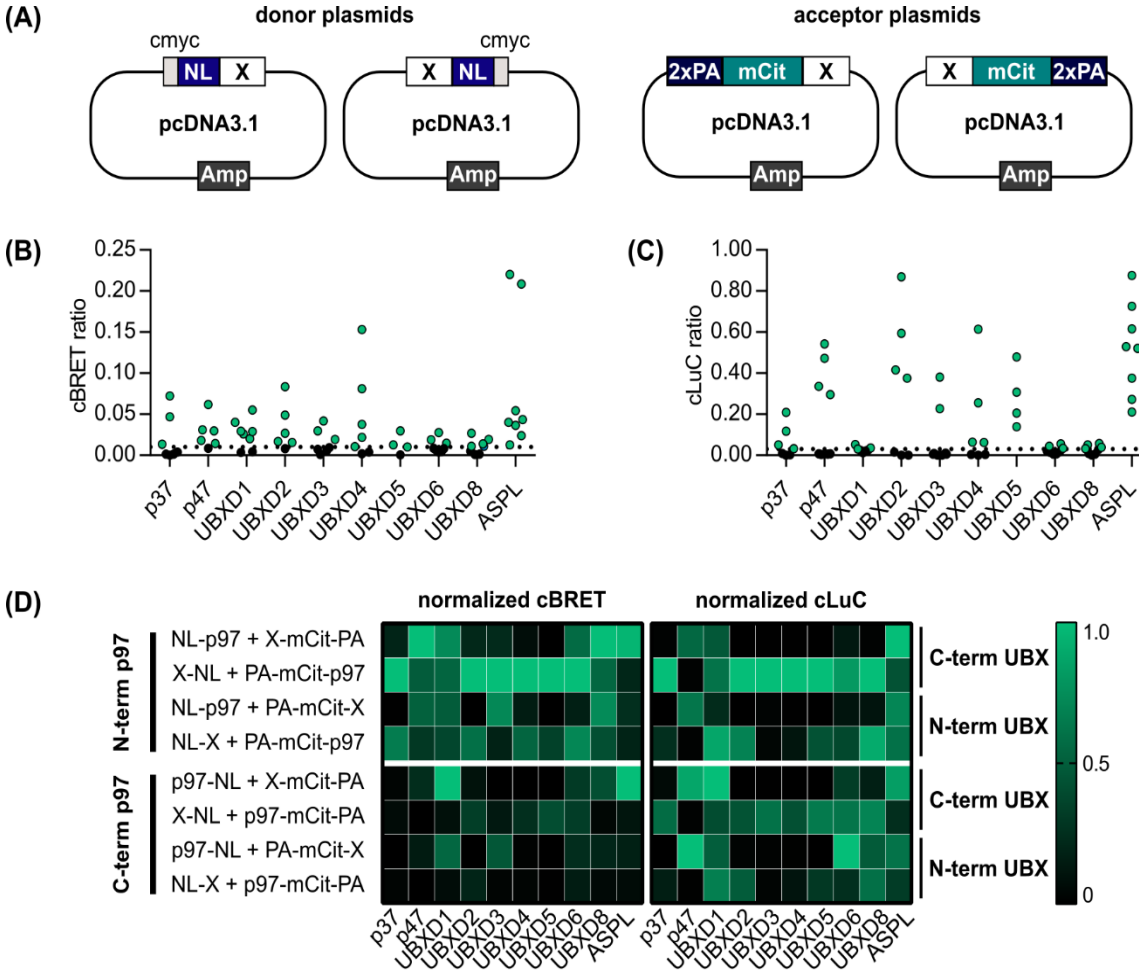


FIG. 1: LuTHy is a valuable tool for the quantification of p97 PPIs.

Scheme of pcDNA3.1 vectors for the expression of **(A)** cmyc-NL or **(B)** 2xPA-mCit tagged variants of p97 and members of the UBX protein (UBXP) family. All proteins of interest were shuttled in all four of the presented plasmids. 10 ng of donor and 100 ng of acceptor plasmid DNA were applied to HEK293 cells for co-expression in LuTHy experiments. cBRET **(B)** and cLuC **(C)** ratios obtained from the binary interactions between p97 and selected UBX proteins. Each dot represents one of total of eight orientations tested for each PPI. Values above the threshold for cBRET (0.01) and cLuc (0.03) are displayed in green. **(D)** cBRET and cLuC data from **(B)** and **(C)** was normalized to the highest cBRET or cLuC value within each individual PPI and plotted in a heat map in order to highlight the best tag orientations. The results presented in this figure were published in Trepte *et al.* 2018.

From the distribution of the values in the heat map one can conclude that, with a few exceptions, the highest ratios were obtained when p97 was applied as the N-terminally tagged acceptor and when the respective UBX protein was tagged at its C-terminus with NL. As already mentioned in the introduction, the interaction between p97 and most of the UBX proteins is mediated by the UBX domain which is present at their C-terminus and enables their binding to the N-terminus of p97[28] (**see FIG. I3B**). This is also the case for ASPL, however, here the best results are obtained with p97 being the donor and ASPL the acceptor due to possible reasons already discussed above. Another interesting example is UBXD1 which binds to the C-terminus of p97 mainly through its PUB domain and in this case the highest LuTHy values were obtained with C-terminally tagged p97. Considering all this, one can conclude that the closer the epitope-tags are to the actual binding domains of the interacting proteins, the higher is the calculated interaction score.

The results from this first set of experiments delivered useful information on the best acceptor/donor combinations and tag orientations for all tested PPIs. Especially helpful were the findings made for the interaction between p97 and ASPL, as this PPI is in the focus of this study. In all follow-up experiments the p97 protein was tagged at its N-terminus, while the ASPL protein was C-terminally tagged in order to utilize the best conditions for the energy transfer-based PPI measurements.

2.1.2 Detecting ASPL binding to p97 at near-endogenous levels of expression

As any other energy transfer-based assay for the detection of PPIs in human cells, LuTHy also makes use of overexpressed fusion constructs for the proteins of interest. However, depending on the relative binding strength of the analyzed interacting proteins and the sensitivity of the assay applied, there is an opportunity to adapt the amounts of DNA used in the transfection reactions in order to reach rates of protein expression near to the endogenous levels. In line with these thoughts, I performed sequential transfections with increasing amounts of pcDNA3.1-ASPL-mCitrine (0-200 ng) and pcDNA3.1-NL-p97 (0-50 ng) in HEK293 cells to test for the DNA concentration required for levels of fusion protein expression similar to or lower than the endogenous ones.

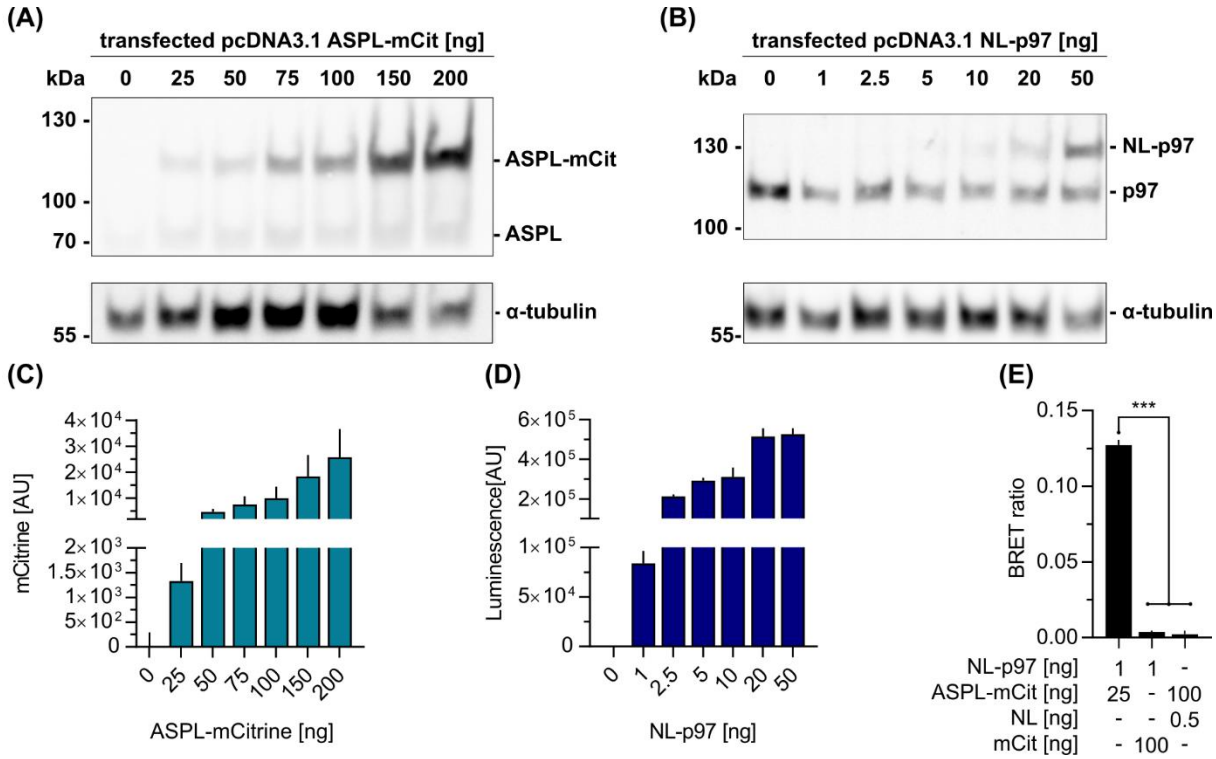


FIG. 2: LuTHy detects ASPL binding to p97 at near-endogenous levels of expression.

(A) Fluorescence and **(B)** luminescence intensity [in AU] measured in HEK293 cells 48 hours after transfection performed with increasing amounts of ASPL-mCitrine-PA [0-200 ng] or NL-p97 [0-50 ng]. The depicted values are means of technical triplicates ± SD. **(C)** BRET ratios acquired from cells co-transfected with pcDNA3.1-NL-p97 [1 ng] and pcDNA3.1-ASPL-mCitrine [25 ng] or the respective negative control combinations containing pcDNA3.1-NL (0.5 ng] or pcDNA3.1-mCitrine [100 ng]. All values are calculated as means of technical triplicates ± SD, $p<0.001$ (One-way ANOVA with Dunett´s multiple comparison test). **(D)** and **(E)** depict the expression of endogenous (70 kDA) and fused constructs of ASPL-mCit (100 kDa) and NL-p97 (130 kDa) in HEK293 cells transfected as described in (A) and (B) analyzed by SDS-PAGE and western blot.

Cell lysates were prepared 48 hours post transfection, applied on an SDS-PAGE and the resulting western blots were developed with an anti-ASPL (**FIG. 2A**) or an anti-p97 antibody (**FIG. 2B**). Comparing the bands of the endogenous ASPL protein (~70 kDa) with the ones resulting from the expression of ASPL-mCitrine (~100 kDa) one can observe that at 25 ng of pcDNA3.1-ASPL-mCitrine the expression levels are comparable and the fluorescent signal is ~1.3 x 10^3 times above the background (**FIG. 2C**). For NL-p97 (~130 kDa) 50 ng of the expression plasmid were needed to get a band for the fusion construct of equal intensity as the one for the endogenous p97 (~110 kDa). Even though there is no visible band on the western blot for NL-p97 when 1 ng of DNA was used for transfection, I could measure luminescence which is ~8.5 x 10^5 times higher than the signal measured in cells transfected with an empty pcDNA3.1 vector (**FIG. 2D**). Following this, I co-transfected HEK293 cells with 1 ng of NL-p97 and 25 ng of ASPL-mCitrine and measured the energy transfer in cells 48 hours post transfection. Under these conditions, the calculated BRET ratio of 0.1272 ± 0.003 was significantly higher than the ratios acquired with the respective negative controls (**FIG. 2E**).

From this experiment I can draw two important conclusions. Firstly, I can gain information about the amount of DNA needed for mimicking the endogenous expression levels of ASPL and p97. Secondly, I successfully demonstrated that the in-cell BRET measurement is a very useful tool for the efficient quantification of high affinity interactions under near endogenous conditions. This provides an opportunity to avoid the disadvantages of overexpression experiments.

2.1.3 Point mutations at conserved residues in ASPL reduce its affinity to p97

For the investigation of the interaction between ASPL and p97 Arumughan and colleagues (2016) created two ASPL mutant variants via site-directed mutagenesis. They demonstrated that the exchange of certain conserved residues within the C-terminus of ASPL reduces its capacity to disassemble p97 hexamers[43].

The residues which Arumughan *et al.*[43] picked for mutagenesis were Asp351 for the single and Pro437-Pro438 for the double ASPL mutant variant. For the aims of my work I decided to make use of these constructs in control tests for the validation of ASPL-dependent binding partners. All three amino acid changes are located within domains, which are crucial for ASPL binding to p97 and its disassembly (**FIG. 3A**). Asp351 is located in the a-helical lariat which is responsible for targeting the area between the N- and D1-domains in p97 and initiating the dissociation of the hexamer. In comparison, the selected proline residues in ASPL form a conserved *cis*-Pro turn-touch motif in the canonical UBX domain which assists the binding to p97 through its docking to a hydrophobic pocket located between the N_n and N_c subdomains at the p97 N-terminus. It was demonstrated that exchanging these amino acids with alanine residues reduces the capability of ASPL to disassemble the p97 hexamer[43] (**FIG. 3B**).

However, their binding affinity towards p97 was not yet experimentally estimated. For this purpose, I tested the effects of the point mutations on the ASPL:p97 interaction by co-transfecting HEK293 cells with plasmids for the expression of NL-p97 and wild-type ASPL or any of the two mutated versions of ASPL, C-terminally tagged with mCitrine-PA. A luminescence scan in the range between 350 and 700 nm in 2 nm intervals demonstrated reduced energy transfer efficiency between the donor, NL-p97, and the mutated ASPL-mCitrine-PA variants, D351A or PP437-438AA, compared to the wild-type protein (**FIG. 3C**). Additionally, donor saturation experiments were performed by co-transfecting HEK293 cells with a constant amount of NL-p97 and increasing concentrations of wild-type or mutated ASPL-mCitrine-PA fusion proteins. The resulting saturation curves and the calculated $BRET_{50}$ values indicated a mutation-dependent decrease in the affinity of ASPL-mCitrine-PA towards NL-p97 (**FIG. 3D, E**). I furthermore confirmed that the mutations did not influence the expression of the mutated ASPL proteins. For this purpose, HEK293 cells expressing either wild-type ASPL or the mutant

variants with an epitope cmyc tag at their N-terminus were lysed and analyzed by western-blotting (**FIG. 3F**). There is no visible difference in the band intensity when comparing wild-type and mutant fusion proteins.

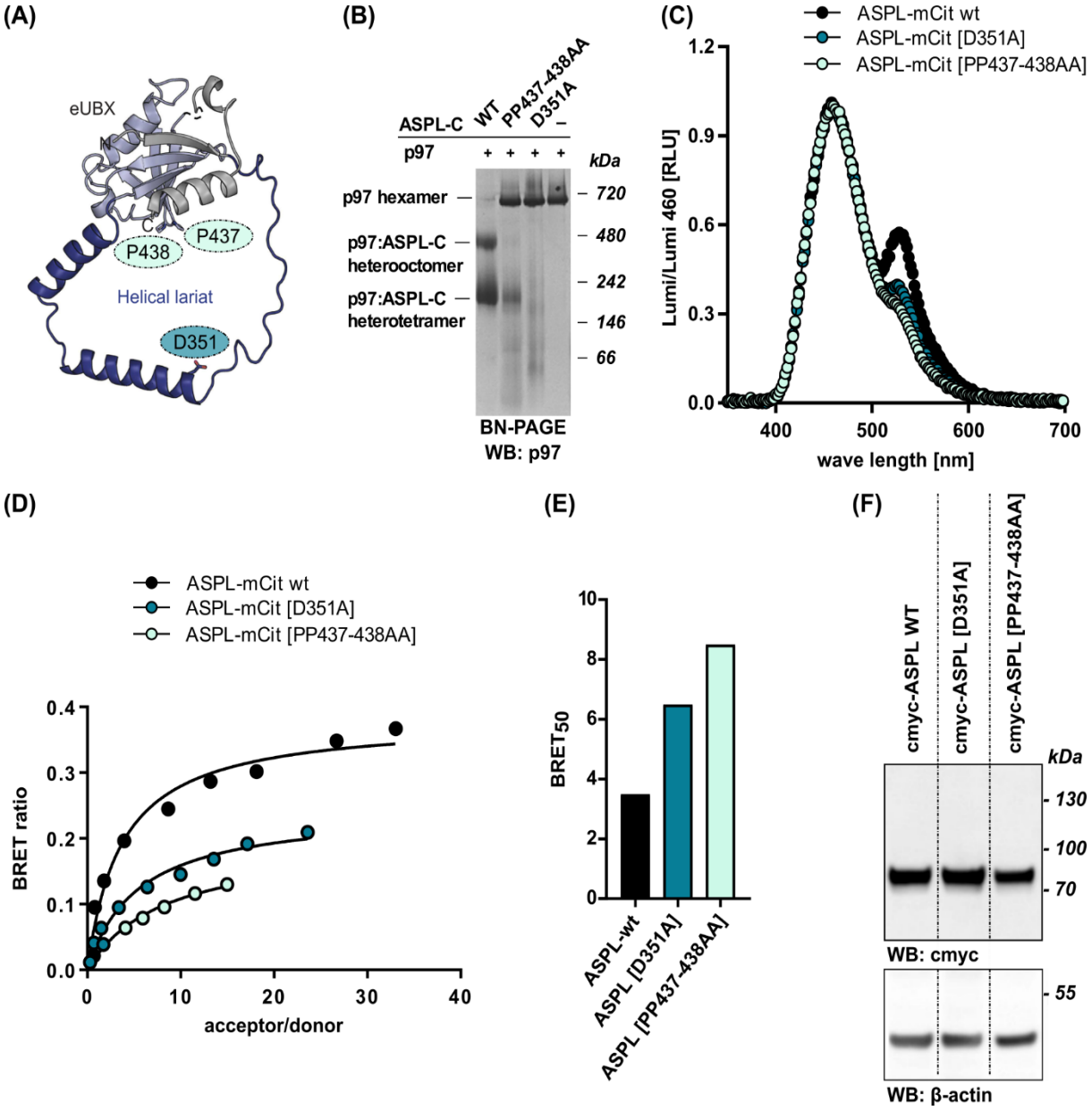


FIG. 3: Investigation of the interactions between p97 and mutant ASPL variants.

(A) Ribbon-type representation of the extended UBX (eUBX) domain of ASPL (313-553 aa) adapted from Arumughan *et al.* (2016). Positions of mutated residues in the single [D351A] and double [PP347-348AA] ASPL mutant versions are highlighted in the structure. **(B)** Blue native (BN) PAGE performed by Arumughan and colleagues (2016) with recombinant ASPL-C (wild-type or mutant variants) and wild-type p97. The western blot was developed with an anti-p97 antibody. **(C)** HEK293 cells co-transfected with 5 ng of pcDNA3.1-NL-p97 and 100 ng of different ASPL-mCitrine variants were subjected to a luminescence scan at the range between 350 and 700 nm with measurements performed in 2 nm intervals. Lowering of the peak at 530 nm is an indication of reduced energy transfer efficiency between NL-p97 and both ASPL-mCitrine mutants. **(D)** BRET ratios plotted against the acceptor/donor ratios obtained by co-transfecting HEK293 cells with 1 ng of pcDNA3.1-NL-p97 and increasing quantities of pcDNA3.1-ASPL WT-mCitrine or mutant variants (0-200 ng). **(E)** Calculated half-maximum BRET ($BRET_{50}$) values of the interaction pairs tested in (D). The higher the $BRET_{50}$ value the lower the relative affinity between the tested proteins. **(F)** HEK293 cells transfected with equal amounts of pDEST26-cmyc-ASPL WT, pDEST26-cmyc-ASPL [D351A] or pDEST26-cmyc-ASPL [PP437-438AA] were lysed and the expression levels of wild-type and mutated ASPL variants were analyzed by SDS-PAGE and western blot. Anti-cmyc antibody was used for detection of cmyc-tagged ASPL proteins.

In summary, I was able to show that the mutations D351A and PP437-438AA in ASPL are able to reduce its affinity to p97 but that they are not sufficient to completely abolish the interaction. In addition, taking the findings by Arumughan *et al.* (2016) under consideration, it can be concluded that the disassembly of the p97 hexamer, which is initiated by its interaction with ASPL, is a two-step process. It requires (1) the association of ASPL and p97 molecules and (2) the conformational conversion of p97 protomers, which is induced by residues in the extended UBX (eUBX) domain in ASPL.

2.1.4 Identification of ASPL binding partners via SILAC

The major goal of the following work was to identify binding partners of the p97:ASPL complex. For this purpose, I concentrated my efforts on the "extended" p97 interactome which includes proteins that specifically interact with p97-bound ASPL (**FIG. 4A**). There are two main reasons why ASPL is a particularly good target for investigating the interactome of the p97:ASPL complex. Firstly, ASPL is a high affinity binding partner of p97 and both proteins form stable oligomeric structures *in vitro* and *in vivo*[43]. Secondly, in a size-exclusion-chromatography (SEC) experiment I observed a very prominent co-migration of both ASPL and p97 (**FIG. 4B**), which let me assume that ASPL in cells is predominantly present in a p97-bound state. However, there is also a population of ASPL molecules which might be incorporated in other structures. This is indicated by the fact that ASPL was found not only in fractions which correspond to the size of the heterotetrameric complex (~440 kD) but also in higher molecular weight fractions (>600 kDa). This suggests that indeed higher order structures are formed *in vivo*, which include the p97 and ASPL complex.

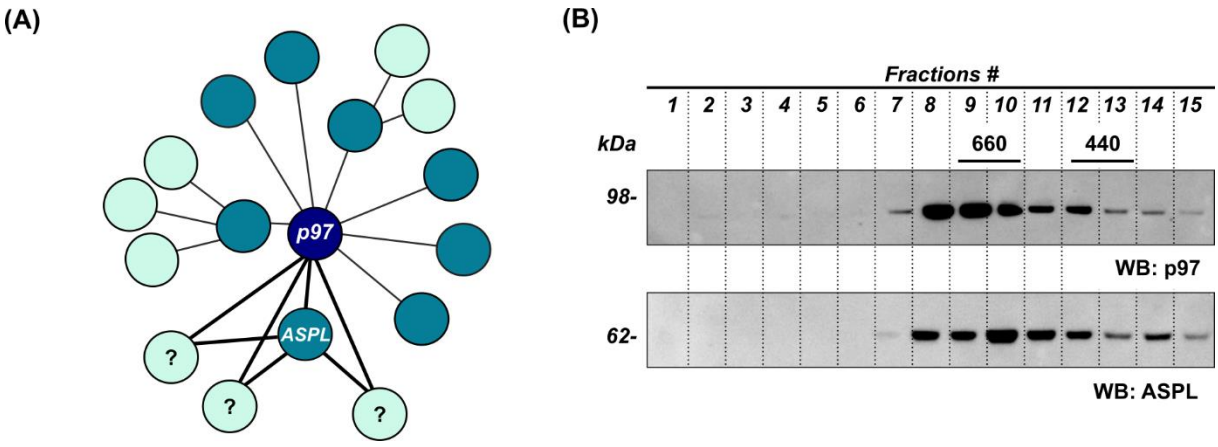


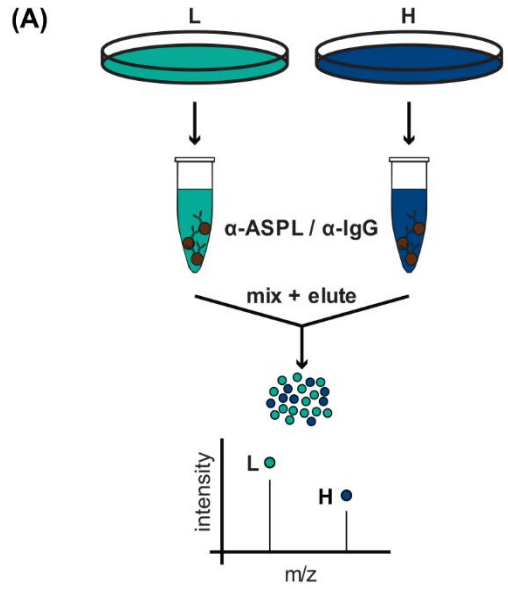
FIG. 4: Investigating the "extended" p97 interactome.

(A) Schematic representation of the "extended" p97 interactome including direct p97 binding partners (dark blue circles) and second-order p97 binders (light blue circles). ASPL was selected as a target for in-depth interactomics studies under the assumption that its interaction partners would also be p97:ASPL binders. **(B)** Size-exclusion chromatography of HEK293 cell lysates analyzed by SDS-PAGE and western blotting after fractionation for the detection of endogenous ASPL and p97 molecules. Fractions #9-10 and #12-13 contain oligomeric structures corresponding to protein complexes with approximate molecular weights of 660 and 440 kDa, respectively.

Taking this under consideration, I presumed that by identifying proteins, which interact with ASPL, the propensity of detecting p97:ASPL-interacting proteins will be increased. Thus, in order to search for ASPL-binding partners a **SILAC** screen (**s**table **i**sotope **l**abeling with **a**mino acids in **c**ell culture) was performed with immunoprecipitated endogenous ASPL from HEK293 cells (**FIG. 5A**). Lysates from cells grown in light and heavy DMEM medium incubated with two specific ASPL antibodies (mouse α-ASPL and rabbit α-ASPL) and respective isogenic IgG control antibodies or beads only, were combined as shown in **FIG. 5B**.

Each detected protein in the SILAC swaps is represented by a mass-to-charge value in both heavy and light fractions resulting in a heavy-to-light (H/L) abundance ratio calculated as a log2fold change (log2FC). The resulting H/L ratios for each reciprocal swap pair were plotted against each other delivering a scatter dot plot in which each dot represents a specific protein (**FIG. 5C-F**). The position of a certain dot and its distance to ASPL correlates with the likelihood with which these two proteins might interact. For instance, as expected ASPL and p97 appeared close to each other on all scatter dot plots, which was an indication for the quality of the performed screen.

Mean and standard deviation of the log2FC H/L ratio were calculated separately for each swap and a threshold was set at a value corresponding to a sum of the mean and standard deviation, marked by a dashed line in each plot in **FIG. 5C** to **FIG. 5F**. In **FIG. 5G** all hits are listed and ranked according to the frequency of their detection in the SILAC swaps (termed here as "counts"). A total number of 26 proteins were identified, including p97 and ASPL. Additional information about the identified hits is listed in **Table S2**.



(B)

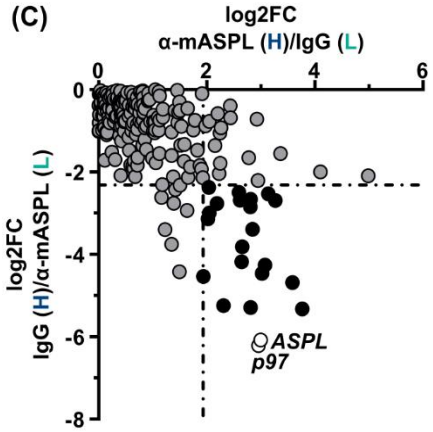
mass spec swaps:

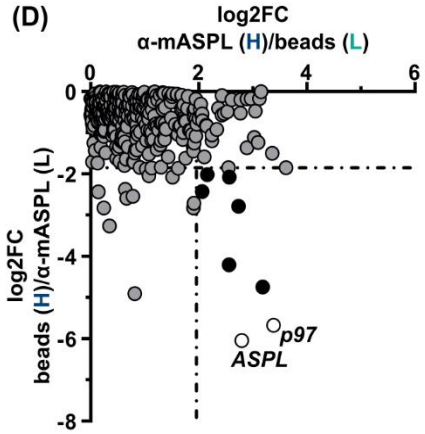
1 IP αASPL (mouse) **HEAVY**/ IP IgG (mouse) **LIGHT**
+
5 IP αASPL (mouse) **LIGHT**/ IP IgG (mouse) **HEAVY**

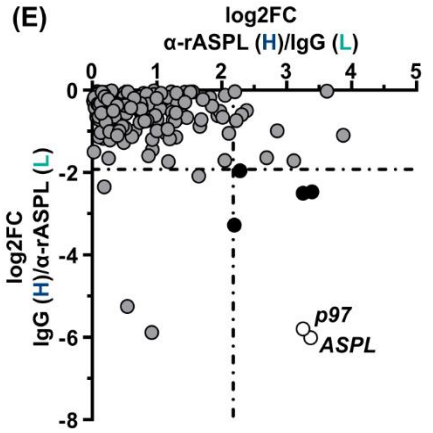
2 IP αASPL (mouse) **HEAVY**/ IP (beads only) **LIGHT**
+
6 IP αASPL (mouse) **LIGHT**/ IP (beads only) **HEAVY**

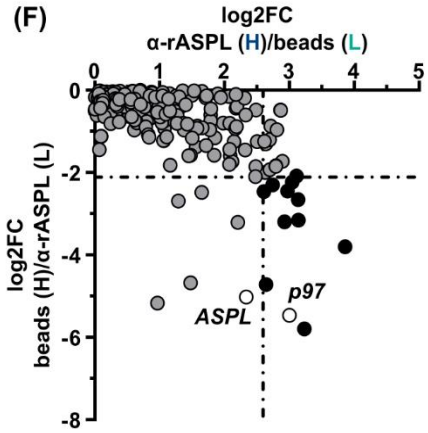
3 IP αASPL (rabbit) **HEAVY**/ IP IgG (rabbit) **LIGHT**
+
7 IP αASPL (rabbit) **LIGHT**/ IP IgG (rabbit) **HEAVY**

4 IP αASPL (rabbit) **HEAVY**/ IP (beads only) **LIGHT**
+
8 IP αASPL (rabbit) **LIGHT**/ IP (beads only) **HEAVY**









(G)

proteins	counts
HSPA4	1
BOLA2	1
LUC7L2	1
CIRBP	1
ARL6IP4	1
SAP18	1
RBM39	1
HSP90A	1
SF3B3	1
RNPS1	1
TRA2B	1
TAF15	1
SNRNP70	1
SFRS3	1
SFRS6	1
SFRS7	1
SNRNP27	1
PNN	1
BOLA1	2
PUF60	2
SRSF10	2
SNRPG	2
U2AF2	2
AGK	3
ASPL	3
SFRS1	3
U2AF1	3
p97	4

FIG. 5: Identification of ASPL binding partners by SILAC.

(A) Sample preparation set up for immunoprecipitation of endogenous ASPL and primary data generation by mass spectrometry. The screen was performed in collaboration with Fabian Hosp in the Selbach lab, MDC. **(B)** Combination of IP samples from cells cultivated in light or heavy medium resulting in eight SILAC swaps. **(C-F)** Log2FC of the H/L abundance ratios of a specific ASPL antibody sample and control IgG sample on the x-axis were plotted versus the corresponding pair on the y-axis for all reciprocal SILAC swap pairs: **(C)** 1+5, **(D)** 2+6, **(E)** 3+7 and **(F)** 4+8. The dashed lines represent the threshold (μ + 1xSD) for both axes. Hits selected as positive are depicted in black, p97 and ASPL are depicted in white. **(G)** List of all 26 identified SILAC hits with detection frequencies.

In order to gain information about the potential cellular functions of the identified ASPL-interacting proteins, I performed a gene ontology (GO) term enrichment analysis using the online **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**), v6.8. **Table S3** summarizes the functional characteristics of all SILAC hits and contains the exact p-values and fold enrichment values for each group. Only the significantly enriched GO-terms ($p<0.01$), containing groups of ten and more proteins, were graphically depicted in a chord plot in which each protein is connected with its specific GO-terms (**FIG. 6**).

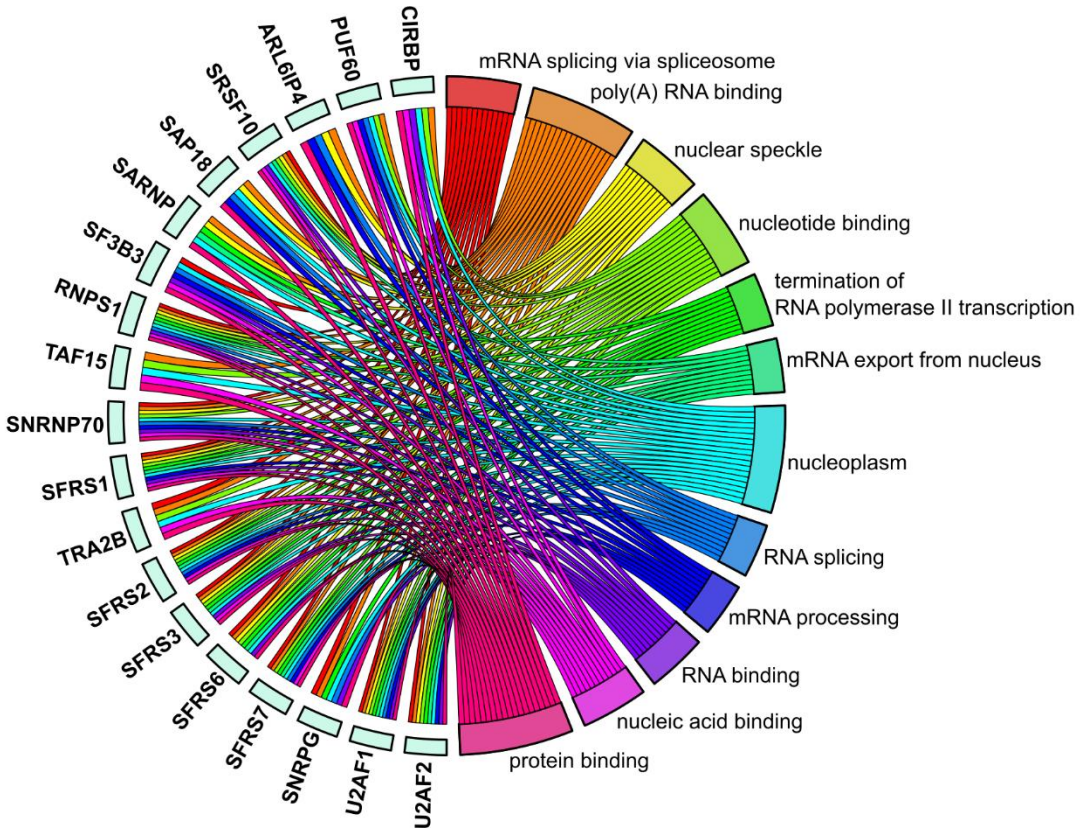


FIG. 6: Enrichment of GO-terms related to RNA processing within the group of SILAC hits.

All proteins identified by SILAC as ASPL interaction partners were subjected to GO-term enrichment analysis via DAVID in the categories "biological process" (BP), "molecular function" (MF) and "cellular component" (CC) using the default "*Homo sapiens*" background list available in DAVID. Selection criteria for the displayed groups were fold enrichment with $p<0.01$ and minimal number of ten proteins per group. The presented genes are connected to at least five distinct GO-terms. The chord plot was generated in "R" with the packages "circlize" and "GOplot".

It became evident that within the group of putative ASPL binding partners there is a prominent enrichment of nuclear proteins with RNA binding properties. These proteins are mainly involved in RNA processing, mRNA maturation and splicing. From the results so far I was able to conclude that the p97:ASPL complex and its associated interaction partners play a potential role in RNA-related pathways in the human cell.

2.1.5 Systematic ASPL interactome validation by LuTHy

As already stated in the introduction and in the previous parts, I was primarily interested in finding binding partners of the p97:ASPL complex and one possibility to achieve this goal is to assess the binding of partner proteins to ASPL molecules, which are in a complex with p97 (**FIG. 4B**). Therefore, identifying ASPL binding partners was the initial step of my investigations and in the following part of my thesis I performed a three-step LuTHy screen for the validation of the SILAC data. **FIG. 7A** represents a schematic overview of the performed experiments and my strategy towards the identification of specific p97:ASPL complex binders.

Firstly, I executed a series of binary validation tests to assess the interactions between the identified SILAC hits and ASPL (**FIG. 7B**). Secondly, I examined the interactions between the same proteins and p97 in the absence (**FIG. 7C**) or presence of ASPL (**FIG. 7D**) in two consecutive experiments. Here, I should emphasize on the amount of the applied expression plasmid for ASPL production, N-terminally fused with an epitope cmyc-tag, at the transfection step. It was intentionally adjusted to 20 ng in order to reach expression levels similar to the ones of the endogenous ASPL protein. The selected amount was extrapolated from the experiment presented in **FIG. 2A**, in which the expression levels of ASPL-mCitrine and endogenous ASPL were compared.

By combining both LuTHy readouts and comparing the results of all three LuTHy data sets, I could demonstrate that among the ASPL interaction partners there are a few proteins which were indeed capable of binding to p97 only in the presence of ASPL (**FIG. 7B-D**). Among all cBRET positive hits 52% interact with ASPL, but only 19% bind to p97 alone. In the presence of additional ASPL molecules the number of positive interactions increased up to 29% (**FIG. 7E**). Interestingly, the LuC readout (**FIG. 7F**) delivered a higher percentage of positive hits in the screen against p97 (28%), but again there was an increase in the number of hits when ASPL was added (38%).

It is noteworthy that when both readouts are taken under consideration, the proteins interacting with p97 alone were also present among the hits of the other two data sets (**Table S4**). At this point one should keep in mind the drastic structural changes taking place within the p97 hexamer upon ASPL binding. Hence, I can conclude that these proteins are rather p97 binders, but not necessarily specific interaction partners of the

p97:ASPL complex and they were therefore excluded from my follow-up analysis. Furthermore, proteins which according to my data bind exclusively to ASPL and were not detected among the hits in the "p97" and "p97 + ASPL" data sets, were not further characterized.

The criteria for selecting a protein as a positive hit from the validation screen included at least one readout (cBRET or cLuc) with values above the threshold in the "ASPL" and "p97 plus ASPL" LuTHy screens. In total four proteins – RBM39, HSP90AA1, PUF60 and U2AF2 – fulfill these requirements and are marked respectively in **Table S4**. I picked U2AF2 (U2 Small Nuclear RNA Auxiliary Factor 2), in order to illustrate the selection process and highlighted it in yellow on the plots in **FIG. 7B** and **FIG. 7D**. There were four positive cBRET ratios, corresponding to four different interaction orientations, calculated for the interaction between U2AF2 and ASPL. None of the orientations rendered positive results when the protein was tested against p97 and two were above the cBRET threshold when tested against p97 in the presence of additional near-endogenous amounts of ASPL.

Based on my data I assume that U2AF2 and the other validated complex binders RBM39, HSP90AA1 and PUF60, interact with p97 in an ASPL-dependent manner. Furthermore, the binding surface, which these proteins potentially target, is likely to be the one that forms upon the conformational re-orientation of p97 protomers upon ASPL binding in p97:ASPL heterotetramers. Considering this, it might appear surprising that the interaction between U2AF2 and wild-type ASPL can be measured without applying additional plasmids for p97 expression. However, this can be easily explained by two facts: (1) the high affinity of ASPL towards p97 and (2) their stable association, leading to the formation of a heteroterameric complex. This would mean that any newly produced ASPL fusion proteins will get rapidly incorporated into a complex with endogenous p97, which is one of the most abundant proteins in human cells[24]. Once this is accomplished and U2AF2 or any of the other validated interacting proteins bind to the newly formed p97:ASPL complexes, the interaction can be measured via LuTHy.

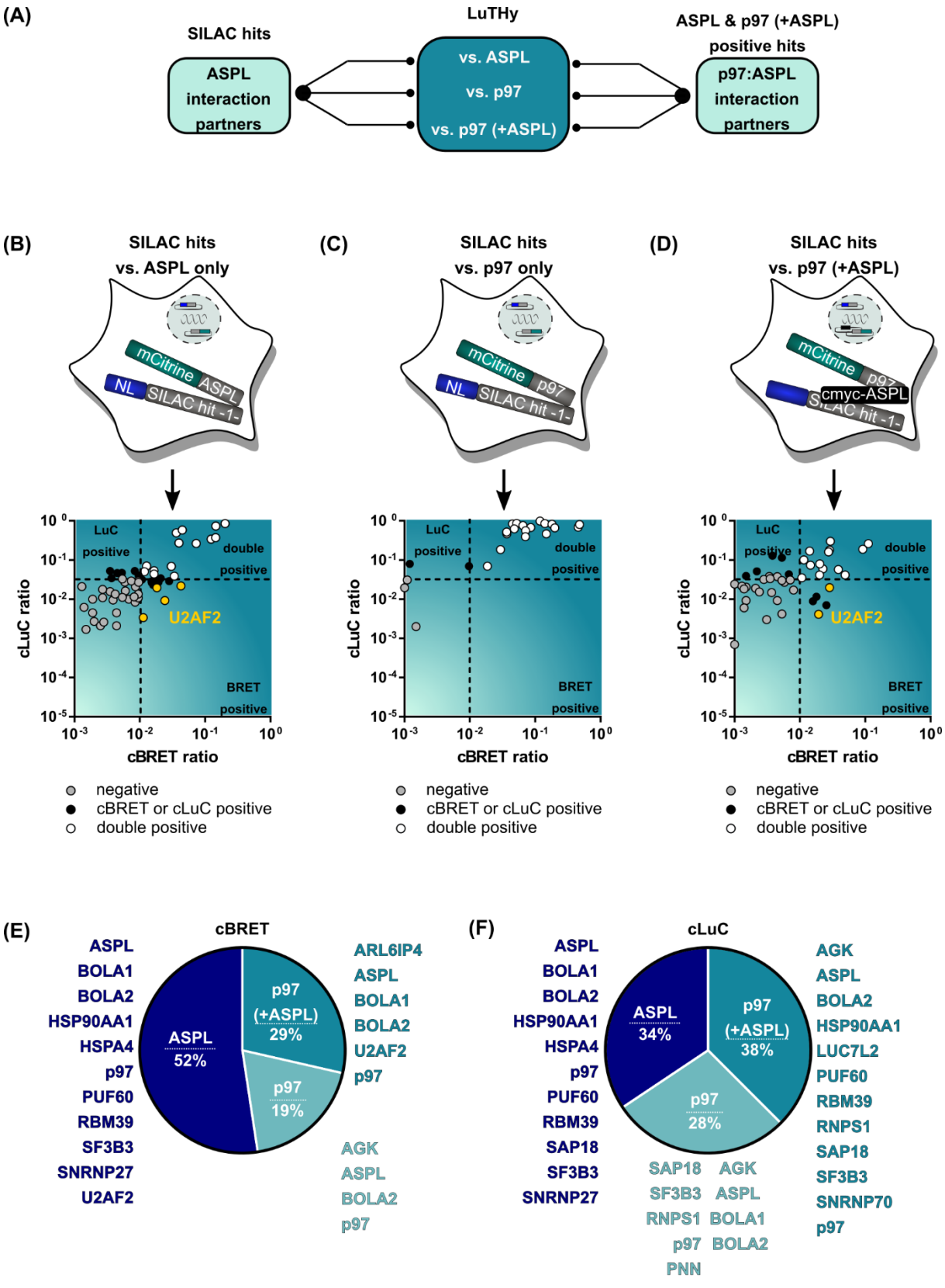

(A)
SILAC hits
ASPL interaction partners
LuTHy
vs. ASPL
vs. p97
vs. p97 (+ASPL)
ASPL & p97 (+ASPL) positive hits
p97:ASPL interaction partners
(B)
SILAC hits vs. ASPL only
mCitrine ASPL
NL SILAC hit -1-
(C)
SILAC hits vs. p97 only
mCitrine p97
NL SILAC hit -1-
(D)
SILAC hits vs. p97 (+ASPL)
mCitrine p97
cmyc-ASPL
SILAC hit -1-
LuC positive
double positive
U2AF2
BRET positive
cLuC ratio
cBRET ratio
negative
cBRET or cLuC positive
double positive
(E)
cBRET
ASPL
BOLA1
BOLA2
HSP90AA1
HSPA4
p97
PUF60
RBM39
SF3B3
SNRNP27
U2AF2
ASPL 52%
p97 (+ASPL) 29%
p97 19%
ARL6IP4
ASPL
BOLA1
BOLA2
U2AF2
p97
AGK
ASPL
BOLA2
p97
(F)
cLuC
ASPL
BOLA1
BOLA2
HSP90AA1
HSPA4
p97
PUF60
RBM39
SAP18
SF3B3
SNRNP27
ASPL 34%
p97 (+ASPL) 38%
p97 28%
SAP18 AGK
SF3B3 ASPL
RNPS1 BOLA1
p97 BOLA2
PNN
AGK
ASPL
BOLA2
HSP90AA1
LUC7L2
PUF60
RBM39
RNPS1
SAP18
SF3B3
SNRNP70
p97

FIG. 7: SILAC validation by LuTHy.

(A) Scheme illustrating the experimental strategy for the identification of p97:ASPL binding partners. From left to right it includes the identification of ASPL binders by SILAC, a LuTHy validation approach as a three-step screen and selection of p97:ASPL complex partners. **(B-D)** Graphical summary of LuTHy results from the binary tests between **(B)** ASPL and all SILAC hits, **(C)** p97 and all SILAC hits and **(D)** p97 versus all SILAC hits in the presence of cmyc-ASPL. In all three plots cLuC and cBRET values are plotted against each other. Each dot represents a mean of two biological replicates for a single interaction direction. Each interaction was tested in all possible directions, unless cloning was not successful. **(E)** Distribution of positively identified proteins by cBRET among all positive hits within this readout for all three screens, vs. ASPL (dark blue), vs. p97 only (turquoise), vs. p97 plus ASPL (light blue). **(F)** Similar to (E) showing the results from the cLuc readout.

2.1.6 p97:ASPL interaction partners are involved in RNA processing

In total, ten proteins were detected as ASPL interaction partners via LuTHy resulting in a 38% validation rate of the SILAC data set. Four of them were identified as specific p97:ASPL complex binders (**Table S4**). HSP90AA1, RBM39, PUF60 and U2AF2 delivered positive LuTHy results in at least one of the assay´s readouts when tested against ASPL and p97 in the presence of additional ASPL molecules, but were not detected in the "p97 only" data set. In order to gain a general impression of the interconnection among these proteins, I performed a search within the STRING database (**S**earch **T**ool for the **R**etrieval of **In**teracting **G**enes/Proteins) (**FIG. 8A**).

Three of these four proteins, PUF60, U2AF2 and RBM39, form a distinct cluster based on existing experimental molecular and PPI evidences[55–57]. They possess DNA and RNA binding properties and are involved in splicing or transcriptional regulation[58]. HSP90AA1, on the other hand, acts as a molecular chaperone whose target proteins are involved in cell cycle control, signal transduction and transcription modulation[59]. The interaction between p97 and ASPL is also curated in the STRING database. However, there is no published information about an associations between ASPL and PUF60, U2AF2 or RBM39 under physiological conditions.

Taking a closer look at the structural characteristics of the proteins PUF60, RBM39 and U2AF2 revealed their similar domain architecture (**FIG. 8B**). Besides the tandem RNA recognition motif (RRM) present in all three proteins, they also possess a U2AF homology motif (UHM) at their C-terminus. This domain enables the association with proteins containing a UHM ligand motif (ULM)[60] (**FIG. 8C**) and promotes the establishment of a

PPI network during the early stages of spliceosome assembly. **FIG. 8D** illustrates the importance of the UHM-ULM recognition for the recruitment of the spliceosomal particle, U2 snRNP, to the 3' splice site and the role of SF3B3 as a scaffolding protein in this process[61].

U2AF2 was selected for in-depth characterization of its association with p97:ASPL due to two main reasons. On the one hand, the signal strength measured between U2AF2 and ASPL or p97, in the presence of additional ASPL, was the highest among the tested interactions. On the other hand, U2AF2 was the only protein identified exclusively by the BRET output, which facilitates the deeper in-cell analysis of the interaction. Furthermore, U2AF2 plays a crucial role as a key component of the early spliceosome assembly[62] which makes it a very relevant hit with great potential in revealing the molecular function of the p97:ASPL complex in human cells.

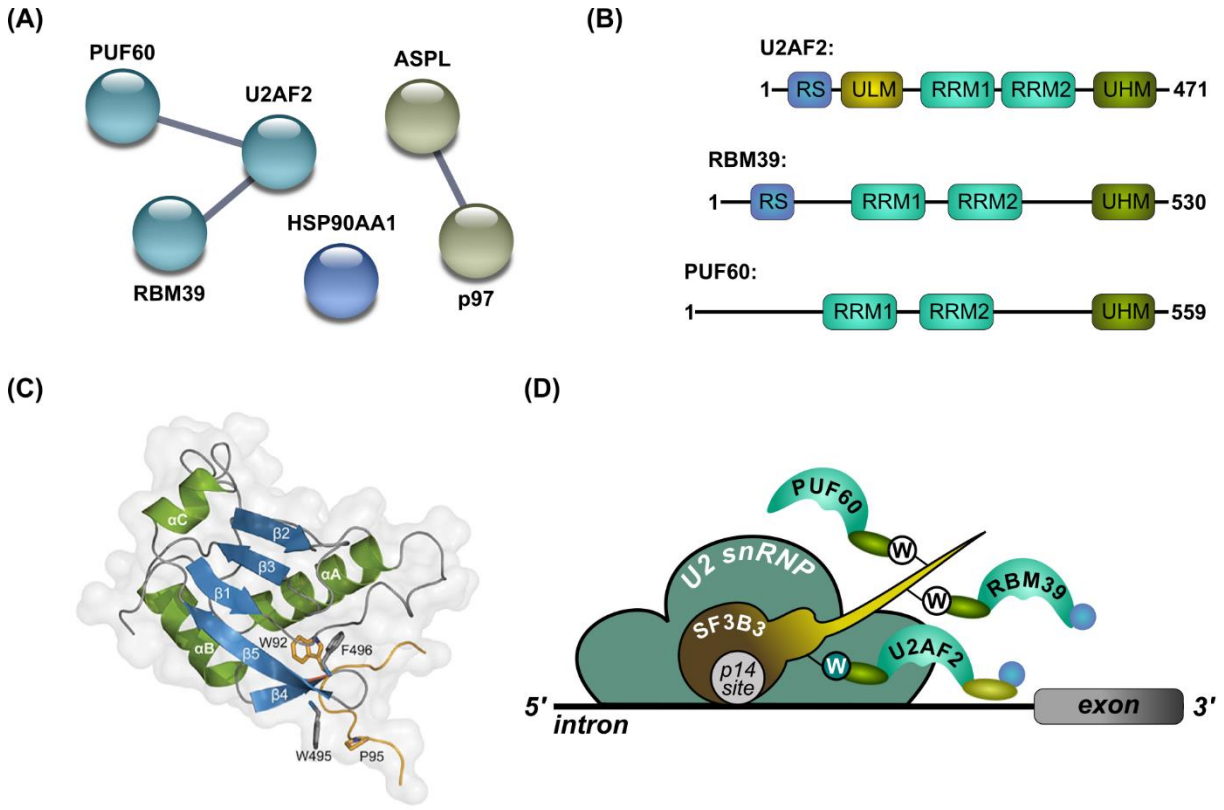


FIG. 8: Characteristics of p97:ASPL binding partners.

(A) Interaction network generated within the STRING database with a medium degree of confidence including putative p97:ASPL interaction partners. **(B)** Domain composition of the selected splicing factors revealing their structural homology. RS = Arg/Ser-rich domain, ULM = UHM ligand motif, RRM = RNA recognition motif, UHM = U2AF homology motif (adapted from Loerch *et al.* 2014) **(C)** Ribbon representation of the UHM domain in RBM39 in its association with U2AF2-ULM, shown in yellow (from Stepanyuk *et al.* 2016). **(D)** Schematic representation of splicing factors interacting with SF3B3 within the U2 snRNP bound to an intron sequence. UHM-ULM mediated binding is marked by Tryptophan residues (W) in ULM domains of SF3B3; the p14 site is required for recognizing the p14 subunit in SF3b14a/p14 and binding to the branch point of pre-mRNA introns (adapted from Loerch *et al.* 2014).

2.1.7 Further evidence confirming the interaction between ASPL and U2AF2

Gaining additional evidence on the interaction between U2AF2 and ASPL would further support my hypothesis for the association of the splicing factor with the p97:ASPL complex. For this purpose, I first investigated the cellular localization of both proteins in HeLa cells by fluorescence imaging. Endogenous U2AF2 was detected exclusively in the nucleus. Due to the unavailability of antibodies for the detection of endogenous ASPL by immunofluorescence microscopy, mCitrine-ASPL was ectopically expressed in

HeLa cells and fluorescence was detected in both the cytosol and the nucleus (**FIG. 9A**).

Besides the cellular co-localization between U2AF2 and mCitrine-ASPL (**FIG. 9A**), I was able to successfully demonstrate their interaction by co-immunoprecipitation. For this purpose, HEK293 cells were lysed and mixed with either a specific anti-U2AF2 antibody or an isogenic IgG immobilized on magnetic beads. Additionally, I applied beads without IgG molecules on their surface to test for any unspecific protein binding ("beads only" control). In this experimental set-up I precipitated endogenous U2AF2 molecules and tested for ASPL binding by immunoblotting (**FIG. 9B**). Indeed, only when the cell lysates were incubated with beads with the specific anti-U2AF2 antibody I could detect U2AF2 and ASPL in the eluate, but bands for neither of these proteins were visible in the eluate of the IgG or "beads only" controls. However, in the flow-through of all three samples a prominent amount of the splicing factor was detected. This was expected for the negative controls, but not for the anti-U2AF2 sample. This is an indication for a non-optimal ratio of antibody molecules to beads leading to a higher number of unbound U2AF2 molecules in the flow-through. Nevertheless, this co-immunoprecipitation demonstrated the interaction between ASPL and U2AF2 at the endogenous level.

In order to confirm the best tag orientation for the interaction between ASPL and U2AF2, which resulted from the validation screens, I once more systematically tested the complete set of directions in follow-up LuTHy tests. The most favorable conditions for measuring resonance energy transfer were given when U2AF2 was expressed as a C-terminal fusion with NanoLuc and mCitrine was tagged to ASPL at its C-terminus (cBRET = 0.038) (**FIG. 9C**). Furthermore, applying U2AF2 tagged with mCitrine and ASPL fused to NanoLuc delivered two additional positive directions with U2AF2-mCitrine and ASPL-NL (0.019) or NL-ASPL (0.024) (**FIG. 9D**). All other reciprocal tag constellations exhibited cBRET values below the cut-off (0.01). The obtained cLuC values were below the threshold of 0.03 (data not shown). These results were in concordance with the outcome of the initial validation screen (**see FIG. 7**, **Table S4**) with the only exception that here the combination U2AF2-NL co-expressed with PA-mCitrine-ASPL rendered a cBRET value just below the threshold (cBRET = 0.009). Additional donor saturation tests were performed with the best two orientations (U2AF2-NL plus ASPL-mCit and ASPL-NL

plus U2AF2-mCit) demonstrating the specificity of the interaction between ASPL and U2AF2 (**FIG. 9E**).

The tag orientation – U2AF2-NL and ASPL-mCitrine-PA – providing the highest BRET ratio was further applied in the following binary tests between the splicing factor and the previously mentioned ASPL mutant variants, D351A (single mutant [s. m.]) and PP437-438AA (double mutant [d. m.]). Testing the ASPL mutants, which exhibit a lower p97 binding affinity and a reduced ability of disassembling p97 hexamers[43], was expected to reduce the BRET signal due to the lower number of p97:ASPL complexes which can be formed. Accordingly, I observed a decline in the cBRET signal with both mutant ASPL versions (**FIG. 9F**). Even more striking was the fact that the reduced binding to U2AF2 corresponded to the observed decreased interaction of ASPL-D351A and ASPL-PP437-438AA with p97 (**see**. **FIG. 3D**).

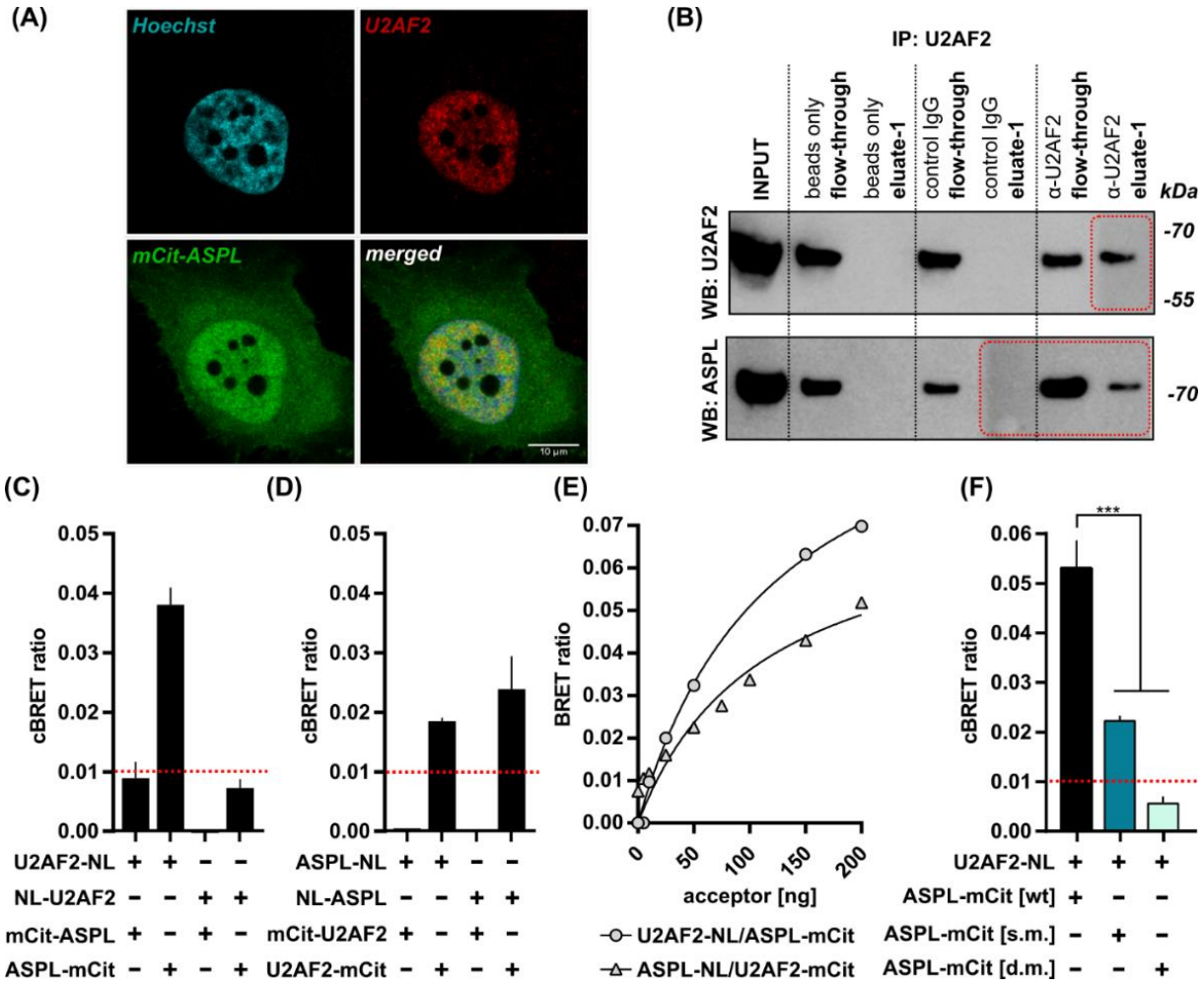


FIG. 9: ASPL and U2AF2 interaction confirmation.

(A) Immunofluorescence microscopy image demonstrating nuclear co-localization between endogenous U2AF2 and mCitrine-ASPL overexpressed in HeLa cells. U2AF2 was detected with a combination of a rabbit anti-U2AF2 and an anti-rabbit Alexa647 conjugated IgG. mCitrine fluorescence was obtained with a GFP filter. Imaging was performed with a 20x objective on a Zeiss Axio Imager. Scare bar = 10 µm. **(B)** Co-immunoprecipitation (Co-IP) of ASPL with U2AF2 from native HEK293 cell lysates. Eluate and flow-through of samples treated with specific anit-U2AF2 antibody, isogenic control IgG or beads only were analyzed by SDS-PAGE and western blotting. Exposure times were 2 min for WB-U2AF2 and 5 min for WB-ASPL. Areas on the western blots surrounded by red boxes were separately exposed for additional 5 and 8 min, respectively. All images were taken via chemiluminescent detection in a Fuji Imager. **(C-D)** Summary of cBRET results from LuTHy experiments performed with HEK293 cells co-expressing ASPL and U2AF2 from LuTHy vectors in eight different combinations. The red dashed lines represent the threshold for cBRET at 0.01. In (C) U2AF2 was applied as a donor and ASPL as an acceptor. In (D), respectively, ASPL as an acceptor and U2AF2 as a donor. All binary tests were performed as four biological replicates, Bars represent mean ± SEM. **(E)** Donor saturation experiments were conducted with the best two orientations: U2AF2-NL vs. ASPL-mCit and ASPL-NL vs. U2AF2-mCit. In both cases 1 ng of donor plasmid was co-transfected in HEK293 cells with increasing amounts of acceptor plasmid ranging from 0 to 200 ng. Each dot represents technical duplicates. A non-linear fit for one site specific binding was performed in GraphPad Prism 7.03. **(F)** U2AF2-NL vs. ASPL-mCitrine wild-type and single [s. m.] or double [d. m.] mutant variants as measured by BRET. The dashed red line marks the threshold for cBRET at 0.01 Bars represent means of biological triplicates ± SEM. One-way ANOVA and Dunnett's multiple comparison test were performed (***$p<0.001$)

Considering these results and the new interaction surface emerging on p97:ASPL molecules[43,63], it seems likely that this surface area of the protein complex interacts with U2AF2. These findings once more indirectly supported my previous observations for ASPL dependency of the association between U2AF2 and p97, which I decided to investigate with further experiments.

2.1.8 p97 and U2AF2 interact exclusively in the presence of wild-type ASPL

With the aim of proving the specificity of the ASPL-dependent effect on the interaction between p97 and U2AF2 I conducted titration experiments. HEK293 cells were co-transfected with constant amounts of NL-p97 and U2AF2-mCitrine along with increasing concentrations of cmyc-epitope-tagged ASPL (**FIG. 10A**). I applied either wild-type ASPL or the mutated ASPL variants, D351A or PP437-438AA, which, as mentioned in the previous sections, bind less efficiently to p97, fail to disassemble p97 hexamers and reduce the binding affinity of ASPL for U2AF2. Notably, only in the test with wild-type ASPL a concentration-dependent increase in the BRET and LuC signals was measured. No such trend could be observed when either of the mutant ASPL variants was applied (**FIG. 10B**). These results further support my hypothesis that ASPL binding to p97 is critical for the interaction with U2AF2.

Next, donor-saturation experiments with fixed amounts of NL-p97 and increasing amounts of U2AF2-mCitrine in the presence of ASPL were executed with the intention to demonstrate the specificity of the tested interaction (**FIG. 10C**). When no ASPL or either of the two mutants, ASPL-D351A or ASPL-PP437-438AA, were used, the obtained BRET ratios delivered flat linear trends. On the contrary, transfecting the cells with wild-type ASPL revealed a strong increase of the BRET signal, resulting in a specific non-linear saturation curve (**FIG. 10D**).

Based on the outcome of these experiments, I can claim that U2AF2 binds to the p97:ASPL complex in an ASPL-dependent manner. The observed effects with both ASPL mutants and the fact that they bind weakly to both U2AF2 and p97 and also do not disassemble the p97 hexamers, is another indication that the unique surface on p97:ASPL heterotetramers[43,63] is critical for the protein-protein interaction with the splicing factor.

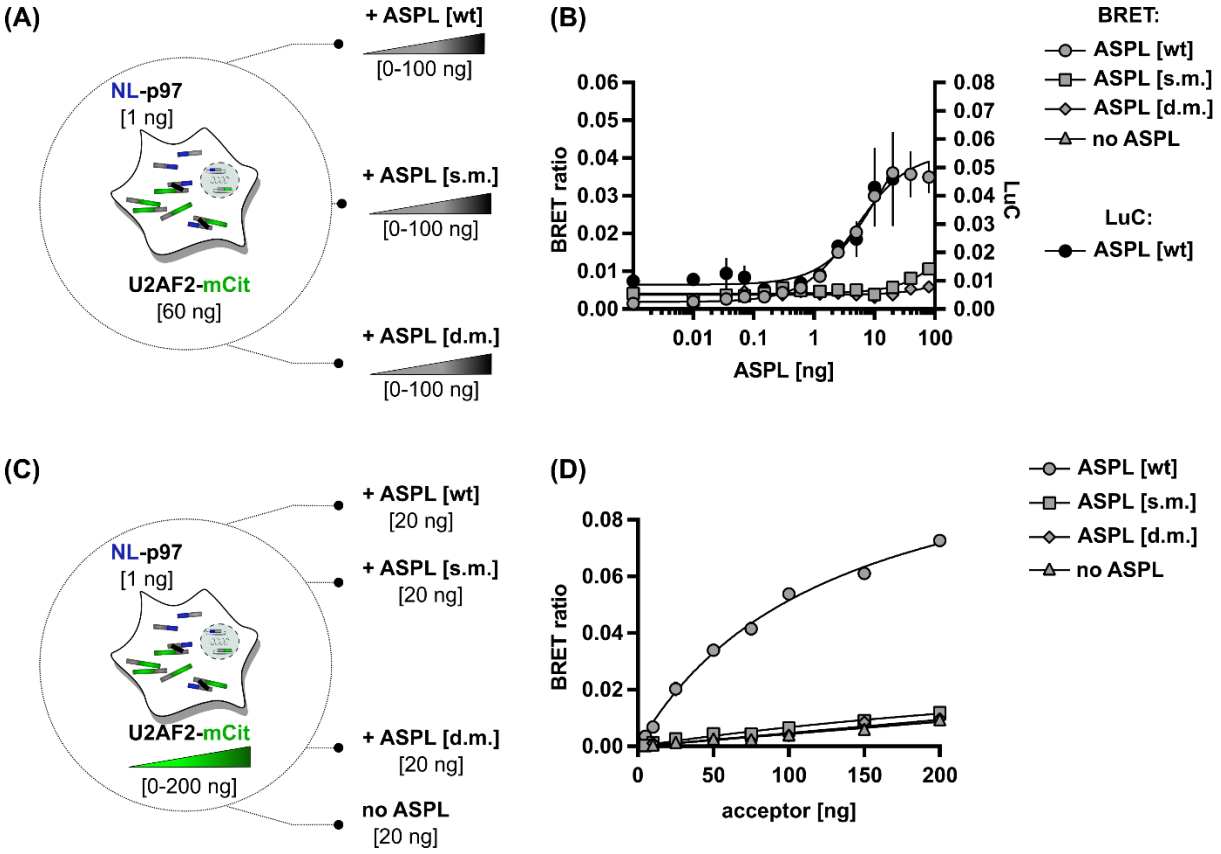


FIG. 10: p97 interacts with U2AF2 in a specific and ASPL-dependent manner.

(A) Scheme of titration experiments in HEK293 cells co-expressing 1 ng of pcDNA3.1-NL-p97, 60 ng of pcDNA3.1-U2AF2-mCitrine and increasing amounts [0-100 ng] of wild-type, single [s. m.] or double mutant [d. m.] pDEST26-cmyc-ASPL variants. **(B)** BRET and LuC measurements of titration experiments described in **(A)**. Values are presented as mean ± SEM and fitted in a non-linear regression function (dose-response-stimulation, log (agonist) vs. response (three parameters)). BRET measurements with wild-type ASPL include five biological replicates, LuC measurements were conducted as biological duplicates and plotted on the right y-axis. **(C)** Schematic overview of donor-saturation experiments performed in HEK293 cells co-transfected with 1 ng of pcDNA3.1-NL-p97, increasing amounts of pcDNA3.1-U2AF2-mCitrine [0-200] ng and 20 ng of pDEST26-cmyc-ASPL variants (wild-type or mutants, same as in (A)). **(D)** Donor-saturation curves for the interaction between U2AF2 and p97 were created with a non-linear regression fit function (one site – specific binding). Values are means of technical triplicates.

2.1.9 U2AF2 interacts with p97:ASPL at the endogenous level

The next effort in my investigations to study the interaction between p97:ASPL and U2AF2 was to test the protein assembly at endogenous conditions. To achieve this goal, I performed a proximity ligation assay (PLA) in HEK293 wild-type cells and used an ASPL knock-out line generated via CRISPR/Cas9 as a negative control (**FIG. 11**).

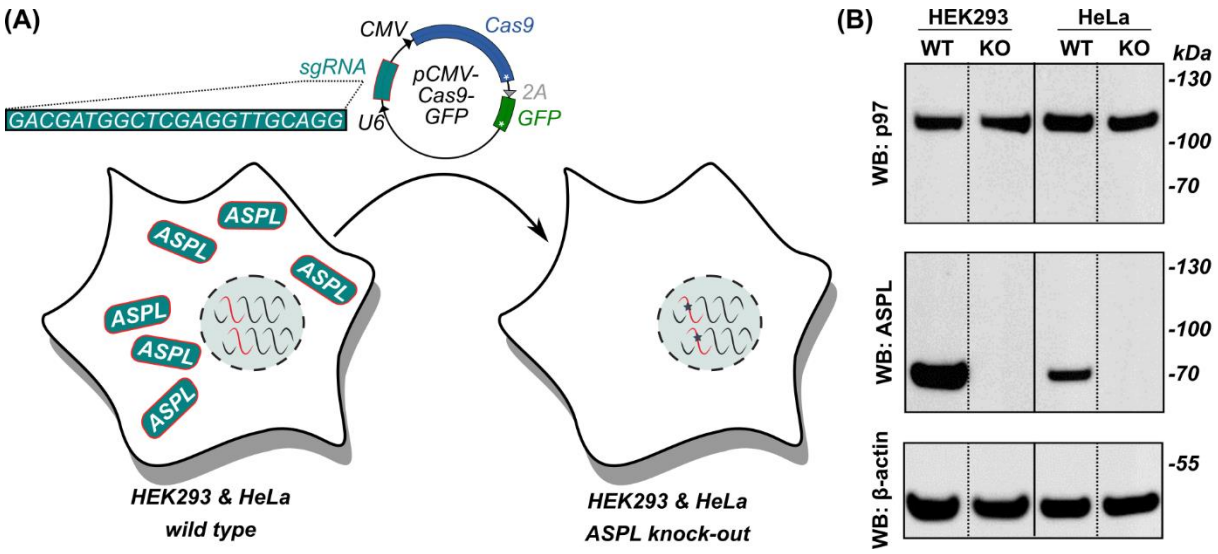


FIG. 11: ASPL deletion in HEK293 and HeLa cells and KO validation.

(A) Expression of endogenous ASPL in HEK293 and HeLa cells was abolished by CRISPR/Cas9 targeting exon 4 of the ASPL gene on chromosome 17. Wild-type cells were transfected with a pCMV plasmid expressing both the sgRNA and Cas9. This cell line was generated and provided by Anup Arumughan and Alexandra Redel in the E. Wanker lab, MDC. **(B)** Cell lysates from wild-type and ASPL-KO HEK293 cells were analyzed for ASPL expression by SDS-PAGE and western blotting. p97 expression is identical in both cell lines, but no ASPL bands can be detected in the KO cells. Used antibodies: a-VCP for p97 detection from ProGen (Cat. No: 65278), dilution 1:5000; a-TUG for ASPL detection from Abcam (Cat. No: ab80659), dilution 1:2000.

According to the experimental design, the fluorescent puncta typical for positive PLA probes should be visible in the wild-type, but not in ASPL knock-out cells as depicted in **FIG. 12A**. This was indeed the case and applying anti-p97 and anti-U2AF2-specific antibodies on both cell lines led to a significantly higher number of interaction spots in the wild-type cells (25.3 ± 3.4) compared to the ASPL knock-out cells (11.9 ± 1.0) (**FIG. 12B**). The fluorescent images in **FIG. 12D-E** clearly demonstrated that the negative control combinations, NC1 and NC2, delivered similarly low signals in both cell lines,

5.1 ± 0.4 in WT and 3.9 ± 1.0 spots/cell in KO cells. These control samples contained respectively specific anti-p97 (NC1) and anti-U2AF2 antibodies (NC2) plus the corresponding isogenic control IgG.

In summary, my studies strongly indicate that the presence of endogenous ASPL is critical for measuring an interaction between the splicing factor U2AF2 and p97 in wild-type HEK293 cells. This supports my hypothesis that ASPL functions as a bridging factor for the interaction between endogenous U2AF2 and p97 in cells.

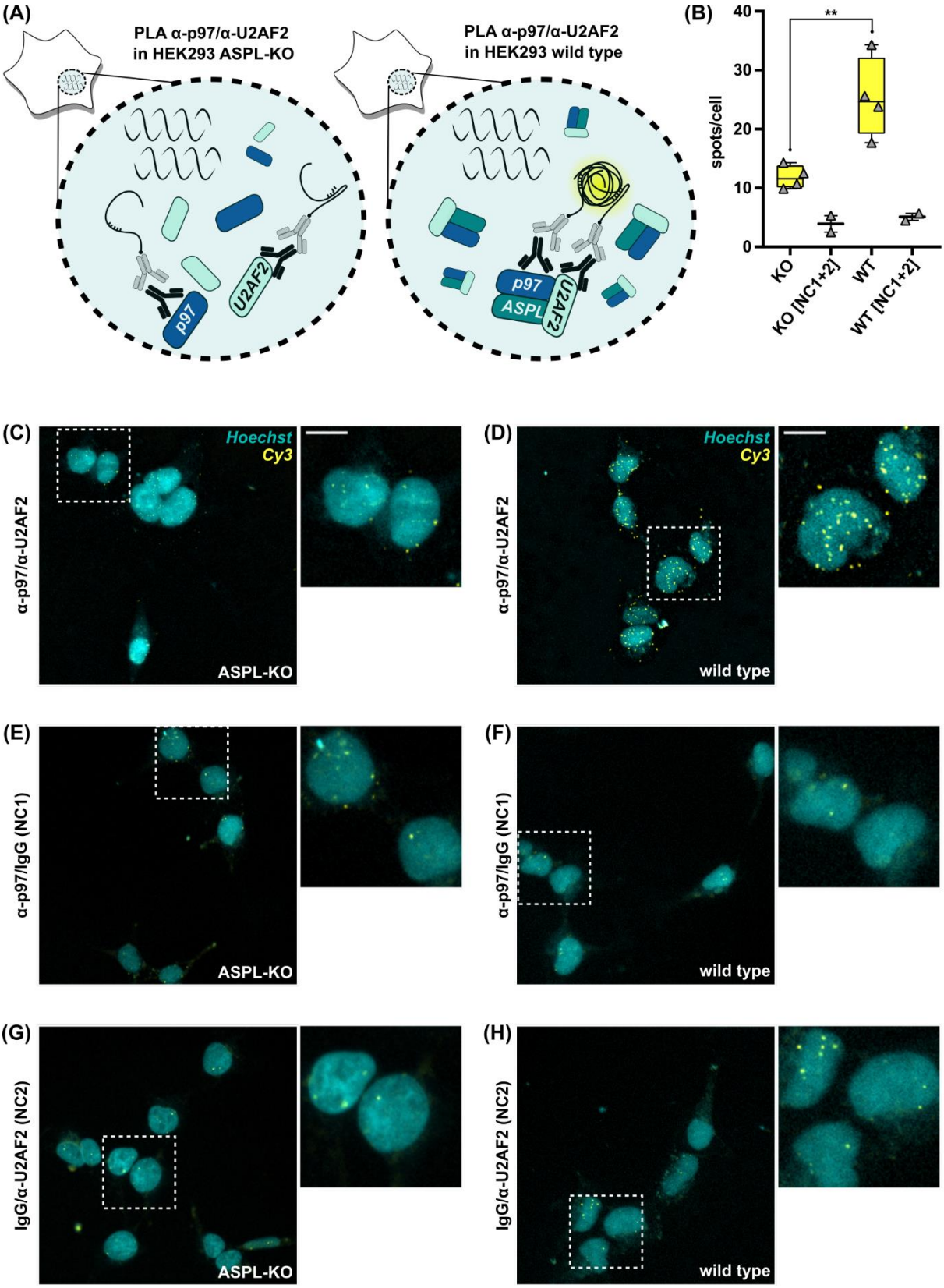

(A)
PLA α-p97/α-U2AF2
in HEK293 ASPL-KO
PLA α-p97/α-U2AF2
in HEK293 wild type
p97
U2AF2
p97
ASPL
U2AF2
(B)
spots/cell
40
30
20
10
0
**
KO
KO [NC1+2]
WT
WT [NC1+2]
(C)
Hoechst
Cy3
α-p97/α-U2AF2
ASPL-KO
(D)
Hoechst
Cy3
α-p97/α-U2AF2
wild type
(E)
α-p97/IgG (NC1)
ASPL-KO
(F)
α-p97/IgG (NC1)
wild type
(G)
IgG/α-U2AF2 (NC2)
ASPL-KO
(H)
IgG/α-U2AF2 (NC2)
wild type

FIG. 12: Quantifying the interaction between endogenous U2AF2 and p97 using PLA.

(A) Schemes depicting the PLA principle for detecting the interaction between p97 and U2AF2 in ASPL knock-out (left) and wild-type HEK293 cells (right). **(B)** Quantification of interaction spots in ASPL-KO and WT cells as analyzed by an unpaired t-test, **p<0.01. Mean values in the boxes represent measurements in KO and WT cells calculated from two biological replicates. A total number of 54 KO and 49 WT cells from four technical replicates were analyzed and depicted as single values in each box. NC1 and NC2 are respectively calculated from one biological replicate each and were combined for quantification. Total number of 38 KO and 34 WT cells was analyzed for NC1 + NC2 quantification. **(C-H)** Representative immunofluorescence microscope images for all PLA antibody combinations applied in both cell lines. Images were obtained with Zeiss Axio Imager (20x dry objective, scale bar = 10 µm).

2.1.10 Identification of domains in U2AF2 required for association with p97:ASPL

To analyze which protein domains in the splicing factor are critical for the interaction with p97:ASPL, four U2AF2 fragments containing different functional parts of the protein were created. Full-length U2AF2 is comprised of five distinct structured elements[60]. An N-terminal arginine-serine-rich domain (RS), responsible for the nuclear localization of the splicing factor[64], is followed by a tryptophan containing peptide named UHM ligand motif (ULM). The ULM domain is essential for mediating protein-protein interactions among different splicing factors as it associates with the U2AF homology motif (UHM) present in several nuclear proteins[58,65]. The ULM-UHM domain interaction facilitates the formation of higher order structures and their assembly at the sites of mRNA splicing[60,66]. U2AF2 contains a UHM domain at its C-terminus, which enables its binding to the splicing factor SF1[67,68]. The N-terminal ULM domain, however, is required for the dimerization of U2AF2 with U2AF1 at the 3' site of a constitutively spliced intron[69]. What facilitates the binding of U2AF2 to pre-mRNA molecules is a tandem RNA recognition motif (RRM) located between the UHM and ULM domains. This is also the only part of the protein whose structure was resolved by X-ray crystallography[70]. **FIG. 13A** illustrates the domain architecture of U2AF2 and its association with U2AF1 and SF1 at an intronic sequence during the very early stages of initiation of the splicing process (adapted from Agrawal *et al.* 2016[70]).

Through PCR and gateway cloning U2AF2 fragments with the following composition were produced: Fragment-1 (FR1) misses the N-terminal RS domain; additionally, in fragment-2 (FR2) the ULM domain was depleted; Fragment-3 (FR3) contains exclusively the RNA recognition motives and fragment-4 (FR4) differs from the full-length protein

by the lack of the C-terminal UHM domain (**FIG. 13B**). All fragments were shuttled in pcDNA3.1 expression plasmid with NanoLuc or PA-mCitrine sequences at the C-terminus. Their expression and cellular localization were validated by western blotting and immunofluorescence microscopy, respectively (**FIG. S2**). The immunofluorescence microscopy images clearly showed that besides FR4, which is present exclusively in the nucleus, all other fragments are also partially present in the cytosol which is consistent with previous findings made by Gama-Carvalho and colleagues (2001)[64] that the RS domain contains a nuclear localization signal (NLS), which influences the localization of the protein in cells.

Firstly, HEK293 cells were consecutively co-transfected with ASPL-mCitrine-PA and either full length U2AF2 or a fragment fused to NanoLuc at the C-terminus. In-cell BRET measurements were performed 72 hours after the transfection and the calculated cBRET values indicated that the C-terminal UHM domain mediates the interaction between the splicing factor and ASPL (**FIG. 13C**). FR1 and FR2 did not significantly differ from full-length U2AF2 and performed similarly well as donors in the energy transfer measurements. In contrast, the binding of the RRM1-RRM2 domain (FR3) to ASPL was significantly reduced. No binding, however, was detected between ASPL-mCitrine-PA and U2AF2-NL-FR4, which lacks the UHM domain. It is noteworthy, that FR3 and FR4 behave in a similar way and do not significantly differ from each other according to the applied statistical test for multiple comparisons.

Subsequently, I tested the interaction between the same U2AF2 fragments tagged with mCitrine at the C-terminus and NL-p97 in the presence of increasing amounts of wild-type ASPL. Surprisingly, all tested fragments exhibited significantly lower cBRET values compared to full-length U2AF2 (**FIG. 13D**). FR3 was not interacting with NL-p97 as no BRET could be measured despite the addition of ASPL to the transfection mix. However, the saturation curves resulting from the titration experiments with FR1, FR2 and FR4 looked similar. When the C-terminal UHM domain in U2AF2 was deleted, a significantly reduced binding towards p97:ASPL was measured. However, deleting the N-terminus (as in FR1 and FR2) also resulted in reduced cBRET values. These observations suggest that both, the N- and C-terminal domains in U2AF2 are critical for the association of the splicing factor with p97:ASPL complex. Additionally, considering the cellular localization of the different fragments (**FIG. S2C**), which is exclusively nuclear

for FR4 in contrast to FR1 and FR2, which are present both in the nucleus and the cytoplasm, the conclusion can be drawn that the RS domain in U2AF2, which functions as an NLS, contributes but is not essential for the interaction with p97.

In this regard, one can assume that there is a complementary interplay between the UHM domain at the C-terminus and the N-terminal region (RS and ULM domains) when U2AF2 associates with p97:ASPL complex. Based on these findings, I suggest a bipartite binding mode, when the splicing factor associates with the heterooligomeric p97:ASPL complex. In this model the C-terminal UHM domain directly interacts with ASPL, while the N-terminal RS-ULM domain in U2AF2 binds to p97. In both cases, the U2AF2 protein binds to an interaction surface, which is exclusively present in p97:ASPL heterotetramers[43,63] but not in p97 hexamers or ASPL (**FIG. 13E**).

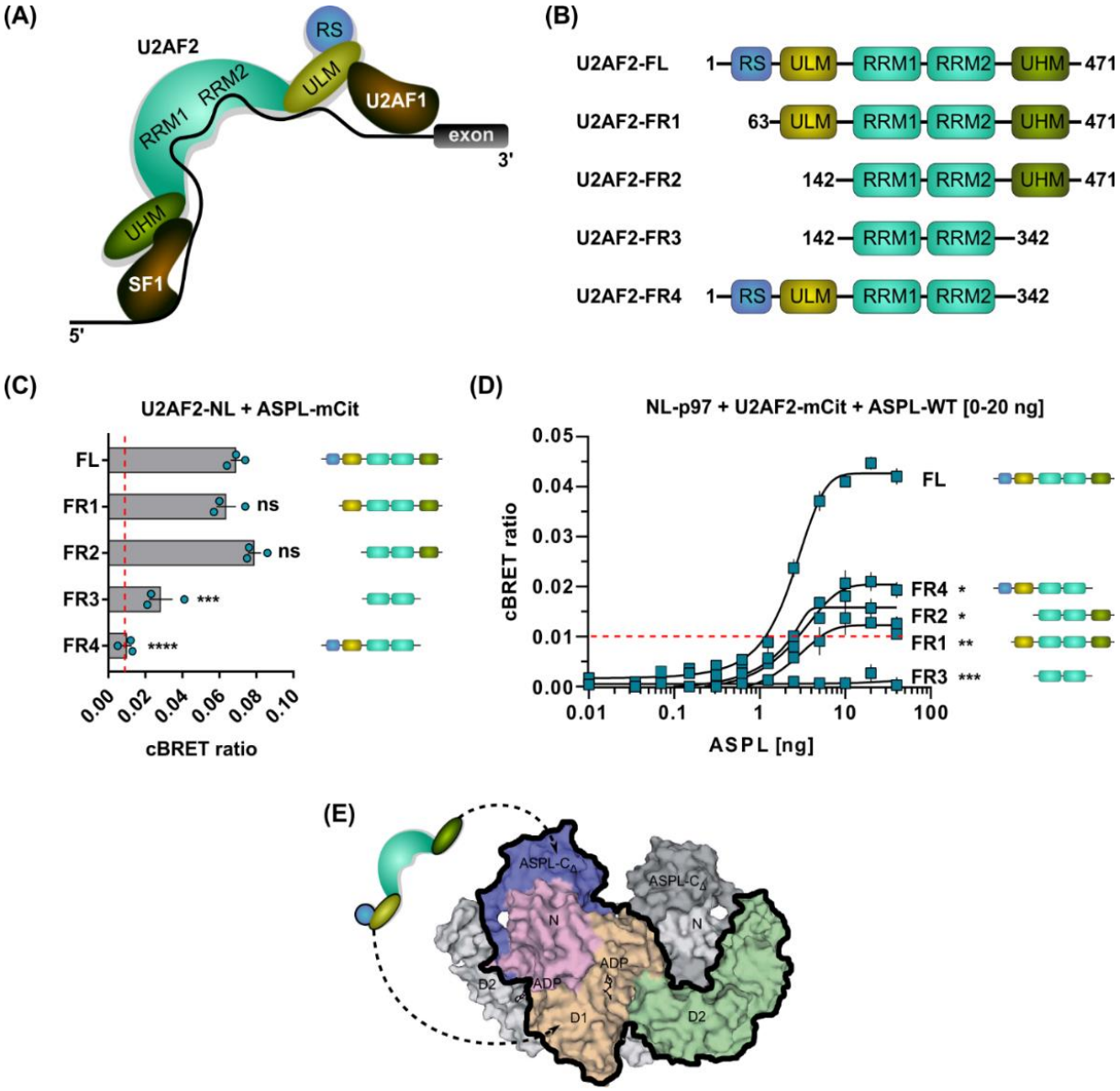


FIG. 13: U2AF2 domain mapping analysis.

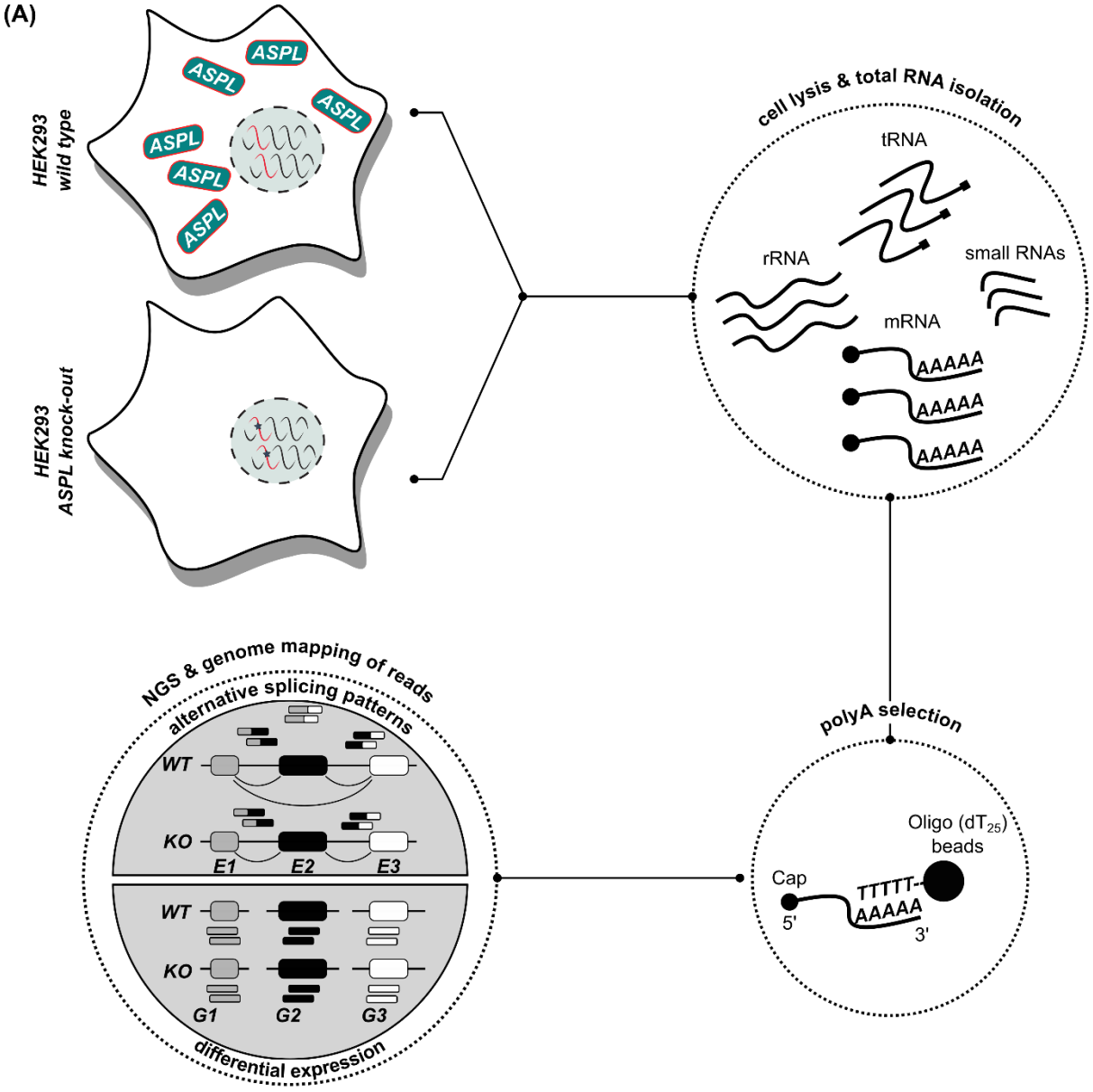
(A) Schematic illustration of U2AF2 binding to splicing factor 1 (SF1), U2 small nuclear auxiliary factor 1 (U2AF1) and the 3'splice site of an intron containing pre-mRNA during splicing initiation (adapted from Agrawal *et al.* 2016) **(B)** Domain composition of full-length U2AF2 and four fragments used for domain mapping. **(C)** Binary BRET tests between ASPL and full-length U2AF2 or truncated variants. Bars represent cBRET values from three biological replicates calculated as mean values ± SEM. One-way ANOVA and a Dunnett´s multiple comparisons test were applied for the statistical analysis (***p<0.001, ****p<0.0001). **(D)** Titration experiments for testing the interaction between U2AF2 and p97:ASPL conducted as biological triplicates. The cBRET ratios represent mean values ± SEM. Statistical analysis was performed as described in (C) (*p<0.05, **p<0.01, ***p<0.001). **(E)** Hypothetic model for bipartite binding between U2AF2 and the p97:ASPL complex. The C-terminal UHM domain is most likely required for binding to ASPL in the heterotetrameric assembly with p97. The attachment of U2AF2 to p97 could be mediated by the N-terminal RS-ULM domains. The border of one of the heterodimeric p97:ASPL units is highlighted with a continuous black line. ASPL-C is shown in purple, the p97-N domain in pink, the p97-D1 and p97-D2 domains, respectively, in yellow and green. For simplicity the second p97:ASPL dimer is indicated in grey in the background.

2.2 Functional characterization of the p97:ASPL complex

2.2.1 Studying the effects of p97:ASPL on the human transcriptome via RNA-Seq

Gaining information about the interaction network of the p97:ASPL complex, was a valuable asset on the way towards its functional characterization. The findings described in the previous chapter and the observed binding of several splicing factors to the complex, encouraged me to take a closer look at the effects of ASPL deletion on the overall splicing events in human cells. For this purpose, the ASPL knock-out HEK293 cell line was used (**see FIG. 11**). These cells together with wild-type HEK293 cells were subjected to an RNA-Seq analysis in order to detect transcripts with altered expression or splicing patterns upon ASPL ablation.

FIG. 14A contains a flow-chart illustrating the strategy for the generation of the RNA-Seq samples from the two cell lines and their subsequent analysis. After cell lysis, total RNA isolation was performed in an identical manner for both samples. Subsequently, mRNA molecules were enriched by polyA selection[71], which eliminates any unwanted RNA molecules, such as rRNA and other RNA species. Following a cDNA synthesis step, the obtained library was sequenced, and the resulting reads mapped to a human reference genome. Two data sets were generated with the tools mentioned in the **Methods** section: One containing a list of all differentially expressed genes upon ASPL-KO and one with all transcripts demonstrating significant changes in the frequencies of detected splicing events (ΔPSI). The top 40 gens in each data set are listed in **Table S5**,**Table S6** and **Table S7**.Additionally, principle-component-analysis (PCA) was conducted using the gene expression characteristics of all four biological replicates for both cell lines in order to validate the sample quality. It revealed two distinct clusters for the wild-type and ASPL-KO samples, as expected (**FIG. 14B**).


(A)
HEK293
wild type
ASPL
HEK293
ASPL knock-out
cell lysis & total RNA isolation
tRNA
rRNA
small RNAs
mRNA
AAAAA
polyA selection
Oligo (dT25)
beads
Cap
5'
TTTTT-
AAAAA
3'
NGS & genome mapping of reads
alternative splicing patterns
WT
KO
E1
E2
E3
G1
G2
G3
differential expression

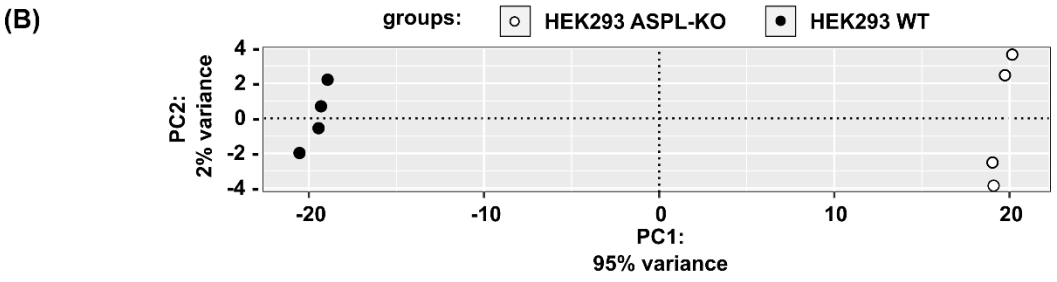

(B)
groups:
HEK293 ASPL-KO
HEK293 WT
PC2:
2% variance
4
2
0
-2
-4
-20
-10
0
10
20
PC1:
95% variance

FIG. 14: Transcriptome analysis of wild-type and ASPL-KO HEK293 cells by RNA-Seq.

(A) Flow-chart depicting the generation of the RNA-Seq data sets. Wild-type and ASPL-KO HEK293 cells were lysed for total RNA isolation. Biological replicates (four in total) were subjected to polyA selection for the enrichment of mRNA molecules prior to cDNA library generation. Next-generation sequencing was performed with an Illumina NextSeq 500 sequencer at the Genomics facility in MDC Berlin in collaboration with Claudia Langnick and Sasha Sauer. The obtained paired-end reads were aligned to the human reference genome GRCh38/hg38 assembly. Alternative splicing and differential expression analysis were performed as described in the **Methods** section. E1, E2 and E3 indicate exon1, exon2 and exon3 G1, G2, G3 stand for gene1, gene2, gene3. **(B)** Principle component analysis (PCA) describing the variance between both cell lines based on four biological replicates for each cell line. Two principle components are sufficient to explain 97% of the variance within these samples; ASPL-KO and WT cells clearly form two distinct clusters. The analysis of the raw NGS data and PCA were kindly performed by Alexander Neumann in the Heyd lab, FU Berlin.

2.2.2 ASPL knock-out affects gene expression and splicing in HEK293 cells

The results from the RNA-Seq analysis convincingly show that knock-out of ASPL has an impact on the human transcriptome. With a single glance at the volcano plot in **FIG. 15A** one can easily notice that there is a large number of genes, which are differentially expressed (both up- and downregulated) in ASPL knock-out compared to wild-type cells. In total, 2,176 transcripts differentially expressed genes were identified, of which 817 (37.5%) were down- and 1,359 (62.5%) upregulated in ASPL-KO cells.

GO-term enrichment analysis was performed separately for both groups of genes using the online tool **PANTHER14.1** (http://pantherdb.org/). For this purpose, I used the human genome as a reference and searched for statistically overrepresented GO-terms (corrected $p<0.05$) in the categories "biological process" and "molecular function".

Interestingly, I found that a significant fraction of the upregulated transcripts in ASPL-KO cells play a role in synaptic functions or regulate ion channel activity, suggesting that these processes are altered upon ASPL depletion (**FIG. 15B**). Since the analyzed samples originate from HEK293 cells, this appears to be a rather surprising finding. Nevertheless, one should keep in mind that according to some published studies HEK293 cells have presumably a neuronal origin and express more than 60 neuron-specific genes[72–74], which can serve as an explanation for the results from the GO-term analysis.

Furthermore, among the downregulated genes there was an enrichment for targets with a role in central molecular processes such as the response to starvation, extracellular matrix organization and even embryonic morphogenesis (**FIG. 15C**). These results indicate that the depletion of ASPL in HEK293 cells affects a broad range of fundamental cellular mechanisms. Therefore, I can postulate that the p97:ASPL complex impacts the transcription of a large number of proteins with distinct and essential molecular functions in human cells.

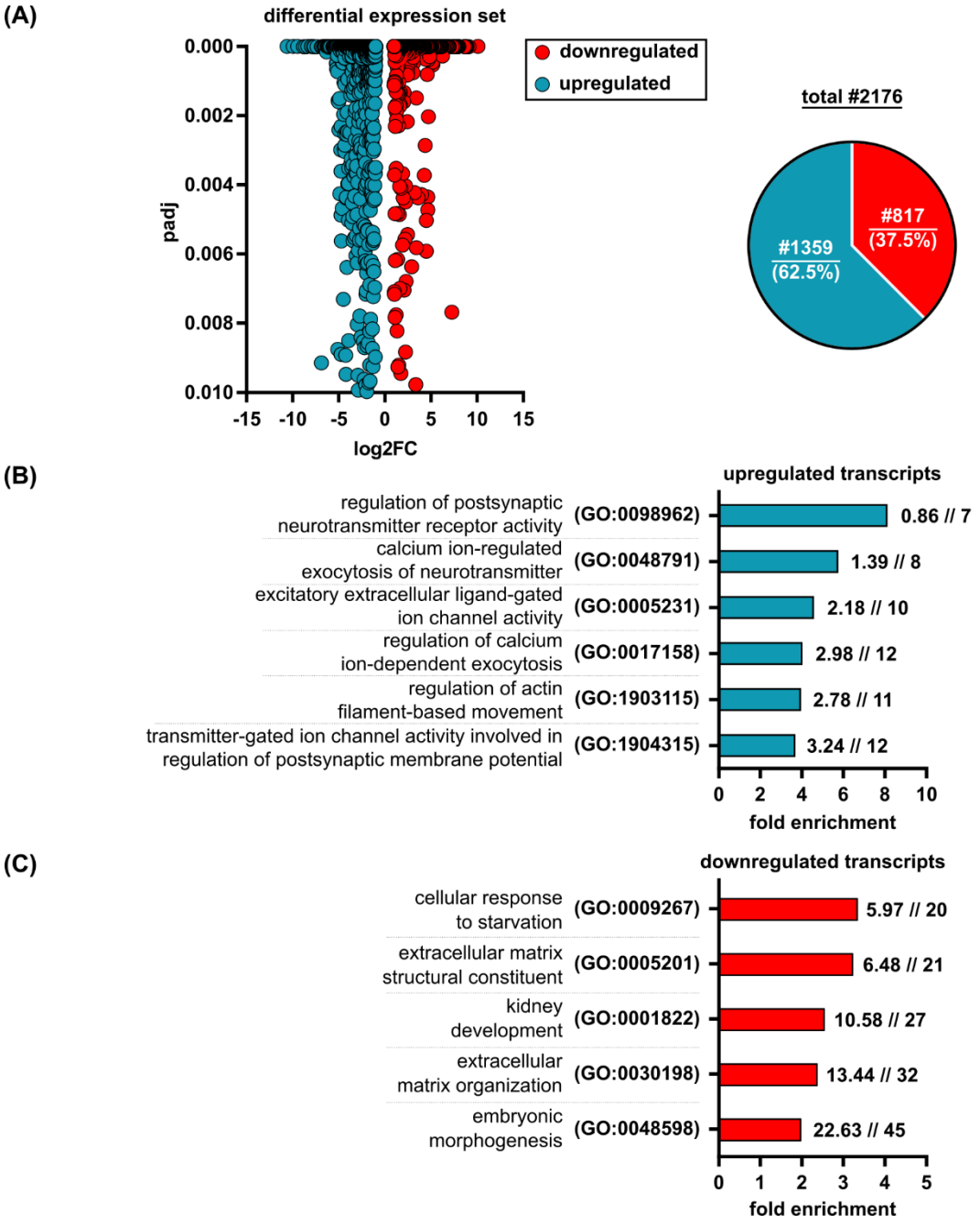


FIG. 15: ASPL-KO alters the expression of a very diverse set of human genes

(A) Volcano plot depicting the distribution of differentially expressed genes upon ASPL knock-out. Selection criteria are log2FC of at least 1 or -1 and padj<0.01. Total number of 2,176 genes exhibit changes in their expression, 1,359 (62.5%) of them get upregulated and 817 (37.5%) downregulated when ASPL expression is abolished. **(B)** GO-term enrichment analysis of all upregulated transcripts was performed with the **PANTHER14.1** classification system. The applied test was a Fisher´s Exact test and a p<0.05 was set as cut-off for the p-values after false discovery rate correction. Only groups with a fold enrichment of 3.5 and higher are displayed. The digits on the right side of the bars represent the number of genes from the respective GO-term group expected to be present in any randomly selected group of genes, followed by the number of genes encountered in the analyzed set. **(C)** Same as (B), for all downregulated transcripts. Fold enrichment threshold for the displayed groups was >1.5.

I followed a similar approach analyzing the enrichment of GO-terms within the list of transcripts which demonstrated reduction or increase in the frequency of detected splicing events (ΔPSI). These were in total 1,329 transcripts, of which 512 (38%) and 837 (62%) showed enhanced and reduced splicing rates, respectively, in ASPL knock-out cells (**FIG. 16A**). Interestingly, I was able to obtain results for overrepresented GO-terms only within the latter group. Similar to the outcome of the analysis of the differentially expressed genes, in this case there the affected pathways were very diverse. Unexpectedly, the GO-term ranked with the highest fold enrichment (8.46) contains genes that play a functional role in myelination in the peripheral nervous system (**FIG. 16B**). This is followed by transcripts encoding proteins that participate in the activation of GTPases, the regulation of transcription, RNA metabolism or biosynthetic processes. From the analysis of the alternative splicing (AS) data set, I conclude that the p97:ASPL complex most likely plays a role in mRNA splicing and influences the processing of a large number of transcripts, which result in proteins with very divergent molecular functions.

In summary, I want to note that there are opposing effects in both data sets which means that p97:ASPL cannot be generally defined either as an activator or as an inhibitor of transcription or splicing. However, similar effects have previously been observed, when important transcription factors such as SOX9[75] and Zfp871[76] were depleted in mammalian cells. This was the first hint that the genes influenced by the heterotetrameric p97:ASPL complex are more likely to possess sequence features[77] that classifies them as targets of the complex rather than their cellular functions. In order to prove this assumption, I decided to investigate the different groups of alternative splicing events in detail and search for specific shared sequence characteristics among the affected transcripts.

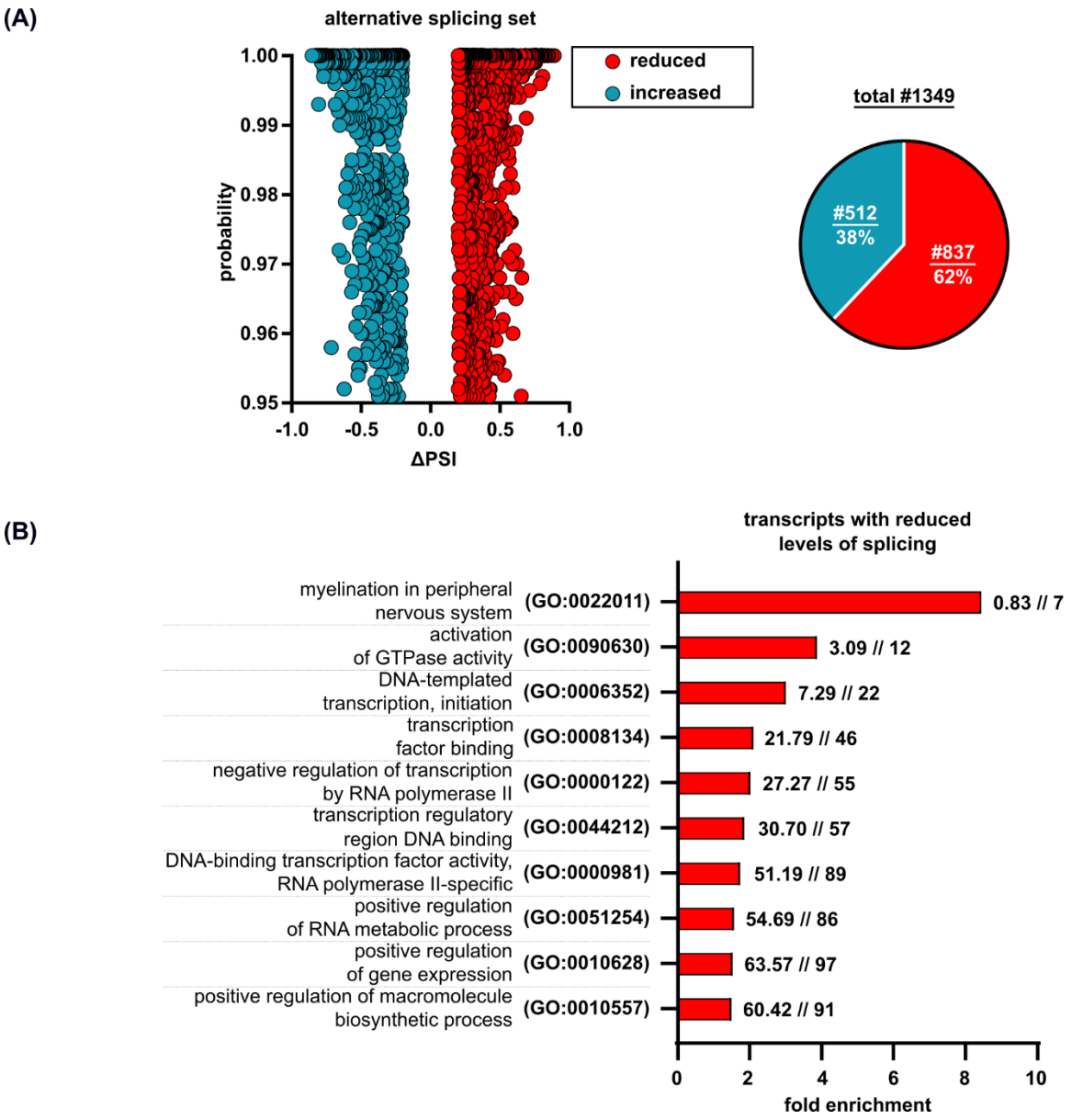


FIG. 16: Alternative splicing of functionally unrelated genes is affected by ASPL-KO.

(A) Volcano plot illustrating the distribution of transcripts showing significant changes in the frequency of splicing events as a result of ASPL-KO (PSI = "percentage spliced in", ΔPSI = PSI(WT) – PSI(KO)). Selection criteria are ΔPSI of at least 0.2 or -0.2 and probability values > 0.95. 1,349 transcripts were modified, 837 demonstrate reduced PSI and 512 increased PSI values in cells in which ASPL expression is abolished compared to wild-type HEK293 cells. **(B)** GO-term enrichment analysis of transcripts demonstrating reduced levels of splicing upon ASPL-KO performed as described in FIG. 15B-C. Cut-off for the fold enrichment of GO-terms to be displayed in the plot was >1.5.

FIG. 17A gives an overview of the eight alternative RNA processing events, which were significantly changed upon ASPL knock-out in HEK293 cells. These are tandem alternative polyadenylation (TE), tandem transcription start site usage (TS), cassette exon usage (CE), intron retention (IR), inclusion of alternative first (AF) or last exon (AL) and alternative donor (AD) or acceptor site usage (AA).

The majority of the alternatively spliced transcripts in ASPL-KO cells are comprised of isoforms resulting from TE events (39%) (**FIG. 17B**). These mRNA molecules differ in the composition and length of the 3' untranslated region[78]. These were followed by a group of genes expressing isoforms from tandem transcription start sites (24%) (**FIG. 17B**). TS events result from the differential binding of transcription factors which define the transcription start site and lead to modified 5' untranslated reagions[79,80]. The third most frequently detected alternative splicing event upon ASPL-KO is the splicing of cassette exons (CE) with 23% (**FIG. 17B**). In this case the matured transcripts either possess, in the case of exon inclusion, or exclude a CE, when it is skipped by the splicing

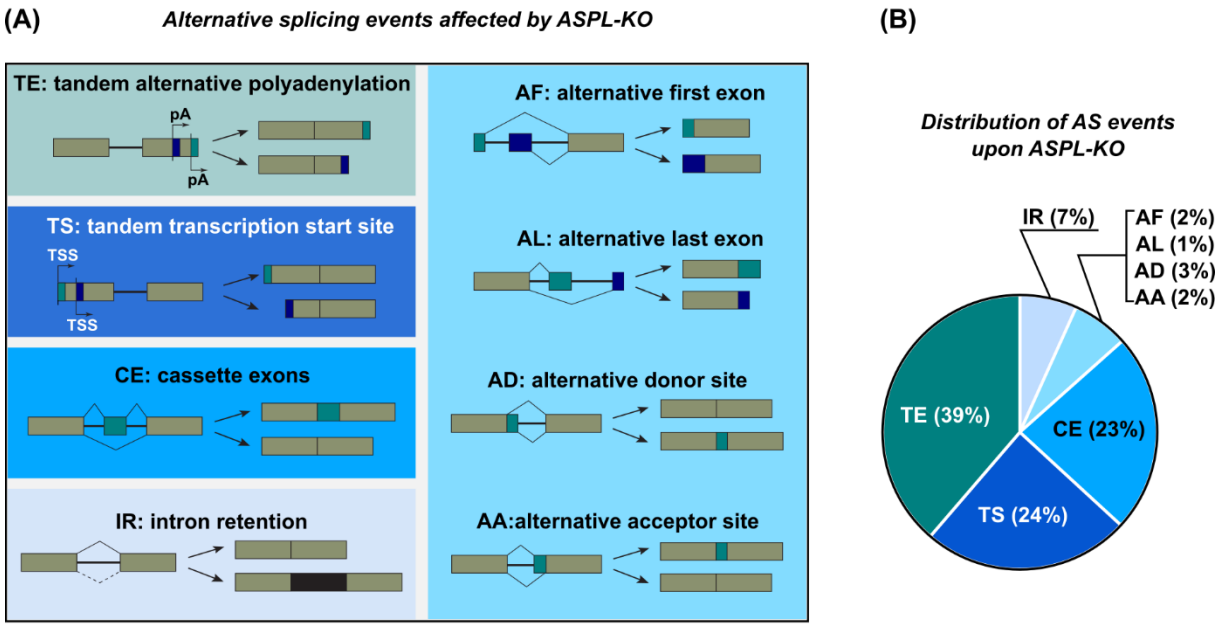


FIG. 17: Distribution of alternative splicing (AS) events perturbed by ASPL deletion.

(A) Occurrence of different RNA processing events in wild-type and ASPL knock-out HEK293 cells was analyzed by "Whippet" (Sterne-Weiler *et al.* 2018). All transcripts selected in the AS list (total 1349) show either significantly reduced or increased splicing when ASPL was deleted. The affected AS events are distributed in eight groups: 522 in TE (tandem alternative polyadenylation), 330 in TS (tandem transcription start site), 316 in CE (cassette exons), 90 in IR (intron retention), 21 in AF (alternative first exon), 11 in AL (alternative last exon), 34 in AD (alternative donor site) and 25 in AA (alternative acceptor site). pA: polyadenylation; TSS: transcription start site. **(B)** Distribution of AS events upon ASPL-KO in percent.

machinery[81,82]. It is noteworthy, that in the human transcriptome exon skipping is the most frequent AS event[83]. On the contrary, alternative first or last exons and exons resulting from the usage of alternative donor (5') or acceptor (3') splice sites are comparably rare in human cells. The least abundant AS event is intron retention (IR)[83,84].

In ASPL-KO cells the distribution of these AS events among the affected transcripts is similar and a rather small number of transcripts are expressed with alternative first (2%) or last exons (1%) (**FIG. 17B**). Similar numbers were obtained by transcripts with modified usage of alternative donor (3%) or acceptor splice sites (2%). Of all genes listed in the AS data set 7% contain retained introns (**FIG. 17B**). From these numbers it can be concluded that the ablation of ASPL expression in human cells alters a large number of different mRNA processing events mainly affecting polyadenylation, the selection of transcription start sites or the splicing of cassette exons.

Prior to any specific downstream investigations of the AS data set, I performed a quality control analysis of the obtained RNA-Seq data by PCR validation of randomly selected splicing events. One example of the validation procedure is presented in **FIG. 18**. I was able to demonstrate the tandem transcription start site usage within the gene RIC8A, a guanine nucleotide exchange factor[85], leading to the expression of two isoforms as depicted in **FIG. 18A**. Isoform 1, containing AUG1, was assigned a ΔPSI score of -0.72 which means that its production is enhanced in ASPL knock-out compared to wild-type HEK293 cells. On the contrary, isoform 2, with start codon AUG2, is predominantly produced in wild-type cells and in cells lacking ASPL this isoform should be reduced. In order to validate this data, I isolated mRNA from wild-type and ASPL-KO cells, reversely transcribed it into cDNA and applied it as a template in a PCR reaction. For amplification a primer pair that should only amplify a product from isoform 1 (~900 bp), but not from isoform 2 was used (**FIG. 18B**). According to the ΔPSI values for isoform 1 a band of the right size should appear mainly in the ASPL-KO sample. This indeed could be observed and only a very weak band was detected on an agarose gel for the wild-type sample (**FIG. 18C**). This successful validation of a specific splicing event indicates the robustness of the initially performed bioinformatics analysis of RNAseq data.

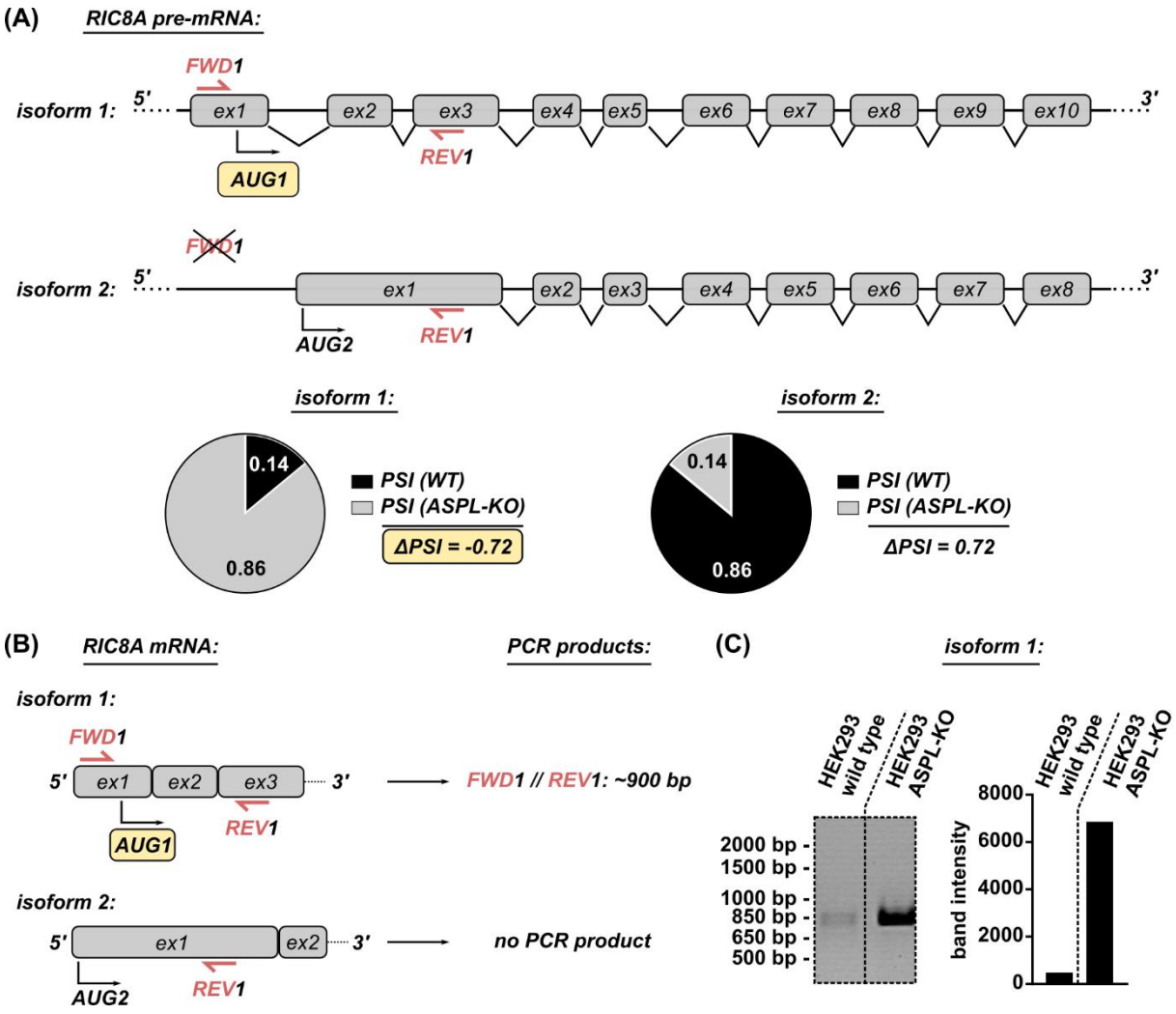


FIG. 18: PCR validation of a specific AS event confirms the quality of the RNA-Seq analysis.

(A) Two RIC8A isoforms resulting from a tandem transcription start site were affected by ASPL knock-out according to the RNA-Seq results. Isoform 1 scored a ΔPSI of -0.72 and isoform 2, respectively, a ΔPSI of 0.72. **(B)** mRNA was isolated from wild-type and ASPL-KO cells, transcribed into cDNA and used as a template in a 40 µL PCR with the primer pair FWD1/REV1. FWD1 can only bind to exon 1 in isoform 1 as its binding site remains within an intronic sequence in isoform 2. Only for isoform 1 an amplicon of ~900 bp should be detected and isoform 2 should not deliver a PCR product. **(C)** Analysis of the products from the validation PCR was conducted by applying 5 µL of both PCR samples on a 0.8% agarose gel. A single prominent band was detected with mRNA from the ASPL-KO sample. Wild-type cells delivered only a weak band of the expected size. This was confirmed by the band intensity analysis performed by Image J.

In my follow-up analysis, I focused on transcripts, which are generated by alternative splicing of cassette exons (CEs) in ASPL-KO cells (**see FIG. 17**). This process influences both the inclusion and the skipping of exons in specific transcripts[81,82]. The splicing of

CEs was selected due to three reasons. First, CEs represent the most abundant alternative splicing event in vertebrates[86]. Furthermore, the strength of the splice sites is important for CE inclusion or skipping[87] which could be the common sequence feature present in all transcripts influenced by ASPL depletion. Finally and most importantly, all three identified p97:ASPL interaction partners – U2AF2, PUF60 and RBM39 – are involved in the recognition of 3' splice sites[88,89], a process which determines the usage of cassette exons[82,90].

The inclusion or skipping of CEs can have an influence on the expression of the affected transcripts depending on the intrinsic features of the resulting isoforms. On the one hand, the inclusion of so-called "poison" exons can lead to a frame shift and the incorporation of a premature termination codon (PTC), which finally induces nonsense-mediated mRNA decay (NMD)[91,92]. On the other hand, the unintended skipping of essential exons can have the same effect on the fate of the transcript by causing a frame shift and the incorporation of a PTC followed by NMD[93]. In both scenarios the lifetime of the emerging mRNA will be perturbed. In order to test if ASPL-KO influences the expression of transcripts with significantly changed rates of CE inclusion or skipping, I plotted the ΔPSI values of those transcripts against their respective log2FC scores. In **FIG. 19A** one can see that there is no obvious correlation between the two data sets which was also supported by the performed Spearman correlation test ($r = 0.03$). This indicates that the observed changes in the expression levels of certain targets upon ASPL-KO are not due to perturbations in splicing.

Despite this, it became evident that in comparison to the abnormally included exons the number of skipped exons is significantly higher in ASPL knock-out cells (**FIG. 19B**). Taking under consideration the fact that CEs are normally defined by weaker splice sites[94–97] and the role the proteins U2AF2, PUF60 and RBM39 play in the recognition of 3' splice sites, I wondered whether the effects of ASPL knock-out on splicing depends on the nucleotide base composition of the 3'splice site. In order to answer this question, the average strength of the 3' splice sites of skipped CEs was compared with the 3' splice site strength of all detected CEs in wild-type HEK293 cells (**FIG. 19C**). This test demonstrated an enrichment of CEs with weaker splice sites among all transcripts containing cassette exons and affected by ASPL-KO. They delivered an average score

of 7.6, while the mean 3' splice site strength in the group of all detected CEs was calculated to be 8.3 (**FIG. 19C**).

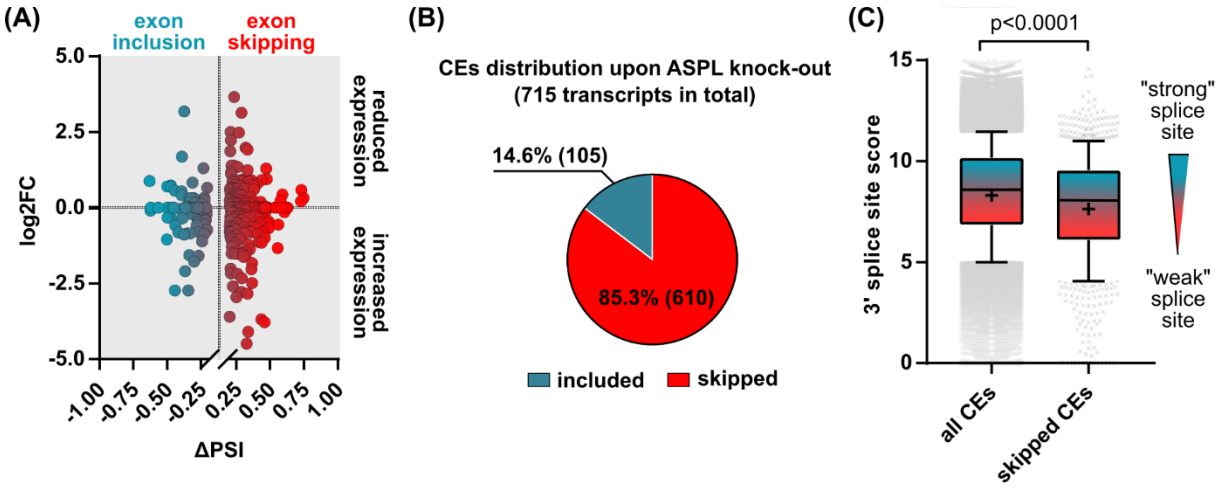


FIG. 19: ASPL-KO enhances exon skipping and perturbs weak 3' splice site recognition.

(A) Log2FC values of all transcripts containing CEs were plotted against their respective ΔPSI scores. Both data sets contain values which are not normally distributed (D'Agostino & Pearson normality test, p<0.0001) and the applied nonparametric Spearman test proved the lack of correlation between them (r = 0.03). Only transcripts were displayed with ΔPSI higher than 0.1 or lower than -0.1 and probability value of 0.95. **(B)** Percentage distribution of skipped and included exons among all significantly changed CE transcripts within the AS data set. **(C)** Comparison between all CEs detected in HEK293 cells (in total 115,183 transcripts) and all skipped CEs upon ASPL-KO based on their 3' splice site strength (ΔPSI>0.1, probability>0.90, in total 903 transcripts). 3' splice site strength was calculated with the MaxEnt Tool Alexander Neumann (Heyd lab, FU Berlin). Box and whisker plots represent the 10-90 percentile of the values for both data sets and the central line corresponds to the respective mean score: 8.3 for all CEs and 7.6 for all skipped CEs. An unpaired t-test was conducted in order to compare both mean values, p>0.0001.

Collectively, based on these results I can assume that ASPL is required for the inclusion of CEs which are under the regulation of 3' splice sites whose strength is below the average for CEs in general. Keeping in mind that ASPL is mainly present in a p97 bound state and specific p97:ASPL interaction partners are part of the machinery responsible for the recognition and regulation of splicing at the 3' splice site, one can now speculate about a specific role of the p97:ASPL complex in this process.

In order to confirm the effect of ASPL on the usage of 3' splice sites with different strengths, I transfected wild-type and ASPL deficient HeLa cells with a representative minigene splicing reporter (pyPY) containing two alternative 3' splice sites of different

strength: a strong 3' splice site with a polypyrimidine tract comprised mainly of pyrimidines (PY site), and a weaker splice site with a degenerate polypyrimidine tract including purines (py site) (**FIG. 20A**). This minigene splicing reporter was designed as a model pre-mRNA splicing substrate and was successfully used in several publications to test for U2AF2 regulated splicing activity[70,98]. After transfecting the cells with the minigene plasmid and incubation for 48 hours, total RNA was isolated and transcribed into cDNA. Finally, it was applied as a template for PCR amplification. The used primer pair specifically amplifies the transcripts resulting from splicing of the minigene. Analyzing the PCR products by gel electrophoresis should deliver three bands: One corresponding to the unspliced transcript (~380 bp), a transcript resulting from the usage of the weak py site (~260 bp) and one emerging from splicing at the strong PY site (~140 bp). In case a splicing inhibitor is used, the observed bands would have a modified intensity. Normally, endogenous U2AF2 would preferentially saturate the PY site, leading predominantly to the detection of the ~140 bp band, while the weak py site is less efficiently used. This means that in comparison to the ~140 bp the ~380 bp should be less abundant in wild-type cells. However, splicing at both sites can be detected in wild-type cells as reported by Pacheco *et al.* (2006)[98]. This is in line with my observations, as I was able to detect splicing taking place at both splice sites in wild-type HeLa cells (**FIG. 20B**). When the cells were treated with the splicing inhibitor pladienolide B (PdB) for 16 hours prior to RNA isolation, however, a reduction in the usage of both sites (~70%) was detected (**FIG. 20C**). Comparably, when expression of ASPL in HeLa cells was depleted, both splice sites were affected. However, the usage of the weak py site appeared to be perturbed to a greater extent. In ASPL-KO cells a reduction of 81% versus 28% for the strong PY site was observed. This suggests that the p97:ASPL protein complex promotes the recognition and cleavage of weak 3' splice sites in mammalian cells.

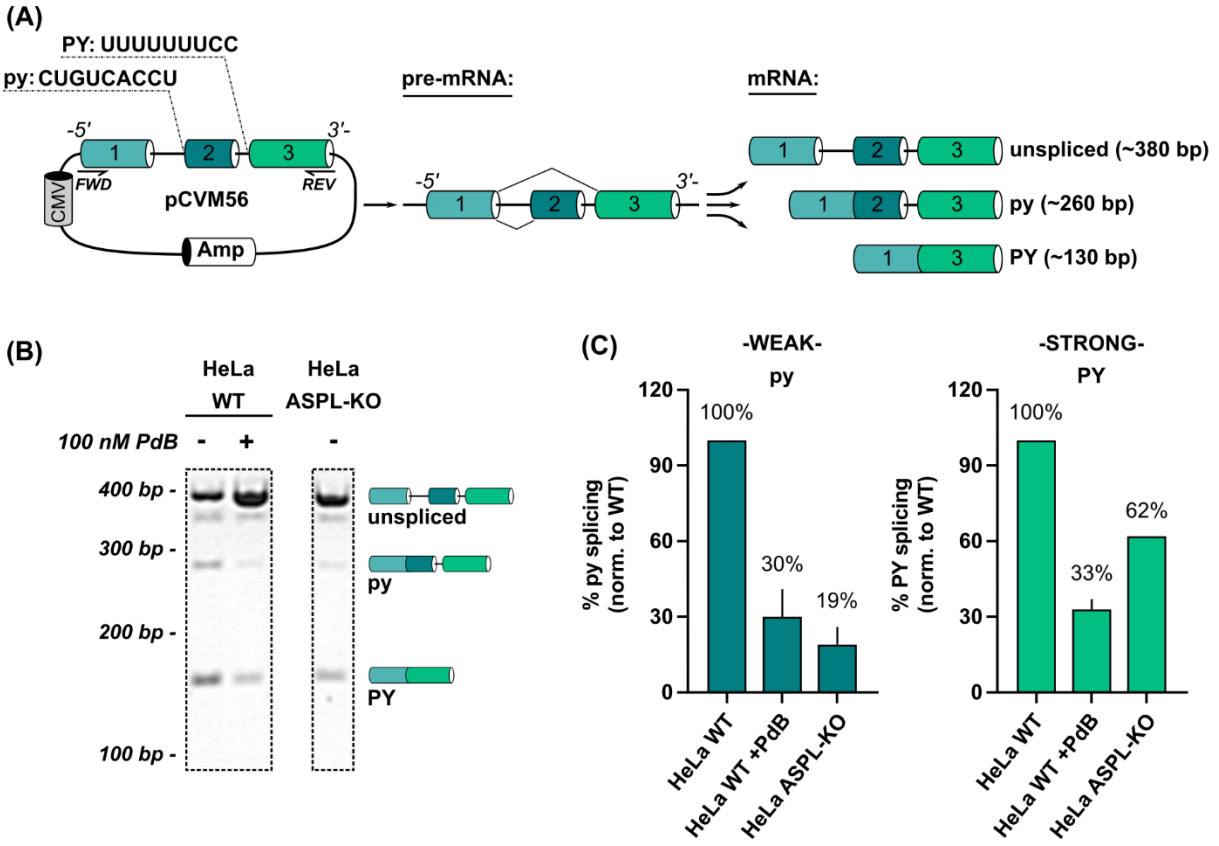


FIG. 20: ASPL knock-out perturbs splicing at weak 3' splice sites in HeLa cells.

(A) Minigene reporter with a strong (PY) and weak (py) 3' splice site was used for transfection of wild-type and ASPL-KO HeLa cells. An additional wild-type sample was treated with 100 nM PdB for 16 hours prior to cell lysis. 48 hours after the transfection total RNA was harvested, transcribed into cDNA and applied as a template in a PCR for the amplification of transcripts resulting from splicing of the minigene. **(B)** PCR products from the three samples were analyzed by gel electrophoresis. Transcripts corresponding to the observed bands are displayed alongside the gel image: unspliced transcript (~380 bp), a transcript resulting from the usage of the weak py (~260 bp) or the strong PY site (~140 bp) **(C)** Quantification of the bands resulting from the py and PY splice site usage in the three samples after normalization to the respective wild-type band. The bars represent mean values of two biological replicates ± SEM. The minigene plasmid was a generous gift from the Carmo-Fonseca Lab (Faculdade de Medicina da Universidade de Lisboa). PdB was kindly provided by Martin Liss (Alumni, Gotthardt lab. MDC Berlin).

So far in this chapter it was pointed out that ASPL influences the expression and splicing of a prominent number of human genes which belong to diverse cellular pathways. Furthermore, in transcripts containing cassette exons the compositions of the 3' splice sites is crucial for the effect of the depletion of ASPL on their splicing.

These observations lead to the conclusion that p97:ASPL might play a role in the initiation of splicing by recruiting critical splicing factors to 3' splice sites. The targets of the heterooligomeric p97:ASPL complex appear to be exons predominantly preceded by weak polypyrimidine tracts.

2.2.3 Effect of p97 point-mutations on the interaction between ASPL and U2AF2

Finally, I investigated the effects of disease-related point mutations in the gene encoding p97 on its association with ASPL and U2AF2. For this analysis, 26 amino acid substitutions exclusively located at the N-terminus in p97 were generated. These residues (**FIG. 21A**) are known to be relevant for the development and progression of the multisystem disorder **IBMPFD** (**I**nclusion **B**ody **M**yopathy with **P**aget´s disease of the bone and **F**rontotemporal **D**ementia)[99]. Mapping these residues in the tertiary structure

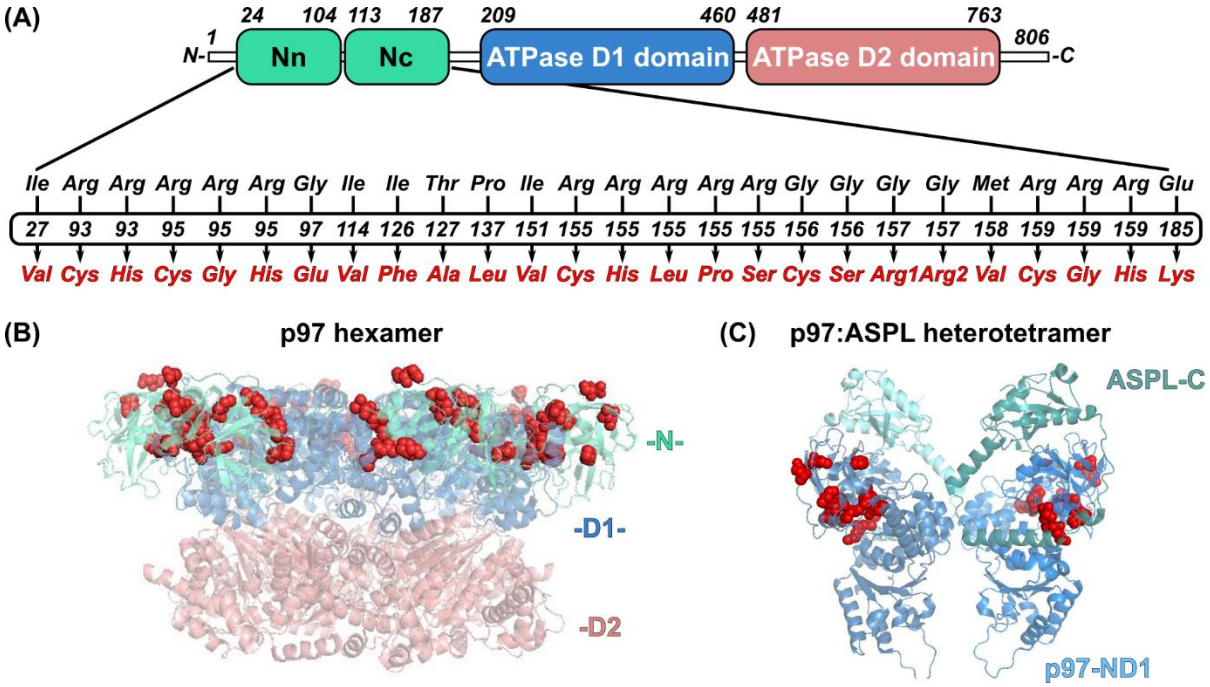


FIG. 21: Disease-relevant point mutations in p97.

(A) Domain structure of a p97 protomer and list of residues at the N-terminus known to be mutated in IBMPFD patients. At position 157 the substituted Arg residue can be expressed from two synonymous codons (Arg1 and Arg2). **(B)** Tertiary structure of the p97 hexamer with mapped point mutations at the N-terminus. **(C)** Same p97 mutations as listed in **(B)** mapped in the tertiary structure of the p97:ASPL heterotetramer here comprised of ASPL-C (313-553 aa) and p97-ND1 (1-480 aa) domains. All point mutations are marked as red spheres in the ribbon representations of the structure of both proteins as obtained by PyMol (PDB accession numbers: 5FTK (p97 hexamer) and 5IFS (p97:ASPL heterotetramer)).

of the p97 hexamer or p97:ASPL heterotetramer illustrated their potential accessibility for interaction partners (**FIG. 21B** and **FIG. 21C**).

In order to quantify the effects of these disease-related mutations on the interaction between ASPL and p97 on the one hand, and p97 and U2AF2 on the other hand, I co-transfected HEK293 cells with plasmids expressing wild-type ASPL-NL with either wild-type PA-mCitrine-p97 or any of the 26 selected mutants. The same procedure was followed for the binary interaction tests between p97 and U2AF2, however, here the p97 variants were applied as a donor (NL-p97) and U2AF2-mCitrine-PA was the acceptor.

Firstly, in-cell BRET was measured in HEK293 cells co-expressing ASPL-NL and wild-type or PA-mCitrine-p97 mutant acceptor variants. From the corrected BRET values, I concluded that each of the mutated residues significantly modulate the binding strength between p97 and ASPL and 25 of the amino acid substitutions exhibited an inhibitory influence on the energy transfer efficiency, suggesting a weakened interaction between the tested proteins (**FIG. 22A**). The only exception was E185K, which delivered an elevated interaction signal.

In a comparable experimental set-up, the impact of the same mutations in p97 on its interaction with the splicing factor, U2AF2, was tested in the presence of wild-type ASPL (**FIG. 22B**). In this case, the amount of the plasmid applied for the expression of ASPL fused to a cmyc epitope tag was similar to the endogenous protein levels (**see FIG. 2A**). Surprisingly, only eleven of the tested amino acid substitutions which affected the association between ASPL and p97 significantly perturb the interaction with U2AF2. Nevertheless, the directions of the observed effects with these mutants which are marked in **FIG. 22B** moderately correlated with the results from the binary tests against ASPL. This statement is based on the Spearman's correlation factor of r = 0.54 obtained with the ΔcBRET values from each data set (**FIG. 22C**). These values are a measure for the magnitude of interaction perturbation caused by each mutated residue and were calculated according to the formula in **FIG. 22C**. In the correlation plot one can clearly see that the mutations with a significant effect on the interaction between p97 and U2AF2 form distinct clusters and also influence the binding of ASPL to p97 in a similar manner. The most prominent example was the increased cBRET ratio for the interaction

between U2AF2 with p97-E185K which is also the only residue that promotes the interaction between p97 and ASPL.

Based on these results one can draw the conclusion that the p97:ASPL complex and its interacting protein U2AF2 are susceptible to disease-causing point mutations in the p97-encoding gene, which may lead to a perturbation of the pre-mRNA processing machinery in cells of IBMPFD patients.

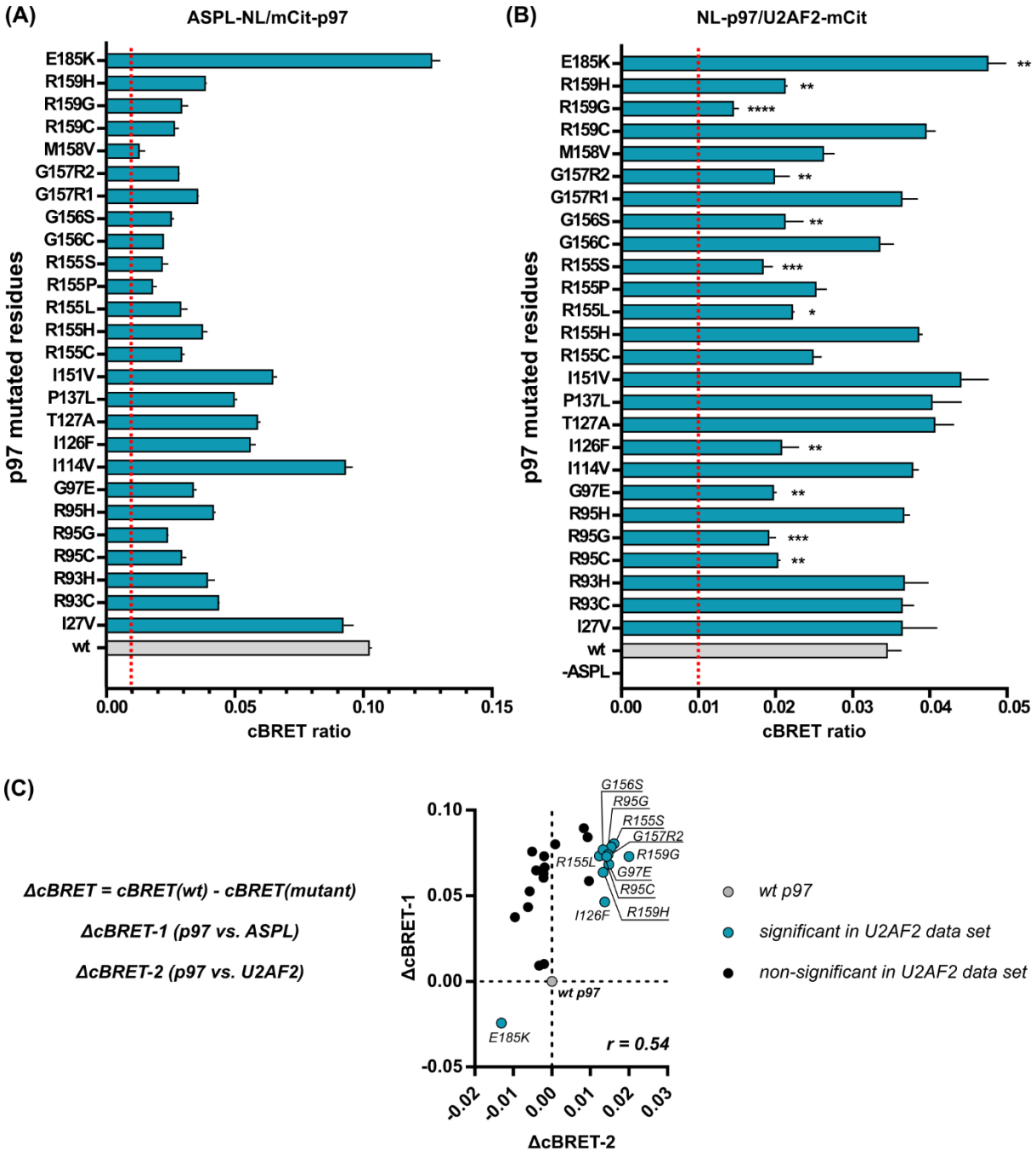


FIG. 22: p97 point mutations perturb the interaction between p97:ASPL and U2AF2.

(A) In-cell BRET measured in HEK293 cells co-transfected with 5 ng of pcDNA3.1-ASPL-NL and 100 ng of pcDNA3.1-mCitrine-p97, either with wild-type or mutant variants. Bars represent mean values of two biological replicates ± SEM. One-way ANOVA and a Dunnett´s multiple comparisons test delivered p<0.0001(****) for all mutants besides p97-I27V (**p<0.01) and p97-I114V (*p<0.05). **(B)** Binding of the same p97 variants to U2AF2 was tested by in-cell BRET in HEK293 cells co-transfected with 5 ng of pcDNA3.1-NL-p97, 100 ng of pcDNA3.1-U2AF2-mCitrine and 20 ng of pDEST26-cmyc-ASPL. A negative control lacking the ASPL expression plasmid was measured as well (-ASPL). Bars represent again a mean ± SEM of two biological replicates. One-way ANOVA and a Dunnett´s multiple comparisons test were applied as well (*p<0.05; **p<0.01, ***p<0.001; ****p<0.0001). **(C)** Delta cBRET (ΔcBRET) was calculated for all PPIs from both data sets by substracting the respective cBRET values from the cBRET obtained with wild-type p97. This calculation was conducted separately in both data sets and the correspodning ΔcBRET values were plotted against each other. A Spearman correlation test delivered a correlation coefficient (r) of 0.54, **p<0.01.

3. Discussion

3.1 LuTHy: a valuable tool for PPI investigations and revealing p97 function

In the initial phase of this study prior to the investigation of the p97:ASPL interactome I conducted three consecutive proof-of-concept experiments. Firstly, a small scale LuTHy screen was performed to test the applicability of the assay for the quantification of PPIs within the p97 interaction network. Additionally, it provided information about the most favorable assay conditions regarding the positions of the tags (N- or C-terminally) and the respective donor/acceptor combinations for measuring the interaction between APSL and p97.

The obtained data revealed that co-expressing p97 as a fusion variant with NanoLuc at the N-terminus and ASPL as a C-terminally tagged fusion protein with mCitrine was the most advantageous set-up for both LuTHy readouts. This was the constellation used in all LuTHy follow-up experiments to study the interaction between ASPL and p97 in this study. As already stated by Trepte *et al.* (2018)[54], the observed effects correlate with the orientation of the interacting domains of both interaction partners, meaning that labeling the proteins of interest close to their binding sites is beneficial for the signal strength. In this case tagging p97 with NanoLuc at either of its termini and ASPL with PA-mCitrine at its C-terminus delivered the highest cBRET ratios (**FIG. 1C** and **FIG. 1D**). When p97 fused to NL at the N-terminus was applied also a high cBRET signal was obtained (cBRET = 0.209 ± 0.003 SEM). This corresponds to the location of the interacting surfaces in the p97-ND1 domain and the extended UBX domain at the C-terminus of ASPL[43]. However, the cBRET ratio measured with p97-NanoLuc and ASPL-mCitrine (0.220 ± 0.005 SEM) was even slightly higher. This observation is likely due to orientation of the molecules within the newly formed oligomeric complex between ASPL and p97[2] (**FIG. I4**), which still allows energy to be efficiently transferred from NanoLuc to mCitrine. Respectively, when applying ASPL as a donor and p97 as an acceptor in the LuTHy set up, the highest cBRET ratio was again measured when ASPL was tagged at the C-terminus and p97 at the N-terminus (0.044 ± 0.001 SEM)) (**FIG. 1C, FIG. S1**). However, this signal was about five-fold weaker compared to the one measured with the reciprocal tag positioning. This is possibly an indication for a

reduced flexibility of the tags which can affect their alignment and impair the resonance energy transfer. It is furthermore noteworthy, that the calculated cLuC values (**FIG. 1C** and **FIG. 1D**) demonstrated the orientation-independent nature of the cell-free measurements, as every tag combination delivered cLuC ratios significantly higher than the threshold. Even the combinations which scored the lowest cBRET values (NL-ASPL with mCit-p97 and NL-ASPL with p97-mCit) performed similarly well compared to their respective counterparts.

The results collected in these first proof-of-concept studies clearly demonstrated the feasibility of the LuTHy assay to detect and quantify the interaction between p97 and members of the UBX protein family. Particularly well one can measure the association between ASPL and p97 in the p97:ASPL complex. Therefore, I decided to further investigate the assay's detection limits and to reduce the amounts of the expression plasmids for ASPL-mCitrine and NanoLuc-97 to levels which result in protein production rates equal to or even lower than the ones from the endogenous loci for both proteins (**FIG. 2D** and **FIG. 2E**). Here, I used a ten times lower amount for the donor plasmid and a fourfold lower amount for the acceptor plasmid compared to the initial tests. The fact that a comparable BRET signal can still be measured demonstrates not only the suitability of BRET for PPI measurements[100], but also clearly shows how the artificiality of a system implementing protein overexpression can be circumvented.

As a final step of the proof-of-concept interaction analysis, I investigated the binding of two ASPL-mCitrine mutant variants – D351A and PP437-438AA – (**FIG. 3A**) to NanoLuc-p97. In two consecutive experiments, (1) the luminescence scan and (2) the donor saturation test, the mutated residues in the ASPL protein variants significantly reduced the affinity towards p97 in BRET measurments (**FIG. 3C** and **FIG. 3D**). The exchange of both Proline residues to Alanine in the *cis*-Pro turn-touch motif had a greater inhibitory impact on the interaction between p97 and ASPL than the single mutation of residue 351 (D351A). With the double ASPL mutant a $BRET_{50}$ value of 8.5 was obtained, it was 6.5 for the single mutant. The $BRET_{50}$ value for the interaction between wild-type ASPL and p97 was estimated to be 3.5. However, Arumughan *et al.* (2016)[43] showed that the single ASPL mutant induces the p97 hexamer disassembly to a lesser extent than the double mutant (**FIG. 3B**). Combining these findings, it can be concluded that the single amino acid exchange in ASPL-D351A, despite its higher

binding capacity towards p97 compared to the double mutant, has a stronger effect on the hexamer disruption process. On the contrary, the PP437-438AA exchange seems rather relevant for the association between ASPL and p97.

In summary, I was able not only to define the best interaction orientation for measuring the association between p97 and ASPL with LuTHy, but also detected their binding at expression levels similar to the ones of the endogenous proteins. Furthermore, I gained valuable information about two ASPL mutant variants and quantified their effects on p97 binding. ASPL-D351A and ASPL-PP437-438AA were later applied in control experiments for the investigation of specific interaction partners of the p97:ASPL complex. They are particularly suitable for that purpose mainly due to their reduced potential in disrupting the p97 hexamer, meaning that proteins which bind to the surface of the p97:ASPL complex should not be able to interact with the ASPL mutants. Keeping this information in mind, I moved to the next step of my investigations and focused on the identification of binding partners of the p97:ASPL heterooligomer.

3.2 p97:ASPL and its binding partners are involved in mRNA processing

The data regarding the ASPL interactome collected through the SILAC analysis (**FIG. 5**) demonstrated a prominent enrichment of proteins involved in splicing and/or regulation of transcription (**FIG. 6**). A list of the 26 identified interaction partners including manually curated information on their cellular activity is presented in **Table S2**. Intriguingly, most of these proteins share distinct structural and functional characteristics which can be interpreted as a hint for interaction specificity towards ASPL. The hypothesis stating that there is a higher chance for a protein to interact with proteins which are similar to other binding partners within its network was recently studied by Kovacs and colleagues[101]. They were able to show how the accuracy of PPI prediction increases notably when it is based on node similarity rather than the "Common Neighbors" principle, which is simply built upon the number of shared interaction partners.

In total 50% of the proteins identified as positive hits in the SILAC data set belong to a group of conserved RNA binding proteins containing a Ser/Arg-rich domain, the SR protein family. In general, it is comprised of twelve proteins defined as canonical SR

proteins[102,103], but depending on the classification criteria there are up to 52 proteins in the human cell, which can be divided in several groups according to their domain composition, specific role in RNA metabolism and ability to translocate to the cytoplasm[104]. However, independently from their unique characteristics, what they have in common is the fact, that they are all essential for either constitutive or alternative pre-mRNA splicing[104,105]. According to the SILAC data, ASPL interacts with five members of the canonical and eight of the broader SR protein family (**Table S2**). It is noteworthy that all canonical SR proteins, detected as positive SILAC hits, are also reported to shuttle between the nucleus and the cytoplasm[104,106,107].

Members of this special subgroup of SR proteins, such as SRSF1, SRSF3, SRSF7 and SRSF10, are particularly unique due to their cytoplasmic functions and have attracted strong scientific interest in the last two decades[102,108–111]. They are not only involved in mRNA export and translation[112,113] or nonsense-mediated decay of transcripts including premature termination codons[114], but can also get incorporated into stress granules upon different stress conditions[110]. Thus, it is not surprising that there is experimental evidence for the role of SR proteins in embryogenesis and early development[108,115]. For instance, Botti *et al.* (2017)[115] demonstrated a clear correlation between differential posttranslational modifications of SR proteins and changes in their shuttling behavior, which modulates the mRNA export of pluripotency factors and hence their expression. Similar observations were made earlier by Sanford and Bruzik (2001)[108] who detected accumulation of SR proteins and RNA polymerase II in the cytosol of *Ascaris lumbricoides* embryos prior to zygotic gene activation and their nuclear translocation upon phosphorylation.

Besides the SR protein family there is no other functionally distinct group among the putative ASPL interaction partners. However, there are few more modulators of splicing or transcription: TAF15[116], PUF60[88], LUC7L2[117] and ARL6IP4[118]. In addition, components of higher molecular weight structures such as SAP18 (histone deacetylase complex)[119], SF3B3 (U2 snRNP)[120], SNRPG (SMN-Sm complex)[121] and SNRNP27 (U4/U6.U5 complex)[120] were found. Finally, there are three proteins with chaperone properties, CIRBP[122], HSPA4[123] and HSP90AA1[124], followed by two less well-known players in the oxidative stress response process, BOLA1 and BOLA2[125,126], and a mitochondrial lipid kinase, AGK[127]. The evident enrichment for splicing factors or related proteins among the

putative ASPL interaction partners was also confirmed by the GO-term analysis, indicating that 23 out of 26 proteins are involved in diverse RNA processing pathways (**FIG. 6**).

In the follow-up validation of the SILAC data with LuTHy assays the interactions between ASPL and ten of the SILAC hits were detected which renders a validation rate of 38%. With very few exceptions all confirmed ASPL binding partners were positive in both LuTHy readouts even though the calculated cBRET and cLuC values demonstrate the prevalence of low affinity interactions within this network (**FIG. 7**). Among the validated targets the majority is still comprised of proteins related to RNA processing including PUF60, RBM39, U2AF2, SF3B3, SNRNP27 and HSP90AA1. However, the remaining four hits, BOLA1, BOLA2, HSPA4 and SAP18, deviate functionally from the rest.

After combining the results from the three sequential LuTHy screens (**FIG. 7B-FIG. 7D**) I was able to distinguish between proteins which interact exclusively with ASPL, but not with p97, and those which bind to p97 in an ASPL-dependent fashion. The latter group is what I defined as specific p97:ASPL complex interaction partners and these were the major focus of my work. Following the described selection criteria, I restricted the number of complex binders to four: U2AF2, PUF60, RBM39 and HSP90AA1 (**FIG. 8A**). There are studies reporting on the interaction between U2AF2 and PUF60[88] along with other recent work providing NMR data in support of the association between U2AF2 and RBM39[14]. Taken together, these findings reveal the existence of a definite PPI network which, I now assume, also includes p97:ASPL as a structural component. Intriguing is the presence of HSP90AA1 which is the stress inducible isoform of the chaperone known as HSP90a[124,128]. Despite the high degree of similarity between HSP90a and the constitutively expressed isoform HSP90β, there is a body of literature suggesting that the functions they exert in the cell are distinguishable[59]. The influence HSP90a has on gene expression regulation, cellular differentiation and development[128-130] is what makes me speculate about its function as a molecular switch within the p97:ASPL PPI network. Particularly interesting are the observations made by Whisenant *et al.* (2015)[131] who detected p97 and HSP90a via SILAC as U2AF2 interaction partners upon CD4 T-cell activation. This is currently the only publication mentioning these three proteins in the same context based on interaction studies.

Furthermore, there are works demonstrating the involvement of HSP90 in splicing of cancer related gene transcripts such as NUMB in hepatocellular carcinoma[132] or AR-V7, a constitutively active truncated version of androgen receptor, in prostate cancer cells[133]. This and the fact that HSP90a accumulates in the nucleus upon stress induction[134] additionally supports the hypothesis that proteins interacting with p97:ASPL might be under the regulation of HSP90a.

The other three binding partners of the p97:ASPL complex, PUF60, RBM39 and U2AF2, seem to share not only structural features, but also operate in a similar fashion when it comes to the early stages of spliceosome assembly and recruitment of the U2 snRNP particle to the 3' splice site of an intronic sequence[58]. On the one hand, there is the interplay between PUF60 and U2AF2 mediated by a specific UHM-ULM domain recognition when it comes to binding to the polypyrimidine (Py) tract at the 3' splice site of pre-mRNAs, which according to Corsini *et al.* (2009)[135] and Hastings *et al.* (2007)[88] takes place non-competitively and simultaneously. On the other hand, there are publications describing the association between RBM39 and U2AF2 which is also enabled by the UHM and ULM domains of both binding partners, respectively[60,136]. For each of the mentioned splicing factors there are indications for their role not only in constitutive but also in alternative splicing. For instance, within the range of pre-mRNAs containing alternatively spliced and RBM39-dependent gene exons, there is an enrichment of transcripts related to G2/M transition, cellular response to DNA damage or endocytosis and these splicing events share ~20% similarity with the ones regulated by U2AF2[137]. Furthermore, the combinatory effect of U2AF2 and PUF60 on alternative splicing strongly depends on their protein levels and can vary between different cell types[88].

Particularly interesting is the role of U2AF2 as an essential splicing factor responsible for the recognition of the Py-tract at the 3' splice site of ~88% of all introns in the human transcriptome[138]. Besides the fact that its proper functioning is of paramount importance for vertebrate development[139], there are a few point mutations within the U2AF2 gene described as relevant for the progression of myeloid malignancies[140] . Since its initial characterization by Ruskin *et al.* (1988)[141] as a prerequisite for binding of the U2 snRNP complex to the pre-mRNA branch point during the process of spliceosome assembly, U2AF2 has been reported to bind to intronic regions beyond

the canonical 3' splice sites. These additional U2AF2 binding events were successfully detected by Shao and colleagues (2014)[138] via CLIP-Seq and their relevance as a mechanism for alternative splicing modulation was confirmed. U2AF2 in its intronic bound state interferes with the recognition of the downstream 3' splice site and depending on its exact position, either up- or downstream of an alternative exon, it may cause exon skipping or inclusion in the matured mRNA. Such position-dependent and polar effects resulting from intron binding seem to be a general mechanism and are relevant for a broader range of splicing regulators[142]. Additionally, some splicing factors, including U2AF2, can recognize and bind to exonic regions which leads to suppression of splicing and exon skipping[142,143]. Noteworthy is also the systematic study by Sutandy and colleagues (2018)[144] who through *in-vitro* iCLIP and machine learning illustrated how U2AF2 binding is highly regulated by other *trans*-acting RNA-binding proteins. However, the molecular mechanisms required for wiring these binding events with the actual outcomes of pre-mRNA maturation remain largely unknown. Taking this under consideration I decided to concentrate my analysis and gather additional information in order to confirm U2AF2 as a new node within the extended p97 PPI network.

3.3 The interaction between U2AF2 and the p97:ASPL complex is specific

The exhaustive analysis of the interaction between U2AF2 and the p97:ASPL complex started with investigations demonstrating the cellular co-localization of the splicing factor and ASPL within the nucleus of HeLa cells (**FIG. 9A**). At this point is essential to mention that even though I did not perform separate p97 localization staining, the presence of p97 in the nucleus of diverse mammalian cells[145] and the role of the p97 N- and C-termini in regulating nucleocytoplasmic shuttling[146] are known. Moreover, interactions between p97 and nuclear proteins involved in DNA damage repair and replication as well as gene expression are well described[145,147,148].

After confirming the presence of ASPL molecules in the nucleus of human cells, I successfully confirmed the interaction between endogenously expressed ASPL and U2AF2 by co-immunoprecipitation experiments (**FIG. 9B**). Furthermore, I performed once again LuTHy binary tests to confirm the conditions enabling the most efficient

energy transfer between ASPL and U2AF2 (**FIG. 9C** and **FIG. 9D**). Additionally, the specificity of this interaction was validated by donor saturation experiments with the two best orientations for efficient BRET measurements: U2AF2-NL/ASPL-mCit-PA and ASPL-NL/U2AF2-mCit-PA (**FIG. 9E**). In both cases I obtained typical saturation curves corresponding to a non-linear regression fit for a one-site specific binding with high confidence. The orientation delivering the highest cBRET ratio – U2AF2-NL and ASPL-mCitrine-PA – was utilized in another LuTHy experiment aiming to demonstrate the direct binding between ASPL and the splicing factor. Here I applied the two ASPL binding mutants described in section **2.1.3** and speculated that their reduced affinity towards p97 will make these molecules more accessible for U2AF2 binding due to lowered steric hindrance. On the contrary, I observed an inhibitory influence on the signal strength corresponding to the amino acid exchanges in the ASPL mutant variants (**FIG. 9F**) meaning that binding to p97 is apparently necessary for their association. Considering the fact that the mutated ASPL versions demonstrate not only perturbed binding features, but also lose a significant proportion of their hexamer remodeling capacity, it can be claimed that the distinct electrostatic composition of the surface of the heterotetrameric complex (A. Arumughan, doctoral thesis, unpublished data)[63] is essential for the recognition by U2AF2 and other interaction partners. These results coincided with the observations showing both the ASPL dependence (**FIG. 10A** and **FIG. 10B**) and the specificity of the interaction between the p97:ASPL complex and U2AF2 (**FIG. 10C** and **FIG. 10D**). The fact that expression of additional wild-type ASPL molecules is required for the establishment of a measurable BRET signal between the splicing factor and p97 supports my hypothesis that ASPL plays a role as a bridging factor in this interaction. However, here I need to clarify that this effect is probably less due to direct scaffolding, but strongly depends on the remodeling features displayed only by wild-type ASPL molecules. Indeed, no increase in BRET can be observed with the mutant ASPL variants either in the titration or in the donor saturation tests. Similarly, if I test for the interaction at endogenous protein level via proximity ligation assay (PLA) in cells in which the expression of ASPL is abolished, I am not able to distinguish it from the negative controls (**FIG. 12B** and **FIG. 12C**). On the contrary, conducting the PLA staining in wild-type HEK293 cells delivers a significant increase in the number of fluorescent puncta indicting an endogenous protein-protein interaction (**FIG. 12B** and **FIG. 12D**). This experiment provided additional evidence for

the existence of an ASPL-dependent specific interaction between p97 and the splicing factor U2AF2.

Declaring U2AF2 as a *bona fide* interaction partner of the p97:ASPL complex included an additional analysis related to the structural features of the splicing factor. In a systematic domain mapping investigation I was able to pinpoint two regions within the U2AF2 molecule, the N-terminal RS-ULM domains and the C-terminal UHM domain, as critical for the association with the p97:ASPL complex (**FIG. 13A-D**). Hence, I suggested a bipartite mode of binding for the splicing factor when it gets incorporated into an oligomeric structure with p97:ASPL (**FIG. 13E**). In order to confine the exact binding interface on the surface of the heterotetramer, additional domain mapping experiments are required with fragments of both p97 and ASPL.

Collectively these observations not only give me confidence for claiming U2AF2 to be recognized as a substrate of the p97:ASPL complex, but also support my initial hypothesis that the p97:ASPL heterotetramer exerts a discrete molecular function, which is different from the one of the p97 hexamer.

3.4 ASPL: more than a repressor of p97 activity

With the experimental evidences collected from the RNA-Seq analysis (**FIG. 14**) I was able to demonstrate the impact of ASPL ablation on the transcriptome of HEK293 cells. Firstly, significant changes in the expression levels for a large number of genes were measured (**FIG. 15A**). Secondly, quantifying the frequency of different splicing events revealed drastic changes upon deletion of ASPL for many transcripts (**FIG. 16A**). In both cases the performed GO-term enrichment analysis revealed groups of transcripts with distinct and very heterogeneous functions (**FIG. 15B**, **15C** and **FIG. 16B**), suggesting that the p97:ASPL complex influences the splicing of a large number of transcripts that encode proteins with many different cellular functions. What at the first glance seems to be surprising with GO-terms related to processes such as ion channels and neurotransmitter activity, tissue development or transcription regulation and RNA metabolism, can be interpreted as a hint for the global influence of the p97:ASPL complex on the human transcriptome. The observed diversity among the identified transcripts is consistent with the polar and position-dependent effects on splicing

regulation resulting from U2AF2 binding to intronic and exonic sequences along the pre-mRNA[149]. This observation additionally supports the suggestion that the p97:ASPL complex in the nucleus is rather a structural scaffold than being actively involved in the process of splicing. Binding to U2AF2 might provide additional binding surfaces for the association with *trans*-acting factors within the network of PPIs defining the context-dependent control of alternative splicing events[142]. In line with this assumption is the presence of a ubiquitin-like domain (UBL) within the ASPL N-terminus which was described as a potential PPI site by Tettamanzi *et al.* (2006)[150]. According to their NMR analysis the UBL domain in ASPL possesses an elevated negative electrostatic potential and an increased backbone mobility which are hallmarks of binding surfaces. Additionally, it lacks the structural features enabling the recognition of polyubiquitinated substrates. If this domain plays a role in the association with U2AF2 or any of the other three p97:ASPL interaction partners is a question that remains to be investigated.

The analysis of the affected transcripts with cassette exons (CEs) which exhibit significant changes in their splicing rates as a result of ASPL depletion provided two very important findings. On the one hand, one could see that the number of CEs which get skipped upon ASPL-KO exceeds the number of predominantly included exons (**FIG. 19A** and **FIG. 19B**). On the other hand, the strength of the 3' splice sites preceding the skipped exons in ASPL-KO cells is lower than the average 3' splice site strength of all cassette exons detected in HEK293 cells (**FIG. 19C**). This observation was confirmed by an experiment with a U2AF2 pre-mRNA substrate performed in HeLa cells. In this case I was able to demonstrate that the weaker 3'splice site (py) in the used minigene was more susceptible to the deletion of ASPL than its stronger counterpart (PY) and its usage was significantly reduced compared to wild-type HeLa cells (**FIG. 20**).

Based on these results I assume that the presence of functional p97:ASPL heterooligomers is advantageous for maintaining the rates of inclusion of those CEs which are under the regulation of weak 3' splice sites. Moreover, I believe that the interaction between p97:ASPL and U2AF2 is an essential driving force for this process. Additional facts which might help this hypothesis to withstand skepticism are some of the inherent characteristics of the p97:ASPL complex itself. First of all, the p97:ASPL complex lacks ubiquitin binding affinity, which was demonstrated by Alexandru *et al.*

(2008)[151]. They performed immunoprecipitation experiments of Flag-tagged ASPL from HEK293 cell extracts and despite sufficient p97 binding no co-precipitation of ubiquitinated substrates was detected. This is a clear indication that the function of the p97:ASPL hetertooligomer in cells is most likely distinct from the canonical chaperone and segregase properties of other well described p97 oligomeric modules[29]. Another evidence which speaks for this theory is the reduced ATPase activity of the purified heterotetramer measured *in-vitro* by Arumughan *et al.* (2016)[43]. Such an inhibition of the ATP turnover was also observed by Meyer and colleagues (1998)[152] when they investigated the binding of p47, another prominent member of the UBX protein family, to p97 hexamers. Despite its 5-fold lowered enzymatic activity, the p97:p47 complex participates in membrane fusion events during the post-mitotic reassembly of the Golgi apparatus[153]. In this process the p97:p47 complex functions rather as a scaffold which recruits further proteins, such as the ubiquitinated Syn5 and the DUB VCIP135, in order to induce the formation of Golgi cisternae[35]. This once again illustrates the multifaceted nature of p97 and supports the hypothesis that the p97:ASPL complex has a unique, specific biological function in cells.

3.5 Disease-causing mutations in p97 affect its binding to ASPL and U2AF2

Speaking about p97 ATPase activity and its functional relevance, I should also mention how disease-causing point mutations in the p97 gene affect its enzymatic properties. The majority of known p97 mutations that cause IBMPFD[154] are located at the N-terminus of the protein (**FIG. 21**) with a few being present in the D1 or D2 domains[99,155]. For some of these single amino acid exchanges it is known that they either elevate the p97 ATP hydrolyzation rate or do not affect the ATPase activity of the enzyme[156,157]. An increase in ATP turnover can be easily connected with energy deficits emerging in high energy-consuming tissues explaining the neurodegenerative defects in IBMPFD disease models[156]. However, in cases, in which p97 ATPase activity remains intact, as shown for p97-R155C by Rycenga *et al.* (2019)[158], the dysfunction of p97 might be mainly caused by perturbed protein-protein interactions. In line with these thoughts applying ATPase inhibitors can be useful only when p97 ATPase activity is increased but otherwise could exacerbate the disease phenotype[158]. Hence, it is crucial to assess the ATP hydrolysis

rate of p97 mutants in a systematic manner. This, however, has so far only been done sporadically for some of the most frequently mutated residues besides R155C, e.g. R155H, R93C or R191Q and A232E[47,156]. A valuable resource of novel targets for alternative therapy development could be gained by exploring the effects of disease-relevant mutations on the p97 PPI network. Perturbations in p97 cofactor binding and their potential contribution to IBMPFD pathology has already been reported[46,159]. Furthermore, Her and colleagues (2016)[160] demonstrated how allosteric ATPase inhibition with small molecules can significantly alter the binding affinity of certain cofactors to p97.

Collectively all these findings motivated me to investigate the impact of all so far described p97 mutants located at its N-terminus on interactions with p97 binding partners (**FIG. 21**). In two consecutive screens I estimated the differences in the calculated BRET ratios for the interactions between ASPL and p97 as well as between U2AF2 and p97 in the presence of additional wild-type ASPL. Significant fluctuations in the interaction profile between ASPL and p97 caused by single amino acid exchanges were detected (**FIG. 22A**). Certainly, these point mutations could also affect the formation of p97:ASPL heterotetramers, however, these effects have not yet been studied. Apart from this, the interaction between p97 and U2AF2 also appeared to be influenced by some of the modified residues in the N-terminus of p97 (**FIG. 22B**). Strikingly, only a fraction of the mutations which reduce p97 binding to ASPL have also an effect on the interaction between p97 and the splicing factor U2AF2. Nevertheless, the effect of the tested residues on the interaction signal correlates moderately between both data sets as shown in **FIG. 22C**.

Hitherto no systematic studies have been reported that have investigated the structural properties of all p97 mutants and the impact of single amino acid exchanges on p97 hexamer formation. Wang and colleagues (2016)[161], however, demonstrated the SUMOylation-dependent nature of p97 hexamer assembly and that this process is influenced by a few point mutations located in the p97 N-terminus, such as R95G, G97E, R155C and R159H. They showed that the tested mutated residues cause reduced SUMOylation of p97 protomers which leads to enhanced p97 monomerization. Hence, one possible explanation for the unaltered interaction between some p97 mutants and U2AF2 might be due to the enrichment of p97 molecules in their monomeric state. In

this case p97 residues might become accessible for U2AF2 binding, which are normally only accessible in the heterotetrameric complex with ASPL. Therefore, cBRET values calculated with these mutants would be similar to the ones obtained between wild-type p97 and U2AF2 despite reduced ASPL binding. Among the p97 mutants tested by Wang *et al.* (2016)[161] and the ones included in the present work, this scenario is plausible only for p97-R155C. This residue modification does not lead to a significant reduction in the cBRET signal measured with U2AF2. However, the other three mutants analyzed by Wang and colleagues – p97-G97E, p97-R95G and p97-R155H – demonstrate decreased binding to U2AF2 in my experiments, even though they enhance the monomerization of p97 protomers. These findings lead me to the conclusion that not the conformation of the p97 protein, but rather the formation of the heterotetrameric complex with ASPL is responsible for the influence of the mutations on the interaction with the splicing factor. Such an assumption would mean that only when the perturbed interaction between p97 mutants and ASPL is additionally accompanied by a drastic inhibition in hexamer disassembly, the effect on the association with U2AF2 can be observed. How the mutations mechanistically influence these processes is a question that awaits to be experimentally explored.

In summary, I was able to measure the effect of disease-causing mutations in p97 on its association with direct interaction partners, such as ASPL. In addition, I also detected perturbations in the interaction between p97:ASPL and the splicing factor U2AF2 – a novel cofactor of this protein complex that so far has not been described as a functional node of the extended p97 interactome. Therefore, these observations are of dual importance: they demonstrate the propagation of such mutation-specific effects within the p97 interaction network and provide evidence for the general feasibility of quantitative PPI detection methods, such as LuTHy, to generate valuable information for the development of alternative therapeutic strategies.

4. Contributions

FIG. 1: Martina Zenkner and Sabrina Golusik executed the PPI screen. Data analysis was performed by Philipp Trepte and included in Trepte *et al.* (2018)[54].

***

FIG. 3A and **3B:** ASPL structure and BN-PAGE originate from Arumughan *et al.* (2016)[43].

***

FIG. 5A and **FIG. 5B:** Immunoprecipitations for SILAC were performed by Alexandra Redel in the E. Wanker lab, MDC. Mass spectrometry analysis was executed by Fabian Host in the Selbach lab, MDC.

***

FIG. 9B: Co-IP experiments were conducted by Oliver Polzer under my supervision during his traineeship in the E. Wanker lab, MDC.

***

FIG. 11A: Generation of HEK293 and HeLa ASPL deficient cell lines was conducted by Anup Arumughan and Alexandra Redel in the E. Wanker lab, MDC.

***

FIG. 13 and **FIG. S4:** The displayed data includes the work performed by Lynn van der Beek for her Bachelor thesis (University of Utrecht) under my supervision.

***

FIG. 14A: Alexandra Redel performed the RNA isolation and sample preparation for RNA-Seq analysis of wild-type and ASPL-KO HEK293 cells in the E. Wanker lab, MDC. Next-generation sequencing was executed in collaboration with the Genomics facility in MDC by Claudia Langnick.

***

FIG. 14B: Alexander Neumann at the Heyd lab (FU Berlin) analyzed the raw sequencing data and tested their quality via PCA. He also provided me with two data sets including information about the differential expression and alternative splicing of all significant hits.

***

FIG. 19C: Analysis of 3' splice site strength was executed by Alexander Neumann in the Heyd lab, FU Berlin.

***

FIG. 20: Maria Carmo-Fonseca (Faculdade de Medicina da Universidade de Lisboa) provided me with the plasmid DNA of the pyPY minigene splicing reporter. I received the splicing inhibitor pladienolide B (PdB) from Martin Liss (Alumni Gotthardt lab, MDC).

***

FIG. 22: Aline Schulz generated the p97 mutant plasmids in the E. Wanker lab and assisted me during the execution of the PPI LuTHy screen.

***

FIG. 22 and **FIG. 7:** For the calculation of cBRET and cLuC values I applied a Python script written by Marcel Sperling (Alumni Imaging Facility, MDC).

5. Materials

5.1 Consumables

5.1.1 List of Antibodies

Protein target	Antibody	Species	Provider	Catalog no.	Clone no.	Application	Dilution/ Concentration
ASPL	anti-TUG (monoclonal)	rabbit	Abcam	ab131217	EPR8616	WB	1:3000
ASPL	anti-TUG (monoclonal)	rabbit	Abcam	ab131217	EPR8616	IP	1:50
p97	anti-p97 (monoclonal)	mouse	Progen	65278	58.13.3	WB	1:5000
p97	anti-VCP (monoclonal)	rabbit	Abcam	ab109240	EPR3307(2)	IP	5-10 µg/reaction
p97	anti-VCP (monoclonal)	mouse	Thermo Scientific	MA3-004	5	IF//PLA	1:200//1:100
U2AF2	anti-U2AF65 (polyclonal)	rabbit	Thermo Scientific	PA5-30442	-	WB	1:5000
U2AF2	anti-U2AF2 (monoclonal)	mouse	Sigma	U4758	clone MC3	IP	5-10 µg/reaction
U2AF2	anti-U2AF65 (polyclonal)	rabbit	Thermo Scientific	PA5-30442	-	IF//PLA	1:500
β-Actin	anti-β-Actin	mouse	Sigma	A5441	Clone AC-15	WB	1:2000
α-Tubulin	anti-α-Tubulin (monoclonal)	mouse	Sigma	T6074	Clone B-5-1-2	WB	1:2000
GAPDH	anti-GAPDH (monoclonal)	mouse	Thermo Scientific	MA5-15738	GA1R	WB	1:5000
Histone H3	anti-Histone H3 (polyclonal)	rabbit	Abcam	ab1791	-	WB	1:2500
Myc epitope tag	anti-myc tag (monoclonal)	mouse	Cell Signaling	9B11	NA	WB	1:1000
Myc epitope tag	anti-myc tag (monoclonal)	mouse	Cell Signaling	9B11	NA	IF	1:8000
mCitrine	anti-GFP (polyclonal)	rabbit	Abcam	ab290	-	WB	1:1000
-	sheep IgG	-	Jackson Immunoresearch	013-000-002	-	IP for LuTHy	10 µg/mL
-	anti-sheep IgG	rabbit	Jackson Immunoresearch	313-005-003	-	IP for LuTHy	3.3 µg/mL
-	anti-mouse HRP-linked IgG	-	Sigma	-	-	WB	1:2000
-	anti-rabbit HRP-linked IgG	-	Cell Signaling	7074	-	WB	1:2000

5.1.2 List of Kits

Kit:	Manifacturer:	Catalog no.:	Application:
BCA Protein Assay Kit	Pierce	23225	determination of total protein concentration
Detergent-Free Total Protein Isolation Kit	NORGEN	30300	total protein isolation from human cells at non-denaturing conditions
DNeasy Blood and Tissue Kit	Qiagen	69504	genomic DNA isolaiton from human cells
Duolink In Situ Detection Reagents Orange	MERCK	DUO92007	Proximity Ligation Assay
Duolink In Situ PLA anti-mouse MINUS Probe	MERCK	DUO92004	Proximity Ligation Assay
Duolink In Situ PLA anti-rabbit PLUS Probe	MERCK	DUO92002	Proximity Ligation Assay
Dynabeads mRNA Purification Kit	Invitrogen	61006	mRNA isolation form total RNA via magentic Oligo (dT) beads
Dynabeads Protein G	Invitrogen	10004D	co-IP via magnetic beads coated with protein G
Gibson Assembly Master Mix	NEB	E2611	fragment DNA based cloning via the Gibson Assembly Method
High Capacity cDNA Reverse Transcription Kit	Applied Biosystems	4368814	RNA reverse transcription into cDNA
MCB Spin PCRapace	Stratec	1020220300	purification of PCR products
Nano-Glo Dual-Luciferase Reporter Assay System	Promega	N1610	NanoLuc and Firefly luminescence detection
Pierce Crosslink Immunoprecipitation Kit	Thermo Scientific	26147	co-IP via agarose beads coated with protein A/G
QIAprep Spin Miniprep Kit	Qiagen	27106 or 27193	plasmid DNA isolation
QIAshredder homogenizer	Qiagen	79654	lysate homogenization during RNA isoaltion
QuickExtract DNA Extraction Solution	Lucigen	QE09050	genomic DNA isolaiton from human cells
RNase free DNase Set	Qiagen	79254	DNA digestion for RNA isolation
Rneasy Set	Qiagen	74106	total RNA isolation from human cells

5.1.3 List of Primers & Plasmids

#	Primer:	Sequence:	Application:
-1-	pyPY-FWD	**TGAGGGGAGGTGAATGAGGAG**	PCR
-2-	pyPY-REV	**TCCACTGGAAAGACCGCGAAG**	PCR
-3-	U2AF2-FL-rev-att_new	**GGGGACAACTTTGTACAAGAAAGTT GGGTACCAGAAGTCCCGGCGGTG**	PCR (U2AF2 fragment generation)
-4-	U2AF2-342-rev-att_new	**GGGGACAACTTTGTACAAGAAAG TTGGGTGATTGATGGTGCTCAGCGT**	PCR (U2AF2 fragment generation)
-5-	U2AF2-63-fwd-att_new	**GGGGACAACTTTGTACAAAAAAGTTG GCACCATGAAACCTTTGACCAGAGGCG**	PCR (U2AF2 fragment generation)
-6-	U2AF2-141-fwd-att_new	**GGGGACAACTTTGTACAAAAAAGTTGGC ACCATGAGCCAGATGACCAGACAAGC**	PCR (U2AF2 fragment generation)
-7-	U2AF2-FL-fwd-att	**GGGGACAACTTTGTACAAAAAA GTTGGCATGTCGGACTTCGACGAGT**	PCR (U2AF2 fragment generation)
-8-	RIC8A-FWD2-new	**ATTCACTAGTTCTCTCTGCAGC**	PCR (RNA-Seq validation)
-9-	RIC8A-iso1-REV	**CTTTCAGCTCCTGAAACAG**	PCR (RNA-Seq validation)
	Plasmid:	**Source/Reference:**	**Identifier/Cat. No.:**
-1-	pDONR221 entry	Thermo Fisher	12536017
-2-	pcDNA3.1	ThermoFisher	V79020
-3-	pcDNA3.1 cmyc-NL-GW	Addgene	113446
-4-	pcDNA3.1 GW-NL-cmyc	Addgene	113447
-5-	pcDNA3.1 2xPA-mCit-GW	Addgene	113448
-6-	pcDNA3.1 GW-mCit-2xPA	Addgene	113449
-7-	pcDNA3.1 2xPA-mCit-NL	Addgene	113444
-8-	pcDNA3.1 2xPA-mCit	Addgene	113443
-9-	pcDNA3.1 NL-cmyc	Addgene	113442
-10-	pCMV U6gRNA-Cas9-2A-GFP	SIGMA	HS0000208208 (ASPL exon-4) HS0000208209 (ASPL exon-4)
-11-	pDest26 cmyc-GW	kind gift from Dr. Katrina Meyer, Selbach lab, MDC	-
-12-	pCMV56 pyPY minigene	kind gift from the Carmo-Fonseca lab, University of Lisbon, Portugal	-

5.1.4 List of Chemicals & Materials

Chemicals:	Provider:	Catalog no.:
1 kb marker	Invitrogen	10787018
Albumin Fraktion V	Roth	8076.3
Benzonase	MERCK	1016540001
Bovine Serum Albumin	Sigma	A2153
BP clonase	Invitrogen	11789
BsrGI	NEB	R3575
Coelenterazine-h	PJK	102182
cOmplete protease inhibitor (EDTA-free)	Roche	37378900
DMEM medium	Gibco, ThermoFisher	42965-039
DMEM medium (phenol red free)	Gibco, ThermoFisher	21063-029
DNase I	Biomatik	A4193
FBS	Gibco, ThermoFisher	10500-064
Fluorescence Mounting Medium	Dako	S3023
Goat Serum	Dako	X0907
Hoechst 33342	Invitrogen	H3570
LR clonase	Invitrogen	11791
Magic Mark XP Western Protein Standard	Novex by Life Technologies	LC5603
PBS	Gibco, ThermoFisher	14190-094
PEI (25 kDa)	Polysciences	23966-1
Pen/Strep	Gibco, ThermoFisher	15140-122
Phusion High-Fidelity DNA Polymerase	NEB	M0530
Pladienolide B	Bioaustralis	BIA-P1411
Salmon Sperm DNA	Invitrogen	15632-011
SDS-PAGE Page Ruler Plus prestained	Thermo Scientific	26621
SeeBlue Plus2 Pre-Stained Protein Standard	Novex by Life Technologies	LC5925
Trypsin	Gibco, ThermoFisher	23500
Western Bright Chemiluminescence Substrate	Biozym	541015
Other materials:	**Provider:**	**Catalog no.:**
NuPAGE Bis-Tris (1 mm, 4-12%) SDS-gels	Invitrogen	NP0322BOX
96-well microtiter plates (for LuTHy-BRET)	Greiner	655983
384-well microtiter plates (for LuTHy-IP)	Greiner	784074
384-well microtiter plates (for BCA)	Nunc	242757

5.1.5 List of Buffers and Solutions

Components marked with * vary among the different Buffer compositions

HEPES-lysis buffer				
	denaturing (for WB)	**denaturing (for LuTHy-IP)**	**mild denaturing (for SEC)**	**native (for NativePAGE)**
components:	**final concentration:**	**final concentration:**	**final concentration:**	**final concentration:**
HEPES/NaOH (mM), pH 7.4	50	50	50	50
NaCl (mM)	150	150	150	150
Glycerin (%)	10	10	-	-
NP-40 (%)	1	1	0.1	-
NaF (mM)	20	20	-	-
$MgCl_2$ (mM)	1.5	1.5	1.5	1.5
EDTA (mM)	1	1	1	1
DNase*	-	-	0.025	0.025
Benzonase*	1:10 000	1:10 000	-	-
PMSF (mM)*	1	1	-	-
Protease inhibitor cocktail*	1:25	1:25	1:25	1:25
Na-deoxycholate (%)*	0.5	0.5	-	-
Glycerol-2-phosphate*	-	25	-	-
Orthovanadate (mM)*	-	1	-	-
Pyrophosphate (mM)*	-	2	-	-
DTT (mM)*	-	1	-	-

5.2 Lab equipment

Devices:	Model/Manifacturer:
Microplate reader (for fluorescence and luminescence measurements)	TECAN Infinite M200
-	TECAN Infinite M1000Pro
-	TECAN Spark
Confocal microscope	SP8 by Leica
Fluorescence microscope	Axio Imager Zeiss
Imaging System	LAS-3000 Fujifilm
Nanodrop 8000	Thermo Scientific
DNA electrophoresis chamber	BioRad
Wet blotting unit	Mini Trans-Blot® Cell, BioRad
Semi-dry blotting unit	Trans-Blot® SD Semi-Dry Transfer Cell, BioRad

6. Methods

6.1 Molecular biology

6.1.1 Plasmid construction

- **Gateway Cloning® and *E. coli* transformation**

All plasmids mentioned in this work which were applied for ectopic gene expression were generated via the Gateway Cloning® technology (commercialized by Invitrogen) unless explicitly stated differently. Entry and destination plasmids DNA were applied in a 1:2 ratio in each LR reaction.

Newly created plasmids were transformed into chemically competent mach1 cells (Life Technologies GmbH) according to the provider´s manual. The amount of the applied cell suspension varied depending on the transformation scale: 5 µL for transformation in single 1.5 mL tubes or 20 µL for 96-well microtiter plates.

- **Site-directed mutagenesis**

Mutations in the p97 sequence were introduced via amplification of the wild-type sequence with primers incorporating the single base changes of interest. Forward and reverse primers were phosphorylated at their 5'-end in order to create a linearized PCR product which was ligated into a circular plasmid.

For the amplification, 100 pg template DNA were applied in a 40 µL PCR reaction, using the Phusion Hot Start DNA Polymerase (NEB) with high-fidelity (HF) buffer and 3% DMSO supplement. The elongation time was calculated with an amplification fidelity of 30 sec per 1 kB DNA. All PCR-products were purified with a PCR clean-up kit (Invitek Molecular) and subsequently 20 fmol were ligated using the T4 DNA ligase according to instructions for sticky-end ligation provided by the manufacturer (Thermo Fisher, Cat No: #EL0016). Mutated p97 sequences were validated by Sanger sequencing (executed by LGC Genomics).

6.1.2 Isolation of plasmid DNA and sequence validation

Plasmid isolation from *E. coli* cells was performed with a QIAprep Spin Mini- or Midiprep Kit depending on the needed DNA yield and following the manufacturer's manual. Sequence validation was performed by test digesting the plasmids isolated from single colonies with appropriately selected restriction enzymes (NEB) and assessing the expected band patterns on a 0.8% TAE agarose gel stained with ethidium bromide (EtBr). Correctly digested samples were further evaluated by Sanger sequencing (executed by LGC Genomics).

6.1.3 Isolation of genomic DNA or total RNA from mammalian cells

High-quality genomic DNA or total RNA samples from cultured mammalian cells were prepared with the DNeasy® Blood & Tissue or respectively the RNeasy® Mini Kit by QIAGEN according to the provider´s instructions.

Prior to total RNA isolation adherent cells were harvested by trypsinization and the cell pellet was resuspended in the appropriate volume of RLT buffer according to the cell number. QIAshredder spin columns were used for lysate homogenization and an additional on-column DNase digestion was performed with the recommended RNase-Free DNase Set. All RNA-isolation steps were performed at room temperature and all used materials and instruments were initially treated with RNaseZap. After a spectrophotometric measurement of the samples' concentration, these were stored at -80°C. RNA integrity was assessed by a 0.8% TAE agarose gel stained with ethidium bromide (EtBr).

6.1.4 Real Time-PCR

Total RNA isolated from cultured mammalian cells was reversely transcribed into cDNA with a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, #4368814) without additional RNase inhibitor and used as a template for real-time PCR applications. Final cDNA concentration was estimated according to the applied amount of total RNA ranging between 0.5 and 2 µg total RNA.

6.1.5 RNA-Seq data generation and analysis

Total RNA, isolated as described above, was subjected to polyA detection for library generation prior sequencing. Paired-end reads were generated in an Illumina NextSeq500 Sequencer. Transcript expression was quantified with the tool "Salmon"[162] and the package "DESeq2"[163] was used for the differential analysis. The method applied for quantitative profiling of alternative splicing (AS) events in our samples was "Whippet"[164].

6.1.6 CRISPR/Cas9 mediated genome editing

- **gRNA design and plasmid generation**

Selection and evaluation of gRNAs for all genes of interest were performed using Benchling (https://www.benchling.com/). A Cas9 expressing plasmid containing a puromycin resistance (pSpCas9(BB)-2A-Puro (PX459) V2.0, a gift from Feng Zhang, Addgene: #62988) was used to insert a preformed gRNA oligo duplex via Golden Gate Assembly according to the protocol published[165].

- **Transfection for knock-out generation**

The generation of an ASPL knock-out cell line was achieved by CRISPR/Cas9 mediated targeting of the endogenous ASPL allele in HEK293 or HeLa cells. The following sgRNA sequences were ordered at SIGMA Aldrich: HS0000208208 (exon 4), HS0000208209 (exon 4), and HS0000208210 (exon 5).

- **Minigene splicing reporter**

An mRNA-splicing substrate widely used as a reporter for U2AF2 splicing activity was a kind gift from the Carmo-Fonseca lab (University of Lisbon, Portugal) (Pacheco et al. 2006). Originally this plasmid was designed and created in the Juan Valcarcel lab (CRG Barcelona). In the present work 1.8 µg of the minigene were applied to wild-type or ASPL-KO HeLa cells for transfection in 6-well plates. An additional condition was tested

in which transfected wild-type HeLa cells were treated with the inhibitor pladienolide B (PdB). It impairs the recruitment of the U2 spliceosome particle to pre-mRNA at the 3' splice site through its binding to the core SF3B complex[166]. The inhibitor was applied to the cells 16 hours prior RNA isolation in a final concentration of 100 nM PdB per well.

The cells were lysed 48 hours post transfection and total RNA isolation and cDNA synthesis were performed as described in **6.1.3** and **6.1.4** . Splicing products resulting from the minigene were amplified in 40 µL PCR reaction using 50 ng of cDNA as template and the following primers: TGAGGGGAGGTGAATGAGGAG (pyPY-FWD) and TCCACTGGAAAGACCGCGAAG (pyPY-REV). The obtained PCR products were analyzed on a 2% agarose gel.

6.1.7 Transfection of mammalian cells

- **Reverse transfection of HEK293 cells**

In all interactomics studies polyethylenimine (PEI) mediated reverse transfection of HEK293 cells was performed with a DNA to PEI ratio of 1:2.75 ng per well in a 96-well microtiter plate (Greiner bio-one: 655983, white). Each transfection reaction was tested in technical triplicates (three wells per a 96-well plate) and consisted of a maximum of 200 ng pcDNA3.1 expression plasmids in 50 µL OptiMEM per well

6.1.8 LuTHy assay

The double-readout bioluminescence-based two-hybrid technology, named LuTHy, developed by Trepte *et al.* (2018) was used to perform all binary interaction tests mentioned in this thesis. These were performed following the experimental procedure described by Trepte *et al.* (2018) with minor corrections, including the number of cells seeded per well in 96-well multitier plates ($3.6x10^4$) and the incubation duration (72 hours). Each interaction of interest was investigated in all eight possible directions resulting from the positioning of the tags (NanoLuc and PA-mCitrine) in the fusion bait and prey constructs. The results from the ASPL interactome validation screen combine the data of two independent biological replicates. The threshold values for cBRET ratio

(0.01) and cLuC (3%) were applied according to the results from the reference set tests obtained by Trepte *et al.* (2018).

The donor-saturation experiments performed in this work were executed by co-transfecting HEK293 cells with a fixed amount of NanoLuc donor plasmid (1 ng) and increasing quantities of PA-mCitrine acceptor plasmid (0, 5, 10, 25, 50, 75, 100, 150 and 200 ng). Acceptor-to-donor ratios were calculated by dividing the fluorescent signals by the luminescent values obtained in these cells and normalizing to the acceptor-to-donor ratio of the PA-mCitrine-NanoLuc tandem construct. Plotting the acceptor-to-donor values against the measured BRET ratios results in donor-saturation curves for the tested interaction and the acquired half-maximum BRET ($BRET_{50}$) serves as an indicator for the relative binding strength between two proteins.

Additionally, HEK293 cells co-transfected with 5 ng of NanoLuc-tagged and 100 ng of PA-mCitrine-fused constructs were subjected to serial luminescent measurements at different wavelengths between 350 and 700 nm in 2 nm intervals. The obtained values for each tested construct were normalized to the luminescent signal at 460 nm which corresponds to the maximum NanoLuc emission. A second peak observed at 530 nm correlates with the mCitrine emission and illustrates the resonance energy transfer efficiency between the proteins of interest. For these experiments a NanoGlo substrate (Promega) in final 1:500 dilution is added to the cells in each well of the 96-well plate prior incubation of 15 min at 37°C. All LuTHy, in-cell BRET, luminescence and fluorescence measurements were conducted in a TECAN plate reader (M1000Pro or M200 for LuC).

6.1.9 Domain mapping via in-cell BRET

The domain mapping experiments described below were executed with 1 ng of donor and 100 ng of acceptor plasmids for full-length proteins and selected fragments. Here in-cell BRET measurements were performed as biological triplicates and statistically analyzed via one-way ANOVA.

6.2 Cell biology

6.2.1 Cell culture

- **HEK293**

Human embryonic kidney cells (HEK293) were cultured in a high-glucose (4.5 g/L) sterile-filtered DMEM medium (Gibco®, ThermoFisher) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin/streptavidin, except for all LuTHy experiments, which required the use of a phenol-red free version of the same DMEM medium.

For the SILAC assay a high-glucose, sterile-filtered DMEM medium (Biosera Ltd.) supplemented with 10% dialyzed FBS and 1% Pen/Strep was applied on the cells either as a light or heavy medium (containing $^{13}C_6{}^{15}N_4$-L-Arg and $^{13}C_6{}^{15}N_2$-L-Lys).

Passaging of the HEK293 cells was performed regularly (every 3-4 days) or when 80% confluence was reached using 0.05% Trypsin-EDTA (ThermoFisher).

- **HeLa**

HeLa cells were cultured in an identical way as HEK293 cells. Cultivation and in-cell experiments were performed at 37°C and 5% CO_2 independent from the used cell line (wild-type or ASPL knock-out).

6.2.2 Cell lysis

Cell lysis was performed with different lysis buffer compositions depending on the desired downstream applications. See section **"Buffers and Solutions"** for exact information on the specific buffer composition. In general, samples were always incubated for 30 min at 4°C under gentle agitation independent from the used lysis buffer.

- **Denaturing and mild-denaturing lysis**

Cells were lysed under denaturing conditions when the protein composition of the lysates was about to be subjected to analysis by western blot or IP within the LuTHy assay. Mild-denaturation was performed by using 0.1 % of the non-ionic detergent Nonidet P-40 (NP-40).

- **Native lysis**

In order to analyze the protein composition of a sample in its native state, cell lysates were prepared by disrupting the cell membrane via mechanical forces without the use of detergents or reducing agents. Here all samples subjected to native lysis were prepared following the procedure from the Detergent-Free Total Protein Isolation Kit by Norgen™. Total protein concentration was determined by the Pierce™ BCA assay (Thermo Scientific) and samples were instantly prepared for a NativePAGE and applied on a gel avoiding their previous freezing. This is crucial for keeping the proteins in their native state.

6.3 Protein biochemistry

6.3.1 NuPAGE and immunoblotting

Protein samples for western blot analysis were prepared by lysing the cells in HEPES-lysis buffer (see **"Buffers and Solutions"** section) for 30 min at 4°C under constant agitation. Cell lysates were centrifuged twice for 5 min at 8000 xg and the total protein concentration was determined from the supernatant by a BCA assay (Pierce™). Equal amounts of protein were loaded on a NuPAGE™ Novex™ 4-12% Bis-Tris precast polyacrylamide gel (Thermo Fisher) and running conditions were adjusted to 80 min and 120 V. The proteins were blotted on a nitrocellulose membrane (0.45 µM, Amersham™ Protran) with the wet blot system by BioRad for 60 min at 100 V or at max. 20 V for 60 min when a semi-dry blotting system was used.

All primary antibodies for the detection of proteins of interest were diluted according to the provider´s recommendation in 3%-milk PBS-Tween (0.05%), applied on the

membranes and incubated for 4 hours at room temperature or overnight at 4°C under constant shaking. The respective secondary HRP-linked antibodies were diluted in 1:2000 ratio in 3%-milk PBS-Tween (0.05%) for 45 min at room temperature under gentle agitation. Finally, the membranes were subjected to chemiluminescent detection using the substrate provided by WesternBright Quantum (Advansta, 12042-D20) and images were taken with a Fujifil LAS-3000 Imager.

6.3.2 NativePAGE

Lysates prepared under native conditions were applied on NativePAGE™ Novex™ 4-12% Bis-Tris precast polyacrylamide gels (ThermoFisher) and run according to the manufacturer´s protocol. Native gels were blotted on a PVDF membrane, pre-activated in 100% p.a. grade EtOH.

6.3.3 Size-Exclusion Chromatography (SEC)

Lysates from HEK293 cells were analyzed by size-exclusion chromatography (SEC) in an ÄKTA purifier system (GE Healthcare) using a pre-calibrated Superose 6 30/100 column. The proteins run as calibration samples were Thyroglobulin (669 kDa), Ferritin (440 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa) and RNase (13.7 kDa).

A total number of 40 x 10^6 cells were lysed using the SEC running buffer, incubated for 30 min at 4°C on a shaker and finally centrifuged for 15 min at 15 000 xg at 4°C. The protein concentration in the collected supernatant was estimated by a BCA assay (Pierce™) and a maximum of 4 mg total protein were loaded on the column. 400 µL fractions were collected on a 96-well plate with a 0.200 flow rate at 4°C. Each two consecutive fractions were pooled together, acetone precipitated and then applied on a NuPAGE™ 4-12% Bis-Tris gel for immunoblot analysis.

6.3.4 In-vivo crosslinking via DSS (disuccinimidyl suberate)

A total number of 40 x 10^6 HEK293 cells was collected via trypsination, washed three consecutive times with PBS and finally treated with 2.5 mM of DSS for 30 min at 37°C under gentle shaking. The reaction was quenched by adding 50 mM TRIS-HCl and incubating for 15 min at room temperature. Following this, the cells were lysed and applied on the Superose 6 30/100 column as described above.

6.3.5 Immunoprecipitation

HEK293 cells cultivated in either heavy or light SILAC DMEM medium were lysed for 60 min on ice with a HEPES-lysis buffer (50 mM, pH 7.4) containing 1% of NP-40, 150 mM NaCl, 2 mM $MgCl_2$, 1U Benzonase, 1 mM EDTA, and protease inhibitor cocktail (Roche, EDTA-free). Endogenous ASPL was immunoprecipitated following the instructions provided with the Cross-Link Immunoprecipitation Kit (Thermo Scientific #26147) with minor corrections, e.g. doubling the amounts of beads, antibodies and buffers used. Two different anti-ASPL antibodies were used: a mouse monoclonal anti-ASPSCR1 (Abnova, Cat. No: H00079058-M01, clone 3D10-1D11) and a rabbit polyclonal anti-TUG (Cell Signaling, Cat. No: #2049). The negative controls included two respective isogenic IgGs and a "beads only" treatment. For each sample 2.5 mg of total protein were applied the Protein A/G agarose beads previously crosslinked with the mentioned antibodies.

Endogenous U2AF2 was immunoprecipitated from wild-type HEK293 cells after 30 min lysis on ice with the buffer provided with the Detergent-Free-Total-Protein Kit from Norgen (#30300). 5 µg of the mouse monoclonal anti-U2AF2 antibody (SIGMA, Cat. No: U4758, clone MC3) were crosslinked on the surface of the Protein A/G agarose beads and incubated with 450 µg of total protein overnight at 4°C.

6.3.6 SILAC assay

To identify new ASPL interaction partners a **SILAC** experiment (**s**table **i**sotope **l**abeling with **a**mino acids in **c**ell culture) was conducted, in which the endogenous protein was immunoprecipitated, as described above, and the sample subjected to analysis by

mass spectrometry. In order to do so, HEK293 cells were previously separated in two fractions and grown respectively in light and heavy (supplemented with $^{13}C_6{}^{15}N_4$-L-Arg and $^{13}C_6{}^{15}N_2$-L-Lys) sterile filtered DMEM medium. After immunoprecipitation, heavy and light samples were combined as depicted in the figure below (**FIG. M1**), resulting in eight different swaps.

mass spec swaps:

1	IP αASPL (mouse) **HEAVY**/ IP IgG (mouse) **LIGHT**	**+**	**5** IP αASPL (mouse) **LIGHT**/ IP IgG (mouse) **HEAVY**
2	IP αASPL (mouse) **HEAVY**/ IP (beads only) **LIGHT**	**+**	**6** IP αASPL (mouse) **LIGHT**/ IP (beads only) **HEAVY**
3	IP αASPL (rabbit) **HEAVY**/ IP IgG (rabbit) **LIGHT**	**+**	**7** IP αASPL (rabbit) **LIGHT**/ IP IgG (rabbit) **HEAVY**
4	IP αASPL (rabbit) **HEAVY**/ IP (beads only) **LIGHT**	**+**	**8** IP αASPL (rabbit) **LIGHT**/ IP (beads only) **HEAVY**

FIG. M1: Sample combinations for SILAC swaps.

Combination of IP samples from cells cultivated in light or heavy medium for SILAC swaps. For graphical representation of the results the values from all reciprocal swap pairs (#1+5, #2+6, #3+7 and #4+8) were plotted against each other (**see FIG. 5**)

Each swap was analyzed by mass spectrometry and for all detected proteins a specific heavy-to-light (H/L) abundance rate is calculated as a log2FC (fold change) from the obtained mass-to-charge ratios. These rates can be either positive or negative, depending on the sample´s composition. For instance, in swaps #1 to #4 the values should be positive as in this case a specific aASPL antibody was applied to the heavy fraction samples. This should result in higher mass-to-charge ratios for the identified proteins compared to the light sample fraction that is immunoprecipitated with an isotype control IgG and would presumably deliver low protein amounts. The resulting heavy-to-light (H/L) abundance rate should be greater than 1 and deliver a positive log2FC value. In samples #5 to #8, on the contrary, the immunoprecipitation of the heavy and light lysates was conducted in a reciprocal manner compared to samples #1 to #4 considering the use of antibodies. This leads to H/L rates lower than 1 and negative values when expressed as a log2FC. Following this procedure facilitates the reduction of the background noise. Ratios differing from the expected pattern were excluded from further analysis steps.

Once obtained, the heavy-to-light (H/L) abundance rates of the proteins from a reciprocal swap pair can be plotted against each other. In this plot each dot

represents a putative interaction partner of the immunoprecipitated target. The closer the distance between any random dot and the one representing the protein of interest, the higher the interaction propensity. A cut off for the log2FC values was calculated within each swap as the respective mean value plus a single standard deviation in order to stringently select hits for further validation (**see FIG. S2**).

6.4 Imaging

6.4.1 Coating of cover slips for microscopy

For fluorescent imaging cells were seeded on cover slips treated with a coating solution (10 µg/mL fibronectine and --- µg/mL poly-L-lysine) for 4 hours at 37°C. Before applying the cells on the pre-coated cover slips, these were washed with autoclaved H_2O and sterile PBS solution in two consecutive steps. The number of cells seeded per cover slip was adjusted according to the different cell lines: 10 000 cells of HEK293 wild-type, 15 000 cells of HEK293 ASPL knock-out, 50 000 cells of HeLa wild-type or HeLa ASPL knock-out.

6.4.2 Proximity Ligation Assay (PLA)

Proximity Ligation Assay (PLA) was applied as an independent PPI validation method for the detection of interactions between endogenously expressed proteins. For this purpose, the components from the PLA kit produced by Duolink® (In Situ Detection Reagents Orange (DUO92007)) were used.

Cells were seeded on pre-coated cover slips 24 hours prior the staining and fixed with a 2% Paraformaldehyde (PFA) solution containing the nuclear dye Hoechst (1:3000 final dilution). This procedure was carried out for 15 min at room temperature. Following three consecutive washing steps with PBS-Tween (phosphate-buffer saline, 0.05% Tween) the fixed cells were permeabilized with 1% BSA-PBS-Tween solution containing 0.1 % Triton for 10 min at room temperature. After one additional washing with PBS-Tween, blocking was performed for 30 min at room temperature with a PLA blocking solution (see **"Buffers and Solutions"** in the "**Materials**" section). All antibodies used for

the detection of binary interactions were diluted according to the provider´s recommendation (see **Antibodies List** in the **Materials** section) in PLA blocking solution. Each interaction test includes three paired IgG combinations: one containing two specific antibodies for the target proteins and two negative controls (NC1 and NC2) containing one specific antibody and one isogenic control IgG, respectively.

The antibody treatment was performed by placing the cover slips in a humidity chamber, applying 50 µL of each antibody solution per cover slip and incubating at 37°C for 60 min. Following three consecutive washing steps with PBS-Tween (0.05%), 40 µL of the diluted MINUS and PLUS PLA probes mixtures (1:5 in PLA probe solution, see **Buffers and Solutions** section) were pipetted per cover slip in the humidity chamber and let to incubate for 60 min at 37°C. Subsequently, a ligation mix was prepared as a fivefold dilution of the PLA ligation stock in autoclaved MiliQ water. After washing the cover slips with 1x PLA washing buffer A (see **Buffers and Solutions** section) two consecutive times for 5 min at room temperature under gentle agitation, the ligase provided in the kit was diluted with the previously prepared ligation mix in a 1:40 ratio and 40 µL were applied on each cover slip. After 30 min of incubation at 37°C the cover slips were washed twice with 1x PLA washing buffer A for 2 min at room temperature under gentle agitation. This step was followed by adding the polymerase from the kit to a pre-diluted amplification stock in a 1:80 ratio and pipetting 40 µL of this reaction mix onto each cover slip in the humidity chamber at 37°C. Here the recommended incubation time accounted 100 min. Note that longer incubation at this step might lead to an increased background signal.

Finally, each cover slip was washed twice in 1x PLA washing buffer B for 10 min at room temperature and once with 0.01x PLA washing buffer B by just shortly dipping the cover slips into the solution. The washed cover slips were embedded on microscope slides in 5 µL mounting medium and let to dry overnight protected from direct light at room temperature. The slides can be kept at 4°C for up to six months.

Interaction spots become visible through the incorporation of fluorescently labeled nucleotides into the newly synthesized DNA during the amplification process. Templates for the amplification are the ligated oligos which emerge when the conjugated antibodies get in close proximity due to the interaction of the proteins of

interest. In case of no interaction the number of detected spots should be drastically reduced and comparable to the background signal obtained with isogenic non-specific IgGs. All images of PLA samples were taken with a Zeiss Axio-Imager using DAPI and Rhodamine filters and analyzed by ImageJ.

6.4.3 Immunofluorescence

Cells used for immunofluorescent staining purposes were seeded on cover slips, fixed and permeabilized identically to the procedures described above. Blocking was performed with 1% BSA-PBS-Tween solution for 30 min at room temperature. Primary antibodies were diluted according to the manufacturer´s instructions in 1% BSA-PBS-Tween and 40 µL of the diluted samples were applied on the cover slips for 1 hour at 37°C. The corresponding secondary antibodies were as well diluted in 1% BSA-PBS-Tween and incubated on the cover slips for 45 min at room temperature. Washings between antibody incubation steps were executed with PBS-Tween (0.05%). Cover slips were mounted on microscopy slides and left to dry as described above.

7. Supplementary information

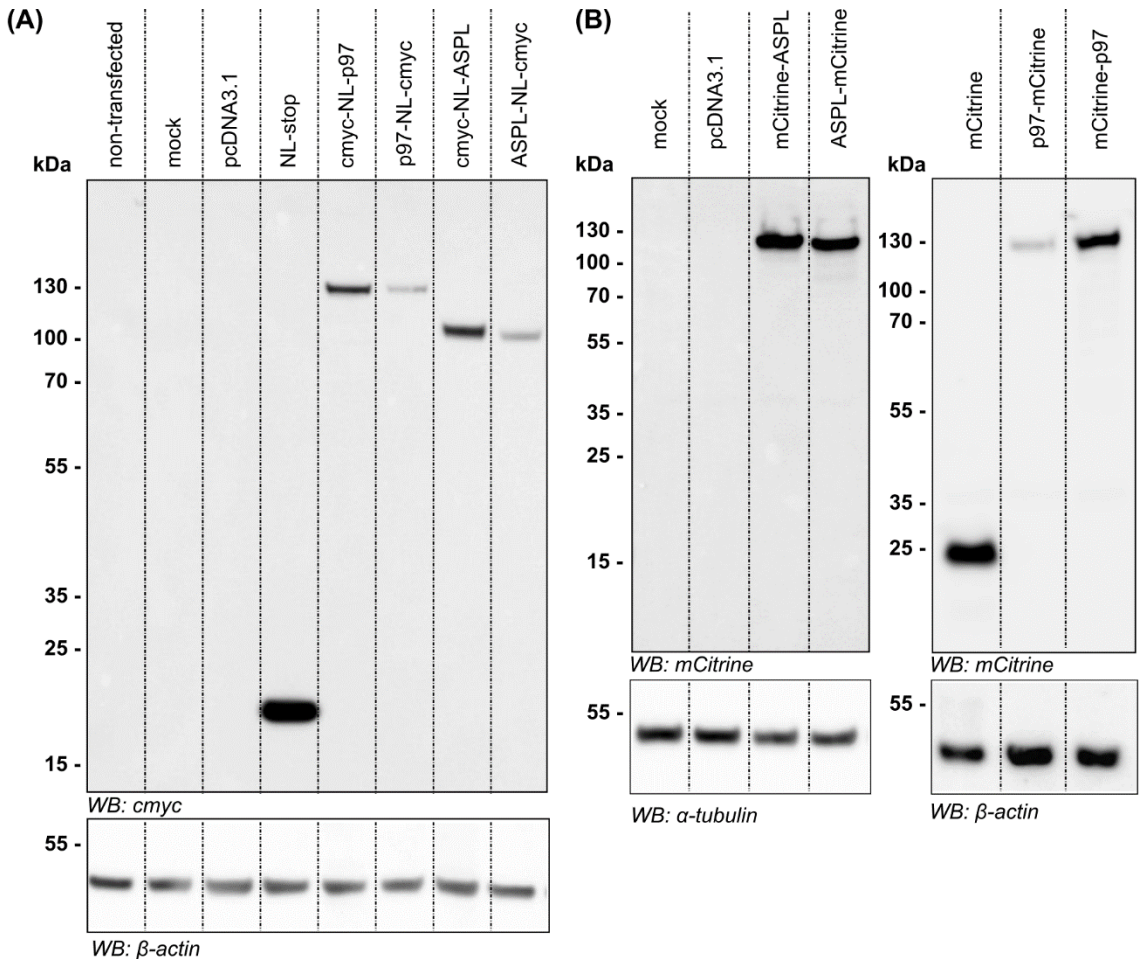
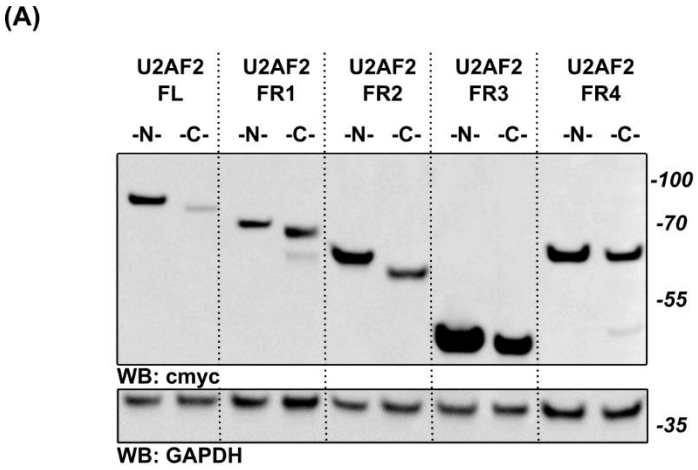


FIG. S1: Expression validation of LuTHy constructs.

ASPL and p97 expression from LuTHy plasmids was validated by western blotting. Wild-type HEK293 cells were transfected individually with the respective plasmids in 6-well plates. 48 hours post transfection the cells were lysed and 20 µg total protein amount were loaded in each lane of an SDS-gel. Following the gel electrophoresis immunoblotting was performed with an antibody detecting the epitope cmyc tag of the NL-constructs and anti-GFP antibody for mCitrine detection. Here the mCitrine plasmids were lacking the double PA-tag. β-actin and a-tubulin were tested as loading controls with respective antibodies. The images were developed by chemiluminescence in a Fuji-Imager and edited by ImageJ.



(B)

constructs	expected molecular weight
U2AF2-NL-cmyc FL	N: 75.5 kDa C: 74.5 kDa
U2AF2-NL-cmyc FR1	N: 67.7 kDa C: 66.6 kDa
U2AF2-NL-cmyc FR2	N: 59.1 kDa C: 58.1 kDa
U2AF2-NL-cmyc FR3	N: 45.2 kDa C: 44.1 kDa
U2AF2-NL-cmyc FR4	N: 61.6 kDa C: 60.6 kDa

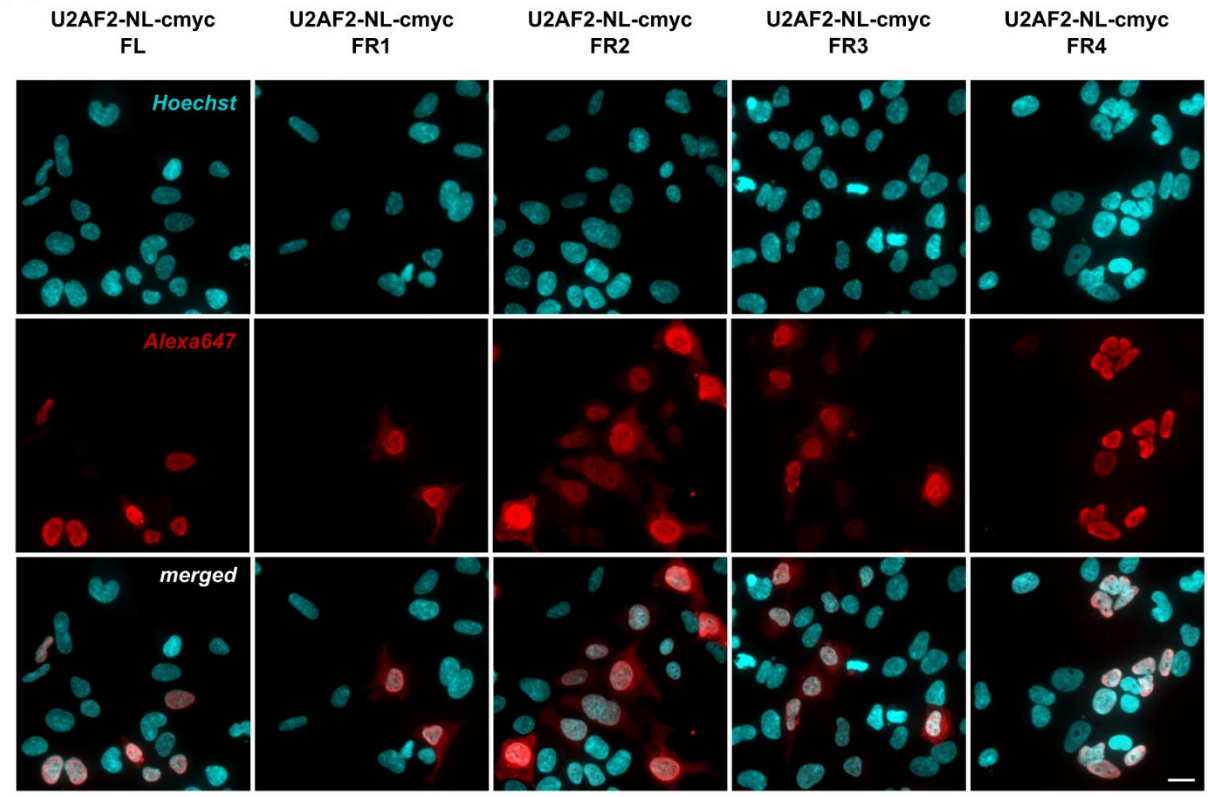


FIG. S2: U2AF2 fragments expression and cellular localization.

(A) Expression of full-length U2AF2 and four fragments as fusion construct with cmyc-NL in wild-type HEK293 cells was validated by immunoblotting with an antibody for the detection of the cmyc epitope tag. 15 µg total protein were loaded per each lane. GAPDH was tested as a loading control. **(B)** List of the tested constructs and the expected molecular weight (in kDa) for N- and C-terminally fused proteins. **(C)** Immunofluorescence performed in wild-type HEK293 cells individually transfected with the depicted constructs. The same mouse anti-myc antibody as applied in the western blot was used as primary antibody and anti-mouse IgG conjugated with Alexa647 was used as secondary antibody. Image acquisition was performed with a Cy5 filter set for Alexa647 signal detection and DAPI filter set for Hoechst signal detection on a Zeiss Axio Imager (20x magnification). Scale bar = 10 µm. All experiments presented in this graph were executed by Lynn van der Beek as part of her bachelor thesis under my supervision.

Table S1: PPIs between p97 and selected UBX proteins quantified by LuTHy.

Negative cBRET or cLuC values were replaced by "0". All values represent means of biological duplicates ± SEM.

interaction:	cBRET	SEM	cBRET (bars)	cLuC	SEM	cLuC (bars)	interaction:	cBRET	SEM	cBRET (bars)	cLuC	SEM	cLuC (bars)
NL-p97/PA-mCit-p37	0,000	0,000		0,000	0,000		NL-p97/PA-mCit-UBXD4	0,022	0,003		0,003	0,000	
NL-p97/p37-mCit-PA	0,014	0,000		0,002	0,000		NL-p97/UBXD4-mCit-PA	0,010	0,001		0,001	0,000	
p97-NL/PA-mCit-p37	0,000	0,000		0,000	0,000		p97-NL/PA-mCit-UBXD4	0,002	0,001		0,004	0,004	
p97-NL/p37-mCit-PA	0,002	0,001		0,010	0,001		p97-NL/UBXD4-mCit-PA	-0,003	0,001		-0,005	0,003	
NL-p37/PA-mCit-p97	0,047	0,000		0,052	0,006		NL-UBXD4/PA-mCit-p97	0,081	0,002		0,065	0,007	
NL-p37/p97-mCit-PA	0,002	0,000		0,032	0,004		NL-UBXD4/p97-mCit-PA	0,004	0,000		0,063	0,009	
p37-NL/PA-mCit-p97	0,073	0,001		0,209	0,034		UBXD4-NL/PA-mCit-p97	0,153	0,000		0,615	0,099	
p37-NL/p97-mCit-PA	0,004	0,001		0,118	0,020		UBXD4-NL/p97-mCit-PA	0,037	0,009		0,256	0,043	
NL-p97/PA-mCit-p47	0,031	0,000		0,336	0,029		NL-p97/PA-mCit-UBXD5	0,000	0,000		0,000	0,000	
NL-p97/p47-mCit-PA	0,062	0,000		0,295	0,013		NL-p97/UBXD5-mCit-PA	0,000	0,000		0,000	0,000	
p97-NL/PA-mCit-p47	0,009	0,001		0,543	0,029		p97-NL/PA-mCit-UBXD5	0,000	0,000		0,000	0,008	
p97-NL/p47-mCit-PA	0,014	0,000		0,472	0,005		p97-NL/UBXD5-mCit-PA	0,000	0,001		0,000	0,007	
NL-p47/PA-mCit-p97	0,018	0,000		0,008	0,003		NL-UBXD5/PA-mCit-p97	0,010	0,001		0,205	0,002	
NL-p47/p97-mCit-PA	0,000	0,001		0,004	0,000		NL-UBXD5/p97-mCit-PA	0,000	0,001		0,139	0,006	
p47-NL/PA-mCit-p97	0,030	0,002		0,008	0,001		UBXD5-NL/PA-mCit-p97	0,030	0,002		0,480	0,001	
p47-NL/p97-mCit-PA	0,000	0,000		0,006	0,001		UBXD5-NL/p97-mCit-PA	0,012	0,001		0,307	0,003	
NL-p97/PA-mCit-UBXD1	0,026	0,001		0,012	0,001		NL-p97/PA-mCit-UBXD6	0,004	0,000		0,006	0,000	
NL-p97/UBXD1-mCit-PA	0,040	0,001		0,024	0,006		NL-p97/UBXD6-mCit-PA	0,015	0,000		0,007	0,003	
p97-NL/PA-mCit-UBXD1	0,029	0,002		0,026	0,005		p97-NL/PA-mCit-UBXD6	0,006	0,002		0,055	0,005	
p97-NL/UBXD1-mCit-PA	0,055	0,004		0,053	0,004		p97-NL/UBXD6-mCit-PA	0,008	0,001		0,017	0,005	
NL-UBXD1/PA-mCit-p97	0,020	0,000		0,046	0,001		NL-UBXD6/PA-mCit-p97	0,019	0,002		0,021	0,003	
NL-UBXD1/p97-mCit-PA	0,003	0,000		0,035	0,000		NL-UBXD6/p97-mCit-PA	0,004	0,001		0,014	0,005	
UBXD1-NL/PA-mCit-p97	0,029	0,001		0,032	0,001		UBXD6-NL/PA-mCit-p97	0,027	0,000		0,044	0,001	
UBXD1-NL/p97-mCit-PA	0,004	0,001		0,021	0,001		UBXD6-NL/p97-mCit-PA	0,009	0,001		0,034	0,005	
NL-p97/PA-mCit-UBXD2	0,000	0,000		0,000	0,000		NL-p97/PA-mCit-UBXD8	0,019	0,001		0,006	0,000	
NL-p97/UBXD2-mCit-PA	0,017	0,001		0,000	0,000		NL-p97/UBXD8-mCit-PA	0,027	0,001		0,001	0,000	
p97-NL/PA-mCit-UBXD2	0,000	0,001		0,017	0,006		p97-NL/PA-mCit-UBXD8	0,005	0,001		0,027	0,001	
p97-NL/UBXD2-mCit-PA	0,008	0,000		0,000	0,006		p97-NL/UBXD8-mCit-PA	0,011	0,000		0,009	0,000	
NL-UBXD2/PA-mCit-p97	0,049	0,001		0,595	0,023		NL-UBXD8/PA-mCit-p97	0,011	0,000		0,051	0,006	
NL-UBXD2/p97-mCit-PA	0,015	0,000		0,416	0,038		NL-UBXD8/p97-mCit-PA	0,001	0,000		0,033	0,001	
UBXD2-NL/PA-mCit-p97	0,083	0,005		0,869	0,059		UBXD8-NL/PA-mCit-p97	0,014	0,001		0,057	0,006	
UBXD2-NL/p97-mCit-PA	0,027	0,002		0,376	0,035		UBXD8-NL/p97-mCit-PA	0,001	0,000		0,039	0,004	
NL-p97/PA-mCit-UBXD3	0,030	0,000		0,006	0,001		NL-p97/PA-mCit-ASPL	0,054	0,000		0,616	0,005	
NL-p97/UBXD3-mCit-PA	0,009	0,000		0,001	0,000		NL-p97/ASPL-mCit-PA	0,209	0,003		0,875	0,070	
p97-NL/PA-mCit-UBXD3	0,019	0,000		0,005	0,002		p97-NL/PA-mCit-ASPL	0,036	0,014		0,528	0,015	
p97-NL/UBXD3-mCit-PA	0,000	0,001		0,002	0,002		p97-NL/ASPL-mCit-PA	0,220	0,005		0,726	0,073	
NL-UBXD3/PA-mCit-p97	0,007	0,000		0,011	0,000		NL-ASPL/PA-mCit-p97	0,040	0,000		0,520	0,018	
NL-UBXD3/p97-mCit-PA	0,001	0,000		0,008	0,001		NL-ASPL/p97-mCit-PA	0,013	0,001		0,272	0,005	
UBXD3-NL/PA-mCit-p97	0,042	0,001		0,381	0,000		ASPL-NL/PA-mCit-p97	0,044	0,001		0,376	0,046	
UBXD3-NL/p97-mCit-PA	0,005	0,000		0,227	0,005		ASPL-NL/p97-mCit-PA	0,024	0,000		0,211	0,017	

Table S2: Proteins identified as ASPL interaction partners in the SILAC data set.

The table includes manually curated information about the function and structural characteristics of the listed hits. The column "Shuttling activity" refers to the reported ability of the proteins to translocate between nucleus and cytoplasm. All proteins validated as ASPL interaction partners via LuTHy are marked with "+" in the last column.

Protein:	UniProt:	Function:	Key domains:	Shuttling activity:	Comments:	Validated in LuTHy
AGK	Q53H12	Acylglycerol kinase, mitochondrial	intermembrain domain	unknown	mitochondrial potein	-
ARL6IP4	Q66PJ3	Alternative splicing modulator		unknown		-
BOLA1	Q9Y3E2	Mitochondrial iron-sulfur (Fe-S) cluster assembly factor		unknown		+
BOLA2	Q9H3K6	Nuclear iron-sulfur (Fe-S) cluster assembly factor		unknown		+
CIRBP	Q14011	Translational activator, mRNA chaperone		+		-
HSP90AA1	P07900	Stress inducable isoform of the molecular chaperone Hsp90	ATPase domain, peptide-binding domain	unknown		+
HSPA4	P34932	Molecular chaperone involved in germ cell progression		unknown		+
LUC7L2	Q9Y383	Putative splicing factor		unknown		-
SAP18	O00422	Component of the histone deacetylase complex		unknown		+
PUF60	Q9UHX1	Putative splicing factor	2xRRM, UHM	unknown	U2AF2 homolog	+
SF3B3	Q15393	Component of U2 snRNP complex	7xULM, HEAT	unknown		+
SNRPG	P62308	Component of the SMN-Sm complex		unknown		-
TAF15	Q92804	RNA polymerase II transcription activiator	QGSY-rich, Gly-rich, RRM, 2xRGG	unknown		-
PNN	Q9H307	Regulation of alternative splicing	RS	unknown	broader SR family	-
RBM39	Q14498	Regulation of alternative splicing, U2AF2 homolog	RRM, RS, UHM	unknown	broader SR family	+
RNPS1	Q15287	Constitutive and alternative splicing regulator	RRM, RS	unknown	broader SR family	-
SFRS1	Q07955	Constitutive and alternative splicing activator	2xRRM, RS	+	canonical SR protein	-
SFRS3	P84103	Constitutive and alternative splicing activator	RRM, RS	+	canonical SR protein	-
SFRS6	Q13247	Constitutive and alternative splicing activator	2xRRM, RS	+	canonical SR protein	-
SFRS7	Q16629	Constitutive and alternative splicing activator	2xRRM, RS, CCHC-type zinc finger	+	canonical SR protein	-
SRSF10	O75494	General splicing repressor	RRM, RS	+	canonical SR protein	-
TRA2B	P62995	Splicing activator	RRM, 2xRS	unknown	broader SR family	-
SNRNP27	Q8WVK2	Associated with U4/U6.U5 complex	RS	unknown	broader SR family	+
SNRNP70	P08621	Constitutive splicing factor	RRM, RS	unknown	broader SR family	-
U2AF1	Q01081	Constitutive splicing factor	RRM, RS, 2xC3H1-type zinc finger	unknown	broader SR family	-
U2AF2	P26368	Constitutive splicing factor	2xRRM, RS, UHM, ULM	+	broader SR family	+

Table S3: GO-term enrichment analysis of SILAC hits.

Besides BOLA2 (**see FIG. 5**), all other proteins were encountered within the displayed GO-terms. Selection criteria for the GO-terms in this chart were a minimum number of at least two assigned proteins and a p-value<0.01 for fold enrichment.

GO-Category	GO-ID // annotation	Count	Proteins	Fold Enrichment	p-value
MF (molecular function)	GO:0005515 // protein binding	24	SFRS6, SFRS7, SNRPG, TRA2B, PUF60, SNRNP70, SAP18, SF3B3, BOLA1, U2AF1, SARNP, U2AF2, TAF15, LUC7L2, SNRNP27, ASPL, RNPS1, HSP90AA1, ARL6IP4, CIRBP, SFRS3, SFRS2, SFRS1, SRSF10	1.8	$3.6x10^{-6}$
CC (cellular compartment)	GO:0005654 // nucleoplasm	22	SFRS6, SFRS7, SNRPG, TRA2B, PUF60, PNN, SNRNP70, SAP18, SF3B3, U2AF1, SARNP, U2AF2, TAF15, SNRNP27, ASPL, RNPS1, HSP90AA1, CIRBP, SFRS3, SFRS2, SFRS1, SRSF10	5.5	$4.6x10^{-14}$
MF (molecular function)	GO:0044822 // poly(A) RNA binding	21	SFRS6, SFRS7, SNRPG, TRA2B, PUF60, PNN, SNRNP70, SAP18, U2AF1, SARNP, U2AF2, TAF15, LUC7L2, HSP90AA1, RNPS1, CIRBP, ARL6IP4, SFRS3, SFRS2, SFRS1, SRSF10	12.6	$2.2x10^{-20}$
BP (biological process)	GO:0000398 // mRNA splicing, via spliceosome	15	SFRS6, SFRS7, SNRPG, TRA2B, PNN, SNRNP70, SF3B3, U2AF1, U2AF2, SNRNP27, RNPS1, SFRS3, SFRS2, SRSF10, SFRS1	47.3	$2.4x10^{-21}$
MF (molecular function)	GO:0000166 // nucleotide binding	15	SFRS6, SFRS7, TRA2B, PUF60, SNRNP70, U2AF1, U2AF2, TAF15, HSP90AA1, RNPS1, CIRBP, SFRS3, SFRS2, SRSF10, SFRS1	29.1	$3.1x10^{-18}$
CC (cellular compartment)	GO:0016607 // nuclear speck	14	SFRS6, U2AF1, SARNP, U2AF2, LUC7L2, PNN, RNPS1, SNRNP70, ARL6IP4, SFRS3, SAP18, SFRS2, SFRS1, SRSF10	48.8	$1.1x10^{-19}$
MF (molecular function)	GO:0003676 // nucleic acid binding	13	SFRS7, TAF15, SNRNP27, TRA2B, PUF60, RNPS1, SNRNP70, CIRBP, SFRS3, SF3B3, SFRS2, SFRS1, SRSF10	8.9	$2.1x10^{-9}$
MF (molecular function)	GO:0003723 // RNA binding	11	SFRS6, U2AF1, U2AF2, SNRPG, RNPS1, SNRNP70, CIRBP, SFRS3, SF3B3, SFRS1, SRSF10	13.6	$1.5x10^{-9}$
BP (biological process)	GO:0006369 // termination of RNA polymerase II transcription	10	SFRS6, U2AF1, SARNP, SFRS7, U2AF2, SNRPG, RNPS1, SFRS3, SFRS2, SFRS1	109.3	$7.4x10^{-17}$
BP (biological process)	GO:0006406 // mRNA export from nucleus	10	SFRS6, U2AF1, SARNP, SFRS7, U2AF2, RNPS1, SFRS3, SFRS2, SFRS1, SRSF10	70.0	$4.9x10^{-15}$
BP (biological process)	GO:0008380 // RNA splicing	10	U2AF1, SFRS7, SNRPG, PUF60, RNPS1, SNRNP70, ARL6IP4, SAP18, SF3B3, SFRS2	42.1	$5.3x10^{-13}$
BP (biological process)	GO:0006397 // mRNA processing	10	U2AF1, SFRS7, U2AF2, PUF60, SNRNP70, ARL6IP4, SAP18, SF3B3, SFRS2, SFRS1	39.1	$1.0x10^{-12}$
BP (biological process)	GO:0031124 // mRNA 3'-end processing	9	SFRS6, U2AF1, SARNP, SFRS7, U2AF2, RNPS1, SFRS3, SFRS2, SFRS1	125.9	$1.6x10^{-15}$
BP (biological process)	GO:0006405 // RNA export from nucleus	9	SFRS6, U2AF1, SARNP, SFRS7, U2AF2, RNPS1, SFRS3, SFRS2, SFRS1	114.5	$3.7x10^{-15}$
BP (biological process)	GO:0048025 // negative regulation of mRNA splicing, via spliceosome	7	SFRS6, SFRS7, U2AF2, TRA2B, RNPS1, SAP18, SRSF10	233.2	$1.7x10^{-13}$
CC (cellular compartment)	GO:0005681 // spliceosomal complex	6	U2AF1, U2AF2, SNRPG, SNRNP70, SF3B3, SFRS2	44.7	$1.6x10^{-7}$

GO-Category	GO-ID // annotation	Count	Proteins	Fold Enrichment	p-value
BP (biological process)	GO:0000381 // regulation of alternative mRNA splicing, via spliceosom	5	SFRS6, TRA2B, RNPS1, SAP18, SFRS2	92.1	1.9×10^{-7}
CC (cellular compartment)	GO:0071013 // catalytic step 2 spliceosome	5	U2AF1, SNRPG, PNN, SF3B3, SFRS1	38.1	7.1×10^{-6}
MF (molecular function)	GO:0003729 // mRNA binding	5	LUC7L2, TRA2B, HSP90AA1, SNRNP70, SFRS1	27.0	2.7×10^{-5}
BP (biological process)	GO:0006376 // mRNA splice site selection	4	SFRS6, LUC7L2, SFRS1, SRSF10	164.6	1.5×10^{-6}
CC (cellular compartment)	GO:0071004 // U2-type prespliceosome	4	U2AF2, SNRPG, LUC7L2, SNRNP70	164.9	1.5×10^{-6}
CC (cellular compartment)	GO:0035145 // exon-exon junction complex	4	PNN, RNPS1, SAP18, SFRS1	127.4	3.5×10^{-6}
MF (molecular function)	GO:0036002 // pre-mRNA binding	3	SFRS6, TRA2B, SFRS2	225.1	6.9×10^{-5}
CC (cellular compartment)	GO:0005685 // U1 snRNP	3	SNRPG, LUC7L2, SNRNP70	110.7	3.0×10^{-4}
BP (biological process)	GO:0000375 // RNA splicing, via transesterification reactions	3	TRA2B, SF3B3, SRSF10	84.0	5.3×10^{-4}
CC (cellular compartment)	GO:0071011 // precatalytic spliceosome	3	SNRPG, SNRNP27, SNRNP70	84.1	5.3×10^{-4}
CC (cellular compartment)	GO:0061574 // ASAP complex	2	RNPS1, SAP18	467.3	4.1×10^{-3}
CC (cellular compartment)	GO:0089701 // U2AF	2	U2AF1, U2AF2	233.6	8.2×10^{-3}
MF (molecular function)	GO:0050733 // RS domain binding	2	SFRS1, SRSF10	225.1	8.5×10^{-3}
CC (cellular compartment)	GO:0000243 // commitment complex	2	U2AF2, SNRNP70	200.3	9.6×10^{-3}
MF (molecular function)	GO:0030628 // pre-mRNA 3'-splice site binding	2	U2AF1, U2AF2	192.9	9.9×10^{-3}
MF (molecular function)	GO:1990446 // U1 snRNP binding	2	SNRPG, SNRNP70	192.9	9.9×10^{-3}

Table S4: Summary of a three-step LuTHy validation of p97:ASPL interaction partners.

All proteins detected via SILAC as ASPL interaction partners were tested vs. ASPL, p97 and p97 in the presence of additional ASPL in three consecutive LuTHy screens. The numbers in parenthesis after cBRET or cLuC indicate the number of interactions directions for the tested interaction above the thresholds for the specified readouts. The validated hits for each of the three data sets are highlighted additionally. The proteins marked in yellow are the ones fulfilling the selection criteria specific interaction partners of the p97:ASPL complex.

	SILAC counts	LuTHy directions					
proteins		vs. ASPL		vs. p97		vs. p97 + ASPL	
HSPA4	1	**cBRET (1) // cLuC (2)**	**HSPA4**	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
BOLA2	1	**cBRET (2) // cLuC (2)**	**BOLA2**	**cBRET (1) // cLuC (2)**	**BOLA2**	**cBRET (2) // cLuC (4)**	**BOLA2**
LUC7L2	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (1)	**LUC7L2**
CIRBP	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
ARL6IP4	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		**cBRET (2) // cLuC (0)**	**ARL6IP4**
SAP18	1	**cBRET (0) // cLuC (2)**	**SAP18**	**cBRET (0) // cLuC (1)**	**SAP18**	**cBRET (0) // cLuC (1)**	**SAP18**
RBM39	1	**cBRET (1) // cLuC (3)**	**RBM39**	cBRET (0) // cLuC (0)		**cBRET (0) // cLuC (2)**	**RBM39**
HSP90AA1	1	**cBRET (1) // cLuC (1)**	**HSP90AA1**	cBRET (0) // cLuC (0)		**cBRET (0) // cLuC (2)**	**HSP90AA1**
SF3B3	1	**cBRET (2) // cLuC (4)**	**SF3B3**	**cBRET (0) // cLuC (1)**	**SF3B3**	**cBRET (0) // cLuC (3)**	**SF3B3**
RNPS1	1	cBRET (0) // cLuC (0)		**cBRET (0) // cLuC (1)**	**RNPS1**	**cBRET (0) // cLuC (2)**	**RNPS1**
TRA2B	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
TAF15	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
SNRNP70	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		**cBRET (0) // cLuC (1)**	**SNRNP70**
SFRS3	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
SFRS6	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
SFRS7	1	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
SNRNP27	1	**cBRET (1) // cLuC (2)**	**SNRPN27**	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
PNN	1	cBRET (0) // cLuC (0)		**cBRET (0) // cLuC (1)**	**PNN**	cBRET (0) // cLuC (0)	
BOLA1	2	**cBRET (1) // cLuC (1)**	**BOLA1**	**cBRET (0) // cLuC (2)**	**BOLA1**	**cBRET (1) // cLuC (0)**	**BOLA1**
PUF60	2	**cBRET (2) // cLuC (2)**	**PUF60**	cBRET (0) // cLuC (0)		**cBRET (0) // cLuC (1)**	**PUF60**
SRSF10	2	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
SNRPG	2	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
U2AF2	2	**cBRET (4) // cLuC (0)**	**U2AF2**	cBRET (0) // cLuC (0)		**cBRET (2) // cLuC (0)**	**U2AF2**
AGK	3	cBRET (0) // cLuC (0)		**cBRET (1) // cLuC (2)**	**AGK**	**cBRET (0) // cLuC (1)**	**AGK**
ASPL	3	**cBRET (3) // cLuC (0)**	**ASPL**	**cBRET (8) // cLuC (8)**	**ASPL**	**cBRET (7) // cLuC (8)**	**ASPL**
SFRS1	3	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
U2AF1	3	cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)		cBRET (0) // cLuC (0)	
p97	4	**cBRET (8) // cLuC (8)**	**p97**	**cBRET (8) // cLuC (8)**	**p97**	**cBRET (5) // cLuC (8)**	**p97**

Table S5: Differential expression modifications in ASPL deficient cells.

Lists of top 40 hits among the down- and upregulated genes upon ASPL-KO in HEK293 cells. Selection criteria were p_{adj}<0.01 and log2FC>1 or <-1.

downregulated				upregulated			
#	gene	log2FC	p_{adj}	#	gene	log2FC	p_{adj}
-1-	HTATIP2	10.093	$2.100x10^{-32}$	-1-	EYA2	-10.583	$4.301x10^{-26}$
-2-	KLHDC9	9.098	$1.867x10^{-17}$	-2-	AL355315.1	-9.934	$1.030x10^{-20}$
-3-	KCNT2	8.920	$7.052x10^{-17}$	-3-	PKP3	-9.463	$1.581x10^{-38}$
-4-	MMP16	8.811	$1.847x10^{-12}$	-4-	GLYATL2	-8.834	$9.716x10^{-17}$
-5-	ZNF572	8.770	$3.882x10^{-16}$	-5-	PAX7	-8.597	$1.482x10^{-15}$
-6-	RIBC2	8.681	$8.470x10^{-16}$	-6-	TMEM200C	-8.344	$8.911x10^{-32}$
-7-	ADGRB3	8.666	$6.504x10^{-62}$	-7-	PXDN	-7.972	$3.983x10^{-19}$
-8-	SERP2	8.616	$1.071x10^{-15}$	-8-	P4HTM	-7.933	$6.895x10^{-51}$
-9-	IAH1	8.034	$2.220x10^{-48}$	-9-	AC004805.1	-7.713	$1.355x10^{-10}$
-10-	RPL23AP47	7.938	$6.479x10^{-13}$	-10-	CDH20	-7.654	$2.289x10^{-12}$
-11-	SLITRK2	7.874	$1.072x10^{-31}$	-11-	AD000671.1	-7.353	$1.319x10^{-10}$
-12-	RAMP1	7.799	$2.037x10^{-12}$	-12-	KNDC1	-6.878	$4.424x10^{-11}$
-13-	B3GALNT1	7.758	$3.243x10^{-31}$	-13-	GCSHP3	-6.873	$0.090x10^{-2}$
-14-	HBQ1	7.717	$1.990x10^{-11}$	-14-	AC108448.2	-6.762	$1.065x10^{-8}$
-15-	GYPC	7.642	$9.212x10^{-14}$	-15-	EVA1A	-6.590	$1.608x10^{-7}$
-16-	HIST1H2BH	7.555	$4.189x10^{-11}$	-16-	LRRN2	-6.568	$3.06x10^{-32}$
-17-	ACSS3	7.517	$3.463x10^{-62}$	-17-	AC126177.7	-6.563	$1.281x10^{-8}$
-18-	ARMCX4	7.508	$4.472x10^{-48}$	-18-	NHS	-6.523	$1.63x10^{-207}$
-19-	STAC	7.461	$4.284x10^{-25}$	-19-	GRIK3	-6.520	$1.296x10^{-12}$
-20-	SPON1	7.451	$0.010x10^{-2}$	-20-	IL18RAP	-6.430	$2.85x10^{-7}$
-21-	ZSCAN18	7.447	$1.300x10^{-99}$	-21-	BORCS7-ASMT	-6.311	0.020x10-2
-22-	MT1G	7.315	$3.112x10^{-11}$	-22-	IL18R1	-6.268	$3.876x10^{-59}$
-23-	AC107075.1	7.266	$0.070x10^{-3}$	-23-	CELF2	-6.223	0.010x10-2
-24-	MGST1	6.971	$4.930x10^{-193}$	-24-	CLSTN2	-6.180	$1.080x10^{-147}$
-25-	OTX2	6.953	$1.261x10^{-12}$	-25-	AL359091.2	-6.179	$3.726x10^{-7}$
-26-	NME5	6.856	$2.321x10^{-11}$	-26-	PGR	-6.098	$1.641x10^{-9}$
-27-	WIF1	6.843	$1.625x10^{-9}$	-27-	VCX3B	-6.090	$1.856x10^{-6}$
-28-	ZNF98	6.780	$2.284x10^{-30}$	-28-	AC000120.2	-6.045	$1.717x10^{-5}$
-29-	PSTPIP2	6.772	$6.100x10^{-176}$	-29-	MAGEB2	-5.992	0.010x10-2
-30-	RPS6KA6	6.648	$7.930x10^{-240}$	-30-	DGKK	-5.895	$6.948x10^{-6}$
-31-	C2orf74	6.551	$1.525x10^{-97}$	-31-	RAB6D	-5.656	$6.250x10^{-7}$
-32-	LDHAP4	6.417	$3.692x10^{-16}$	-32-	PTPRO	-5.590	$2.629x10^{-17}$
-33-	TRIM17	6.410	$3.146x10^{-8}$	-33-	EN1	-5.571	$6.765x10^{-17}$
-34-	TAL1	6.386	$1.760x10^{-57}$	-34-	TMIE	-5.556	$1.904x10^{-6}$
-35-	AC139677.1	6.252	0.010x10-2	-35-	MAP1LC3BP1	-5.538	$4.254x10^{-5}$
-36-	TCEAL7	6.216	$2.161x10^{-40}$	-36-	RUNX3	-5.527	$1.490x10^{-102}$
-37-	ZNF492	6.194	$1.748x10^{-43}$	-37-	C15orf62	-5.525	$1.400x10^{-28}$
-38-	AC004080.3	6.097	$6.820x10^{-6}$	-38-	IRF8	-5.487	$4.724x10^{-5}$
-39-	GUSBP3	6.096	$3.224x10^{-6}$	-39-	PLXNA4	-5.472	$9.044x10^{-27}$
-40-	IDUA	6.044	$2.083x10^{-6}$	-40-	SLC25A48	-5.466	$8.552x10^{-18}$

Table S6: Alternative splicing perturbations in ASPL deficient cell lines (A).

List of top 40 hits among the transcripts demonstrating significant reduction in the frequency of alternative splicing events upon ASPL-KO in HEK293 cells. Selection criteria were **probability>0.950** and **ΔPSI>0.200**. The "Type" column gives information about the type of the detected AS event. TS: tandem transcription start site; TE: tandem alternative polyadenylation.

#	gene	ΔPSI	PSI_{WT}	PSI_{KO}	coordinates	Strand	Type	Probability
-1-	**GGT1**	0.891	0.914	0.023	22:24628969-24628975	+	TE	1.000
-2-	**SLC10A7**	0.871	0.874	0.003	4:146256491-146256520	-	TE	1.000
-3-	**PFN1P2**	0.851	0.930	0.079	1:120433656-120434037	-	TS	1.000
-4-	**JPH3**	0.849	0.884	0.035	16:87698143-87698156	+	TE	1.000
-5-	**STAT4**	0.844	0.937	0.093	2:191029576-191029579	-	TE	1.000
-6-	**POLM**	0.842	0.908	0.066	7:44082515-44082540	-	TS	1.000
-7-	**APPL2**	0.829	0.911	0.082	12:105233155-105233450	-	TS	1.000
-8-	**SLC51A**	0.823	0.902	0.079	3:196233381-196233423	+	TE	1.000
-9-	**HIST1H4H**	0.820	0.916	0.096	6:26285113-26285509	-	TS	1.000
-10-	**SENP3**	0.819	0.921	0.102	17:7566034-7566051	+	TE	1.000
-11-	**TUT1**	0.807	0.811	0.004	11:62575048-62576244	-	TE	0.997
-12-	**SMG1P7**	0.805	0.890	0.084	16:70219581-70219795	-	TE	1.000
-13-	**C12orf57**	0.805	0.893	0.088	12:6944653-6945066	+	TE	1.000
-14-	**BID**	0.804	0.898	0.095	22:17774490-17774492	-	TS	1.000
-15-	**CAMKMT**	0.803	0.894	0.090	2:44361950-44361963	+	TS	1.000
-16-	**ECHDC2**	0.803	0.885	0.082	1:52921553-52921703	-	TS	1.000
-17-	**RTKN**	0.801	0.897	0.096	2:74441899-74441899	-	TS	1.000
-18-	**HHIPL2**	0.799	0.934	0.135	1:222522260-222522808	-	TE	1.000
-19-	**MRTFA**	0.799	0.901	0.102	22:40636686-40636687	-	TS	1.000
-20-	**AC092279.1**	0.799	0.960	0.161	19:24033445-24033448	+	TS	1.000
-21-	**H2AFY**	0.797	0.892	0.095	5:135399230-135399622	-	TS	1.000
-22-	**GSTCD**	0.795	0.899	0.104	4:105708784-105708786	+	TS	1.000
-23-	**CEP68**	0.793	0.937	0.144	2:65074485-65075251	+	TE	1.000
-24-	**STK33**	0.791	0.907	0.116	11:8594156-8594228	-	TS	1.000
-25-	**RNPC3**	0.789	0.846	0.058	1:103525697-103525823	+	TS	0.996
-26-	**BTN3A2**	0.780	0.902	0.122	6:26375702-26375737	+	TE	1.000
-27-	**TMEM91**	0.772	0.865	0.093	19:41383874-41384060	+	TE	1.000
-28-	**ENO3**	0.769	0.885	0.116	17:4956978-4957118	+	TE	1.000
-29-	**AC090197.1**	0.767	0.867	0.100	8:23336208-23337457	+	TS	1.000
-30-	**HIVEP3**	0.764	0.769	0.004	1:41510230-41511263	-	TE	1.000
-31-	**PFKFB1**	0.763	0.854	0.091	X:54933387-54933462	-	TE	1.000
-32-	**TPM2**	0.761	0.881	0.120	9:35682926-35682931	-	TE	1.000
-33-	**IL15**	0.761	0.888	0.127	4:141733436-141733458	+	TE	1.000
-34-	**IFI27**	0.760	0.866	0.106	14:94112134-94112194	+	TE	1.000
-35-	**RANGRF**	0.760	0.873	0.113	17:8288654-8288670	+	TS	1.000
-36-	**AL391825.1**	0.760	0.883	0.123	1:161890833-161890939	+	TS	1.000
-37-	**SNAI3-AS1**	0.758	0.882	0.124	16:88686791-88687180	+	TE	1.000
-38-	**BCL11B**	0.758	0.864	0.106	14:99271229-99271485	-	TS	1.000
-39-	**PCDHGB8P**	0.757	0.880	0.123	5:141428259-141429158	+	TE	1.000
-40-	**FAH**	0.756	0.961	0.205	15:80186130-80186340	+	TE	0.999

Table S7: Alternative splicing perturbations in ASPL deficient cell lines (B).

List of top 40 hits among the transcripts demonstrating significant increase in the frequency of alternative splicing events upon ASPL-KO in HEK293 cells. Selection criteria were **probability>0.950** and **ΔPSI<-0.200**. The "Type" column gives information about the type of the detected AS event. TS: tandem transcription start site; TE: tandem alternative polyadenylation.

#	gene	ΔPSI	PSI_{WT}	PSI_{KO}	coordinates	Strand	Type	Probability
-1-	**PFN1P2**	-0.852	0.068	0.920	1:120434038-120434052	-	TS	1.000
-2-	**STAT4**	-0.841	0.064	0.905	2:191029580-191029866	-	TE	1.000
-3-	**APPL2**	-0.830	0.088	0.917	12:105233091-105233154	-	TS	1.000
-4-	**AC012615.6**	-0.829	0.064	0.892	19:1822561-1822692	+	TS	1.000
-5-	**GAL3ST1**	-0.828	0.074	0.902	22:30554647-30555497	-	TE	1.000
-6-	**SLC51A**	-0.826	0.092	0.918	3:196233063-196233380	+	TE	1.000
-7-	**SENP3**	-0.825	0.077	0.902	17:7565765-7566033	+	TE	1.000
-8-	**CPEB1**	-0.820	0.094	0.915	15:82571634-82571759	-	TS	1.000
-9-	**HIST1H4H**	-0.820	0.084	0.904	6:26285510-26285534	-	TS	1.000
-10-	**C12orf57**	-0.810	0.105	0.915	12:6945067-6945098	+	TE	1.000
-11-	**BID**	-0.809	0.101	0.910	22:17774413-17774489	-	TS	1.000
-12-	**SMG1P7**	-0.808	0.109	0.918	16:70219574-70219580	-	TE	1.000
-13-	**TUT1**	-0.807	0.189	0.996	11:62575045-62575047	-	TE	0.993
-14-	**HHIPL2**	-0.803	0.064	0.867	1:222522809-222522887	-	TE	1.000
-15-	**MRTFA**	-0.802	0.097	0.899	22:40636688-40636691	-	TS	1.000
-16-	**CAMKMT**	-0.801	0.109	0.910	2:44361964-44362145	+	TS	1.000
-17-	**ECHDC2**	-0.801	0.116	0.917	1:52921768-52921785	-	TS	1.000
-18-	**AC092279.1**	-0.800	0.038	0.838	19:24033449-24033621	+	TS	1.000
-19-	**GSTCD**	-0.798	0.103	0.901	4:105708787-105709016	+	TS	1.000
-20-	**H2AFY**	-0.790	0.112	0.902	5:135399205-135399229	-	TS	1.000
-21-	**CEP68**	-0.787	0.064	0.851	2:65074405-65074484	+	TE	1.000
-22-	**IZUMO4**	-0.786	0.072	0.858	19:2096943-2096997	+	TS	1.000
-23-	**RNPC3**	-0.785	0.155	0.940	1:103525691-103525696	+	TS	0.999
-24-	**FAM13A**	-0.785	0.077	0.862	4:88823232-88823258	-	TS	1.000
-25-	**BTN3A2**	-0.783	0.096	0.879	6:26375515-26375701	+	TE	1.000
-26-	**STK33**	-0.781	0.095	0.876	11:8594229-8594259	-	TS	1.000
-27-	**ENO3**	-0.772	0.113	0.886	17:4957119-4957127	+	TE	1.000
-28-	**AC090197.1**	-0.772	0.126	0.898	8:23336171-23336207	+	TS	1.000
-29-	**TRAF3IP2-AS1**	-0.771	0.140	0.911	6:111503512-111503608	+	TE	0.997
-30-	**TMEM91**	-0.767	0.131	0.898	19:41384061-41384080	+	TE	1.000
-31-	**SNAI3-AS1**	-0.765	0.118	0.883	16:88687181-88687186	+	TE	0.998
-32-	**AL391825.1**	-0.765	0.113	0.878	1:161890940-161891680	+	TS	1.000
-33-	**HIVEP3**	-0.763	0.233	0.996	1:41506365-41510229	-	TE	1.000
-34-	**ZNF224**	-0.763	0.003	0.766	19:44106534-44108326	+	TE	1.000
-35-	**TPM2**	-0.763	0.121	0.884	9:35682937-35683241	-	TE	1.000
-36-	**BCL11B**	-0.760	0.132	0.893	14:99271161-99271228	-	TS	0.999
-37-	**IFI27**	-0.757	0.139	0.896	14:94111774-94112133	+	TE	1.000
-38-	**GRB2**	-0.756	0.114	0.870	17:75405654-75405681	-	TS	1.000
-39-	**IL15**	-0.756	0.115	0.870	4:141732738-141733435	+	TE	1.000
-40-	**EPN1**	-0.752	0.132	0.884	19:55675226-55675244	+	TS	1.000

8. References

1. Xia, K., Fu, Z., Hou, L. & Han, J.-D. J. Impacts of protein–protein interaction domains on organism and network complexity. *Genome Res.* **18**, 1500–1508 (2008).

2. Harper, J. W. & Bennett, E. J. Proteome complexity and the forces that drive proteome imbalance. *Nature* **537**, 328–338 (2016).

3. Liu, Y. *et al.* Impact of Alternative Splicing on the Human Proteome. *Cell Rep.* **20**, 1229–1241 (2017).

4. Spirin, V. & Mirny, L. A. Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12123–12128 (2003).

5. Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **92**, 291–294 (1998).

6. Sahni, N. *et al.* Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell* **161**, 647–660 (2015).

7. Sahni, N. *et al.* Edgotype: a fundamental link between genotype and phenotype. *Curr. Opin. Genet. Dev.* **23**, 649–657 (2013).

8. Sakai, Y. *et al.* Protein Interactome Reveals Converging Molecular Pathways Among Autism Disorders. *Sci. Transl. Med.* **3**, 86ra49-86ra49 (2011).

9. Brown, E. A. *et al.* Clustering the autisms using glutamate synapse protein interaction networks from cortical and hippocampal tissue of seven mouse models. *Mol. Autism* **9**, 48 (2018).

10. Karunakaran, K. B., Chaparala, S. & Ganapathiraju, M. K. *Potentially repurposable drugs for schizophrenia identified from its interactome*. http://biorxiv.org/lookup/doi/10.1101/442640 (2018) doi:10.1101/442640.

11. Singh, R., Letai, A. & Sarosiek, K. Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. *Nat. Rev. Mol. Cell Biol.* **20**, 175–193 (2019).

12. Campbell, K. J. & Tait, S. W. G. Targeting BCL-2 regulated apoptosis in cancer. *Open Biol.* **8**, 180002 (2018).

13. Souers, A. J. *et al.* ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **19**, 202–208 (2013).

14. Stumpf, M. P. H. *et al.* Estimating the size of the human interactome. *Proc. Natl. Acad. Sci.* **105**, 6959–6964 (2008).

15. Amaral, L. A. N. A truer measure of our ignorance. *Proc. Natl. Acad. Sci.* **105**, 6795–6796 (2008).

16. Szilagyi, A. & Zhang, Y. Template-based structure modeling of protein–protein interactions. *Curr. Opin. Struct. Biol.* **24**, 10–23 (2014).

17. Stelzl, U. *et al.* A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell* **122**, 957–968 (2005).

18. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **45**, D408–D414 (2017).

19. Collins, F. S. The Human Genome Project: Lessons from Large-Scale Biology. *Science* **300**, 286–290 (2003).

20. Rual, J.-F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005).

21. Baek, G. H. *et al.* Cdc48: A Swiss Army Knife of Cell Biology. *Journal of Amino Acids* https://www.hindawi.com/journals/jaa/2013/183421/ (2013) doi:10.1155/2013/183421.

22. Woodman, P. G. p97, a protein coping with multiple identities. *J. Cell Sci.* **116**, 4283–4290 (2003).

23. Barthelme, D. & Sauer, R. T. Origin and functional evolution of the Cdc48/p97/VCP AAA+ protein unfolding and remodeling machine. *J. Mol. Biol.* **428**, 1861–1869 (2016).

24. Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M. & Mann, M. A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines. *Mol. Cell. Proteomics MCP* **11**, (2012).

25. DeLaBarre, B., Christianson, J. C., Kopito, R. R. & Brunger, A. T. Central Pore Residues Mediate the p97/VCP Activity Required for ERAD. *Mol. Cell* **22**, 451–462 (2006).

26. Bodnar, N. O. & Rapoport, T. A. Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex. *Cell* **169**, 722-735.e9 (2017).

27. van den Boom, J. & Meyer, H. VCP/p97-Mediated Unfolding as a Principle in Protein Homeostasis and Signaling. *Mol. Cell* **69**, 182–194 (2018).

28. Hänzelmann, P. & Schindelin, H. The Interplay of Cofactor Interactions and Post-translational Modifications in the Regulation of the AAA+ ATPase p97. *Front. Mol. Biosci.* **4**, (2017).

29. Ye, Y., Tang, W. K., Zhang, T. & Xia, D. A Mighty 'Protein Extractor' of the Cell: Structure and Function of the p97/CDC48 ATPase. *Front. Mol. Biosci.* **4**, 39 (2017).

30. Chou, T.-F. *et al.* Specific Inhibition of p97/VCP ATPase and Kinetic Analysis Demonstrate Interaction between D1 and D2 ATPase Domains. *J. Mol. Biol.* **426**, 2886–2899 (2014).

31. Hänzelmann, P. & Schindelin, H. Structural Basis of ATP Hydrolysis and Intersubunit Signaling in the AAA+ ATPase p97. *Structure* **24**, 127–139 (2016).

32. Schuller, J. M., Beck, F., Lössl, P., Heck, A. J. R. & Förster, F. Nucleotide-dependent conformational changes of the AAA+ ATPase p97 revisited. *FEBS Lett.* **590**, 595–604 (2016).

33. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat. Cell Biol.* **14**, 117–123 (2012).

34. Meyer, H. & Weihl, C. C. The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis. *J. Cell Sci.* **127**, 3877–3883 (2014).

35. Stach, L. & Freemont, P. S. The AAA+ ATPase p97, a cellular multitool. *Biochem. J.* **474**, 2953–2976 (2017).

36. Kloppsteck, P., Ewens, C. A., Förster, A., Zhang, X. & Freemont, P. S. Regulation of p97 in the ubiquitin–proteasome system by the UBX protein-family. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1823**, 125–129 (2012).

37. Rezvani, K. UBXD Proteins: A Family of Proteins with Diverse Functions in Cancer. *Int. J. Mol. Sci.* **17**, 1724 (2016).

38. Ladanyi, M. *et al.* The der(17)t(X;17)(p11;q25) of human alveolar soft part sarcoma fuses the TFE3 transcription factor gene to ASPL , a novel gene at 17q25. *Oncogene* **20**, 48–57 (2001).

39. Aisner, S. C., Beebe, K., Blacksin, M., Mirani, N. & Hameed, M. Primary alveolar soft part sarcoma of fibula demonstrating ASPL–TFE3 fusion: a case report and review of the literature. *Skeletal Radiol.* **37**, 1047–1051 (2008).

40. Sood, S. *et al.* Imaging features of primary and metastatic alveolar soft part sarcoma: single institute experience in 25 patients. *Br. J. Radiol.* **87**, (2014).

41. Ishiguro, N. & Yoshida, H. ASPL-TFE3 Oncoprotein Regulates Cell Cycle Progression and Induces Cellular Senescence by Up-Regulating p21. *Neoplasia N. Y. N* **18**, 626–635 (2016).

42. Tsuda, M. *et al.* TFE3 fusions activate MET signaling by transcriptional up-regulation, defining another class of tumors as candidates for therapeutic MET inhibition. *Cancer Res.* **67**, 919–929 (2007).

43. Arumughan, A. *et al.* Quantitative interaction mapping reveals an extended UBX domain in ASPL that disrupts functional p97 hexamers. *Nat. Commun.* **7**, 13047 (2016).

44. Watts, G. D. J. *et al.* Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat. Genet.* **36**, 377–381 (2004).

45. Zhang, X. *et al.* Altered cofactor regulation with disease-associated p97/VCP mutations. *Proc. Natl. Acad. Sci.* **112**, E1705–E1714 (2015).

46. Ritz, D. *et al.* Endolysosomal sorting of ubiquitylated caveolin-1 is regulated by VCP and UBXD1 and impaired by VCP disease mutations. *Nat. Cell Biol.* **13**, 1116–1123 (2011).

47. Rijal, R. *et al.* Mutant p97 exhibits species-specific changes of its ATPase activity and compromises the UBXD9-mediated monomerisation of p97 hexamers. *Eur. J. Cell Biol.* **95**, 195–207 (2016).

48. Anderson, D. J. *et al.* Targeting the AAA ATPase p97 as an Approach to Treat Cancer through Disruption of Protein Homeostasis. *Cancer Cell* **28**, 653–665 (2015).

49. Chapman, E., Fry, A. N. & Kang, M. The complexities of p97 function in health and disease. *Mol BioSyst* **7**, 700–710 (2011).

50. Franz, A., Ackermann, L. & Hoppe, T. Create and preserve: Proteostasis in development and aging is governed by Cdc48/p97/VCP. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1843**, 205–215 (2014).

51. Buntru, A., Trepte, P., Klockmeier, K., Schnoegl, S. & Wanker, E. E. Current Approaches Toward Quantitative Mapping of the Interactome. *Front. Genet.* **7**, (2016).

52. Miura, K. An Overview of Current Methods to Confirm Protein-Protein Interactions. *Protein Pept. Lett.* **25**, 728–733 (2018).

53. Choi, S. G. *et al.* Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.* **10**, 1–13 (2019).

54. Trepte, P. *et al.* LuTHy: a double-readout bioluminescence-based two-hybrid technology for quantitative mapping of protein–protein interactions in mammalian cells. *Mol. Syst. Biol.* **14**, (2018).

55. Ellis, J. D., Llères, D., Denegri, M., Lamond, A. I. & Cáceres, J. F. Spatial mapping of splicing factor complexes involved in exon and intron definition. *J. Cell Biol.* **181**, 921–934 (2008).

56. Zhang, W. J. & Wu, J. Y. Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol. Cell. Biol.* **16**, 5400–5408 (1996).

57. Cass, D. M. & Berglund, J. A. The SF3b155 N-Terminal Domain Is a Scaffold Important for Splicing. *Biochemistry* **45**, 10092–10101 (2006).

58. Loerch, S. & Kielkopf, C. L. Unmasking the U2AF homology motif family: a bona fide protein–protein interaction motif in disguise. *RNA* **22**, 1795–1807 (2016).

59. Subbarao Sreedhar, A., Kalmár, É., Csermely, P. & Shen, Y.-F. Hsp90 isoforms: functions, expression and clinical importance. *FEBS Lett.* **562**, 11–15 (2004).

60. Stepanyuk, G. A. *et al.* UHM–ULM interactions in the RBM39–U2AF65 splicing-factor complex. *Acta Crystallogr. Sect. Struct. Biol.* **72**, 497–511 (2016).

61. Loerch, S., Maucuer, A., Manceau, V., Green, M. R. & Kielkopf, C. L. Cancer-relevant Splicing Factor CAPERa Engages the Essential Splicing Factor SF3b155 in a Specific Ternary Complex. *J. Biol. Chem.* **289**, 17325–17337 (2014).

62. Lin, C.-L., Taggart, A. J. & Fairbrother, W. G. RNA structure in splicing: An evolutionary perspective. *RNA Biol.* **13**, 766–771 (2016).

63. Arumughan, A. A quantitative AAA+ ATPase Interactome Reveals an Evolutionary Conserved Mechanism of ASPL-Mediated Remodeling of p97 Hexamers. (FU Berlin, 2015).

64. Gama-Carvalho, M., Carvalho, M. P., Kehlenbach, A., Valcárcel, J. & Carmo-Fonseca, M. Nucleocytoplasmic Shuttling of Heterodimeric Splicing Factor U2AF. *J. Biol. Chem.* **276**, 13104–13112 (2001).

65. Kielkopf, C. L., Lücke, S. & Green, M. R. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.* **18**, 1513–1526 (2004).

66. Tari, M. *et al.* U2AF65 assemblies drive sequence-specific splice site recognition. *EMBO Rep.* **20**, e47604 (2019).

67. Manceau, V. *et al.* Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF56. *FEBS J.* **273**, 577–587 (2006).

68. Rain, J. C., Rafi, Z., Rhani, Z., Legrain, P. & Krämer, A. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. *RNA* **4**, 551–565 (1998).

69. Kielkopf, C. L., Rodionova, N. A., Green, M. R. & Burley, S. K. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**, 595–605 (2001).

70. Agrawal, A. A. *et al.* An extended U2AF65–RNA-binding domain recognizes the 3′ splice site signal. *Nat. Commun.* **7**, 10950 (2016).

71. Zhao, S., Zhang, Y., Gamini, R., Zhang, B. & Schack, D. von. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Sci. Rep.* **8**, 1–12 (2018).

72. Shaw, G., Morse, S., Ararat, M. & Graham, F. L. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J.* **16**, 869–871 (2002).

73. Thomas, P. & Smart, T. G. HEK293 cell line: A vehicle for the expression of recombinant proteins. *J. Pharmacol. Toxicol. Methods* **51**, 187–200 (2005).

74. Stepanenko, A. A. & Dmitrenko, V. V. HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* **569**, 182–190 (2015).

75. Girardot, M. *et al.* SOX9 has distinct regulatory roles in alternative splicing and transcription. *Nucleic Acids Res.* **46**, 9106–9118 (2018).

76. Han, H. *et al.* Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Mol. Cell* **65**, 539-553.e7 (2017).

77. Chen, S.-Y., Li, C., Jia, X. & Lai, S.-J. Sequence and Evolutionary Features for the Alternatively Spliced Exons of Eukaryotic Genes. *Int. J. Mol. Sci.* **20**, 3834 (2019).

78. Akman, H. B. & Erson-Bensan, A. E. Alternative polyadenylation and its impact on cellular processes. *MicroRNA Shariqah United Arab Emir.* **3**, 2–9 (2014).

79. Reyes, A. & Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* **46**, 582–592 (2018).

80. de Klerk, E. & 't Hoen, P. A. C. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet.* **31**, 128–139 (2015).

81. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed. Rep.* **3**, 152–158 (2015).

82. Cui, Y., Cai, M. & Stanley, H. E. Comparative Analysis and Classification of Cassette Exons and Constitutive Exons. *BioMed Res. Int.* **2017**, 1–8 (2017).

83. Sugnet, C. W., Kent, W. J., Ares Jr., M. & Haussler, D. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 66–77 (2004).

84. Sammeth, M., Foissac, S. & Guigó, R. A General Definition and Nomenclature for Alternative Splicing Events. *PLOS Comput. Biol.* **4**, e1000147 (2008).

85. Papasergi, M. M., Patel, B. R. & Tall, G. G. The G Protein a Chaperone Ric-8 as a Potential Therapeutic Target. *Mol. Pharmacol.* **87**, 52–63 (2015).

86. Chen, S.-Y., Li, C., Jia, X. & Lai, S.-J. Sequence and Evolutionary Features for the Alternatively Spliced Exons of Eukaryotic Genes. *Int. J. Mol. Sci.* **20**, (2019).

87. Ye, Z. *et al.* Computational analysis reveals a correlation of exon-skipping events with splicing, transcription and epigenetic factors. *Nucleic Acids Res.* **42**, 2856–2869 (2014).

88. Hastings, M. L., Allemand, E., Duelli, D. M., Myers, M. P. & Krainer, A. R. Control of Pre-mRNA Splicing by the General Splicing Factors PUF60 and U2AF65. *PLoS ONE* **2**, e538 (2007).

89. Královičová, J. *et al.* PUF60-activated exons uncover altered 3′ splice-site selection by germline missense mutations in a single RRM. *Nucleic Acids Res.* **46**, 6166–6187 (2018).

90. Kim, E., Goren, A. & Ast, G. Alternative splicing: current perspectives. *BioEssays* **30**, 38–47 (2008).

91. Hentze, M. W. & Kulozik, A. E. A Perfect Message: RNA Surveillance and Nonsense-Mediated Decay. *Cell* **96**, 307–310 (1999).

92. Frischmeyer, P. A. & Dietz, H. C. Nonsense-Mediated mRNA Decay in Health and Disease. *Hum. Mol. Genet.* **8**, 1893–1900 (1999).

93. Pervouchine, D. *et al.* Novel autoregulatory cases of alternative splicing coupled with nonsense-mediated mRNA decay. *bioRxiv* 464404 (2018) doi:10.1101/464404.

94. Clark, F. Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum. Mol. Genet.* **11**, 451–464 (2002).

95. Smith, L. D., Lucas, C. M. & Eperon, I. C. Intron Retention in the Alternatively Spliced Region of RON Results from Weak 3' Splice Site Recognition. *PLoS ONE* **8**, e77208 (2013).

96. Abramowicz, A. & Gos, M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* **59**, 253–268 (2018).

97. Koren, E., Lev-Maor, G. & Ast, G. The Emergence of Alternative 3' and 5' Splice Site Exons from Constitutive Exons. *PLOS Comput. Biol.* **3**, e95 (2007).

98. Pacheco, T. R., Coelho, M. B., Desterro, J. M. P., Mollet, I. & Carmo-Fonseca, M. In Vivo Requirement of the Small Subunit of U2AF for Recognition of a Weak 3' Splice Site. *Mol. Cell. Biol.* **26**, 8183–8190 (2006).

99. Tang, W. K. & Xia, D. Mutations in the Human AAA+ Chaperone p97 and Related Diseases. *Front. Mol. Biosci.* **3**, (2016).

100. Boute, N., Jockers, R. & Issad, T. The use of resonance energy transfer in high-throughput screening: BRET versus FRET. *Trends Pharmacol. Sci.* **23**, 351–354 (2002).

101. Kovács, I. A. *et al.* Network-based prediction of protein interactions. *Nat. Commun.* **10**, 1240 (2019).

102. Howard, J. M. & Sanford, J. R. The RNAissance family: SR proteins as multifaceted regulators of gene expression: The RNAissance family. *Wiley Interdiscip. Rev. RNA* **6**, 93–110 (2015).

103. Manley, J. L. & Krainer, A. R. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* **24**, 1073–1074 (2010).

104. Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**, 15–27 (2009).

105. Shepard, P. J. & Hertel, K. J. The SR protein family. *Genome Biol.* **10**, 242 (2009).

106. Jeong, Sunjoo. SR Proteins: Binders, Regulators, and Connectors of RNA. *Mol. Cells* **40**, 1–9 (2017).

107. Hammarskjold, M.-L. & Rekosh, D. SR proteins: To shuttle or not to shuttle, that is the question. *J. Cell Biol.* **216**, 1875–1877 (2017).

108. Sanford, J. R. & Bruzik, J. P. Regulation of SR protein localization during development. *Proc. Natl. Acad. Sci.* **98**, 10184–10189 (2001).

109. Long, Y. *et al.* Distinct mechanisms govern the phosphorylation of different SR protein splicing factors. *J. Biol. Chem.* **294**, 1312–1327 (2019).

110. Twyffels, L., Gueydan, C. & Kruys, V. Shuttling SR proteins: more than splicing factors: Shuttling SR proteins. *FEBS J.* **278**, 3246–3255 (2011).

111. Sapra, A. K. *et al.* SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo. *Mol. Cell* **34**, 179–190 (2009).

112. Sanford, J. R. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* **18**, 755–768 (2004).

113. Müller-McNicoll, M. *et al.* SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.* **30**, 553–566 (2016).

114. Singh, G. *et al.* The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus. *Cell* **151**, 750–764 (2012).

115. Botti, V. *et al.* Cellular differentiation state modulates the mRNA export activity of SR proteins. *J. Cell Biol.* **216**, 1993–2009 (2017).

116. Ballarino, M. *et al.* TAF15 is important for cellular proliferation and regulates the expression of a subset of cell cycle genes through miRNAs. *Oncogene* **32**, 4646–4655 (2013).

117. Hershberger, C. E. *et al.* The Role of LUC7L2 in Splicing and MDS. *Blood* **128**, 5504–5504 (2016).

118. Sasahara, K. *et al.* Molecular cloning and expression analysis of a putative nuclear protein, SR-25. *Biochem. Biophys. Res. Commun.* **269**, 444–450 (2000).

119. McCallum, S. A. *et al.* Structure of SAP18: A Ubiquitin Fold in Histone Deacetylase Complex Assembly. *Biochemistry* **45**, 11974–11982 (2006).

120. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).

121. Kroiss, M. *et al.* Evolution of an RNP assembly system: A minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster. *Proc. Natl. Acad. Sci.* **105**, 10045–10050 (2008).

122. Liao, Y., Tong, L., Tang, L. & Wu, S. The role of cold-inducible RNA binding protein in cell stress response. *Int. J. Cancer* **141**, 2164–2173 (2017).

123. Kaneko, Y., Kimura, T., Kishishita, M., Noda, Y. & Fujita, J. Cloning of apg-2 encoding a novel member of heat shock protein 110 family. *Gene* **189**, 19–24 (1997).

124. Zuehlke, A. D., Beebe, K., Neckers, L. & Prince, T. Regulation and function of the human HSP90AA1 gene. *Gene* **570**, 8–16 (2015).

125. Li, H. & Outten, C. E. Monothiol CGFS Glutaredoxins and BolA-like Proteins: [2Fe-2S] Binding Partners in Iron Homeostasis. *Biochemistry* **51**, 4377–4389 (2012).

126. Kasai, T. *et al.* Solution structure of a BolA-like protein from Mus musculus. *Protein Sci. Publ. Protein Soc.* **13**, 545–548 (2004).

127. Bektas, M. *et al.* A novel acylglycerol kinase that produces lysophosphatidic acid modulates cross talk with EGFR in prostate cancer cells. *J. Cell Biol.* **169**, 801–811 (2005).

128. Krone, P. H. & Sass, J. B. Hsp 90a and Hsp 90β Genes Are Present in the Zebrafish and Are Differentially Regulated in Developing Embryos. *Biochem. Biophys. Res. Commun.* **204**, 746–752 (1994).

129. Lele, Z. *et al.* Disruption of Zebrafish Somite Development by Pharmacologic Inhibition of Hsp90. *Dev. Biol.* **210**, 56–70 (1999).

130. Sidera, K., Samiotaki, M., Yfanti, E., Panayotou, G. & Patsavoudi, E. Involvement of Cell Surface HSP90 in Cell Migration Reveals a Novel Role in the Developing Nervous System. *J. Biol. Chem.* **279**, 45379–45388 (2004).

131. Whisenant, T. C. *et al.* The Activation-Induced Assembly of an RNA/Protein Interactome Centered on the Splicing Factor U2AF2 Regulates Gene Expression in Human CD4 T Cells. *PLOS ONE* **10**, e0144409 (2015).

132. Lu, Y. Y. *et al.* Molecular chaperone HSP90 regulation of the alternative splicing of numb as a clue for HCC progression. *J. Clin. Oncol.* **32**, e15114–e15114 (2014).

133. Ferraldeschi, R. *et al.* Second-Generation HSP90 Inhibitor Onalespib Blocks mRNA Splicing of Androgen Receptor Variant 7 in Prostate Cancer Cells. *Cancer Res.* **76**, 2731–2742 (2016).

134. Langer, T. Intracellular localization of the 90 kDA heat shock protein (HSP90a) determined by expression of a EGFP–HSP90a-fusion protein in unstressed and heat stressed 3T3 cells. *Cell Biol. Int.* **27**, 47–52 (2003).

135. Corsini, L. *et al.* U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat. Struct. Mol. Biol.* **14**, 620–629 (2007).

136. Huang, S.-C. *et al.* Protein 4.1R Exon 16 3′ Splice Site Activation Requires Coordination among TIA1, Pcbp1, and RBM39 during Terminal Erythropoiesis. *Mol. Cell. Biol.* **37**, e00446-16, /mcb/37/9/e00446-16.atom (2017).

137. Mai, S. *et al.* Global regulation of alternative RNA splicing by the SR-rich protein RBM39. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1859**, 1014–1024 (2016).

138. Shao, C. *et al.* Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.* **21**, 997–1005 (2014).

139. Golling, G. *et al.* Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat. Genet.* **31**, 135–140 (2002).

140. Maciejewski, J. P. & Padgett, R. A. Defects in spliceosomal machinery: a new pathway of leukaemogenesis. *Br. J. Haematol.* **158**, 165–173 (2012).

141. Ruskin, B., Zamore, P. D. & Green, M. R. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 207–219 (1988).

142. Fu, X.-D. & Ares, M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* **15**, 689–701 (2014).

143. Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J. & Fairbrother, W. G. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci.* **108**, 11093–11098 (2011).

144. Sutandy, F. X. R. *et al.* In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome Res.* **28**, 699–713 (2018).

145. Partridge, J. J., Lopreiato, J. O., Latterich, M. & Indig, F. E. DNA Damage Modulates Nucleolar Interaction of the Werner Protein with the AAA ATPase p97/VCP. *Mol. Biol. Cell* **14**, 4221–4229 (2003).

146. Song, C. *et al.* Nucleocytoplasmic shuttling of valosin-containing protein (VCP/p97) regulated by its N domain and C-terminal region. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1853**, 222–232 (2015).

147. Koike, M. *et al.* Valosin-containing Protein (VCP) in Novel Feedback Machinery between Abnormal Protein Accumulation and Transcriptional Suppression. *J. Biol. Chem.* **285**, 21736–21749 (2010).

148. Xia, D., Tang, W. K. & Ye, Y. Structure and function of the AAA+ ATPase p97/Cdc48p. *Gene* **583**, 64–77 (2016).

149. Wu, T. & Fu, X.-D. Genomic functions of U2AF in constitutive and regulated splicing. *RNA Biol.* **12**, 479–485 (2015).

150. Tettamanzi, M. C., Yu, C., Bogan, J. S. & Hodsdon, M. E. Solution structure and backbone dynamics of an N-terminal ubiquitin-like domain in the GLUT4-regulating protein, TUG. *Protein Sci.* **15**, 498–508 (2006).

151. Alexandru, G. *et al.* UBXD7 Binds Multiple Ubiquitin Ligases and Implicates p97 in HIF1a Turnover. *Cell* **134**, 804–816 (2008).

152. Meyer, H. H., Kondo, H. & Warren, G. The p47 co-factor regulates the ATPase activity of the membrane fusion protein, p97. *FEBS Lett.* **437**, 255–257 (1998).

153. Uchiyama, K. & Kondo, H. p97/p47-Mediated biogenesis of Golgi and ER. *J. Biochem. (Tokyo)* **137**, 115–119 (2005).

154. Kimonis, V. Inclusion Body Myopathy with Paget Disease of Bone and/or Frontotemporal Dementia. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).

155. Nalbandian, A. *et al.* The Homozygote VCPR155H/R155H Mouse Model Exhibits Accelerated Human VCP-Associated Disease Pathology. *PLoS ONE* **7**, e46308 (2012).

156. Chang, Y.-C. *et al.* Pathogenic VCP/TER94 Alleles Are Dominant Actives and Contribute to Neurodegeneration by Altering Cellular ATP Level in a Drosophila IBMPFD Model. *PLoS Genet.* **7**, e1001288 (2011).

157. Halawani, D. Hereditary Inclusion Body Myopathy-Linked p97/VCP Mutations in the NH2 Domain and the D1 Ring Modulate p97/VCP ATPase Activity and D2 Ring Conformation. *Mol. Cell. Biol.* **29**, 4484–4494 (2009).

158. Rycenga, H. B., Wolfe, K. B., Yeh, E. S. & Long, D. T. Uncoupling of p97 ATPase activity has a dominant negative effect on protein extraction. *Sci. Rep.* **9**, 10329 (2019).

159. Zhang, X. *et al.* Altered cofactor regulation with disease-associated p97/VCP mutations. *Proc. Natl. Acad. Sci.* **112**, E1705–E1714 (2015).

160. Her, N.-G. *et al.* p97 Composition Changes Caused by Allosteric Inhibition Are Suppressed by an On-Target Mechanism that Increases the Enzyme's ATPase Activity. *Cell Chem. Biol.* **23**, 517–528 (2016).

161. Wang, T. *et al.* Pathogenic Mutations in the Valosin-containing Protein/p97(VCP) N-domain Inhibit the SUMOylation of VCP and Lead to Impaired Stress Response. *J. Biol. Chem.* **291**, 14373–14384 (2016).

162. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* **14**, 417–419 (2017).

163. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

164. Sterne-Weiler, T., Weatheritt, R. J., Best, A. J., Ha, K. C. H. & Blencowe, B. J. Efficient and Accurate Quantitative Profiling of Alternative Splicing Patterns of Any Complexity on a Laptop. *Mol. Cell* **72**, 187-200.e6 (2018).

165. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

166. Cretu, C. *et al.* Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds. *Mol. Cell* **70**, 265-273.e8 (2018).

9. Acknowledgements

Completing this work was a mentally challenging task which I would not have been able to accomplish without the support of a great number of people. To me this experience was comparable to a long marathon run with no close sight of a finish line… and I do not even like jogging.

Here I want to express my gratitude to those who made me keep running and seeking that finish line.

***

First of all, I would like to thank my supervisor, Erich E. Wanker, who not only thought me what persistence and scientific passion are, but he always managed to find the right advice to guide me through any of the obstacles I was facing during my project. He never stopped believing in my potential and every conversation we had was encouraging and motivating.

***

Philipp Trepte, there are not enough ways to tell how grateful I am, that we were able to work together, and I had the chance to learn from you. You have showed me what it takes to pursue scientific perfection and find a way to solve a problem. This is the best lesson an unexperienced scientist can ask for. You were and still are a great mentor and friend!

***

Anup Arumughan, you are the person without whose hard work, I would not have had the opportunity to get involved in this exciting project. Your professionalism and knowledge were a great inspiration to me!

Anne Ast, thank you so much for every conversation and discussion we had on all the different topics. You had given me the best advises at the times I needed them most and thought me the importance of preserving mental balance.

Christopher Secker, I am more than happy to have been sharing with you not only an office and a lab bench, but so many fruitful discussions. I have always admired your creativity… and envied you table tennis skills!

***

Gizem Inak, thank you for all your moral support through numerous office decorations, shared fun stories and good times at the Gesundbrunnen Center! Your positive energy has cheered me up on multiple occasions.

***

Maria Lucia Pigazzini, thank you for making "Under Pressure" the official miniprep hymn, sharing your knowledge on heat shock experiments with me and convincing me that submerging a sealed microtiter plate in a water bath should shock me less than the cells in that plate… Thank you for all the conversations we had on science and life! I am lucky and grateful to have met such an exceptional colleague and a true friend!

***

Sasa Petrovic, our mutual admiration for p97 and volleyball has helped me find a great new friend and colleague, for which I am extremely grateful!

***

Lynn van der Beek, thank you for your hard work, persistence and patience under my supervision during your Bachelor thesis!

***

Alexander Neumann, thank you so much for contributing to my project with your bioinformatics skills and knowledge! Our collaboration was without no exaggeration crucial for my progress!

***

Oliver Polzer, I was tremendously lucky to have such an exceptionally talented student under my supervision! Thank you for your devotion and high levels of motivation during your traineeship! It was a pleasure working with you!

Lydia Brusendorf, Hannah Niederlechner, Sabrina Golusik, Martina Zenkner, Stephanie Beetz and Nancy Neuendorf, thank you so much for support on all technical matters in the lab! I would not have managed to finish this work without your help!

***

Marta Dominguez Martinez, thank you for all the Palé Alé evenings we had and I am looking forward to the many to come!

***

Sigrid Schnögl and Erika Pisch, thank you for your support on all possible administration matters! Without it I literally would have been lost!

***

All members of the Wanker lab, thank you for the acceptance through the years! I appreciate how lucky I am to have had the opportunity of working in such a friendly and supportive team!

***

Kalinka Patleeva, thank you so much for the initial push you gave me in pursuing a scientific career! Your teaching was a great source of inspiration and support through the years!

***

Katalin Vasileva, thank you for being my partner in crime ever since the very beginning! Thank you for the shared hours in the library and all the delicious meals we had! Thank you for knowing me so well!

***

Marcel Sperling, I will never be able to express in just a few lines, how grateful I am that you were by my side during that journey! Thank you for being the person who would not let me quit and assure me I have the strength to achieve my goals! Thank you for knowing how to fix me, how to get me out of the hours of self-doubt and simply for being there!

My family and especially my mother, thank you for letting me make my own mistakes, take risks and make bad choices, walk my own path and follow my goals! You are that never ending source of energy that helps me come further! Thank you for being my best friend!