TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl Für Experimentelle Genetik

Molecular dissection of SON's interactome reveals the molecular link to splicing

Matteo Gorza

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende: Univ.-Prof. A. Schnieke, Ph.D

Prüfer der Dissertation:
1. apl.Prof. Dr. J. Adamski
2. Univ.-Prof. Dr. B. Küster
3. Univ.-Prof. Dr. M. Ueffing
(Eberhard-Karls-Universität Tübingen)

Die Dissertation wurde am 15.10.2013 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 21.07.2014 angenommen.

Frangar, non flectar

Table of contents

Table of Contents

Table of Contents

Abbreviations

AP	Affinity purification
APS	Ammonium Persulphate
BSA	Bovine serum albumin
cDNA	Complementary DNA
Da	Dalton
DAPI	4',6-diamidino-2-phenylindole
ddH2O	Ultra-pure water
ddNTPs	Dideoxynucleotides
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
eIP	Immunoprecipitation of the endogenous protein
ER	Endoplasmatic reticulum
Exp	Experiment
FBS	Fetal bovine serum
GFP	Green fluorescent protein
G-patch	Glycine-rich domain
GST	Glutathione-S-transferase
HEK293T	Human embryonic kidney cells
HeLa	Human cervical cancer cells
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
IgG	Immunoglobulin G
IPTG	Isopropyl-β-D-thiogalactoside
KD	Knock-down

LB-	Luria-Bertani
LC-MS/MS	Liquid chromatography mass spectrometry analysis
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein particle
MS	Mass spectrometry.
MS/MS	Tandem mass spectrometry
NaCl	Sodium chloride
NLS	Nuclear localization signal
NP40	Nonidet P-40
IGCs	Interchromatine granule clusters
dsRBD	Double-stranded RNA-binding domain
PBS	Phosphate buffered saline
PEI	Polyethylenimine
PES	Polyethersulfon
pre-mRNA	Immature mRNA
PVDF	Polyvinylidene difluoride
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SF-AP	Single Flag-affinity purification
SF-TAP	Strep-Flag-tandem affinity purification
SILAC	Stable isotope labelling by amino acids in cell culture
siRNA	Short interfering RNA
snRNPs	Small nuclear ribonucleoproteins
SONc	C-terminal portion of SON (amino acids 2079-2426)
SONc1	Domain 1 of SONc (amino acids 2079-2234)
SONc12	Combination of SONc1 and SONc2
SONc2	Domain 2 of SONc (amino acids 2319-2346)
SONc23	Combination of SONc2 and SONc3
SONc3	Domain 3 of SONc (amino acids 2261-2426)
TBS	Tris-buffered saline

TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
v/v	Volume per volume
w/v	Weight per volume
Y2H	Yeast two-hybrid

Summary

In the recent years, protein-protein interaction studies have revealed the link between impaired cellular mechanisms and protein complexes involved. Given that proteins are no longer intended as independent entities but as members of a macro protein complex, many methods have been developed to identify the single components of these protein complexes. The aim of this study was to study more in detail the function of the SON DNA-binding protein, pointing particular attention to its newly described role in mRNA splicing. In order to identify new interacting partners and cellular pathways involving the SON protein, different methods were applied.

Four methods to perform protein interaction analysis have been selected, considering carefully the strength, the drawbacks and the limitations of each technique. Three of these methods were intended to be used for protein complexes isolation, performing affinity purification with or without tag. In this group were included the SF-tag affinity purification, the GST pull-down and the immunoprecipitation of the endogenous protein complexes. These techniques, based on the binding by affinity of the protein of interest to a resin or antibody, allowed the identification of proteins belonging to the SON complexes. The fourth technique used, the yeast two-hybrid, differently from the previous techniques is suitable to determine exclusively binary interacting partners of SON. For each technique have been performed preliminary experiments to evaluate the consistence of the produced SON baits with the downstream applications. These pilot experiments identified many nuclear and ribonucleoproteins. These findings were coherent with the proprieties already known for SON, the nuclear localization and the presence of RNA-binding domains in its structure. Subsequently to the establishment of the experimental protocols, the analysis of SON protein complexes and dissection of its role were performed, at that time still unknown. During the course of this study, the C-terminal fragment of SON was reported as protein linked to splicing activity. This fragment was therefore selected to perform SF-tag affinity purification, GST pull-down and yeast two-hybrid. Immunoprecipitation of endogenous SON protein was also performed. Protein complex identifications were performed by mass spectrometric analysis and, if applicable, in combination with the SILAC labeling and quantification. By applying these four different methods, it could be demonstrated the meaningful combination of these different techniques, obtaining valuable

information on the nature of the single interactions. Additionally, upon the combination of the SONc sequence dissection in discrete sub-domains and their cellular localization, was possible to newly identify an atypical nuclear localization signal (NLS) at the very end of the SON sequence.

Protein complex analysis identified many proteins involved in splicing machineries. In particular the splicing-related proteins were restricted to two major splicing sub-machineries, the spliceosome C and the mRNP/mRNP-granule complexes. Further network analysis on the identified interactors of SON revealed high interconnectivity between SON and the detected splicing sub-machineries, providing evidences of the key role of SON protein in splicing. Following SON protein knock-down, protein complexes formed by representative member of the two previously mentioned splicing sub-machineries (DDX5 and DHX9), presented significant variations on complex composition. More in detail, for both DDX5 and DHX9 protein complexes, the SON protein depletion resulted in an altered content of proteins connected with the cytoskeleton and the mRNA binding. These findings are supporting the role of SON in splicing through the functional organization of the spliceosome C and the mRNP/mRNP-granule complexes.

Zusammenfassung

Protein-Protein Interaktionsstudien haben in den letzten Jahren den Zusammenhang zwischen Störungen Zellulärer Funktionen und den damit assoziierten Proteinkomplexen aufgezeigt. In Anbetracht der Erkenntnis dass Proteine nicht mehr als unabhängige Einheiten sondern als Teil von Makroproteinkomplexen zu betrachten sind, wurden viele Methoden entwickelt um die einzelnen Komponenten von Proteinkomplexen zu identifizieren. Das Ziel dieser Studie war die Funktion des SON-DNA-binding proteins genauer zu untersuchen, insbesondere in Hinblick auf seine kürzlich beschriebene Rolle beim mRNA splicing. Um neue Interaktionspartner und zelluläre Pathways in denen das SON Protein involviert ist zu finden, wurden unterschiedliche Methoden angewendet. Vier Interaktionsanalysemethoden wurden unter Berücksichtigung der Stärken, Schwächen und Limitationen der einzelnen Methoden ausgewählt. Drei dieser Methoden beruhen auf der Isolation von Proteinkomplexen und beinhalten Affinitätsaufreinigung mit oder ohne Tags. Diese Methoden umfassen die SF-tag Affinitätsaufreinigung, GST pull-down und

Immunopräzipitation des endogenen Proteinkomplexes. Durch die Anwendung dieser Methoden, welche auf der spezifischen Bindung des untersuchten Proteins an eine Affinitätsmatrix oder einen Antikörper beruhen, wurden Proteine der SON-Komplexe identifiziert. Im Gegensatz dazu ist die vierten Technik, das yeast-two-hybrid, ausschließlich dazu geeignet Proteine mit binären Interaktionen zu SON zu finden. Für jede Technik wurde die Kompatibilität der generierten SON baits mit der jeweiligen downstream Anwendung durch Vorexperimente bestätigt. In dieser Pilotstudie wurden bereits viele nukleäre Proteine und Ribonukleoproteine als Interaktionspartner identifiziert. Diese Ergebnisse passen zu den bereits bekannten Eigenschaften von SON – insbesondere seiner nukleären Lokalisation und den RNA bindenden Domänen in seiner Struktur. Nach der Etablierung der Techniken wurden die SON-Komplexe einer feineren Untersuchung unterzogen um die Funktion des Proteins aufzuklären, welche zu dieser Zeit noch unbekannt war. Während des Verlaufs der Studie zeigt eine andere Gruppe dass der C-terminale Teil von SON splicing beeinflussen kann. Daher wurde dieses Fragment mit SF-tag Affinitätsaufreinigung, GST pull-down und yeast-two-hybrid untersucht. Zusätzlich wurde das endogene Protein mit Immunopräzipitation analysiert. Identifizierung der Komplexproteine erfolgt durch massenspektrometrische Analyse in Kombination mit Quantifizierung durch SILAC labeling wenn möglich. Durch die Anwendung aller vier Techniken konnte gezeigt werden dass diese sich sinnvoll ergänzen und in Kombination wertvolle Informationen über die Natur der einzelnen Interaktionen liefern. Darüber hinaus konnte durch die Trunkation des C-terminalen Teils von SON in diskrete Subdomänen und deren Lokalisation ein atypisches nukleäres Lokalisationssignal (NLS) am äußersten C-terminalen Ende identifiziert werden. Durch die Proteinkomplex Analyse wurden zahlreiche Proteine des splicing Apparates identifiziert. Tatsächlich waren diese splicing assoziierten Proteine Bestandteil zweier bedeutender splicing Unterkomplexe – dem spliceosome C und der mRNP/mRNP-granule Komplexe. Weiterführende Netzwerkanalyse der identifizierten Interaktoren zeigte einen hohen Grad der Vernetzung zwischen SON und den erwähnten splicing Unterkomplexen. Diese Ergebnisse implizieren eine Schlüsselrolle für SON im splicing. In SON knock-down Experimenten fanden wir eine wesentliche Änderung in der Zusammensetzung der Komplexe welche von zwei repräsentativen Proteinen der erwähnten splicing Unterkomplexe (DDX5 und DHX9) gebildet werden. Sowohl für den DDX5 als auch für den DHX9 Komplex manifestierte

sich dies in der Abnahme der Menge von Zytoskelett- und mRNA bindender Proteine nach SON Depletion. Diese Ergebnisse unterstützen die Rolle von SON im splicing durch die funktionelle Organisation des spliceosome C und der mRNP/mRNP-granule Komplexe

1 Introduction

The development of molecular biology opened a new insight on correlation between disease and mechanism underlying. Since decades animal models have been used to identify the roles that specific proteins could have in disease occurrence, providing a powerful tool for medical research. The basic principles of animal models in specific pathologies are the removal/inactivation (knock-out/-down) or the addition (knock-in) of specific gene products, including wild-type, mutant or truncated forms of the protein of interest [1-2]. These models have been helpful for morphological and physiological studies, like the analysis of the structure of a specific organ or the evaluation of the preserved function of a cellular structure after genetic manipulation [3]. The major drawback of animal models (in particular mice models) is their generation time, since many months are required from the design of the suitable gene targeting strategy until the generation of homozygous/heterozygous animals [4-5].

Although the single-protein targeted approach has been largely used in the past, resulting in the production of precious drugs against diseases, like cancer (Imatinib against leukemia [6] and Trastuzumab against breast cancer [7]), recently the focus has been moved to protein complexes.

With the development of new research branches such as proteomics and metabolomics, it appeared more evident that proteins are not self-standing entities, but parts of sub-cellular machineries involved in different and various functions. Proteomics in particular has the purpose to describe the proteins involved in a cell's function, including cross interactions, regulations and localization [8]. Whereas the transcriptome (mRNA, rRNA, tRNA, and other non-coding RNA) can give just a picture of the current cellular status and explaining differential expression patterns, proteins are the real effectors of cellular activities. Changes in posttranslational modifications (e.g. phosphorylation), subcellular localization and binding partners are making the proteome independent from the transcriptome profiling [9-10]

Protein complexes can be intended as functional modules dependent by the single protein-protein interactions. This modular organization can easily explain how a variation at the protein level could interfere with complex function, since modules are tightly interconnected and not self-dependent. Furthermore, a single protein can belong to

multiple complexes or connecting them, producing a more elaborate network overview. For instance the super-complex involved in DNA control during mitosis is formed by modules assembling the mitotic spindle, checking the correct chromosomal alignment and promoting the cell-cycle progression [11]

1.1 Identification and characterization of protein networks

Protein networks are the simplified picture of sophisticated protein-protein interactions (PPIs) and dynamics occurring within protein complexes. Protein networks represent all proteins involved in a complex (or more) and the different connections (interactions) between them. The most important part for a correct and informative protein network is the selection of the appropriate dataset of PPIs: at the protein level, yeast two-hybrid and mammalian protein complexes purification techniques are the most used. The information obtained by these two techniques are drastically different but complementary. The yeast two-hybrid method allows the screening of binary interactions between the protein of interest (bait) against tissue-derived complete libraries of proteins (preys) in a relative short time. The major drawbacks are that parts of posttranslational modification do not occur (some phosphorylations and glycosylation), possibly resulting in the loss of interaction partners, and the chimerical constructs could create false positives. Even if the yeast two-hybrid data are including only binary binders, these findings could be useful when drawing a protein complex related to a protein of interest.

In vivo mammalian protein complexes are more closely to the cellular physiological environment and the most frequently used methods are affinity-based protein complex purifications that will be described later.

As reported by Deszo [12] , among the interactors of a target protein, more than half of them represent essential proteins. These are defined as essential when confirmed as part of an interactome using different baits belonging to the same complex. From this analysis, the proteins forming a functional complex are producing the essentiality of a network despite the single protein function and interaction. The full proteome of a cell can be distinguished in "role-dependent categories" according to the functions and roles of the single protein components (e.g. protein synthesis or RNA metabolism), and subsequently networks could be used to predict an unknown protein function according to the neighbor counting index

(NC). This index is a ratio between the interacting partners of a protein with a given function against the total number of partners detected [13]. The bases of NC are easy to apply and are based on the syllogism as protein A interacts mainly with proteins involved in a specific role-group, protein A will almost certainly belong and act at the level of that group.

1.2 Isolation and identification of protein complex components

1.2.1 Isolation and identification of proteins acting as binary binders

The first method used in research to determine interacting partners of a specific protein was the yeast two-hybrid technique. This protein analysis approach has been developed in 1989 by Fields and Song [14]. The basis of this technique is the use of the GAL4 protein expressed in yeast that acts as transcriptional activator for the target gene β-galactosidase, involved in the galactose metabolism. This protein is composed of 2 domains, one binding the DNA at the promoter for the β-galactosidase gene (GAL4-binding domain) and a domain involved in the activation of the transcription (GAL4-activating domain). These domains have been cloned in separate expression vectors (pBD-GAL4 and pAD-GAL4) and can be fused to DNA sequences encoding for other proteins or part of them. The mechanism of this approach is rather simple: the 2 proteins of interest are co-expressed in yeast presenting each a complementary portion of the GAL4 transcription factor. In case of a physical interaction between the 2 proteins of interest, the GAL4 protein is reconstituted and the target gene is expressed and its presence confirmed with a colorimetric assay (figure 1) [15]. New improvements on the two-hybrid technique consist of yeast strains lacking of four essential amino acids (tryptophan and leucine genes are missing, histidine and alanine genes are GAL4-promoter dependent) and GAL4 domain-carrier vectors encoding for tryptophan (on the pBD-GAL4 vector) and leucine (on the pAD-GAL4 vector) [16]. Efficient co-expression of both GAL4-domain vectors and the occurrence of direct and strong interaction between bait and pray proteins, promotes the production of the lacking amino acids. A high stringent media missing all the 4 listed amino acids is allowing the growth of only the clones where the interaction is significantly taking place [17-20].

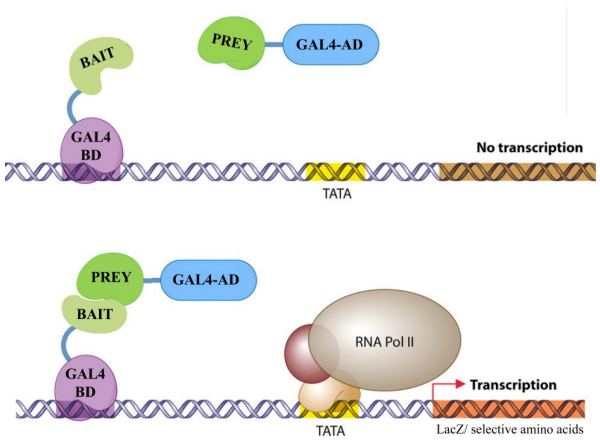

Figure 1: Schematic depiction of the yeast two-hybrid technique. In the upper part of the figure the two GAL4-domains fused to the bait of interest and a random pray protein are shown. Subsequent to the physical interaction between the two proteins present in the yeast cell, the GAL4 transcription factor is reconstituted and activates target genes expression. In our approach, the β-galactosidase, the histidine and the alanine genes were our reporter and selection markers respectively.

1.2.2 Isolation and purification of protein complexes

1.2.2.1 Purification via tag

In the last years many different tags have been developed and commercialized. The most used can be divided in small-length tags made by polypeptides (e.g. Flag, Strep II, c-myc and HA tag) [21-24] or protein-fused tags (e.g. GST [25] and MBP [26]) (Table 1).

The principle of tag usage consists of the construction of a DNA sequence containing the protein of interest fused to one of the previously mentioned tags on the N or C-terminus, in order to obtain an easier purification of the complexes containing the target protein. The location of the tag has to be decided cautiously as the tag must not interfere with the correct folding, localization and function of the protein. Besides, some tags (e.g. GST) are more suitable for protein expression in prokaryotic systems as E. coli and special attention must be payed to the expression system of choice. The tagged protein is expressed in the chosen system and purified through the direct interaction of the tag with the antibody/resin that has affinity for it. Given that the purification could carry some non-

specific binders (false positive), it is preferable to use higher affinity-binding resins in lower amounts in order to decrease the presence of non-specific binders [27].

Tag	Size (aa)	Sequence	Binding resin	Elution
Flag	8	DYKDDDDK	Anti-FLAG M2 MAb agarose	Peptide
Strep II	8	WSHPQFEK	Strep-Tactin–Sepharose	Desthiobiotin
c-myc	10	EQKLISEEDL	Anti-myc MAb agarose	Peptide
HA	9	YPYDVPDYA	Anti-HA MAb agarose	Peptide
GST	218	(ref 23)	GSH–Sepharose	Glutathione

Table 1: Examples of commercially available tags for affinity purification. In this table, length and sequences of the tags available in the lab are shown in the second and third column, respectively. The fourth column displays the affinity-binding or antibody-coupled resins used for specific binding of tagged proteins and the fifth column describes the method used for elution. For the elution of Flag, HA and c-myc tagged proteins, the peptide used is the same as reported in the third column. (Adapted from [27])

1.2.2.1.1 GST pull-down

Glutathione-S-transferase (GST) fusion proteins have had a range of applications since their introduction as tools for synthesis of recombinant proteins in bacteria. Typically, GST pull-down experiments are used to identify interactions between a bait protein and unknown targets and to confirm suspected interactions between the bait protein and a known protein. The bait protein is a GST-fused protein, whose coding sequence is cloned into an isopropyl-β-D-thiogalactoside (IPTG)-inducible expression vector [28-29]. This fusion protein is expressed in bacteria and purified by affinity with glutathione-agarose beads (table 1). This approach allows the study of interactions occurring between the protein of interest and a pool of proteins obtained from every tissue of choice, mostly selected depending on the role, localization and physiological aspect of the bait protein. The protein complexes of interest are recovered and characterized as shown in figure 2. The recent developments of mass spectrometry analysis on protein complexes were combined with this technique, providing valuable technical advantages for high-throughput protein discovery.

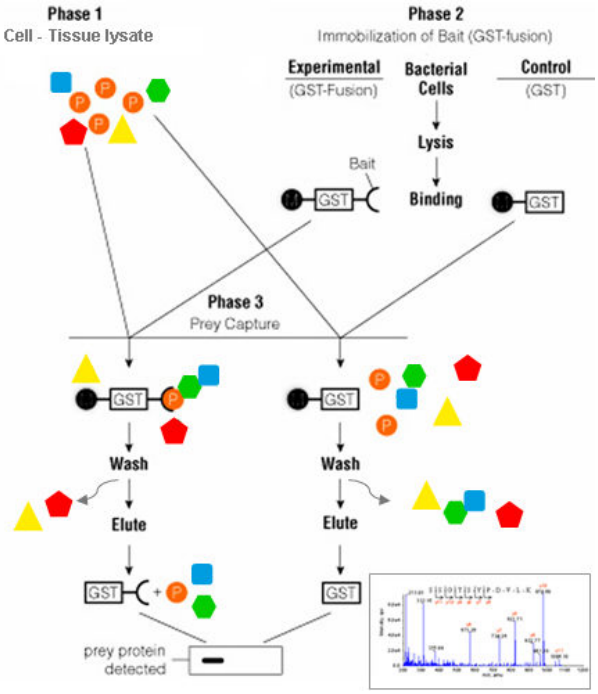

Figure 2: Schematic principle of GST pull-down. After production and purification of GST-fused baits (phase 2), these are incubated with a cell lysate of interest (phase 1). After incubation and washing steps (phase 3), only proteins specifically interacting with the prey (direct or indirect) are eluted and the interactors can be identified either by western blot or mass spectrometry.

1.2.2.1.2 Tandem affinity purification with Strep-FLAG tags (SF-TAP)

The technological development of the instruments used for mass spectrometric sample analysis has increased the sensitivity, requiring consequently lower amount of proteins. Nevertheless, such increase of sensitivity had the drawback to identify higher amounts of non-specific binders and contaminants, neutralizing the obtained technical improvements. An important breakthrough was the possibility to combine different tags to perform two-step purifications reducing drastically the contaminants carryover [30-31].

The Strep-Tactin FLAG tandem affinity purification (SF-TAP) purification is a quick and efficient protein complex enrichment method that can be sufficient for quantitative approaches or directed identification of complex components. This method is based on the pairing of two tags (Strep and FLAG) used for two subsequent purification and enrichment procedures (figure 3). A new TAP tag was produced combining 2 sequences of StrepII tag and one FLAG tag (SF-TAP tag) [32]. The SF-TAP with the reduced size of the tag (4.6 KDa) was shown to be suitable for protein complex purification and due to the medium

binding affinity of the tags, elution is allowed by antigen competition without using a protease (TEV protease used for first generation TAP tag) [33].

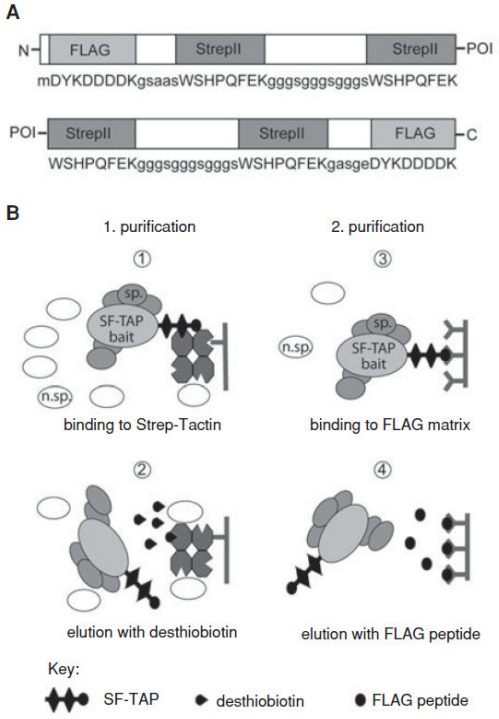

Figure 3: Schematic representation of the Strep/FLAG tandem affinity purification. (A) Amino acid sequence of the SF-TAP tag for N-terminal (upper) or C-terminal tagging (lower) of the protein of interest (POI). (B) Schematic representation of the tandem affinity purification obtained with StrepII tag (number 1 and 2) and with FLAG tag (number 3 and 4). Figure obtained from [33].

1.2.2.2 Immunoprecipitation

Additionally, to tag purification methods, the immunoprecipitation is a valuable method to identify interactors of endogenous target proteins using a specific antibody against the bait. The main advantage of this technique is the possibility to study protein complexes under physiological conditions directly in tissues of interest. Furthermore, different tissues could provide different protein complex patterns, increasing the information about the target protein's function in different organs [34]. However, this method requires the production and purification of a high-affinity antibody that sometimes is not available. In contrast to the exogenous tagged protein approach, immunoprecipitation can only be applied to the full-length endogenous protein (or its variants in patients), preventing complex analysis based on domain dissection or target mutations.

1.2.2.3 Domain-based analysis of protein complexes.

The methods previously described, when involving a full-length protein, are certainly high informative of complexes in which the target protein is involved, but they are underestimating the protein modularity. Proteins are composed of motifs and functional domains that participate in protein-protein interactions and specific functions. Although these domains alone cannot be considered as independent entities, their structure can be analyzed to identify domain-domain interactions among proteins or nucleic acids [35].
In case of a protein with unknown function, the full sequence can be dissected into sub-sequences including domains identified through *in-silico* analysis, and studied separately. As previously mentioned for the yeast two-hybrid technique, these domain-based studies could provide precious information about the nature of the interactions, narrowing down the minimal domain responsible for binding.

1.3 Quantitative complex analysis

As described before, the techniques available for protein complex purification are suitable to identify proteins interacting (directly or not) with the protein of interest. However, to determine the significance of the proteins identified as belonging to a complex, a quantitative/semi-quantitative approach is required. The main aim of this approach is to quantify enriched proteins specifically belonging to the complex against non-specific binders to resin or antibodies. In order to obtain a significant protein network based on protein complexes analysis, mass-spectrometry analyses were combined with protein labeling techniques.

1.3.1 Protein labeling

The increase of sensitivity and the development of high-throughput analysis through mass-spectrometry (MS) machineries stimulated the improvement of new methods for protein labeling. The labeling techniques available nowadays are mostly based on the modification of amino acids mass through the substitution of physiological amino acids with exogenously provided isotope-labeled ones or the chemical attachment of isotopic residues to amino acid side chains. These modifications are producing, at peptide level, mass differences easily detectable during MS analysis. These differences can be quantified

according to the intensity of peptide signals for each condition and the enrichment factor can be then determined considering the ratio between the different samples. The labeling techniques differ mostly according to the labeling target, whether it's happening *in vivo* or *in vitro*. Principle of the SILAC system (stable isotope labeling by amino acids in cell culture) is the addition of heavy-/light-isotope labeled amino acid to the cell culture growth media. To obtain a high number of labeled proteins, essential and abundant amino acids such as arginine and lysine are used. The modifications consist in the substitution of physiological ^{12}C and ^{14}N atoms with ^{13}C and ^{15}N isotopes respectively and combinations of them. The naturally essential amino acids $^{12}C_6$ $^{14}N_2$-L-Lysine / $^{12}C_6$ $^{14}N_4$-L-Arginine, are replaced with $^{13}C_6$ $^{15}N_2$-L-Lysine / $^{13}C_6$ $^{15}N_4$-L-Arginine in heavy conditions and D_4-L-Lysine / $^{13}C_6$-$^{14}N_4$-L-Arginine in medium conditions (chemical formulas reported from http://www.silantes.com/). The peptides resulting from the MS analysis will present a mass shift, in Dalton (Da), dependent by the amount of lysine and arginine residues contained: $Lys^{+\ 8\ Da}$/$Arg^{+\ 10\ Da}$ and $Lys^{+\ 4\ Da}$/$Arg^{+\ 6\ Da}$ respectively in heavy and medium conditions. In cell culture the unlabeled amino acids of the endogenous proteins are substituted, after a minimal amount of cell passages, with isotope-labeled ones, providing a homogenous protein labeling. The main advantage of this technique is the possibility to combine different samples in very early experimental steps and keep the experimental error low [36]. Additional labeling methods can be selected in that cases an *in vivo* labeling is not applicable and must be performed after cell lysis. Often used post-lysis labeling methods are ICAT (isotope coded affinity tag) and ICPL (isotope coded protein label). ICAT [37] is a hybrid combination of labeling and purifications of peptides. The labeling substrate is composed by a reacting group interacting with free cysteines residues, a linker (heavy or light) and a biotin. After labeling the proteins are digested, the peptides isolated by affinity to avidine and analyzed by MS (figure 4, left panel). ICPL [38] consists in the labeling of primary amines (N-Term and lysine) of the proteins, less rare than the cysteine residues for ICAT, with *N*-nicotinoyloxy-succinimides in light or heavy conditions (figure 4, right panel).

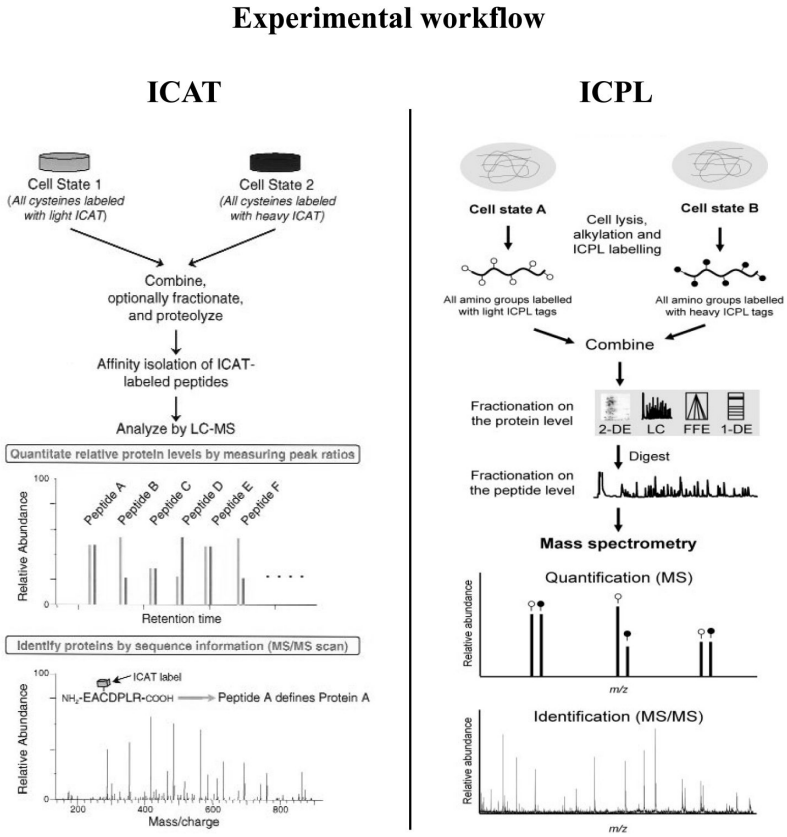

Figure 4: Schematic workflow of ICAT and ICPL labeling techniques. Here are represented the workflows of labeling and quantitative mass spectrometry analysis of protein complexes. ICAT (left, adapted from [37]) is based on the exclusive labeling of cysteines and subsequent purification of labeled peptides. In the ICPL (right, adapted from [38]), differently from the ICAT protocol, the labeling occurs to primary amines (N-term and lysine) and the labeled peptides are not purified prior mass spectrometry analysis.

1.3.2 Protein complex analysis with SILAC labeled samples

The application of the SILAC technique for protein complex analysis allows the identification of protein complexes enrichment via affinity purification (with or without tag). To determine more stringently the amount and the identity of non-specific binders, an unrelated tagged protein or the empty vector containing the tag can be used as negative controls. As lately will be described and according to our preliminary experimental evidences, has been shown that the unrelated tagged protein is the best solution as negative control compared to the tag alone for overexpression-based approaches. The protein composition of target complexes labeled and purified as described before, is compared to control protein complexes purified using an unrelated protein (e.g. GFP protein) or with non-specific antibodies (control IgGs).

In this study the protein complexes identification methods introduced previously (chapter 1.2.2.3) have been combined, when possible, with the SILAC cell-labeling in order to obtain additional quantitative interaction data. The SILAC method also provided a list of interactors surely belonging to the non-specific binders [39].

1.4 Complexes and diseases

The analysis of protein complexes could provide valuable prediction of the role of the protein of interest or the molecular mechanism in which this protein is involved. Furthermore, an extended knowledge about the protein network formed by the protein of interest, allows a better understanding of the impact of a functional impairment of the target protein to the whole network. By analyzing a network based on protein interactions it is evident that each protein is connected to one or more partners, and these "connections" (edges) are dependent on the function assigned [40]. However, the overall network connections can result altered by pathological mutations targeting the protein of interest, suggesting related loss- or gain- of function.

Protein network and disease-dependent deregulation can be analyzed with "disease-to-protein", or "protein-to-disease" approaches. In the disease-to-protein approach the starting point is the identification of the pathway involved in a known disease. Pathways are dissected into protein modules and the modules containing the protein of unknown function are specifically analyzed. Considering the connections and mode of interaction occurring to the protein of interest, a putative function can be provided. In the protein-to-disease approach the newly described protein and its interacting partners are studied to identify functional networks and pathways containing it (figure 5A). Mutations occurring to the protein of interest (*) could produce: 1) physical loss of protein and the interactions taking place with it are lost and interacting partners could not associate to the same protein complex (figure 5B, protein X and Y); 2) protein loss-of function and loss of an edge (figure 5C, loss of the edge between X and *); 3) protein gain of function producing a new edge with possible deregulated functions (figure 5D, new edge between * and Z) [41-42].

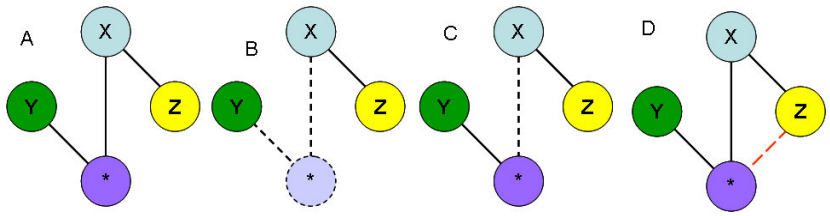

Figure 5: Effects of a functional impaired protein on the network. In this sketch the effects that a mutation or functional impairment of a protein of interest could produce at network level are summarized. (A) The protein of interest (purple circle and marked with an asterisk) shows interconnectivity with other proteins in normal conditions. Subsequently to a mutational event, protein * expression can be lost and affects the interconnectivity between the proteins belonging to the same network (B, dashed lines). Additionally, mutation on protein * can cause a loss-of function (C, dashed line) or a gain-of function (D, red-dashed line). Adaptation from [41]

1.5 Identification and characterization of the SON protein

First evidences of the protein existence of SON have been detected in 2004 [43] analyzing the protein components of interchromatine granule clusters (IGC). These sub-nuclear structures, also referred as nuclear speckles, are containing proteins involved in gene expression. Most of the splicing factors and subunits of RNA polymerase II have been detected in these organelles. The spatial and temporal structure of the IGCs is tightly dependent by the state of the cell and by gene expression as these structures are storage environments for splicing factors. The overlap between IGCs and spliceosome protein components suggest that splicing factors are recruited as complexes from IGCs to sites of active transcription. Performing IGCs purification and proteome analysis with mass spectrometry, have been identified 360 proteins and among these the SON protein (indicated as KIAA1019). SON was further identified as member of high-salt insoluble nuclear fractions together with many nuclear proteins known to be involved in mRNA maturation [44].

SON is a large protein of 2426 amino acids presenting unique features among IGCs proteins. Its structure is composed by 5 main domains (figure 6): a tandem repeat domain, contains 6 unique amino acid repeats spanning most of the N-terminal portion (amino acids 334-1493); a DNA-binding domain, responsible for the binding of HBV genome [45]; a high repetitive arginine/serine-rich sequence (RS) domain (amino acids 1925-2039), which might play a role in mRNA-processing [46] and characteristic of almost all of the IGCs proteins; a glycine-rich domain (G-patch, amino acids 2305-2351) and a double-

stranded RNA-binding domain (DSRM or dsRBD, amino acids 2371-2426) involved in mRNA binding [47-48]. The last three domains are classical features for proteins involved in RNA processing [43-50].

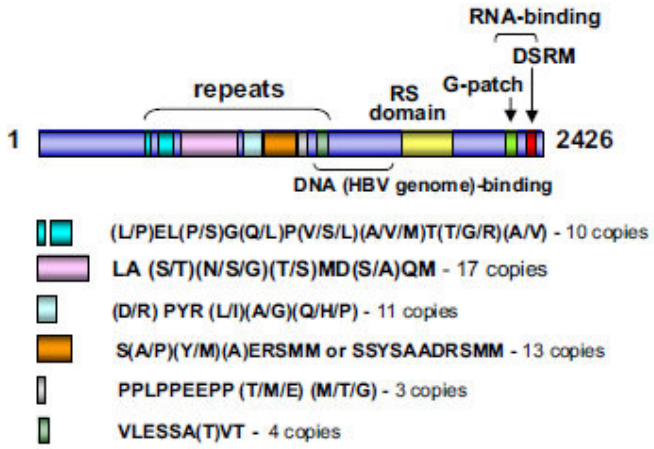

Figure 6: Modular structure of the SON protein. In this picture, the structure of the SON protein is shown focusing on the domains constituting the protein. The 6 sequence repeats present in the first half of the sequence are reported in the lower part including the amino acidic sequences and the amount of copies. Picture adapted from [50])

1.5.1 SON is correlated with leukemia onset and progression

The first indirect functional description of SON has been provided in 2008 when SON was identified as direct interacting partner of the leukemogenic protein AML1-ETO [51]. This protein has been associated with leukocytes deregulated proliferation leading to acute myeloid leukemia (AML) even if its expression in cells is not a key feature for leukemia onset, involving additional hits for the disease development. Has been reported the identification of a splice variant (AML1-ETO9a) correlated with early onset and faster progression of the disease due to the absence of the regulating zinc-chelating domains 3 and 4 (NHR3 and NHR4). SON has been detected as direct interactor of the NHR4 domain and indicates a role in cell-proliferation control, showing that the lack of this interaction leads to a higher proliferation rate [52-54]. Besides the described interaction-dependent function, it has been further shown that depletion of SON from leukemia cells is correlated with significant reduced proliferation rate and cell viability. Nevertheless, cell proliferation was restored after transfection of the N-terminal portion (amino acids 1-140) in SON-depleted cells. Whether this interaction could provide precious information about

AML1-ETO induced leukemia and SON as a possible target for chemotherapy is still under debate.

1.5.2 SON is required for cell cycle progression and correct localization of splicing factors

SON proliferation ability has been tested further in different cell lines, showing that SON is not involved in tumorigenic processes but it is essential for cell physiological growth and proliferation [55]. SON-depleted cells were studied more in detail, showing an accumulation of these cells during metaphase due to a cell-cycle arrest. This incomplete mitosis has been described to happen because of an inefficient assembly of the mitotic spindle occurring in SON-depleted cells. Interestingly, failure of the cell-cycle mitosis is not only dependent by the spindle formation but also by several checkpoints involved in genome stability control and by specific protein expression (figure 7, upper panel). Combining all these previous findings, Sharma et al. associated a clear phenotype cell change upon SON depletion. Many splicing factors co-localizing with SON in the IGCs [44] appear to be significantly re-distributed after SON deprivation. In particular the pattern of the splicing factors U1-70K, SF2/ASF, MAGOH, MLN51 and SC35 changed from punctuate to doughnut-shaped distribution (figure 7, lower panel), but these proteins were not redistributing in other organelles. Since the localization and assembly of nuclear speckles have been shown to be a self-associating and self-assembly mechanism, these findings suggest a role for SON mainly as scaffold protein needed for correct speckles organization [50-56].

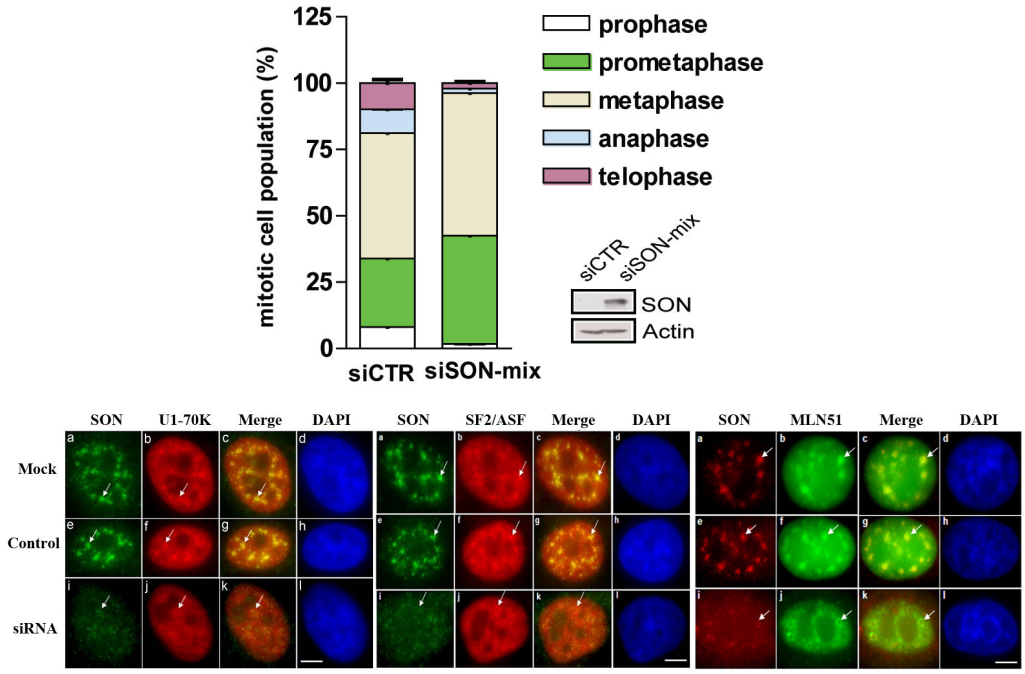

Figure 7: Effects of SON depletion in cell-cycle and splicing factor distribution in HeLa cell line. Upper panel: SON depleted cells are accumulating in the metaphase (and prometaphase) indicating that SON is responsible for the progression of mitosis in cells (adapted from [55]). Lower panel: upon SON depletion, different splicing factors are redistributing in a doughnut-like shape, altering the structure and function of nuclear speckles (modified from [50], scale bar 5µm).

1.5.3 SON mediates the processing of pre-mRNA into mature mRNA

According to the previously reported roles of SON it has been questioned whether SON is not only associated with splicing factors but also takes actively part in the splicing process, an event taking place directly in the sites of active transcription and involved in the maturation and modification of the nascent RNA. The immature mRNA (pre-mRNA) is a sequence composed by coding exons separated by not coding introns, and the splicing machinery is performing fine intron removals in order to produce a mature mRNA that will be used lately by the ribosomes. This mRNA maturation is performed by the spliceosome, peculiar nuclear machinery composed by many small nuclear ribonucleoproteins (snRNPs) and auxiliary proteins [57]. The first step of the splicing involves the U1 snRNP complex able to recognize the 5´splicing site sequence of the intron, normally presenting the GU dinucleotide sequence. The second step involves the U2 snRNP complex involved in the binding of the A nucleotide marking specifically the

branch site. The third step, mediated by the U2AF factor, identifies the 3´splicing site sequence of the intron normally presenting the AG dinucleotide. Besides the classical GU-AG recognition pattern a more seldom recognition pattern formed by AU-AC dinucleotide sequences has been described also [58]. The last steps of the mRNA maturation, involving the U4, U6 and U5 snRNPs complexes, are responsible of the complete removal of the intron and the junction of the adjacent exons. This process is repeated until every single intron has been removed from the initial pre-mRNA sequence [59].

Although the link between splicing, cell proliferation and survival is known since decades to happen in human [60] and in *S. cerevisiae* models [61], the specific association between splicing and cancer has received less attention [62]. The cell-cycle progression is a process finely regulated by precise timing and protein levels. The balance between apoptotic and anti-apoptotic regulators is modulated by the availability of the corresponding mRNAs. In case of unbalanced amounts of these modulators, cells could tend towards early death or unregulated proliferation [49, 63-64]. The splicing defects occurring to oncogenes like p53 [65], BRCA1 [66] and KIT [67], normally involved in the physiological cell-cycle progress, are producing deregulated proteins acting directly on the cell-cycle or modifying the action of cell-cycle control proteins. As an example, the splicing factor SF2/ASF has been shown up-regulated in many cancers and its involvement in cancer onset is due to its splicing effect on the BIN1 protein, a tumor suppressor of c-myc [68-69].

Comparing mRNA expression profiles on HeLa cells with the same cells depleted of the SON protein, it has been remarkably noticed that mRNA levels change especially for genes involved in cell-cycle and cancer development. More in detail, the genes that showed a significant down-regulation after SON depletion, are mainly involved in DNA repair (e.g. Exo1 and FANCG), in cell-cycle and microtubules dynamics (e.g. TUBG1) or to both of them. Further experiments showed that this decrease expression is correlated with aberrant splicing products of the above genes, implicating the decay of these mRNAs as main reason for the lower levels detected (figure 8) [70-71].

SON depletion causes the inclusion of an intron in the mini genes tested, revealing an active role of SON in recognition of weak splice sites. These sites have a lower sequence-derived affinity for the SR proteins responsible for recognition of intron boundaries [72]. Using deletion mutants it has been possible to identify the C-terminal portion of SON as minimal essential sequence for correct splicing mediated by SON full-length

physiologically. All these findings are pointing toward a functional role of SON in splicing-derived cell proliferation, probably similar to the SF2/ASF pathological role described before.

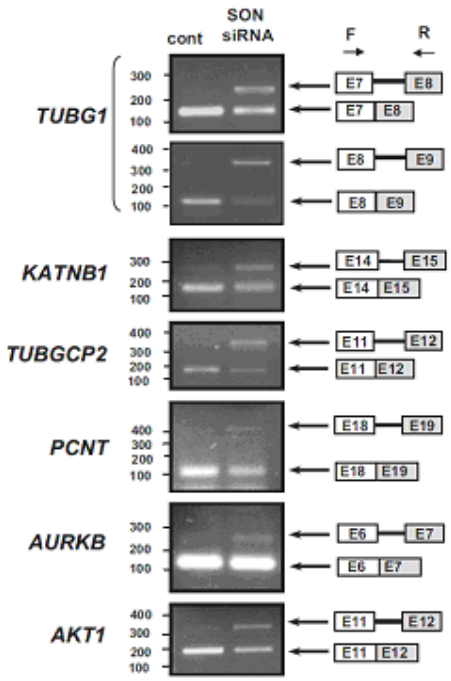

Figure 8: SON depletion impairs splicing causing intron retention. The picture shows the effects of SON depletion on different genes (adapted from [70]). In the figure the genes studied are indicated, the exons (and intron) cloned in the mini genes and the target mRNA obtained after splicing (sequences amplified with PCR). E.g.: for the gene KATNB1, the exons 14 and 15 are cloned including the intron 14. Upon SON depletion the intron 14 is not efficiently removed as in the control and the sequence is appearing longer.

1.6 Aim of the study

The field of protein complex analysis has rapidly developed in the last decades, providing new methodologies for higher sensitive studies. Nevertheless, older interaction techniques (such as the yeast two-hybrid) have been modified to satisfy increasing standards for protein-protein interaction identification, remaining still eligible methods. The identification of protein members involved in the particular cellular machinery provided important new insight on the organization and modulation of specific functional protein complexes, within the last years. However, most of the recent efforts are pointing to explain downstream effects of mutation or depletion of proteins of interest as independent entity, without considering this protein as a member of a bigger and more complex module [46, 60, 71]. In the case of the SON protein, upon its depletion, effects on splicing and splicing factor organization have been discovered and extensively described. Nevertheless, the mechanism underlying these finding has been neither satisfactorily described nor identified.

The aim of this work was to explore, identify and characterize the interacting partners of SON and the cellular machineries wherein it is involved. In this study, according to the recent findings of the involvement of the C-terminal portion of SON in splicing [70], interaction studies on the C-terminal portion of SON (SONc), corresponding to the amino acids 2079-2426 (figure 9) were also included.

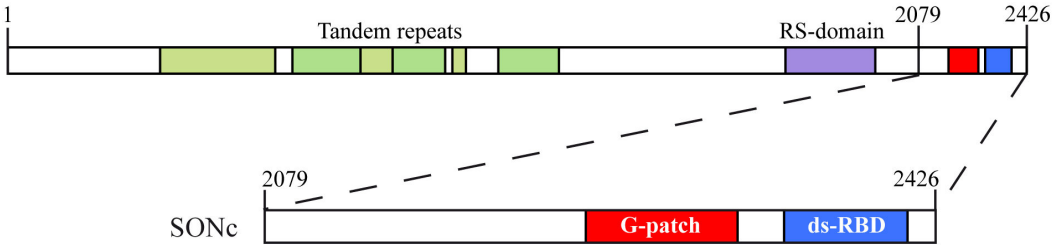

Figure 9: Representation of the SONc fragment. In the figure is shown the SONc portion used in the study, composed by the amino acids 2079-2426. The SONc fragment contains exclusively the G-patch domain (red box) and the double strand RNA-binding domain (dsRBD) (blue box). Numbers are indicating the amino acids used as N and C ends of SONc.

Since different techniques for interaction studies are presenting different peculiar advantages and drawbacks, their combination offers a possibility to obtain a more meaningful and comprehensive overview of the mechanism SON is involved in.

Objective #1

The first aim of this work was to identify interacting partners of the SON protein and the SONc portion applying the different techniques available in the lab. The yeast two-hybrid technique was used for SONc studies to determine binary binding partners. The tag-affinity based purification (SF-AP) and the GST pull-down techniques were used to identify SONc interacting partners in cell culture system and in human tissue. Last the immunoprecipitation of the endogenous SON protein was intended to be used to identify physiological interactors in a native-state cell culture system. The last three techniques were additional combined with qualitative (for the Y2H) or quantitative (for the SF-AP and eIP) mass spectrometric analysis.

Objective #2

Subsequent to the identification of SON and SONc interactors, the second aim of this study was the identification of the minimal protein sequence of SONc responsible for binding to the newly identified interacting partners. The yeast two-hybrid, the SF-AP and the GST pull-down techniques were used for this approach.

Objective #3

The third aim of this study was the identification of a putative role of the SON protein according to the interacting partners identified. Network based analysis of the interacting partners obtained from the objective #1, were used to determine, by similarity, a putative mechanism of SON in cellular machineries in order to correlate the interaction datasets to the previously observed SON effects on cells.

Objective #4

The last part of this study was pointed to determine effects of SON depletion in cell culture on the molecular complexes identified containing SON from the objective #1. The main focus was to understand mechanisms that in case of SON depletion lead to splicing defects.

2. Methods

2.1 Molecular biology

Lab equipment	Supplier
Cycler PTC-225	MJ Research, BioRad, Munich,Germany
GFL Incubator Shaker for E. coli	GFL, Burgwedel, Germany
Incubator for E. coli	Memmert, Schwabach, Germany
UV transiluminator UVT-40M	Herolab, Wiesloch, Germany
Water bath HRB 4 digital	IKA Labortechnik, Staufen, Germany
SubCell GT chambers	BioRad, Munich, Germany
10 cm culture dish	Nunc, Wiesbaden, Germany

Reagents	Supplier
Big Dye Terminator v3.1 Sequencing Kit	Applied Biosystems, Foster City, USA
Gateway BP Clonase II enzyme mix	Invitrogen, Carlsbad, USA
Gateway LR Clonase II enzyme mix	Invitrogen, Carlsbad, USA
Phusion High-fidelity PCR Kit	New England Biolabs, Ipswich
PureYield Plasmid Midiprep System	Promega, Mannheim, Germany
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
QIAquick PCR purification Kit	Qiagen, Hilden, Germany
QIAprep Spin Plasmid Miniprep Kit	Qiagen, Hilden, Germany
Bacto typtone	Roth, Karlsruhe, Germany
Yeast extract	Roth, Karlsruhe, Germany
Quick ligation kit	New England Biolabs, Ipswich

E. Coli strains	Supplier
DH5α	Invitrogen, Carlsbad, USA
XL1-Blue supercompetent cells	Stratagene, La Jolla, USA

Sequencing primers

Target vector	Name	Sequence
pDEST15	GST-forward	5´ CCACGTTTGGTGGTGGCGAC 3´
	GST-reverse	5´ CACCGTCATCACCGAAACG 3´
	T7-forward	5´ TAATACGACTCACTATAGGG 3´
pDEST53	GFP-forward	5´ ACATTGAAGATGGATCCGTT 3´
	GFP-reverse	5´ ACAATGCGATGCAATTTCCT 3´
pcDNA3-SF-TAP	SF-TAP-forward	5´ GCGGTAGGCGTGTACGGTGGG 3´
	SF-TAP-reverse	5´ GGGCAAACAACAGATGGCTGGC 3´
pAD-GAL4_2.1/DEST	-193	5´ CATGAATAATGAAATCACGGCTAGT 3´
	7848	5´ CTATCTATTCGATGATGAAG 3´
	7849	5´ CTCTGCAGTAATACGACTCAC 3´
pBD-GAL4_Cam/DEST	pBD-forward	5´ GATTGGCTTCAGTGGAGAC 3´
	pBD-reverse	5´ GCCTCGAGCTCAGTCATCCTC 3´
pDONOR201	pDONOR201-for	5´ TCGCGTTAACGCTAGCATGGATCTC 3´
	pDONOR201-rev	5´ GTAACATCAGAGATTTTGAGACAC 3´

PCR primers

Name	Used for construct	Sequence
GW_SON_1F	SONc, SONc1, SONc12	5´ AAAAAGCAGGCTTCCAAAGTAAAGAAGATGAT 3´
GW_SON_1R	SONc1	5´AAGAAAGCTGGGTCTACTTTTTAATCCAGGCTTG 3´
GW_SON_2F	SONc2, SONc23	5´ AAAAAGCAGGCTTCATTGATGCCTGGGCTCAG 3´
GW_SON_2R	SONc2, SONc12	5´AAGAAAGCTGGGTCTATGACAGATCTTTCATTGC 3´
GW_SON_3F	SONc3	5´ AAAAAGCAGGCTTCCAAAAGAGGTCTGGGAAC 3´
GW_SON_3R	SONc, SONc3, SONc23	5´ AAGAAAGCTGGGTCTATCTACGTGAGGCACT 3´
attB1	pDONOR	5' GGGGACAAGTTTGTACAAAAAAGCAGGCTTC 3´
attB2	pDONOR	5' GGGGACCACTTTGTACAAGAAAGCTGGGTTTTA 3´

All oligonucleotides were purchased from Metabion (Martinsried, Germany).

Vector used

Name	System	TAG	Promoter	Antibiotic resistance	Size (bp)
pDONOR201	E. Coli		T7	Kan/Chp	4471
pDEST15	E. Coli	N-GST	T7	Amp	7013
pDEST53	Mammal	N-GFP	CMV	Amp/Neo	7767
pDEST-N-SF-TAP	Mammal	N-StrepFlag	CMV	Amp/Neo	7274
pAD-GAL4_2.1/DEST	Yeast	N-GAL4-AD	ADH1	Chp	7620
pBD-GAL4_Cam/DEST	Yeast	N-GAL4-BD	ADH1	Chp	6485

Legend: Kan (Kanamycin), Amp (Ampicillin), Chp (Chloramphenicol), Neo(Neomycin)

Solution used	Composition
LB (Luria-Bertani) medium	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) sodium chloride
LB-plates	LB-medium, 1.5% w/v agar, supplemented with appropriate antibiotic
SOC medium	2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% sodium chloride, 20 mM glucose
TFB1 buffer	30 mM potassium acetate, 100mM rubidium chloride, 10 mM calcium chloride, 50 mM magnesium chloride, 15% (v/v) glycerol, pH adjusted to 5.8
TFB2 buffer	10 mM MOPS [3-(N75morpholino) propanesulfonic acid], 75 mM calcium chloride, 10 mM rubidium chloride, 15% (v/v) glycerol, pH adjusted to 6.5
6X loading buffer	0,25% bromphenol blue, 40% (w/v) sucrose
TAE buffer	40 mM tris-acetate, 1 mM EDTA (ethylenediaminetetraacetat) pH 8

2.1.1 General handling of competent bacterial cells

2.1.1.1 Generation of chemically competent Escherichia coli (*E. coli*)

E. coli of the strain DH5α were cultured in liquid media over-night in 2.5 ml LB-medium without antibiotics. Afterwards the culture was diluted 1:100 in LB-medium, supplemented with 20 mM magnesium sulfate. The cultured bacteria were then grown, until an optical density of 0.4-0.6 (OD600) was reached; and collected by centrifugation at 5,000 g for 5 minutes at 4°C. The bacterial pellet was resuspended in TFB1 buffer and incubated for 5 minutes at 4°C and centrifuged for 5 minutes at 5,000 g at 4°C. The bacteria were then resuspended in TFB2 buffer, incubated for 15 minutes on ice and subsequently split into 100 µl aliquots. The aliquots were frozen in liquid nitrogen and stored at -80°C for several months without transformation efficiency loss.

2.1.1.2 Thermic transformation of *E. coli*

In order to transform a desired plasmid into competent *E. coli*, the bacteria were subjected to thermic transformation. This method destabilizes the bacterial membrane and forms holes that can be used for a passive flux of the plasmid inside the cell. *E. coli* were thawed on ice and 10 to 30 ng of the desired plasmid DNA were added to 50 µl suspension of competent *E. coli*. The bacteria-DNA mixture was incubated for 30 minutes on ice and subsequently incubated for 45 seconds at 42°C to produce a thermic shock and initiate the

uptake of the DNA. Thereafter bacteria were cooled down on ice for 5 minutes. 200 µl of SOC medium were added and the bacteria were recovered under constant agitation for one hour at 37°C. Then the final suspension was plated out on LB-plates containing the appropriate antibiotic and incubated for at least 8 hours at 37°C.

2.1.1.3 *E. coli* cultures

According to the method to be applied for the bacterial cells, the *E. coli* have been cultured either as liquid cultures or as solid cultures.
Solid cultures are usually used to isolate single clones containing a specific plasmid DNA after cloning. The bacteria solution obtained after transformation (see following chapters for more extensive explanations) is poured on the surface of appropriate LB-plates and incubated overnight at 37°C. This procedure allows to pick, culture and analyze single clones and the plates can be stored at 4°C for one week for further usage.
Liquid cultures are used to perform clone expansion to obtain a larger amount of plasmid. Amount of media and cells to use for plasmid preparation must be chosen according to the vector features, since a low-copy plasmid will produce a lower amount compared to a high-copy plasmid.

2.1.1.4 Small and large-scale plasmid DNA preparation

Small scale plasmid preparation was done with the Plasmid Miniprep Kit, while large scale plasmid preparation was performed with the Plasmid Midiprep kit.
For small scale preparations, 4 ml overnight culture was used, for preparative purposes 100 ml. The preparations were done according to the manufacturer's instructions. All kits are based on alkaline lysis of bacteria and binding of the DNA to a silica resin. For the Miniprep kit, the DNA was eluted from the silica resin by addition of water (and incubation for 1 minute); for the Midiprep kit, the DNA was eluted and precipitated according to the manufacturer's instructions before it was resuspended in 100 µl sterile ddH_2O. The concentration and purity of DNA preparations were determined by photometric determination of the absorbance at 260 and 280 nm. Therefore, the DNA preparation was diluted 1 to 50 or 1 to 100 in HPLC grade water. A minimal 260/280 nm ratio of 1.6 indicated a pure DNA preparation. The concentration was calculated as follows: DNA concentration [µg/µl] = absorption at 260 nm x 50 x dilution factor.

2.1.1.5 Generation of cryo-stocks

In order to preserve a bacterial clone of interest, 500 µl of an overnight culture was mixed with the same amount of a 100% (v/v) of sterile glycerol. The combined bacteria-glycerol solution bacteria-glycerol can be stored at -80°C for several months without losing the bacteria efficiency to produce copies of the plasmid of interest.

2.1.2 Cloning of SON-cDNA sequences for bacterial and mammalian expression

Overexpression of ectopic proteins is often applied to overcome protein detection issues or in dissection studies of the single domains constituting the protein of interest.
The plasmid choice has to be performed according to the tag that has to be fused to the protein of interest and the cell type where this protein has to be expressed. Mammalian cells and bacteria have different machineries for gene transcription and protein expression. As shown in the beginning of this chapter, different plasmids have been used which are able to produce proteins only in a precise cellular system. In order to obtain a suitable vector containing the tag and the gene of interest (or part of it) suitable for downstream applications, the cDNA sequence of interest has to be produced, fused to an appropriate tag and inserted in specific vectors that will be used for gene expression.

2.1.2.1 Gateway cloning

The GATEWAY cloning technology is based on the recombination system of the phage λ which uses attachment (*attB*) sites in the target DNA to integrate its DNA into the target DNA by recombination. In the GATEWAY system, these *attB* sites are added to the DNA sequence of interest by a PCR reaction before the fragment is recombined into the entry vector by the first reaction, the BP reaction. This vector serves as universal donor vector from which the sequence of interest can easily be transferred into any GATEWAY compatible expression vector by the LR reaction. The main drawback of this system is the relevant length of the final primers that will insert 10-12 amino acids between tag and gene of interest. To circumvent possible problems with long primers, the attachment of the *attB* sequences to the sequence of interest was done in two PCR reactions. In the first reaction were used primers specific for the sequence of interest containing at the ends part of the *attB* sequence (table2). The second reaction used primers completing the attB sequence and pairing with the previous primer sequence used.

Plasmid containing the target sequence	50 ng
5 X High-fidelity Buffer	5 µl
Gateway specific primer forward (10 µM)	1,3 µl
Gateway specific primer reverse (10 µM)	1,3 µl
dNTPs (10 µM)	0,5 µl
Phusion Tag Polymerase (2 U/ µl)	0,5 µl
ddH_2O	Up to 25 µl

Step 1: denaturation of the double strand template	98°C 2 min
Step 2: denaturation of the double strand template	96°C 30 sec
Step 3: annealing of the primer	55°C 40 sec
Step 4: elongation of the fragment by polymerase	72°C 1 min / 1Kb sequence length
Step 5: repetition of steps 2 to 4	16 times
Step 6: final elongation	72°C 5 min

Table 2: Gateway PCR solution composition and PCR program. In the upper part of the table the reagents and respective amounts needed for the PCR amplification of the gateway-compatible insert are listed; in the lower part of the table the steps (including time and temperature) used for the DNA amplification are listed including a short description.

The second reaction was performed as the first PCR with following modifications: as template 5 µl of the first PCR are used, the generic primers for the *attB* sequence (mixture of attB1 and attB2 primers) were used and the number of cycles was increased to 25 (table 3).

The amplified fragment was separated by agarose gel electrophoresis and extracted before subjected to the BP reaction.

Entry vector DONR201	90 ng
Purified PCR product	3 µl
BP Clonase II mix	1 µl
ddH_2O	Up to 5 µl

Table 3: Solution for BP specific recombination. In this table are reported the reagents and respective amounts used for the BP recombination needed to recombine the gateway PCR insert previously obtained inside the pDONOR201 accepting vector through attB sites recombination.

The reaction was incubated at 25°C for 2 hours before 0.5 µl Proteinase K was added and incubated for 10 minutes at 37°C in order to inactivate the clonase enzyme. The reaction mixture was then transformed into DH5α and amplified in liquid cultures before the plasmids were prepared by using the MiniPrep kit. The insert was checked by sequencing.
Verified constructs were used in the LR reaction to transfer the coding sequence into expression constructs. For this reaction the following amount were used:

pDONR201 containing the insert	90 ng
pDEST vector	90 ng
LR Clonase II mix	1 µl
ddH_2O	Up to 5 µl

Table 4: Solution for LR specific recombination. In this table are reported the reagents and respective amounts used for the LR recombination needed to recombine the pDONOR201 containing the insert of interest into a suitable pDEST expression vector through attL sites recombination.

The reaction was incubated for 2 hours at 25°C before 0.5 µl Proteinase K was added. The mixture was further incubated for 10 minutes. The plasmids were transformed into DH5α and amplified in liquid cultures. The plasmids were prepared using the MiniPrep kit and first verified by restriction digest and agarose gel analysis, and afterwards by sequencing.

2.1.3.2 DNA sequencing

Since the high processivity of DNA polymerase could insert a not complementary nucleotide or an extra nucleotide during the DNA copy, a deeper analysis of the produced plasmid´s sequence is a relevant issue. Furthermore, when a protein has to be fused to a tag, it is important to prove that the two DNA sequences (tag and protein of interest) are joint correctly in-frame.
After the transformation of competent bacteria with the cloned gene, the plasmids are collected and their sequence analyzed using the BigDye-Terminator v3.1 Sequencing Kit. In this kit, the dNTP mixture includes 4 dideoxynucleotides (ddNTPs) labeled with different fluorescent dyes, one for each nucleotide. The specific fragments will be amplified until a ddNTP is randomly included, disrupting the polymerase amplification (table 5). The mixture of fragments obtained, are separated by capillary electrophoresis and the fluorescence detected. The sequence analyzer correlates the time of run of the fragment to

its length and also correlates the size to the fluorescence detected. The final output is a highly reliable DNA sequence.

For the sequencing reaction the following amounts were used

DNA template	100 ng
5X sequencing buffer	2 µl
Specific sequence primer (5 µM)	1 µl
BigDye kit	2 µl
ddH_2O	up to 10 µl

The sequencing reaction was performed in a thermo cycler with the following program

Step 1: denaturation of the double strand template	96°C 2 min
Step 2: denaturation of the double strand template	96°C 30 s
Step 3: annealing of the primer	50°C 15 s
Step 4: elongation of the fragment by polymerase	60°C 4 min
Step 5: repetition of steps 2 to 4	30 times
Step 6: final elongation	60°C 4 min

Table 5: PCR solution composition and PCR program for DNA sequence. In the upper part of the table the reagents and respective amounts needed for the DNA sequencing PCR solution are listed; in the lower part of the table the steps (including time and temperature) used for the DNA sequencing PCR are listed including a short description.

The obtained DNA fragments were precipitated by addition of 40 µl 80% (v/v) ethanol, incubated at -20°C for 30 minutes and centrifuged for 15 minutes at 16,000 x g. The supernatant was removed and the pellet washed with 200 µl 70% ethanol, centrifuged again for 10 minutes at the same speed and the supernatant removed. The pellet was air dried and resuspended in 60µl HPLC grade water before the fragments were analyzed on an automated sequencer. The resulting spectra were analyzed by using Vector NTI Suite 10 software package.

2.1.3.3 List of generated plasmids containing SONc

In order to perform the experiments planned before, I produced a wide range of expression vectors containing SONc and portions of it (table6). Here the methods used and the list of plasmids produced for each of it are listed.

Purpose	Plasmid
Gateway cloning template	pDONOR201-SONc1 - pDONOR201-SONc2 pDONOR201-SONc3 - pDONOR201-SONc12 pDONOR201-SONc23 - pDONOR201-SONc
SF-AP purification	SF-AP-SONc3 - SF-AP-SONc23 - SF-AP-SONc
Yeast-two hybrid	pBD-SONc1 - pBD-SONc2 - pBD-SONc3 pBD-SONc12 - pBD-SONc23 - pBD-SONc
GST-pull-down	GST-SONc1 - GST-SONc2 - GST-SONc3 GST-SONc12 - GST-SONc23 - GST-SONc
Sub-cellular localization	GFP-SONc1 - GFP -SONc2 - GFP -SONc3 GFP -SONc12 - GFP -SONc23 - GFP -SONc

Table 6: List of produced plasmids containing SONc and its domains. Here the plasmids that have been cloned in order to perform all experiments needed in this study are listed. SONc: C-terminal portion of SON

2.2 Mammalian cell culture

Lab equipment	Supplier
CO2 incubator	Sanyo, Munich, Germany
Rotanta 460 centrifuge	Hettich, Newport Pagnell, UK
Laminar flow	BDK, Sonnenbühl-Genkingen, Germany
Liquid Nitrogen Tank	Chronos Messer, Sulzbach, Germany
Zoom Stereomicroscop Nikon SMZ645	Nikon, Amstelveen, Netherlands
Cell culture plates 6-well/24-well/10cm/14cm	Nunc, Wiesbaden, Germany
Falcon conical tubes 15/50 mL	BD Bioscience, Heidelberg, Germany

Reagents	Supplier
DMEM (Dulbecco´s modified Eagle medium)	Invitrogen, Carlsbad, USA
FBS (Fetal bovine serum)	Invitrogen, Carlsbad, USA
Penicillin/Streptomycin 100X	Invitrogen, Carlsbad, USA
PBS	Invitrogen, Carlsbad, USA
Trypsin/EDTA	Invitrogen, Carlsbad, USA
L-Glutamine	PAA, Pasching, Austria
DMSO	Invitrogen, Carlsbad, USA
Polyethylenimine (PEI)	Sigma-Aldrich, Taufkirchen, Germany
SILAC DMEM	PAA, Pasching, Austria
FBS dialyzed	PAA, Pasching, Austria
$^{13}C_6$ $^{15}N_4$-L-Arginine (R10)	Silantes, Munich, Germany
$^{13}C_6$-$^{14}N_4$-L-Arginine (R6)	Silantes, Munich, Germany
$^{12}C_6$-$^{14}N_4$-L-Arginine (R0)	Silantes, Munich, Germany
$^{13}C_6$ $^{15}N_2$-L-Lysine (K8)	Silantes, Munich, Germany
4.4.5.5-D_4-L-Lysine (K4)	Silantes, Munich, Germany
$^{12}C_6{}^{14}N_2$-L-Lysine (K0)	Silantes, Munich, Germany
L-Proline	Silantes, Munich, Germany
SON silencing siRNA (ID 143161)	Invitrogen, Carlsbad, USA
Silencer negative control (ID140050)	Invitrogen, Carlsbad, USA
Oligofectamine	Invitrogen, Carlsbad, USA

Cell-line	Description	Supplier
HEK293T	Human embryonic kidney cells	Walter Kolch, Glasgow, UK
HeLa	Human cervical cancer cells	DSMZ, Braunschweig, Germany

2.2.1 Growth and maintenance of mammalian cells

2.2.1.1 Routine culture

HeLa and HEK293T cells were grown in growth medium (Dulbecco's modified eagle medium (DMEM) containing 10% fetal bovine serum (FBS), 50 units/ml Penicillin and 0.05 mg/ml Streptomycin) in 10 cm cell culture dishes at 37° C and 5% CO2 in a cell culture incubator. The FBS was heat inactivated by incubation at 55°C for 30 minutes before use. Cells were split two times a week at a ratio of 1:10 for HeLa cells or 1:15 for HEK293T. Therefore, the medium was removed, cells washed once with PBS and incubated with 1 ml of trypsin/EDTA solution (0.5 mg/ml trypsin, 0.22 mg/ml EDTA in PBS) per 10 cm dish for 5 minutes. Then 4 ml of DMEM medium was added and the cell suspension was centrifuged for 3 minutes at 700 g. Cell pellets were resuspended in 5 ml of growth medium and the appropriate amount of the suspension was added to 9 ml fresh medium in fresh cell culture dishes.

2.2.1.2 Growth and maintenance of SILAC cultures

The SILAC technique is widely applied in protein quantification approaches and consists in the stable isotope labeling of proteins using amino acids containing carbon and nitrous isotopes [35]. The labeling of cells in culture take place substituting some amino acids with isotope labeled ones in culture media (for more references see chapter 1.3.1). After a minimum of five cell divisions, nearly all the proteins are labeled with the isotopic supplemented amino acids. In contrast to normal growth medium, the SILAC DMEM is L-lysine and L-arginine deficient and these amino acids were replaced with isotopic labeled ones added subsequently to the media. For double labeling, $^{13}C_6$ $^{15}N_2$-L-Lysine and $^{13}C_6$ $^{15}N_4$-L-Arginine were used for the heavy condition and normal amino acids for the light condition; for triple labeling experiments, normal amino acids were used in the light condition, 4.4.5.5-D_4-L-Lysine and $^{13}C_6$-$^{14}N_4$-L-Arginine in the medium condition and $^{13}C_6$ $^{15}N_2$-L-Lysine and $^{13}C_6$ $^{15}N_4$-L-Arginine in the heavy condition. All SILAC amino acids were added to a final concentration of 0.1 mg/ml. Proline was added to a final concentration of 0.23 mg/ml to prevent arginine to proline conversion, which could impair the quantification. After addition of 50 units/ml Penicillin and 0.05 mg/ml Streptomycin,

the medium was sterile filtered. Afterwards dialyzed FBS was added to a final concentration of 10% (v/v). Heavy, medium and light media were prepared in the same way in parallel.

2.2.2 Generation of cryo stocks

Confluent cell culture dishes were trypsinized, washed once with PBS and centrifuged for 3 minutes at 700 g. The pellets were therefore resuspended in a mixture of 90% FBS and 10% DMSO. The cryo stocks were incubated at 4°C for one hour, afterwards frozen at -20°C for one hour and at -80°C overnight before they were transferred into liquid nitrogen for long-term storage

2.2.3 Transient transfection of HEK293T cells

HEK293T cells were transfected using the polyethylenimine (PEI) transfection method [73]. The day before transfection, the cells were seeded onto fresh cell culture plates in order to obtain an 80% confluency at the day of transfection. For this procedure, the PEI solution (100 µg of PEI in 1 liter of 100 mM sodium chloride, adjusted to pH 7.4) was incubated at 80°C overnight and filtered sterile. Target DNA was diluted in PEI solution as indicated in table 7; the mixture was mixed by vortexing for 10 seconds and incubated for at least 10 minutes at room temperature and subsequently added dropwise to the cells.

Culture format	DNA (µg)	Volume of PEI reagent (µl)
24-well plate	0.25	25
6-well plate	0.8	66
10 cm dish	3	300
14 cm dish	8	1000

Table 7: Reagents and volumes used for PEI transfections. The table shows the reagent volumes used for the transfection of cell lines using the PEI transfection method for different cell culture size formats.

2.2.4 SON silencing

In order to perform analysis of SON depletion from HeLa cells, the SON gene was silenced with commercial siRNA, as previously described [70]. The SON specific siRNA is formed by a double strand 21 nucleotides long RNA already provided annealed. According to the manufacturer indications, Oligofectamine was selected for siRNA transfection in HeLa cells. These cells were cultured in normal DMEM without antibiotics until reaching a

confluence of 30-50%. Oligofectamine was equilibrated in DMEM without FBS and antibiotics (DMEM-/-) for 5 minutes (Mix A from table 8) and in the meantime the siRNA was diluted in DMEM -/- (Mix B from table 8). Afterwards the two solutions were combined and incubated for 20 minutes at room temperature. siRNA-oligofectamine complexes were lately added to the cells and incubated for 72 hours. The suitable silencing time is highly dependent by the silencing ability of specific siRNA and by the half-life of the target protein. In the following table the amounts used according to the cell-culture plate used based on the manufacturer protocol are listed.

Target siRNA	**Sequence (5´- 3´)**	**Molecular weight (Da)**
SON (sense)	GCAUUUGGCCCAUCUGAGAtt	6646.2
SON (antisense)	UCUCAGAUGGGCCAAAUGCtc	6654.2

siRNA silencing with Oligofectamine

	Mix A		**Mix B**	
	siRNA	DMEM -/-	Oligofectamine	DMEM -/-
24-weel plate (single well)	60 pmoles	47 µl	3 µl	9 µl
6-well plate (single well)	300 pmoles	235 µl	15 µl	45 µl
10 cm dish	2500 pmoles	1974 µl	120 µl	378 µl

Table 8: Solutions used for SON´s silencing using siRNA. In this table the amounts of reagents used for different culture dishes size are listed.

2.3 Protein chemistry

Lab equipment	Supplier
Microcentrifuge 5415 D	Eppendorf, Hamburg, Germany
Rotating wheel Rotator SB3	Bibby Scientific, Staffordshire, UK
Ultrospec 3300 pro spectrophotometer	GE Healthcare, Buckinghamshire, UK
Power Supply Power Pac 3000	BioRad, Munich, Germany
XCell SureLock Mini-Cell	Invitrogen, Carlsbad, USA
Mini Trans-Blot cell	BioRad, Munich, Germany
Mini Protean 3 for SDS-PAGE	BioRad, Munich, Germany
Hybond PVDF membranes	GE-Helthcare, Freiburg, Germany
Agfa Curix 60 Developer	Agfa, Cologne, Germany
Microcon Centrifugal Filter Units	Millipore, Billerica, USA

Reagents	Supplier
Bradford solution	Bio-Rad Laboratories, mu
Ponceau S powder	Sigma-Aldrich, Taufkirchen, Germany
ECL Western Blotting Systems	GE-Helthcare, Freiburg, Germany
Hyperfilm ECL	GE-Helthcare, Freiburg, Germany
Illustra microspin column	GE-Helthcare, Freiburg, Germany
Protease inhibitor cocktail	Roche, Switzerland
NuPAGE NOVEX Bis-Tris gels	Invitrogen, Carlsbad, USA

Antibody (anti-)	Species	Dilution	Supplier
SON	Rabbit, polyclonal	1-2000	Sigma
FLAG	Mouse, monoclonal	1-5000	Sigma
DHX9	Mouse, monoclonal	1-1000	Abcam
hnRNPH1	Rabbit, polyclonal	1-1000	Abcam
DHX15	Rabbit, polyclonal	1-5000	Abcam
DDX5	Mouse, monoclonal	1-1000	Santa Cruz

Table 9: Antibodies and relative dilutions used for western-blot analysis. The antibodies used for western-blot detection of specific proteins are here listed together with their working dilution. Legend: Sigma (Sigma-Aldrich, Taufkirchen, Germany), Abcam (Abcam, Cambridge, UK), Santa Cruz (Santa Cruz Biotechnology, Santa Cruz, USA)

Solutions

NP40-Lysis buffer	50 mM Tris-HCl, pH 7.4, 150 mM sodium chloride, 0,5% NP-40, protease inhibitor cocktail
Washing buffer	50 mM Tris-HCl, pH 7.4, 150 mM sodium chloride, 0,2% NP-40, protease inhibitor cocktail
Running buffer 1X	192 mM Glycine, 25 mM Tris, 0,1% (w/v) SDS
Blotting buffer	150 mM Glycine, 20 mM Tris, 20% (v/v) Methanol, 0,02% (w/v) SDS
5-fold Laemmli buffer	5% (w/v) SDS, 250 mM Tris-HCl pH 6.8, 50% (v/v) glycerol, 500 mM β-mercaptoethanol, 0.025% (w/v) bromphenol blue
Ponceau S staining solution	0,1% (w/v) Ponceau S, 5% (v/v) Acetic acid
TBS-T	30 mM Tris-HCl, pH 7,4, 150 mM Sodium chloride, 0,1% (v/v) Tween20
Coomassie fixation solution	50% (v/v) methanol, 12% (v/v) acetic acid
Coomassie staining solution	50% (v/v) methanol, 10% (v/v) acetic acid, 0.1% (w/v) Coomassie brilliant blue R-250
Preserving solution	20% (v/v) ethanol, 2% (v/v) glycerol

2.3.1 Cell harvesting and generation of protein extracts

In order to produce cell lysates suitable for our applications, cells have to be harvested from the culture dish containing them. Culture media was removed and cells were collected in ice-cold PBS. The cell pellet was obtained centrifuging those 3 minutes at 1000 x g. The intact cell pellet was resuspended in NP40-lysis buffer; 2 ml were used for 14 cm dish and 1 ml for 10 cm dish. The lysate was incubated for 60 minutes at 4°C under constant agitation followed by centrifugation at 4°C for 15 minutes at 15000 x g. The supernatant was collected in clean tubes and used for downstream experiments while the pellet was discarded.

2.3.1.1 Bradford determination of protein concentration

The Bradford method is based on complex formation of proteins with the Coomassie brilliant blue G-250 dye [74]. Binding of the Coomassie dye leads to an absorbance shift from 465 nm in its unbound form to 595 nm in its complexed form. The reaction is mainly based on interaction of the anionic form of the Coomassie dye with arginine and only slightly on its interaction with other basic (histidine, lysine) or aromatic residues (tryptophan, tyrosine, phenylalanine). Whereas the color response is not totally linear, by

the use of a standard curve it is possible to achieve exact protein concentration determinations with very high sensitivity (1 μg protein/ml reaction) [75]. The reaction is not influenced by the presence of reducing reagents but is sensitive to detergents like sodium dodecyl sulphate (SDS) or strong basic reagents.
For the protein determination, the Bradford reagent stock solution was diluted in water. A standard curve was prepared with bovine serum albumin (BSA) in a concentration range from 1 mg/ml to 15 mg/ml. Lysis buffer was added to the BSA standard dilutions so that the same volume of lysis buffer was present in each standard as in the samples of unknown protein concentration. The same amount of lysis buffer (without BSA) was used as a blank reference. Each standard concentrations and sample was incubated for one minute with 1 ml of the diluted Bradford reagent solution at room temperature. The absorptions were measured in the spectrophotometer at 595 nm and the protein concentrations were calculated according to the standard curve.

2.3.1.2 Protein concentration

In order to analyze protein lysate, a fundamental point is the protein concentration. A diluted sample presents a ratio amount/volume unfavorable for protein run and analysis. These problems can reflect in a tryptic digestion in a low amount of protein and less probable in the disability to detect all the proteins of interest. Sample concentration can be obtained with specific equipment or via chemical reactions. For this study I performed exclusively protein concentration via centrifugal units, the microcon columns. These columns are formed by a reservoirs connected by a PES (polyethersulfon) porous membrane with specific proteic cut-off. For all the experiments has been used the 3 kDa cut-off membrane. The sample to be centrifuged is transferred into the upper reservoir and the column is centrifuged at 10000 x g at 4°C until the desired volume is reached. The proteins or salts with a mass lower than 3 kDa are flowing through the membrane and are collected in the lower reservoir. Once the desired volume is reached, the concentrated sample can be transferred to a clean tube and used for the desired application.

2.3.2 Protein separation on SDS-PAGE

Protein samples were separated according to their molecular weight by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) for downstream analysis like

Western blotting or protein staining. Before the proteins were separated through application of an electric field in a polyacrylamide gel, Laemmli buffer containing SDS was added. SDS binds to the proteins and adds negative charges, which masks the intrinsic charge of the proteins. This leads to a charge-to-mass ratio that is almost constant for all proteins. In combination with the reduction of disulfide bonds by a reducing agent and heat, proteins are separated by their chain length proportional to their mass. The discontinuous gel system used in this study consisted of a separating gel, overlaid by a stacking gel. By the use of a stacking gel, which is more acidic compared to the separating gel and consists of a lower acrylamide concentration, the proteins are concentrated in a sharp lane before the separation takes place. For this study were used both self-casted gels and ready-to use NuPAGE gels, the last ones were preferred especially for pre-fractionation prior to mass spectrometry. Gel casting was performed by the use of the Mini Protean 3 system. The volumes of solutions used are shown in table 10. The solutions were mixed and filled between a 1 mm spacer plate and a short glass plate. After polymerization, the separating gel was overlaid with the stacking gel. A comb of 1 mm thickness with 10 or 15 wells was stuck between the glass plates before the gel was polymerized. The casted gels were placed in a Mini Protean 3 chamber and the chamber was filled with running buffer. The NuPAGE gels were placed in the NuPAGE chamber and the chamber was filled with MES electrophoresis buffer. For both systems, the comb was removed and the wells were rinsed with electrophoresis buffer. Before the protein lysates were loaded in the wells, 5-fold Laemmli buffer was added to the samples in a 1 to 5 ratio and the samples were incubated at 70°C for 5 minutes. The gel run was started with a constant potential of 80 Volts, followed by an increase to 130 Volts after the protein-corresponding band reached the separating part of the gel. The gel run was stopped when the bromophenol blue front reached the end of the separating gel or when the separation length needed was reached.

Separating gel (1 x Mini Protean 3 gel – 1mm spacers)

	Separating gel 10%	Separating gel 12%	Separating gel 15%
ddH_2O	2 ml	2 ml	2 ml
30% (w/v) Acrylamide:bisacrylamide)	1,65 ml	2 ml	3 ml
1.5 M Tris-HCl, pH 8.8	1,25 ml	1,25 ml	1,25 ml
SDS (20% w/v)	50 µl	50 µl	50 µl
APS (10% w/v)	25 µl	25 µl	25 µl
TEMED	10 µl	10 µl	10 µl

Stacking gel (1 x Mini Protean 3 gel – 1mm spacers)

ddH_2O	1,2 ml
30% (w/v) Acrylamide:bisacrylamide (37.5:1)	240 µl
1 M Tris-HCl, pH 6.8	240 µl
SDS (20% w/v)	17 µl
APS (10% w/v)	8,5 µl
TEMED	3,5 µl
Bromophenol blue (0,1% w/v)	17 µl

Table 10: Reagents and volumes used for casting SDS-PAGE gels. The tables show the reagents and volumes used for casting SDS-PAGE gels of different acrylamide concentrations and for casting the stacking gel. The volumes indicated are for one gel with 1 mm thickness. APS = Ammonium Persulphate; TEMED = Tetramethylethylenediamine.

2.3.2.1 Staining of SDS-PAGE gels

SDS-PAGE gels can be stained with different methods, such as Coomassie, colloidal Coomassie or silver staining. The choice of the staining to be used is highly dependent on the application and the sensitivity required detecting the protein bands. Although the Coomassie staining method is less sensitive when compared to colloidal Coomassie or silver staining, this staining was chosen for prefractionation experiments because of its optimal compatibility to mass spectrometric analysis.

The Coomassie staining is based on the binding of the Coomassie brilliant blue R-250 dye to proteins, presenting a detection limit around 100 ng proteins per band. After gel electrophoresis, the gel was placed in Coomassie fixation solution for 20 minutes in order to remove the SDS from the electrophoretic run and to equilibrate the gel prior the staining. Therefore the fixation solution was removed and the gel was incubated in

Coomassie staining solution for 5 minutes up to 1 hour, until the desired staining is obtained. The staining solution was then removed and fixation solution was added to remove background staining. The solution was exchanged several times until no background staining remained.

2.3.2.2 Drying gels

The stained gels were equilibrated for 30 minutes in preservation solution, placed between two cellophane foils (removing carefully air bubbles) and air dried.

2.3.2.3 Digitalizing gels and films

Stained gels and films were digitalized using a GS-710 calibrated imaging densitometer and Adobe Photoshop CS. Scans were performed at a resolution of minimum of 400 dpi.

2.3.3 Western blot analysis

For the detection of individual proteins in complex mixtures, the sample was separated by SDS-PAGE and proteins were transferred onto Hybond-P polyvinylidene difluoride (PVDF) membranes. This method enables the detection of proteins of interest by specific antibodies. For this purpose, the gel was incubated for 5 minutes in blotting buffer while the PVDF membrane was activated in methanol by incubation for 1 minute and subsequently moved in blotting buffer. The transfer was done in a wet blotting apparatus (Mini Trans-Blot® Electrophoretic Transfer Cell). Two filter papers were wetted in blotting buffer and laid on a sponge on the anode side, subsequently the PVDF membrane was positioned over them. The gel was laid air bubble free on the membrane and three filter papers, wetted with blotting buffer, on top of the gel. Possible air bubbles were removed and the "transfer sandwich" closed. The transfer was performed for 90 minutes with 100 V constant voltage, in a cold cabinet to keep the buffer temperature low. After the transfer, the membrane was washed in ddH_2O and incubated with Ponceau S solution for 5 minutes. This step allows a preliminary quality check of loading and running of the sample. The membrane was then washed 2 times for 10 minutes in TBS-T and incubated in blocking buffer for 1 hour. Then the membrane was incubated overnight with the primary antibody, diluted in blocking buffer as previously indicated (table 9). After incubation overnight at 4°C under constant agitation, the membranes were washed three times for 10

minutes with TBS-T followed by incubation with the specie-specific secondary antibody, coupled with horseradish peroxidase (HRP). The secondary antibodies were diluted 1:7500 in blocking buffer. After one hour incubation at room temperature under constant agitation, the membranes were washed three times for 15 minutes with TBS-T before the HRP was detected with ECL plus reagent. Hyperfilm ECL films, illuminated by incubation on the membranes, detected the chemiluminescent signals. Depending on the signal strength, the incubation was performed in a range time between 5 seconds and one hour.

2.4 Protein complex analysis

2.4.1 Yeast-two hybrid

Reagents	**Composition**
YPAD	5% w/v Difco YPD broth, 40 mg/l adenine hemisulfate salt
1.1x TE/LiAc	11% v/v 10X TE, 11% v/v 1M LiAc
PEG/LiAc	80% PEG3350 (50% w/v), 10% v/v 10X TE, 10% v/v 1M LiAc
Resuspension buffer	50 mM MgSO4, 10 mM Tris pH 8.0
SD medium	6.7 g/l Difco yeast nitrogen base without amino acids, 182.2 g/l D-sorbitol, (pH 5.8), 40 ml/l sterile 50% (w/v) D-(+)-glucose, appropriate amino acids supplement (dropout)
SD-L/-W/-WL/-WLH/-WLHA	SD medium with amino acids supplement (dropout) missing specifically leucine (-L), triptophane (-W), hystidine (-H) or alanine (-A)

In this system, a bait protein is expressed as a fusion to the GAL4-DNA-binding domain (BD-GAL4, figure 10, left plasmid), and this is used to screen a "prey" library of cDNAs fused to the GAL4 activation domain (AD-GAL4, figure 10, right plasmid), by co-expression in yeast. The use of a sensitive (plus or minus) growth selection medium helps to discriminate yeast clones in which interactions take place from false positives.

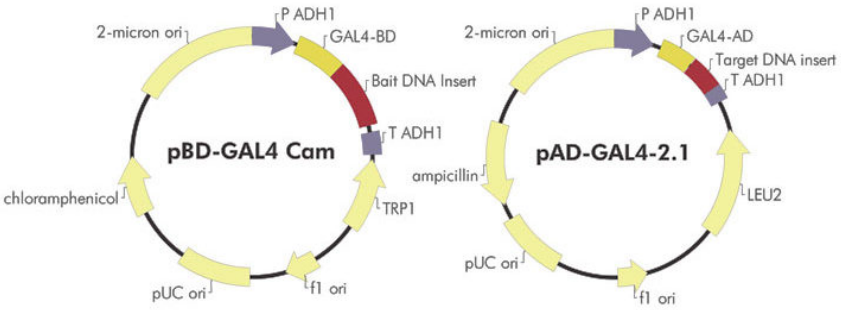

Figure 10: pBD-GAL4 and pAD-GAL4 vector maps. Here are shown the maps of the two plasmids used for yeast-two hybrid experiments, including the area of cloning of bait and prey. LEU2: leucine; TRP1: tryptophan. Adapted from the Invitrogen website

2.4.1.1 Generation of a yeast two-hybrid mating library

2.4.1.1.1 Preparation of competent yeast cells

One fresh colony (maximum 3 days old, 2–3 mm diameter) of PJ69-4alfa cells was inoculated into 3 mL of YPAD medium and incubated at 30°C for 8 hours under constant agitation. 50 µl of the grown culture were inoculated in 250 ml of YPAD medium and grown overnight at 30°C. Cells were grown until the OD600 of the culture reached 0.15-0.30 and then centrifuged at 900 X g 5 minutes. The cell pellet was resuspended in 100 ml YPAD and incubated again at 30°C to final OD600 of 0.4-0.5 and afterwards centrifuged as before and resuspended in 1.2 ml 1.1X Te/LiAc solution.

2.4.1.1.2 Transformation of competent yeast cells

1-10 µg of plasmid DNA (cDNA library in pAD-Gal4) and 200 µg denatured herring testes carrier DNA (BD Bioscience) were combined in a sterile microcentrifuge tube and gently vortexed. 600 µl TE/LiAc solution containing competent cells were added to the DNA mixture and gently vortexed. 2.5 mL PEG/LiAc solution were additionally combined with cells/DNA mixture, vortexed and incubated in agitation at 30°C for 45 minutes.
160 µl of DMSO were added to the mixture, gently vortexed and incubated at 42°C for 20 minutes before centrifuging and resuspending the cells in 40 ml of sodium chloride solution (0.9% w/v).

2.4.1.1.3 Evaluation of the transformation efficiency and titration

Several dilutions of the transformed cells were plated onto 10 cm SD-L plates and the rest of the yeast suspension was plated onto 120 x 15-cm SD-L plates and incubated at 30°C for 2-4 days.
Analyzing the 10 cm dishes, colonies were counted and the transformation efficiency (number of transformants per microgram of DNA), titer (number of yeast cells per milliliter) and total number of transformed yeast cells were calculated.

2.4.1.1.4 Preparation of cryostocks of cDNA libraries

Plates were stored at 4°C for at least 1 hour before scraping the cells with 10 ml cold Milli-Q water for each 15-cm plate. All the fractions were pool together and pelleted at 900 x g for 5 minutes at 4°C. 500 ml resuspension buffer supplemented with 5% v/v glycerol was added before re-centrifuging the cells. The last steps were repeated increasing the glycerol percentage to 10% (in 100 ml buffer) and finally to 65% (in 25 ml volume). The cell suspension was aliquoted in 0.5 ml tubes and snap-frozen in liquid nitrogen and stored at -80°C.

2.4.1.2 Yeast two-hybrid library screening by mating

2.4.1.2.1 Mating

PJ69-4A yeast cells were transformed with the bait plasmid construct (cDNA of interest in pBD-GAL4). The transformed cells were plated onto a SD –W plate and incubated for 2–3 days at 30°C. Additionally, intrinsic transcriptional activation (autoactivation) by the bait plasmid was tested by streaking three individual colonies on SD –W, SD –WH, and SD –WHA plates, followed by incubation for 2–3 days at 30°C. Growth on SD –WH and SD –WHA plates indicates autoactivation, disqualifying the bait for use in the screening procedure.
Once obtained a bait that is not autoactivacting, one colony from the SD-W plate was inoculated in 3 ml SD –W liquid medium and incubated overnight at 30°C with shaking (230–250 rpm).
The overnight culture was then inoculated in 100 ml SD –W liquid medium and grown to OD600=1 (corresponding to 25 Mio cells per ml). One vial of frozen mating library was

thawed on ice and 4 Mio library cells were let recovering in 20 mL YPAD for 10 min at 30°C with shaking (230–250 rpm). 8×10^8 bait cells were added and mixed briefly. The cell mixture was centrifuged for 5 min at 900 × g, the pellet resuspended in 2 mL YPAD and plated onto four 15-cm YPAD plates. The mating process occurred incubating the cells for 4 hours at 30°C before collecting the cells by scraping with 10 ml 0.9% NaCl. Fractions were all pooled and mixed well. Cells were centrifuged, resuspended in 10 ml 0.9% NaCl and poured onto 20 SD-WLHA plates. After 3-14 days the colonies were appearing.

2.4.1.2.2 Determination of the mating efficiency

A small aliquot from the mating process was used to prepare dilutions in 0.9% NaCl and plating each dilution onto a 9-cm SD –WL plate. Each plate was incubated at 30°C for 2-4 days and the colonies counted to determine the mating efficiency.

2.4.1.2.3 Selection and validation of positive clones

The colonies grown onto SD-WLHA plates were streaked onto fresh –WLHA plates in a numbered and gridded pattern, followed by incubation for 2–3 days at 30°C. All grown colonies were inoculated in 3 ml SD-WLHA medium and grew for 2 days at 30°C with shaking. DNA was extracted from the yeast following the protocol used for bacteria, with one modification: cell pellets were incubated in P1 buffer supplemented with Lyticase (40kU/100 ml buffer) for 3 hours at 37°C.
The DNAs obtained were transformed back on dh5alfa competent bacteria to increase the DNA yield and purity; the bacteria were then grown on ampicillin plates, the DNAs were extracted as described previously and subjected to sequencing. DNA sequences were aligned and searched against the NCBI DNA database (NCBI Nucleotide Blast) to detect proteins coded by the selected DNAs.

2.4.2 GST pull-down

2.4.2.1 Cloning of GST fused protein

The pDONR201 containing SONc (or portion of it) was recombined with pDEST-15 vector (N-terminal GST) according to the LR reaction method described previously. After

sequencing to confirm the correct fusion tag-protein, BL21 (DE3) competent bacteria cells were transformed as described and plated onto ampicillin plate.

2.4.2.2 Expression and purification of GST-fused protein

Lab equipment	**Supplier**
DIGITAL Sonifier® S-250D	Branson Ultrasonics Corporation, Danbury, USA
Slide-A-Lyzer dyalsis cassette	Thermo Fisher Scientific, Bonn, Germany
Glutathione Sepharose 4B	GE-Helthcare, Freiburg, Germany

Bacterial strains used	**Supplier**
BL21 DE3 E. Coli	Invitrogen, Carlsbad, USA

Solution	**Composition**
STE	50mM Tris pH 7.4, 50mM NaCl, 1mM EDTA pH 8
STE-N	STE, 0.5% v/v NP-40
GST elution buffer	100mM GSH (glutathione reduced form), 200mM Tris pH 8

2.4.2.2.1 Small scale production and purity control

Five colonies of BL21 cells obtained after transfection with GST-SONc (or domains) vector, were inoculated in 2 ml LB-medium supplemented with ampicillin and grew overnight at 37°C with shaking. The day after 160 µl of the culture were diluted in 8 ml LB-medium-Amp, to a final dilution of 1:50 in double (induced and not induced) and the rest frozen as described before. Cells were grown at 37°C with shaking until OD600=0.6, afterwards IPTG was added to 1 mM final concentration to the "induced" culture and shaken at 30°C. 2 ml induced/not induced aliquots were collected every hour, centrifuged and the pellets kept on ice. Cells pellet were resuspended in STE buffer, 10 mM DTT, Lysozime (final concentration 0.4 mg/ml) and incubated 1 hour at 4°C under gentle agitation. N-Lauroylsarcosine was then added to a final concentration of 1% v/v and the lysate was sonicated (30% amplitude) until the solution became fluid and clear. Samples were then centrifuged and the supernatants incubated with 20 µl of slurry glutathione-agarose resin previously washed 3 times with STE buffer. After 1 hour under gentle agitation at 4°C, the resin of each sample was washed 3 times with 500 µl STE-N. Protein samples were eluted with 50 µl 2-fold Laemlli buffer and resolved on SDS-PAGE to analyze level of protein induction of the different clones, suitable induction time and purity

of the preparation. The percentage of NP-40 on the washing steps could be increased to reduce the non-specifics.

2.4.2.2.2 Large scale purification of GST-fused protein

From the small-scale GST-fused protein production, one clone was selected for expression and purity level. This clone was inoculated in 20 ml LB-medium-Amp (added of ampicillin) and grown with shaking at 37°C overnight. The following day 8 ml of culture were diluted in 400 ml LB-medium-Amp and grown at 37°C until OD600=0.6 was reached. Afterwards the culture was supplemented with IPTG (final concentration 1mM) and shaken at 30°C for the optimal amount of time determined previously with the small scale tests. Cells were then pelleted, resuspended in 8 ml STE buffer supplemented with 10mM DTT and lysozyme (0.4mg/ml final concentration) and incubated at 4°C under gentle agitation for 1 hour. N-Lauroylsarcosine was then added to a final concentration of 1% v/v and the lysate was sonicated (30% amplitude) until the solution became fluid and clear. Samples were then centrifuged and the supernatants were incubated at 4°C with 300 µl of glutathione-agarose resin slurry previously washed 5 times with 1 ml STE buffer. After 3 hours under gentle agitation the resin was recovered and washed extensively with STE-N. The GST-fused protein was eluted 4 times with 1 bead volume of GST elution buffer and subsequently pooled together. The eluate was quantified according to the Bradford method and purified using a dialysis membrane (3.5 KDa cut-off) in 1X TBS with 0.05% v/v Triton X-100 overnight at 4°C with gentle steering. The following day the sample was collected, re-quantified as before and resolved onto a SDS-PAGE to check amounts and purity of the preparation. The GST-fused protein preparation was then supplemented with sterile glycerol (final concentration 50% v/v) and stored at -20°C.

2.4.2.3 GST pull-down of protein complexes

The GST pull-down was performed modifying the protocol previously reported [76]. 150 µl of glutathione-agarose resin slurry were washed 3 times with 1 ml STE buffer and incubated with 20 mg of GST protein (prepared as the bait protein) for 1 hour at 4°C under gentle agitation. The resin was washed 3 times with STE-N and incubated with 20 mg of cell or tissue lysate for 2 hours at 4°C under agitation. Subsequently the sample was centrifuged 3 minutes at 3000 x g and the supernatant collected, obtaining the GST-pre-

cleared lysate. 50 µl of slurry glutathione-agarose resin were washed 3 times with 1 ml STE buffer and incubated with an equimolar amount of GST-fused protein and GST alone (5-10 µg) for 1 hour at 4°C under gentle agitation. Resins were then washed 3 times with STE-N and incubated with 5 mg of previously pre-cleared lysate each. After 1 hour at 4°C under gentle agitation, the resin was washed 3 times with STE-N and the protein complexes were eluted with 2-beads volume Laemmli 2-fold buffer.

2.4.3 Affinity purification using SF-AP-fused proteins

Reagents	**Supplier**
Anti-FLAG affinity gel	Sigma-Aldrich, Taufkirchen, Germany

Solutions	
NP40-Lysis buffer	50 mM Tris-HCl, pH 7.4, 150 mM sodium chloride, 0,5% NP-40, protease inhibitor cocktail
Washing buffer	50 mM Tris-HCl, pH 7.4, 150 mM sodium chloride, 0,2% NP-40, protease inhibitor cocktail
2-fold Laemmli buffer	2% (w/v) SDS, 100 mM Tris-HCl pH 6.8, 20% (v/v) glycerol, 200 mM β-mercaptoethanol, 0.01% (w/v) bromphenol blue

HEK293T cells were transfected with SF-GFP vector (as control) or with SF-SONc (or fragments of it) as described before with PEI solution. After 48 hours cells were collected, lysed in NP-40 lysis buffer, and quantified with Bradford solution. For the single-step affinity purification, 25 µl of Anti-FLAG affinity gel (agarose resin coupled with anti-Flag-M2 antibody) were washed 3 times with washing buffer (at least 10 beads volume) and incubated with 1 milligram of cell lysate form cells expressing the SF-AP tagged proteins. Depending on the downstream application, the abundance of the bait protein and complex components as well as the amount of cell lysate used for the purification, the concentration varied between 1 and 5 milligrams. The mixture was incubated for 1 hour at 4°C under gentle agitation. Then the beads with bound protein complexes were centrifuged for 1 minute at 3000 x g, the supernatant was discarded and the beads were transferred into micro-spin columns. The beads were then washed three times with 500 µl of washing buffer to remove unbound proteins. 50 µl of 2-fold Laemmli buffer was added to the beads and mixed at 900 rpm/min at 70°C. Eluted proteins were collected inserting the micro-spin

column in a clean eppendorf tube and centrifuging it at 8000 x g. Samples were used fresh or stored at -20°C for a maximum of one month.

2.4.4 Immunoprecipitation of endogenous SON

The immunoprecipitation of the endogenous protein (eIP) is performed by incubating the specific antibody with the cell lysate, to permit the interaction of antibody and specific protein. The complexes are recovered by adding protein A or G sepharose resin according to the animal in which the antibody was raised.

For the immunoprecipitation of endogenous SON and its interactors, HeLa cells were lyzed in NP-40 lysis buffer and processed as described in the previous chapter. 2 milligrams of HeLa total lysate were incubated with 5 micrograms of antibody (specific for SON or rabbit IgG as control) and incubated 1 hour at 4°C under gentle agitation. 50 microliters of sepharose-Protein-A beads (Santa Cruz, USA) were transferred into a micro-spin column, washed two times with washing buffer (at least 10 beads volumes) and combined with the previous antibody-lysate solution for an additional hour at 4°C under gentle agitation. At the end of the incubation, the samples were transferred to a micro-spin column and the beads were washed 3 times with at least 10 beads volumes of washing buffer. 50 µl of 2-fold Laemmli buffer was added to the beads and mixed at 900 rpm/min at 70°C. Eluted proteins were collected inserting the micro-spin column into a clean eppendorf tube and centrifuging it at 8000 x g. Samples were used fresh or stored at -20°C for a maximum of one month.

2.4.5 Network generation

In order to describe more efficiently the interactions detected with the previously described methods, I created a network of the identified interactors. In a network-based visualization, not only the newly identified interactions are depicted, but also the interactions already described in literature are represented. For this study, the Java-based network generator Pathway Palette hosted by the Harvard University [77] and available at http://blaispathways.dfci.harvard.edu/Palette.html was used. Upon data upload, Pathway Palette generates a protein-protein interaction (PPI) network based on information curated by BioGRID [78] and HPRD [79] databases. In order to have a more comprehensive PPI network, three specific data sources from BioGRID and three from HPRD have been

selected. In particular I selected databases created from reported binary interactions (BioGRID: HTP/Pairwise and HPRD (+mouse): yeast 2-hybrid, databases from *in-vivo* detected PPIs (BioGRID: Low-Throughput and HPRD (+mouse): *in-vivo*) and databases from in vitro experiments, e.g. pull downs (BioGRID: HTP/Complexes and HPRD (+mouse): *in-vitro*).

Additionally, Pathway Palette enables the user to filter, prioritize, and color-code edges based on the underlying data source and on the visualization intentions.

Pathway enrichment studies were performed using the DAVID web-based tool (http://david.abcc.ncifcrf.gov/) based on the Kyoto encyclopedia of genes and genomes (KEGG) database [80]. Pathway identification has been also checked for significance including Bonferroni correction, Benjamini index and false discovery rate (FDR) calculation.

2.5 Mass spectrometry

2.5.1 Sample preparation

The sample preparation for mass spectrometry investigation has to be carefully planned according to the protocol used to purify the protein complexes and the sample complexity to accomplish the highest analysis quality. Given that all the elutions were performed in Laemmli buffer, the samples deriving from the previously described approaches are composed by hundreds of proteins. However, the samples are containing also the bait proteins and immunoglobulins that in the experiments performed represent around 20-30% of the overall sample protein content.

The protein complexes isolated with the methods described before were subjected to gel pre-fractionation and subjected to proteolytic cleavage, producing peptides more suitable for mass spectrometry analysis. The pre-fractionation was used to remove the known GST baits or the immunoglobulins form the SF-AP and eIP approaches, producing less complex samples and enriched for SON´s interacting partners.

2.5.2 In-gel tryptic digestion

The samples to be used for mass spectrometry investigation were loaded on a precast SDS-PAGE gel and run for 2-3 cm. Gels were Coomassie stained and the bands corresponding

to GST-fused baits or to the immunoglobulins, were excised and analyzed separately to confirm the protein identities. Afterwards the gel was destained 1 hour in fixation solution and extensively washed with ddH_2O to remove interfering substances like methanol and acetic acid. Subsequently the gel lanes were excised with a clean scalpel, partitioned in bands, cut to plugs of 1 mm^3 and the pieces were transferred onto low-protein binding 1.5 ml tubes. Then all gel plugs were incubated for 15 minutes at 60°C in 100 µl of 5 mM dithiothreitol (DTT), cooled down to room temperature and incubated in 100 µl of 25 mM 2-iodacetamide for 15 minutes in the dark. Afterwards the gel plugs were washed twice for 15 minutes with 40% acetonitrile and dehydrated by incubation in 100% acetonitrile for 5 minutes. After removing the acetonitrile solution and air-drying the plugs for 15-30 minutes, 30 µl of trypsin solution (10 ng/µl sequencing grade trypsin in 50 mM ammonium bicarbonate) were added and the tryptic digest was performed at 37°C overnight. The resulting peptides were acidified by addition of 3 µl of 0.5% TFA (trifluoroacetic acid) and the supernatant was transferred into a fresh low-protein binding 1.5 ml eppendorf tube. Additional peptide extraction steps were performed adding 100 µl of 50% acetonitrile, 0.1% TFA to the gel plugs and incubated for 15 minutes under agitation. The second supernatant was pooled with the first and 100 µl of 99.9% acetonitrile, 0.1% TFA were added to the gel plugs. After 15 minutes of incubation, the third supernatant was pooled with the first two and the supernatants were dried in a Speedvac drier. Samples were stored at -20°C.

2.5.3 Protein complex analysis by mass spectrometry

The peptide mixture derived from isolated protein complexes obtained from the previously described method were subjected to mass spectrometry analysis in collaboration with Dr. Karsten Boldt from the University of Tübingen.

Briefly, liquid chromatography mass spectrometry analysis (LC-MS/MS) was performed on an Ultimate3000 nano RSLC system (Thermo Fisher Scientific, Waltham, USA) coupled to a LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) by a nano spray ion source. Tryptic peptide mixtures were automatically injected and separated by a linear gradient from 5% to 40% of buffer B in buffer A (2% acetonitrile, 0.1% formic acid in HPLC grade water) in buffer A (0.1% formic acid in HPLC grade water) at a flow rate of 300 nl/min over 90 minutes. Remaining peptides were eluted by a short gradient from 40%

to 100% buffer B in 5 minutes. The eluted peptides were analyzed by the LTQ Orbitrap Velos mass spectrometer. From the high resolution MS pre-scan with a mass range of 300 to 1500, the ten most intense peptide ions were selected for fragment analysis in the linear ion trap if they exceeded an intensity of at least 500 counts and if they were at least doubly charged. The normalized collision energy for CID was set to a value of 35 and the resulting fragments were detected with normal resolution in the linear ion trap. The lock mass option was activated and the background signal with a mass of 445.12002 was used as lock mass. Every ion selected for fragmentation, was excluded for 20 seconds by dynamic exclusion. For SILAC experiments, all acquired spectra were processed and analyzed using the MaxQuant software (version 1.0.13.13) and the human specific IPI database version 3.52 (http://www.maxquant.org/) in combination with Mascot ((Matrix Science, version 2.2). Cysteine carbamidomethylation was selected as fixed modification, methionine oxidation and protein acetylations were allowed as variable modifications. The peptide and protein false discovery rates were set to 1%. Contaminants like keratins were removed. Proteins, identified and quantified by at least two unique peptides were considered for further analysis. The significance values were determined by the Perseus tool (part of MaxQuant) using significance A.

2.5.4 SILAC approach and quantification

As introduced in chapter 1.4, perturbation of levels or functions of a specific protein are correlated with a network alteration. Since decades, the comparison of different proteomes has been performed running cell protein extracts on two-dimensional gels, in order to estimate the network impact of a target protein. This technique provides a good resolution of the proteome and the different stained-patterns can be compared by overlaying, and specific protein differences could be identified by mass spectrometry [81-82]. Although the intensity of the peaks obtained from the peptide spectra can be quantified (calculating the underneath area), the mass spectrometry approach is not per se a quantitative method considering the different ionization yields of different compounds and other interfering factors.

To circumvent this limitation, many different methods have been developed for relative or absolute quantification in order to make peptide quantification possible.. A precise absolute quantification can be made comparing the peak signals of an isotope-labeled peptide (used

as standard for quantification) and the peptide obtained by proteolysis from a given sample of interest. The resulting ratio, knowing the amount of the isotope-labeled peptide, is providing a precise quantification of the peptide of interest in the sample analyzed [83-84]. Nevertheless this approach, to be economically affordable, cannot be used to quantify a broad range of proteins of interest.

The relative quantification of proteins with the SILAC (stable isotope labeling by amino acids in cell culture) technique is the key method to study variation in proteomes of two different systems [85]. The main limitation of this technique is the possibility to label only tissues of small animals (e.g. worms up to mice). An application for e.g. human tissue is not suitable as the isotopic-labeled amino acids have to be incorporated into the proteome of the cell by active substitution [36]. The labeling is obtained modifying the carbon and nitrogen atoms of the amino acids supplied to the media with heavier isotopes. As described in the introduction (chapter 1.3.1) the ^{12}C has been substituted with ^{13}C producing an increment of 1 Da for each carbon used and the ^{15}N has been substituted with ^{14}N increasing the mass of 1 Da for each nitrogen used. For example in the heavy SILAC conditions not isotopic lysine ($^{12}C_6$ $^{14}N_2$-L-Lysine) and arginine ($^{12}C_6$ $^{14}N_4$-L-Arginine) are substituted with isotopic lysine ($^{13}C_6$ $^{15}N_2$-L-Lysine) and arginine ($^{13}C_6$ $^{15}N_2$-L-Lysine), producing a mass difference of 8 Da and 10 Da respectively. The choice to use these two isotopic amino acids derives by the tryptic proprieties of the trypsin enzyme: it cleaves at lysine or arginine level, assuring the presence of at least one isotopic amino acid per tryptic peptide.

The differently labeled samples can be combined in a very early stage, avoiding sample loss or handling artifacts. The relative quantification is performed comparing the spectra from the same peptide (with different labeling) and in particular comparing the peak areas. The ratio obtained between the peak area of the heavy peptide and the one from the light peptide is representing the different amounts of peptide in the two conditions and, by logical extension, the amount differences of the protein of interest can be evaluated (figure 11) [36].

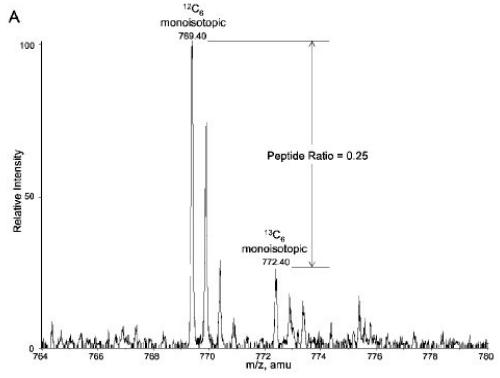

Figure 11: SILAC quantification of differently labeled SILAC pairs. In this figure, extracted from [85], a pair of SILAC peptides sharing the same sequence is shown. The left peak is showing the non-labeled peptide (normal 12C6 atoms) while the right peak represent the labeled peptide (isotopic 13C6 atoms). The ratio shown is obtained calculating the areas under the selected peaks. This ratio can be used to determine the relative amount of peptide (and protein) in the two different samples analyzed.

2.6 Immunocytochemistry

Immunocytochemistry (ICC) is a common technique used to visualize the protein of interest inside a cell or tissue. ICC is used to detect the sub-compartment of the cell where the target protein is localized and to co-localize the protein of interest with other putative interacting partners. The basis of the technique is to permit the formation of a target protein-antibody complex and detecting this complex with the aid of a fluorescent secondary antibody. The coverslips are then mounted on glass slides and analyzed with a microscope able to excite and detect the fluorescence from the secondary antibodies.

Lab equipment	**Supplier**
Axioscope2 Zeiss	Carl Zeiss, Jena, Germany
Axiocam HRc	Carl Zeiss, Jena, Germany
Fluorsafe mounting solution	Merck, Darmstadt, Germany

Solutions	**Composition**
Paraformaldehyde (PFA)	4% (w/v) of paraformaldehyde in PBS
PBS-TX	0,1% (v/v) Triton X-100 in PBS
Blocking solution	1% (v/v) NGS (natural goat serum), 0,2% (w/v) BSA in PBS
Primary and secondary antibody solution	Blocking solution with addition of antibodies of choice

Primary antibody (anti-)	Dilution	Species	Supplier
SON	1:2000	Rabbit, polyclonal	Sigma
DHX9	1-500	Mouse, monoclonal	Abcam
DDX5	1-500	Mouse, monoclonal	Santa Cruz

Secondary antibody	Dilution	Species	Supplier
Mouse Alexa fluor 488	1:2000	Goat, polyclonal	Invitrogen
Mouse Alexa fluor 568	1:2000	Goat, polyclonal	Invitrogen
Rabbit Alexa fluor 488	1:2000	Goat, polyclonal	Invitrogen
Rabbit Alexa fluor 568	1:2000	Goat, polyclonal	Invitrogen

Table 11: Antibodies and relative dilutions used for immunohistochemistry analysis. The antibodies used for immunohistochemistry detection of specific targets are here listed together with their working dilution. Legend: Sigma (Sigma-Aldrich, Taufkirchen, Germany), Abcam (Abcam, Cambridge, UK), Santa Cruz (Santa Cruz Biotechnology, Santa Cruz, USA) Invitrogen (Invitrogen, Carlsbad, USA)

2.6.1 Preparation and fixation of cells

HEK293T and HeLa cells were seeded on 12 mm glass coverslips positioned in a 24-well plate. The cells were normally cultured until suitable confluence was reached. To study an exogenous over-expressed protein, HEK293T cells were grown to a 50% confluence and transfected with the vectors of interest with PEI, as described previously. To study endogenous proteins, HEK293T and HeLa cells were grown until 40-50% of confluence was reached in order to better visualize isolated cells. To study effects of siRNA silencing in cells, HeLa were cultured as described in the siRNA chapter.

In order to detect the protein/proteins of interest, the whole cell´s proteome has to be immobilized (fixation). The ice-cold methanol fixation is used to precipitate the proteins through alcoholic dehydration. Cell morphology is preserved and the epitopes are not blocked by fixation, however aromatic rings belonging to fluorescent tag-proteins (GFP) are minimally disrupted, causing a lower fluorescence signal.

2.6.2 Staining

After the PBS washings of the coverslips to remove any trace of the fixating reagent, these were first incubated with PBS-TX for 10 minutes, in order to permeabilize the cell membranes. Afterwards cells were washed 3 times with PBS and incubated in blocking solution for 1 hour to block all the non-specific binding sites for antibodies. Cells were then washed once with PBS and incubated with primary antibody solution at 4°C overnight

according to the dilutions previously indicated (table 11). After the incubation with primary antibodies, cells are washed extensively with PBS and incubated with secondary antibody solution for 1 hour at room temperature. To the secondary solution, DAPI (4',6-diamidino-2-phenylindole) can be added to a final concentration of 0,1 μg/ml in order to visualize the cell nuclei. The secondary antibodies must be selected according to the animal where the primary antibodies were raised, in order to obtain a specific signal for each primary antibody. The coverslips were then washed 3 times for 10 minutes with PBS, rinsed with ddH_2O and mounted on glass slides using the Fluorsafe mounting solution. The mounted coverslips were left at room temperature protected from light for at least 2 hours prior to acquire images.

2.6.3 Image acquisition

After the solidification of the mounting solution, the coverslips were analyzed with fluorescent microscope. The pictures were acquired at 1300 x 1300 pixel resolution (2 Megapixel) through the Axiovision program provided with the microscope. All acquisitions were made with 40x and 60x magnification. Coverslips, after image acquisition, were stored at 4°C for long-term storage.

3 RESULTS

3.1 The C-terminal portion of SON confers nuclear targeting

According to current literature, SON´s function is tightly correlated to the correct subcellular localization. In particular, SON has been characterized in the nuclear environment in physiological conditions and exerts its function as splicing modulator here. Additionally an occasional altered localization has been reported, in particular a cytoplasmatic accumulation subsequent to a nuclear deprivation, in cell lines derived from leukemic donors [54]. These findings are pointing to a specific role of SON in cell survival and cell cycle progression highly dependent on its cellular compartmentalization. In order to coherently undertake studies on the interactome of the SON protein, it is therefore necessary to investigate whether the truncated form of the full length SON, the SONc fragment, still localizes to nuclei.

Therefore, the GFP tag was fused to the N-terminal portion of SONc and the cellular distribution of the chimerical protein was analyzed, in comparison to the GFP protein alone, in HEK293T cells. The GFP-SONc signal was distributed exclusively in the nucleus, presenting a punctuate pattern (figure 12, e and f). The localization of endogenous SON was also detected by the use of a SON specific antibody (figure 12, c and g). The GFP-SONc pattern partially but not fully resembles the endogenous SON distribution (figure 12, h)

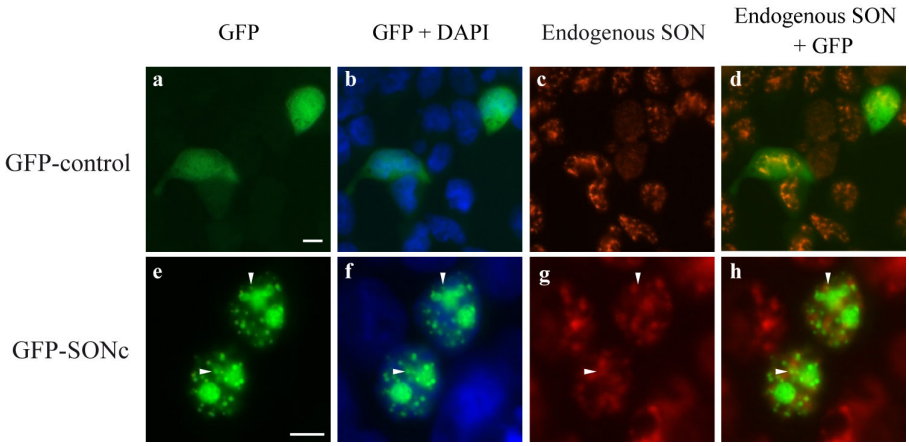

Figure 12: The C-terminal portion of SON is responsible for the nuclear localization. GFP-SONc has been produced and transfected in HEK 293T cells in comparison to GFP control protein. GFP-SONc shows a nuclear distribution (e-f) compared to the GFP control (a-b). GFP-SONc additionally shows a particular punctate staining detected in the nuclei (f). Endogenous SON antibody was used to evaluate potential co-localization (c and g). GFP-SONc presents a minor co-localization with the endogenous SON protein indicated with the white arrows (h).
Scale bar: 12.5 microns

3.2 Critical choice of protein-protein interaction methods

The fundamental criterion to perform a reliable and comprehensive network analysis is the choice of the most suitable technique that will be used to performed complex property studies. As will be discussed later, among the several techniques available four independent and partially overlapping approaches to perform interaction studies have been selected: the SF-AP, the yeast-two hybrid, the GST pull-down and the co-immunoprecipitation. These approaches were selected carefully comparing the interaction data benefits and taking in consideration the related drawbacks, such the organisms producing the bait protein, the occurring post-translational modification and the tissues used for the interaction studies. The combination of all these techniques has never been reported before in literature and represented an important challenge to face.

3.3 Identification of SONc interactors through affinity purification using strep-FLAG tagged SONc

3.3.1 Evaluation of SF-TAG single or two-step purification

The tandem affinity purification technique is based on a two-step purification using two different tags fused together in order to reduce mutual non-specific binders. In particular, the SF-tag is made by the fusion of two StrepII tags and a FLAG peptide sequence [33], and the resulting protein complexes are purer than in single step approaches and easier to analyze by mass spectrometry. However, the second step of purification can reduce significantly the identification of weak or transiently bound proteins belonging to the purified complexes. Therefore, the single step purification has been chosen to identify the interactors of SONc

3.3.2 Optimization of the C-terminal portion of SON expression and purification

The SONc cDNA sequence has been cloned with the SF-tag at the N-terminus and it was transfected in HEK293T cells. As shown in figure 13 the tagged SONc could be detected 48 hours after transfection with the anti-FLAG antibody not presenting any degradation pattern, aggregates formation nor evident cell toxicity (here evaluated by cell death).

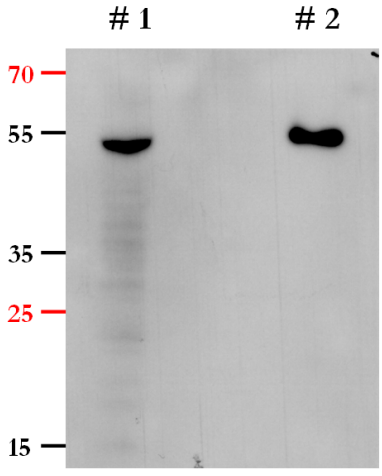

Figure 13: Expression and purification of SF-SONc. In the figure the expression levels of SF-SONc after transfection in HEK293T cells (lane 1) and after FLAG purification (lane 2) are shown. SF-SONc transfected cells were lysed in NP-40 lysis buffer and the detection of SF-SONc in the whole cell lysate was performed using the anti-FLAG antibody. The SF-SONc band is expected to be around 52 KDa.

3.3.3 Quantitative analysis of SF-SONc interacting partners through the SILAC approach

The preliminary data obtained from pilot experiments of SF-AP on SONc protein, provided qualitative information about the interacting partners of SONc. However, our interest was to increase the quality of our data by providing an additional quantitative dimension to the SONc's interactome. Given that the SF-AP procedure has been designed to be performed in cell culture systems, protein labeling can be easily performed using the SILAC technique (Stable Isotope Labeling by Amino acids in Cell culture), based on the incorporation of normal or stable isotope labelled arginine and lysine amino acids into newly built proteins. For this quantitative analysis the amino acids in the media were substituted either with normal light (Arg0, Lys0) or heavy (Arg8, Lys10) amino acids. As final output, specific interactors of SONc will result significantly enriched in the purified complexes containing the SF-SONc protein compared to the negative control.

Additionally, it has been critically discussed what the most suitable negative control for these experiments is, thereby having to choose between an empty vector and the same vector used for SONc containing an unrelated protein. We concluded that an empty vector carrying only the tag has a cellular impact not comparable with the tagged protein of interest, and the cell machinery responsible for protein synthesis is not overloaded as in the case of the SF-SONc [86]. According to that I choose to fuse the SF-tag to the N-terminal of the GFP protein, obtaining both a suitable control for protein complexes purification and a transfection efficiency control. SF-SONc and SF-GFP plasmids were transfected into HEK293T cells (previously labeled with heavy and light SILAC amino acids) and a single-step FLAG-affinity purification (SF-AP) was performed for both samples. After the purification, the protein complexes were eluted with Laemmli buffer. These eluates were prefractionated and in-gel digested and later subjected to mass spectrometric analysis and the proteins were identified and quantified with MaxQuant software [87]. In order to minimize artifacts due to the differential SILAC media, the experiments were replicated inverting the labeled amino acids for the SONc and GFP control. Only those proteins which were identified with at least 2 unique peptides with a significance over 95% ($p<0.05$) were considered to be specific interactors. As a result we detected 48 potential interactors of SONc due to their significant enrichment together with SF-SONc compared to the GFP

control (table 12). Among these newly identified interactors, twelve were reported to belong to the splicing machinery (table 13). Three of these proteins have been reported to be directly involved in the splicing processes, DHX9 [88], DDX5 [89] and IGF2BP1 [90-91]. Further gene ontology analysis of the identified proteins, identified 4 main splicing sub-components involved: 1) the mRNP complex, represented by four proteins (DBPA [92], DHX9, IGF2BP1 and PABPC1 [93]), which is involved in the binding of nascent mRNA [94]; 2) the mRNP-granule complex, here represented by the same four proteins identified for the mRNP complex plus five ribosomal proteins (RPS3 [95], RSP3A [96], RPS4X [97], RPS8 [98] and RPS9 [99]), which is a late step of maturation and translation of mRNA; 3) the spliceosome C complex, represented by PABPC1, helicases (DDX5) and heterogeneous ribonucleoproteins (HNRNPH1 [100-101] and HNRNPA1 [102]), which is one of the transitory complexes during the splicing of introns from the newly transcribed template [103]; 4) the major coding-region determinant of instability complex (CRD), here identified with DHX9 and IGF2BP1, which is a quality control complex responsible for degradation of undesired mRNA [104].

These findings are underlining the cross-connectivity occurring between the two major identified complexes (spliceosome C and mRNP-/mRNP granule-complex) and are pointing toward a functional role of the SON protein in both splicing sub-complexes and most probably as a scaffold adapter for them. Further confirmation of this hypothetical role can be identified by the phenotypic observations following SON depletion [55, 70-71].

UniProt ID	Gene name	Identified protein	SONc/GFP ratio	Unique Peptides	SONc/GFP Significance
P18583	SON	SON DNA binding protein	4,597474	498	8,37E-05
		Nuclear proteins			
O43143	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	6,038698	869	5,68E-06
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3	3,9063	3	0,000349976
Q08211	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9	3,821362	276	0,000420204
Q9NZI8	IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1	2,458512	63	0,00999192
P16989	DBPA	DNA-binding protein A	2,881872	12	0,003557788
O15226	NKRF	NF-kappa-B-repressing factor	2,753177	112	0,004847628
P05455	SSB	Lupus La protein	2,234661	69	0,017537264
P11940	PABPC1	Polyadenylate-binding protein 1	2,074977	74	0,026354278
P17844	DDX5	Probable ATP-dependent RNA helicase DDX5	2,036425	67	0,029095975
Q9BVP2	GNL3	Guanine nucleotide-binding protein-like 3	2,010966	24	0,031064894
Ribonucleoproteins					
Q13151	HNRNPA0	Heterogeneous nuclear ribonucleoprotein A0	3,1151	3	0,002052142
Q9BUJ2	HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like 1	3,079954	128	0,002227649
Q5JTH9	RRP12	RRP12-like protein	3,221317	30	0,001604352
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H	2,896938	144	0,003432154
P61247	RPS3A	40S ribosomal protein S3a	2,673829	48	0,005878092
P09651	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	2,122024	13	0,023363317
		Not nuclear proteins			
Ribonucleoproteins					
P40429	RPL13A	60S ribosomal protein L13a	5,873938	5	7,59E-06
P84098	RPL19	60S ribosomal protein L19	4,557521	20	9,07E-05
P46776	RPL27A	60S ribosomal protein L27a	4,516796	13	9,85E-05
P62241	RPS8	40S ribosomal protein S8	3,001646	20	0,002677499
P26373	RPL13	60S ribosomal protein L13	2,912401	8	0,00330802
P62277	RPS13	40S ribosomal protein S13	2,819736	25	0,004128829
P62847	RPS24	40S ribosomal protein S24	2,80672	6	0,004260111
P46781	RPS9	40S ribosomal protein S9	2,661146	38	0,006062847
P15880	RPS2	40S ribosomal protein S2	2,478432	69	0,009509044
P27635	RPL10	60S ribosomal protein L10	2,413409	10	0,011181479
P62269	RPS18	40S ribosomal protein S18	2,386603	31	0,01195709
P62701	RPS4X	40S ribosomal protein S4, X isoform	2,375916	49	0,012281636
P08865	RPSA	40S ribosomal protein SA	2,358143	113	0,01284168
P60866	RPS20	40S ribosomal protein S20	2,336174	14	0,013570672
P62263	RPS14	40S ribosomal protein S14	2,253744	8	0,016709333
P46782	RPS5	40S ribosomal protein S5	2,230021	28	0,017744908
P62249	RPS16	40S ribosomal protein S16	2,19318	49	0,019485921
P23396	RPS3	40S ribosomal protein S3	2,189348	139	0,019676872
P62857	RPS28	40S ribosomal protein S28	2,187849	13	0,019752049
P05386	RPLP1	60S acidic ribosomal protein P1	2,1425	5	0,022172535
P62851	RPS25	40S ribosomal protein S25	2,117793	11	0,023617443
P18621	RPL17	60S ribosomal protein L17	2,1003	3	0,024698579
P18077	RPL35A	60S ribosomal protein L35a	2,0287	2	0,029679538
Miscellaneous					
Q9H6S0	YTHDC2	Probable ATP-dependent RNA helicase	3,560073	86	0,000745209

UniProt ID	Gene name	Identified protein	SONc/GFP ratio	Unique Peptides	SONc/GFP Significance
Q9UPY3	DICER1	Endoribonuclease Dicer	3,383808	313	0,001106785
Q6P158	DHX57	Putative ATP-dependent RNA helicase DHX57	2,973177	97	0,002863738
P25786	PSMA1	Proteasome subunit alpha type-1	2,657313	4	0,006119872
Q6PKG0	LARP1	La-related protein 1	2,323321	57	0,014016851
Q15084	PDIA6	Protein disulfide-isomerase A6	2,271624	24	0,015970124
P35520	CBS	Cystathionine beta-synthase	2,203514	5	0,018980515
Q7L2E3	DHX30	Putative ATP-dependent RNA helicase DHX30	2,13726	147	0,022471217
Q13867	BLMH	Bleomycin hydrolase	2,031661	7	0,029454457

Table 12: Interactors list from the SF-AP approach. In this table the proteins identified with the single-step SF-tag affinity purification technique using SF-SONc fusion protein expressed in SILAC labeled HEK293T cells are listed. The proteins identified have been sorted in nuclear (including the ribosomal and ribonucleoprotein sub-groups) and not nuclear (including the ribosomal, the ribonucleoprotein and miscellaneous sub-groups).. In the last 2 columns the number of peptides used for quantification and the significance of the ratio SONc/GFP identified are indicated.

UniProt ID	Identified protein	Function	Splicing sub-machineries
P16989	DBPA	Posttranscriptional regulation of mRNA	mRNP, mRNP-G
Q08211	DHX9	Transcription regulation	mRNP, mRNP-G, CRD
Q9NZI8	IGF2BP1	mRNA stabilization and cancer promoter	mRNP, mRNP-G,CRD
P11940	PABPC1	Regulation of mRNA transcription	mRNP, mRNP-G, Spl.C
P23396	RPS3	Ribosome biogenesis and DNA repair	mRNP-G
P61247	RPS3A	Promotes cell proliferation	mRNP-G
P62701	RPS4X	Promotes DNA repair X-linked	mRNP-G
P62241	RPS8	Regulation of translation	mRNP-G
P46781	RPS9	Regulation of translation and cancer growth arrest	mRNP-G
P17844	DDX5	mRNA maturation and export	Spl.C
P09651	HNRNPA1	Post-translational control of mRNA	Spl.C
P31943	HNRNPH1	Post-translational control of mRNA	Spl.C

Table 13: Interacting partners of SF-SONc belong to 4 major splicing sub-machineries. Analyzing the overall proteins identified with SONc, we identified among them 12 interacting partners belonging to the splicing machinery and in particular to 4 specific sub-machineries: the mRNP complex (mRNP), the mRNP granule-complex (mRNP-G), the major coding-region determinant of instability complex (CRD) and the spliceosome C (Spl. C).

3.4 Identification of SONc interactors through the yeast-two hybrid technique

3.4.1 Optimization of the SONc bait expression

The C-terminal portion of SON was N-terminal fused to the pBD domain of the GAL4 transcription factor and expressed in PJ69-4A competent yeast cells using 1 μg of bait

DNA. Self-activation of the bait was tested on SD -W plates allowing the cells to grow and the resulting colonies were further cultured on SD –WH and SD -WHA plates. After 2 days of incubation, no colonies appeared on the SD –WH/-WHA plates confirming that the SONc bait was not self-activating and the construct was suitable for proceeding with the yeast-two hybrid screening.

3.4.2 pBD-SONc interacting partners identified two major splicing sub-complexes

As described in the methods section, the yeast cells transformed with the pBD-SONc were mated with yeast cells containing oligo-dT cDNA libraries produced form placental and retinal tissues of human origin. In order to obtain only strong interactions and to reduce the false positive occurrence, highly stringent conditions of growth were applied for the screening, using SD –WLHA plates. 480 colonies were collected and their cDNA, after extraction, was back transformed in DH5α E. coli bacteria under selective antibiotics in order to amplify the cDNA of the pray proteins. The resulting nucleotide sequences were blasted with nucleotide-BLAST (NCBI) and false positive removed. A total of 32 proteins were identified as binary interactors of SONc (table 14) and 3 of them have been linked to the splicing machinery: STAU1 [105-106], ILF3 [107] and HNRNPH1 (table 15). Each of these proteins, differently from SF-AP analysis, is belonging only to a single complex: STAU1 and ILF3 have been described in the mRNP-/mRNP granule-complex, while HNRNPH1 belongs to the spliceosome C. According to these findings, SONc is directly interacting with fundamental and structural proteins involved in the formation of the two complexes, as reported for the scaffolding role of STAU1 [106]

The re-identification of HNRNPH1 as SONc interacting partner confirms two major findings: 1) the interaction between the two proteins occurs through a binary interaction and physical contact; 2) this interaction, already identified with the SF-AP approach, proof that different proteomic techniques can be reasonably combined, obtaining precious information (binary interaction) which could be misread or ignored using only a single approach.

UniProt ID	Gene name	Identified protein	Independent clones
		Nuclear proteins	
P18583	SON	SON DNA binding protein	18
Q15633	TARBP2	TAR (HIV-1) RNA binding protein 2	15
Q12906	ILF3	Interleukin enhancer binding factor 3	9
Q569K4	ZNF385B	Zinc finger protein 385B	3
P63165	SUMO1	SMT3 suppressor of mif two 3 homolog 1 (SUMO1)	2
P63279	UBE2I	Ubiquitin-conjugating enzyme E2I	2
Q8NDW4	ZNF248	Zinc finger protein 248	2
O75928	PIAS2	Protein inhibitor of activated STAT, 2	1
Q96PM9	ZNF385A	Zinc finger protein 385A	1
Q9UBB5	MBD2	Methyl-CpG binding domain protein 2	1
Q9H898	ZMAT4	Zinc finger, matrin type 4	1
Q9HCU9	BRMS1	Breast cancer metastasis suppressor 1	1
O95833	CLIC3	Chloride intracellular channel 3	1
Q08117	AES	Amino-terminal enhancer of split	1
Q12873	CHD3	Chromodomain helicase DNA binding protein 3	1
Ribonucleoprotein			
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H1	3
		Not nuclear proteins	
O75569	PRKRA	Interferon-inducible double stranded RNA dependent activator	35
P14902	INDO	Indoleamine-pyrrole 2,3 dioxygenase	4
P09912	IFI6	Interferon, alpha-inducible protein 6	3
P02747	C1QC	Complement component 1, q subcomponent, C chain	2
P07711	CTSL1	Cathepsin L1	2
Q7Z5U7	CTSC	Cathepsin C	2
O75843	AP1G2	Adaptor-related protein complex 1, gamma 2 subunit	1
P07858	CTSB	Cathepsin B	1
P14174	MIF	Macrophage migration inhibitory factor	1
Q14152	EIF3A	Eukaryotic translation initiation factor 3, subunit A	1
O95793	STAU1	Staufen, RNA binding protein, homolog 1	1
Q86YQ8	CPNE8	Copine VIII	1
Q8N1I0	DOCK4	Dedicator of cytokinesis 4	1
Q96PE1	GPR124	G protein-coupled receptor 124	1
Q96RW7	HMCN1	Hemicentin 1	1
Q9NX74	DUS2L	Dihydrouridine synthase 2-like, SMM1 homolog	1

Table 14: Interactors list from yeast two-hybrid approach. In this table the proteins identified with the yeast two hybrid technique using pBD-SONc fusion protein against retinal and placental cDNA libraries from human origin are listed. The proteins identified have been sorted in nuclear (including the ribonucleoprotein sub-group) and not nuclear proteins. In the last column the numbers of independent clones identifying the specific interactors are indicated.

UniProt ID	Identified protein	Function	Splicing sub-machineries
O95793	STAU1	mRNA stability and regulation	mRNP
Q12906	ILF3	mRNA stability and translation control	mRNP, mRNP-G
P31943	HNRNPH1	Post-translational control of mRNA	Spl.C

Table 15 Interacting partners of pBD-SONc belong to 3 major splicing sub-machineries. Analyzing the overall proteins identified with pBD-SONc, we identified 3 interacting partners belonging to the splicing machinery and in particular to the mRNP complex (mRNP), the mRNP granule-complex (mRNP-G) and the spliceosome C (Spl. C).

3.5 Identification of SONc interactors by GST pull down

3.5.1 Optimization of the SONc expression and purification

According to the previous approaches, the sequence of the glutathione S-transferase (GST) protein was fused to the N-terminal end of SONc in the pDEST-15 vector, suitable for BL21 (DE3) bacteria expression of the protein (T7 promoter). Given that changes in temperature, cell concentration, amount of inducing IPTG and time of induction can highly influence the production and stability of the protein, expression tests were performed. Finding the best conditions has been a complicated task because modifying one induction parameter requires the modification of a second or even a third parameter. For example, it has been observed that the reduction of the induction temperature from 37°C to 30°C was producing a protein with lower degradation pattern. However, in order to achieve a protein amount comparable to the 37°C condition, the time of induction needed to be doubled. After finding the suitable induction conditions, the purification procedure for the GST-SONc protein has been carefully considered. The first step involves the lysozyme, an N-acetylmuramide glycanhydrolase, used to digest the peptidoglycans that are forming the cell wall of bacteria. The second step requires a detergent able to permeabilize the membranes and to solubilize the overall protein content. The choice of the right detergent can dramatically influence the solubility and the extraction amounts of the protein of interest. Results from several tests showed that the detergent Sarcosyl (N-Lauroylsarcosine) at a concentration of 1% was able to extract the most of the GST-SONc protein and maintained it stable in the soluble phase (figure 14).

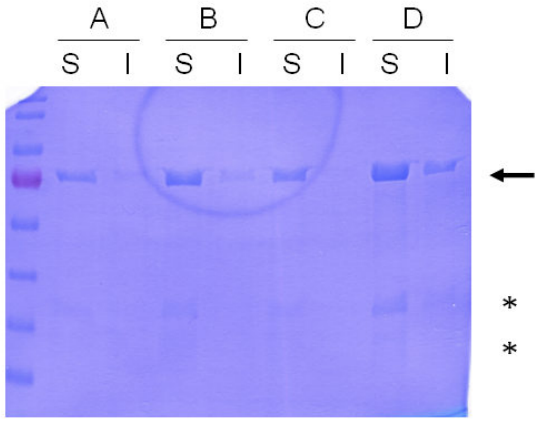

Figure 14: Induction and purification of GST-SONc. In the figure the different yields of GST-SONc protein extracted and obtained by glutathione sepharose purification are shown. Different extractions were performed for GST-SONc maintaining constant the composition of the buffer, changing only the solvents used for the lysis. STE buffer was added (**A**) with 0,5% NP-40 (v/v), (**B**) with 1% Triton X-100 (v/v), (**C**) with the combination of 0,5% NP-40 (v/v) and 1% Triton X-100 (v/v) or (**D**) with 1% (v/v) Sarcosyl. Soluble proteins (S) and insoluble proteins (I) following lysis are loaded for comparison. GST-SONc purified protein is indicated by the arrow (molecular weight close to 76 KDa) while bands indicating a protein degradation are indicated with asterisks (*).

3.5.2 GST-SONc interacting partners in placental tissue.

As previously described, the GST-SONc fused protein was purified from bacteria and quantified through Bradford colorimetric assay and band-intensity comparison on a Coomassie blue-stained gel. The freshly frozen placental tissue collected after abdominal baby delivery, was lysed in NP-40 buffer (see materials and methods). Equal amounts of placenta lysates were incubated with 5 µg of GST protein, as negative control, or 5µg GST-SONc. After the elution of the proteins interacting with SONc, the complexes were subject to mass spectrometric analysis and proteins were identified with the Scaffold software. Coherently to the previous methods, the interacting partners were accepted only when identified in at least 2 of 3 independent experiments, by at least 2 unique peptides and with a minimum of 95% identification probability ($p<0.05$). From the list of interacting partners obtained, the proteins identified also with the GST-alone control were removed and considered as non-specific binders. Fourteen putative SONc´s interactors have been identified (table 16). Among the interactors identified, two proteins were already been identified as SONc interactors with the SF-AP approach: DHX9, a significant member of the splicing machinery (table 17), and DHX15. These findings are increasing the number

(and the significance) of the commonly identified proteins using different approaches and consequently confirming the correctness of the method combination hypothesis.

UniProt ID	Gene name	Identified protein	Total peptides		
			Exp 1	Exp 2	Exp 3
P18583	SON	SON DNA binding protein	22	21	20
		Nuclear proteins			
O43143	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	32	32	31
P62826	RAN	GTP-binding nuclear protein Ran	4		2
Q08211	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9	3		2
Ribonucleoproteins and ribosomal proteins					
Q08ES8	RPL11	60S ribosomal protein L11	2		2
		Not nuclear proteins			
Ribosomal proteins					
P62888	RPL30	60S ribosomal protein L30	3	2	3
Miscellaneous					
P10301	RRAS	Related RAS viral (r-ras) oncogene homolog	3		4
P18085	ARF4	ADP-ribosylation factor 4	2		2
P48047	ATP5O	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	2		3
P49411	TUFM	Elongation factor Tu	3	3	3
P84077	ARF1	ADP-ribosylation factor 1		2	2
Q13576	IQGAP2	IQ motif containing GTPase activating protein 2		2	2
Q8IZ29	TUBB2C	Tubulin, beta 2C	4	3	4
Q9H0U4	RAB1B	Ras-related protein Rab-1B	2	3	
Q9Y6Q1	CAPN6	Calpain 6		3	2

Table 16: Interactors list from the GST pull down approach. In this table the protein identified with the GST pull down technique using GST-SONc fusion protein against placental tissue lysate from human origin are listed. From the original list all the proteins that were also identified with the GST tag alone have been removed. In the last 3 columns the numbers of unique peptides identified for each protein in 3 independent experiments with an identification probability over 95% ($p < 0.05$) are indicated.

UniProt ID	Identified protein	Function	Splicing sub-machineries	Methods
Q08211	DHX9	Transcription regulation	mRNP, mRNP-G, CDR	SF-AP, GST

Table 17: One interacting partner of GST-SONc belongs to the splicing machinery. Analyzing the overall proteins identified with GST-SONc, the DHX9 protein was reported as member of the mRNP complex (mRNP), the mRNP granule-complex (mRNP-G) and the major coding-region determinant of instability complex (CRD).

3.6 Identification of interacting partners of endogenous SON in HeLa cells by co-immunoprecipitation

3.6.1 Basic requirements for immunoprecipitation of endogenous SON

The analysis of protein complexes by the immunoprecipitation of the endogenous protein is the method of choice if a suitable antibody is available. This technique allows the direct study of endogenous and physiological protein complexes containing a protein of interest since interacting partners occurring due to the protein over-expression or by incorrect folding are excluded. In order to perform any sort of study on endogenous proteins, two main prerequisites are necessary: the availability of a suitable antibody and a at least moderate expression of the protein of interest. The first condition is one of the most challenging for protein studies since this requires the production of an antibody able to recognize the tertiary structure of a target protein and its eventual post-translational modifications (figure 15). The second condition is a less stringent prerequisite since in case of a lowly expressed protein, starting material can be increased. Nevertheless, abundantly expressed proteins require less starting material, reducing drastically the amount of potential non-specific binders.

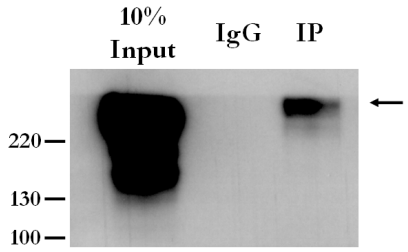

Figure 15: Immunoprecipitation test of endogenous SON. In this figure the western blot of the immunoprecipitation of endogenous SON from HeLa lysate is shown. The Sigma Prestige antibody for SON was the only commercially available antibody able to recognize and immunoprecipitate endogenous SON protein efficiently. In the figure SON levels in the input from HeLa cells (first lane), the rabbit generic IgGs as control (second lane) and the specific SON antibody (third lane) are compared. From this figure the functionality of the SON antibody and the lack of SON´s band in the IgGs control can be appreciated.

3.6.2 Quantitative analysis of endogenous complexes of endogenous SON by the SILAC approach

After the identification of the complexes interacting with the SONc protein, the interest for complex analysis moved to decipher protein complexes formed by the complete

endogenous SON protein. The immunoprecipitation was performed with HeLa cells since these cells are showing moderate levels of SON protein expression, satisfying the pre-requisites [42-43, 50]. The SILAC labeling approach was applied, as described for the SF-AP experiments (chapter 3.3.3), to obtain quantitative data about the enriched complexes formed. As reported before, a label switch was performed to reduce SILAC media-derived artifacts. Equal amounts of HeLa cell extracts were incubated with 1 microgram of specific SON antibody or with control rabbit IgGs. Eluates were combined and subject to mass spectrometric analysis. MaxQuant software was used to identify and quantify the interacting partners of SON. Twenty-eight potential interactors of the endogenous SON were identified according to their significant enrichment compared to the control and enrichment significance >95%, $p<0.05$). Interacting partners were considered significant only if identified in at least 2 of 4 independent experiments, with at least 2 unique peptides (table 18). Among the potential SON interacting partners 4 nuclear proteins have been reported previously as SON interactors [55], RBM25, RBM8A [108], PRPF8 [109], snRNP200 [110], the latest 3 have been correlated with the mRNA splicing. These 4 detected proteins are increasing the confidence in the feasibility of our approach. Three of these four proteins (RBM8A, PRPF8 and snRNP200) and three additional proteins (the SRRM2 and the heterogeneous ribonucleoproteins HNRNPC and HNRNPM) have been all linked to the splicing machinery and are belonging to the spliceosome C complex. (table 19). In particular, the RBM8A protein has been identified as member of the spliceosome C and of the mRNA splicing dependent exon-junction-complex (EJC) [108].

Noticeably the DHX15 protein has been identified also with this method, further confirming the feasibility of the combination of approaches since it was previously identified with both the SF-affinity purification and the GST pull-down methods.

UniProt ID	Gene Name	Identified protein	Ratio SON/IgG Normalized			Unique Peptides	SON-IP Significance
			Exp1	Exp2	Exp3		
P18583	SON	SON protein	35,387	27,869	24,842	25	1,24E-35
		Nuclear proteins					
Q9UBK9	UXT	Protein UXT	20,207	17,515	n. def.	4	1,61E-27
P19388	POLR2E	DNA-directed RNA polymerases I, II, and III subunit RPABC1	18,219	8,8511	8,8428	5	2,46E-20
P01100	FOS	Proto-oncogene protein c-fos	26,413	2,9539	0,054048	1	1,50E-17
O94763	RMP	Unconventional prefoldin RPB5 interactor	7,6252	6,2076	n. def.	4	3,34E-13
Q96A72	MAGOHB	Protein mago nashi homolog 2	6,2361	5,083	n. def.	1	5,16E-11
O43143	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	5,9779	3,8066	2,9583	19	2,87E-08
Q9NR30	DDX21	Nucleolar RNA helicase 2	3,9714	3,7	n. def.	6	2,11E-07
Q9Y5S9	RBM8A	RNA-binding protein 8A	4,939	4,0406	1,6305	5	9,42E-07
P0C0S8	HIST1H2AA	Histone H2A type 1-A;	3,2077	5,0501	1,1704	1	7,15E-06
Q99877	HIST1H2BN	Histone H2B	3,1284	4,8752	1,0811	3	1,30E-05
P62805	HIST2H4	Histone H4	3,0387	4,95	0,89833	7	1,85E-05
P84243	H3F3B	Histone H3.3	3,0426	4,5987	1,008	3	2,81E-05
Q9UQ35	SRRM2	Serine/arginine repetitive matrix protein 2	3,8113	2,2279	1,1303	12	0,0004027
O75643	SNRNP200	U5 small nuclear ribonucleoprotein 200 kDa helicase	3,3581	2,7027	0,83136	19	0,000667
P63167	DYNLL1	Dynein light chain 1	2,2203	2,0777	n. def.	2	0,0015007
P31944	CASP14	Caspase-14	2,5453	3,0916	0,48822	8	0,002695
Ribonucleoproteins							
P07910	HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	4,1231	4,7049	1,2225	13	2,44E-06
P49756	RBM25	RNA-binding protein 25	5,2182	4,3107	1,9413	6	2,25E-07
Q15365	PCBP1	Poly(rC)-binding protein 1	3,1338	2,3808	2,3537	6	0,0001136
Q6P2Q9	PRPF8	Pre-mRNA-processing-splicing factor 8	3,1442	2,9075	n. def.	11	1,32E-05
Q14978	NOLC1	Nucleolar phosphoprotein p130	2,3758	2,4774	n. def.	5	0,0003296
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	2,4609	2,2155	1,6941	41	0,001725
		Not nuclear proteins					
Miscellaneous							
Q9H6T3	RPAP3	RNA polymerase II-associated protein 3	17,767	10,921	n. def.	7	4,69E-23
O15212	PFDN6	Prefoldin subunit 6	15,118	5,7231	3,7476	5	3,13E-15
Q9UHV9	PFDN2	Prefoldin subunit 2	10,702	4,4969	3,3462	4	5,98E-12
P08670	VIM	Vimentin	3,7843	4,6417	1,1585	29	5,45E-06
P84090	ERH	Enhancer of rudimentary homolog	4,2784	3,7901	1,242	4	8,78E-06
Q01469	FABP5	Fatty acid-binding protein	2,2341	2,054	3,08	12	0,0002808

Table 18: Interactors list from immunoprecipitation of the endogenous SON. In this table the proteins identified with the immunoprecipitation of endogenous SON in SILAC labeled HeLa cells are listed.. In the column labeled "SON/IgG ratio normalized" the ratios identified for each protein in the 3 independent experiments used for quantification are listed. In the last 2 columns the number of peptides used for quantification and the significance of the ratio SON/IgG identified are indicated. Not available ratios are indicated with "n. def.".

UniProt ID	Identified protein	Function	Splicing sub-machineries
O75643	snRNP200	mRNA splicing	Spl.C
P07910	HNRNPC	Formation of 40-s hnRNP particles	Spl.C
P52272	HNRNPM	Splicing regulation	Spl.C
Q6P2Q9	PRPF8	Spliceosome formation and mRNA splicing]	Spl.C
Q9UQ35	SRRM2	Spliceosome formation and mRNA splicing	Spl.C
Q9Y5S9	RBM8A	Exon-exon junction complex formation	Spl.C-EJC

Table 19: Interacting partners of endogenous SON belong to 2 major splicing sub-machineries. Analyzing the overall proteins interacting with the endogenous SON, we identified 6 proteins all belonging to the spliceosome C (Spl. C). In particular, for the RBM8A protein the major coding-region determinant of instability complex (CRD) was identified.

3.7 Combination of protein interactions from different experimental data sets.

The first aim of this study was to identify SON and SONc interacting partners with different techniques. However, great expectation was given to the possibility to combine and integrate different methods, and related datasets, in order to obtain a more comprehensive description of SON´s interactome. The combination of different techniques can appear laborious and not worth of it, but in case of newly described proteins this effort is providing much more information than a single approach. Particular attention was given to the differences of the baits used. More in detail the impact of the different cell systems used: a simple eukaryotic system (yeast), a more complex eukaryotic system (human cell lines) and a prokaryotic system (*E. coli*). Nevertheless, including suitable and stringent controls and considering differences at the post-translational level within the systems, the obtained data sets can be compared with a reasonable confidence. On the other hand using a single approach for protein complexes studies, it would have not been possible to discriminate the nature of the protein interactions (binary or indirect interactors) and the impact that a tag can produce at the protein-protein interaction level.
Combining the previously produced interaction datasets from the different approaches (SF-AP, yeast two-hybrid, GST pull-down and immunoprecipitation), a total of 117 unique proteins have been identified. Interestingly only 3 of them have been detected in at least 2 different approaches. However these proteins have been reported to be involved in mRNA splicing, in particular the DHX9 (belonging to the mRNP/mRNP-granule complexes) and the HNRNPH1 (belonging to the spliceosome C) proteins (table 20).

Uniprot ID	Gene name	Identified protein	Identifying techniques
O43143	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	SF-AP, GST, eIP
Q08211	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9	SF-AP, GST
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H1	SF-AP, Y2H

Table 20: Combination of the different approaches allows the identification of three common interactors. After the collection of interactions datasets obtained from different techniques, we combined critically all the identified interactors of SON and SONc. We were able to identify 3 proteins identified with more than one method. In the last column the methods used to identify these proteins as SON´s interacting partners are indicated.

3.7.1 The SON protein interacts mainly with nuclear, ribosomal and ribonucleoproteins.

Resolving the sub-cellular and organelle localization of the interacting partners of the protein of interest (the SON protein in this study) is providing precious information about the putative function or mechanism in which this protein could be involved. SON´s nuclear localization has been extensively described [43, 55, 71] and expectations were pointing towards a prevalent identification of nuclear interactors. However, differently from expectations but still meaningful enough, 53 identified proteins were described as nuclear proteins (45% of the total). Analyzing these proteins more in detail showed that 15 nuclear interactors of SON have been identified with SF-AP (32% of total SF-AP identified proteins), 15 interactors with yeast-two hybrid technique (47% of total Y2H identified proteins), 4 nuclear proteins were identified with GST pull-down (29% of total GST pull-down identified proteins), and 22 proteins performing the immunoprecipitation of endogenous SON (79% of endogenous complex identified proteins) (table 21). These findings, according to the high percentage of nuclear protein identified in each approach, are showing that the different constructs of SONc are still able to interact with proteins that are physically present in the same sub-cellular compartment where the endogenous SON is localized.

More in detail, among the overall identified interacting partners of SON (and SONc), 37 proteins (32% of the total identified interactors) have been described as ribonucleoproteins and nucleic acid-associated proteins. In particular, 13 of them are nuclear interacting partners of SON (one-third of total nuclear interactors identified). This group includes both the ribonucleoproteins and the ribosomal proteins. The first ones are involved in a wide range of cellular processes such as export, association and translation of mature mRNA. The second ones are presenting the mRNA to be translated to the ribosomes for

protein production. In particular some ribosomal proteins have been detected at nuclear level, involved in the export of mature mRNA from the nucleus to the ribosome [103-104]. All together these findings are pointing toward a putative role of SON as scaffold for mRNA-processing factors and for loading mature mRNA on nucleus-to-ribosome transporters.

Uniprot ID	Gene name	Identified protein	Identifying techniques
Nuclear proteins			
Q08117	AES	Amino-terminal enhancer of split	Y2H
Q9HCU9	BRMS1	Breast cancer metastasis suppressor 1	Y2H
P31944	CASP14	Caspase-14	eIP
Q12873	CHD3	Chromodomain helicase DNA binding protein 3	Y2H
O95833	CLIC3	Chloride intracellular channel 3	Y2H
P16989	DBPA	DNA-binding protein A	SF-AP
Q9NR30	DDX21	Nucleolar RNA helicase 2	eIP
P17844	DDX5	Probable ATP-dependent RNA helicase DDX5	SF-AP
O43143	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	GST, eIP, SF-AP
Q08211	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9	GST
P63167	DYNLL1	Dynein light chain 1	eIP
P01100	FOS	Proto-oncogene protein c-fos	eIP
Q9BVP2	GNL3	Guanine nucleotide-binding protein-like 3	SF-AP
P84243	H3F3B	Histone H3.3	eIP
P0C0S8	HIST1H2AA	Histone H2A type 1-A;	eIP
Q99877	HIST1H2BN	Histone H2B	eIP
P62805	HIST2H4	Histone H4	eIP
Q9NZI8	IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1	SF-AP
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3	SF-AP
Q12906	ILF3	Interleukin enhancer binding factor 3	Y2H
Q96A72	MAGOHB	Protein mago nashi homolog 2	eIP
Q9UBB5	MBD2	Methyl-CpG binding domain protein 2	Y2H
O15226	NKRF	NF-kappa-B-repressing factor	SF-AP
P11940	PABPC1	Polyadenylate-binding protein 1	SF-AP
O75928	PIAS2	Protein inhibitor of activated STAT, 2	Y2H
P19388	POLR2E	DNA-directed RNA polymerases I, II, and III subunit RPABC1	eIP
P62826	RAN	GTP-binding nuclear protein Ran	GST
Q9Y5S9	RBM8A	RNA-binding protein 8A	eIP
O94763	RMP	Unconventional prefoldin RPB5 interactor	eIP
O75643	SNRNP200	U5 small nuclear ribonucleoprotein 200 kDa helicase	eIP
Q9UQ35	SRRM2	Serine/arginine repetitive matrix protein 2	eIP
P05455	SSB	Lupus La protein	SF-AP
P63165	SUMO1	SMT3 suppressor of mif two 3 homolog 1 (SUMO1)	Y2H
Q15633	TARBP2	TAR (HIV-1) RNA binding protein 2	Y2H
P63279	UBE2I	Ubiquitin-conjugating enzyme E2I	Y2H
Q9UBK9	UXT	Protein UXT	eIP
Q9H898	ZMAT4	Zinc finger, matrin type 4	Y2H
Q8NDW4	ZNF248	Zinc finger protein 248	Y2H
Q96PM9	ZNF385A	Zinc finger protein 385A	Y2H
Q569K4	ZNF385B	Zinc finger protein 385B	Y2H
Nuclear ribonucleoproteins			
Q13151	HNRNPA0	Heterogeneous nuclear ribonucleoprotein A0	SF-AP
P09651	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	SF-AP
P07910	HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	eIP
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H	Y2H, SF-AP
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	eIP
Q9BUJ2	HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like 1	SF-AP
Q14978	NOLC1	Nucleolar phosphoprotein p130	eIP
Q15365	PCBP1	Poly(rC)-binding protein 1	eIP
Q6P2Q9	PRPF8	Pre-mRNA-processing-splicing factor 8	eIP
P49756	RBM25	RNA-binding protein 25	eIP
Q08ES8	RPL11	60S ribosomal protein L11	GST
P61247	RPS3A	40S ribosomal protein S3a	SF-AP
Q5JTH9	RRP12	RRP12-like protein	SF-AP

Uniprot ID	Gene name	Identified protein	Identifying techniques
Not nuclear ribonucleoproteins			
P27635	RPL10	60S ribosomal protein L10	SF-AP
P26373	RPL13	60S ribosomal protein L13	SF-AP
P40429	RPL13A	60S ribosomal protein L13a	SF-AP
P18621	RPL17	60S ribosomal protein L17	SF-AP
P84098	RPL19	60S ribosomal protein L19	SF-AP
P46776	RPL27A	60S ribosomal protein L27a	SF-AP
P62888	RPL30	60S ribosomal protein L30	GST
P18077	RPL35A	60S ribosomal protein L35a	SF-AP
P05386	RPLP1	60S acidic ribosomal protein P1	SF-AP
P62277	RPS13	40S ribosomal protein S13	SF-AP
P62263	RPS14	40S ribosomal protein S14	SF-AP
P62249	RPS16	40S ribosomal protein S16	SF-AP
P62269	RPS18	40S ribosomal protein S18	SF-AP
P15880	RPS2	40S ribosomal protein S2	SF-AP
P60866	RPS20	40S ribosomal protein S20	SF-AP
P62847	RPS24	40S ribosomal protein S24	SF-AP
P62851	RPS25	40S ribosomal protein S25	SF-AP
P62857	RPS28	40S ribosomal protein S28	SF-AP
P23396	RPS3	40S ribosomal protein S3	SF-AP
P62701	RPS4X	40S ribosomal protein S4, X isoform	SF-AP
P46782	RPS5	40S ribosomal protein S5	SF-AP
P62241	RPS8	40S ribosomal protein S8	SF-AP
P46781	RPS9	40S ribosomal protein S9	SF-AP
P08865	RPSA	40S ribosomal protein SA	SF-AP

Table 21: List of total interacting partners characterized as nuclear or ribonucleoproteins. In this table all the proteins identified with SF-AP, Y2H, GST pull-down and endogenous IP of SON are listed. In particular it was given attention to the proteins indicated in databases as nuclear proteins or as ribonucleoproteins, proteins involved in RNA binding. In the last column the methods responsible for identification are indicated.

3.7.2 Splicing modules found as interacting entities of the SON protein show high interconnectivity.

Detailed analysis of the list of the total interacting partners of SON identified many proteins involved in mRNA processing, maturation, export and translation. Among them, 20 proteins have been reported in cellular machineries and pathways responsible for mRNA splicing, mRNA stabilization and regulation [88-90, 91-93, 95-102, 105-113]. In particular proteins belonging to the mRNP-/mRNP granule-complex, the spliceosome C complex, the mRNA splicing dependent EJC complex and the major coding-region determinant of instability (CRD) complex were identified (table 22). In order to simplify the analysis of the datasets, the mRNP complex and mRNP-granule complex have been considered as a single macro-complex (the mRNP-/mRNP granule-complex) since the function and the structures (protein-mRNA complex) of these complexes are similar. Analyzing the splicing-related interactors and the sub-complex containing them, it appeared clear that there are some peculiar properties concerning the interconnectivity of the SON

binding partners: the first class of proteins are belonging to more than one splicing sub-complex such DHX9 and IGF2BP1 (mRNP-/mRNP granule-complex and CRD complex) and PABPC1 (mRNP-/mRNP granule-complex and spliceosome C complex); the second class of proteins are instead belonging to a single complex such DDX5 (spliceosome C) and DBPA (mRNP-/mRNP granule-complex).

UniProt ID	Identified protein	Function	Splicing sub-machineries	Methods
O95793	STAU1	mRNA stability and regulation [105-106]	mRNP	Y2H
P16989	DBPA	Posttranscriptional regulation of mRNA [92]	mRNP, mRNP-G	SF-AP
Q12906	ILF3	mRNA stability and translation control [107]	mRNP, mRNP-G	Y2H
Q08211	DHX9	Transcription regulation [88]	mRNP, mRNP-G, CRD	SF-AP, GST
Q9NZI8	IGF2BP1	mRNA stabilization and cancer promoter [90, 91]	mRNP, mRNP-G,CRD	SF-AP
P11940	PABPC1	Regulation of mRNA transcription [93]	mRNP, mRNP-G, Spl.C	SF-AP
P09651	HNRNPA1	Post-translational control of mRNA [102]	mRNP, mRNP-G, Spl.C	SF-AP
P23396	RPS3	Ribosome biogenesis and DNA repair [95]	mRNP-G	SF-AP
P61247	RPS3A	Promotes cell proliferation [96]	mRNP-G	SF-AP
P62701	RPS4X	Promotes DNA repair X-linked [97]	mRNP-G	SF-AP
P62241	RPS8	Regulation of translation [98]	mRNP-G	SF-AP
P46781	RPS9	Regulation of translation and cancer growth arrest [99]	mRNP-G	SF-AP
O75643	snRNP200	mRNA splicing [110]	Spl.C	eIP
P17844	DDX5	mRNA maturation and export [88-89]	Spl.C	SF-AP
P07910	HNRNPC	Formation of 40-s hnRNP particles [111]	Spl.C	eIP
P31943	HNRNPH1	Post-translational control of mRNA [100-101]	Spl.C	SF-AP, Y2H
P52272	HNRNPM	Splicing regulation [112]	Spl.C	eIP
Q6P2Q9	PRPF8	Spliceosome formation and mRNA splicing [109]	Spl.C	eIP
Q9UQ35	SRRM2	Spliceosome formation and mRNA splicing [113]	Spl.C	eIP
Q9Y5S9	RBM8A	Exon-exon junction complex formation [108]	Spl.C, EJC	eIP

Table 22: Twenty SON interactors belong to 4 major splicing sub-machineries. Analyzing the overall proteins identified with SONc and endogenous SON, 20 interacting partners have been identified as splicing machinery members and in particular belonging to 4 specific sub-machineries: the mRNP complex (mRNP), the mRNP granule-complex (mRNP-G), the major coding-region determinant of instability complex (CRD) and the spliceosome C (Spl. C). In the last column the methods that identified the protein as specific SON´s interacting partner: yeast two-hybrid (Y2H), GST pull-down (GST), SF-tag affinity purification (SF-AP) and endogenous immunoprecipitation (eIP) are depicted.

However, it is important to note that these splicing modules are not isolated from each other, but on the contrary, they are highly interconnected since the mRNA production and metabolism is their common role. Network analysis, and in particular the interconnectivity between the splicing proteins identified (table 24), was performed using the web tool Pathway Palette (http://blaispathways.dfci.harvard.edu/Palette.html). The protein-protein interaction network was produced recovering all the known interactions among the target proteins searching in the BioGRID and in the HPRD databases. From the listed databases

all the interactions detected with pairing techniques (yeast two-hybrid or isolated recombinant proteins interaction), with *in-vitro* experiments (like GST pull-down) and with *in-vivo* (endogenous immunoprecipitation or immunohistochemistry on tissues) were included. The resulting network (figure 16) showed not only the equal distribution of the interactors of the SON protein between the mRNP-/mRNP granule-complex and the spliceosome C, but also the high interconnectivity between different members of different complexes. Noteworthy are PABPC1 and HNRNPA1, belonging at the same time to the mRNP-/mRNP granule-complex and the spliceosome C complex, therefore representing key members for protein-mediated interconnection between complexes. Given that SON has interacting partners from both mentioned complexes, a connection role between these splicing sub-machineries could be assumed for the SON protein.

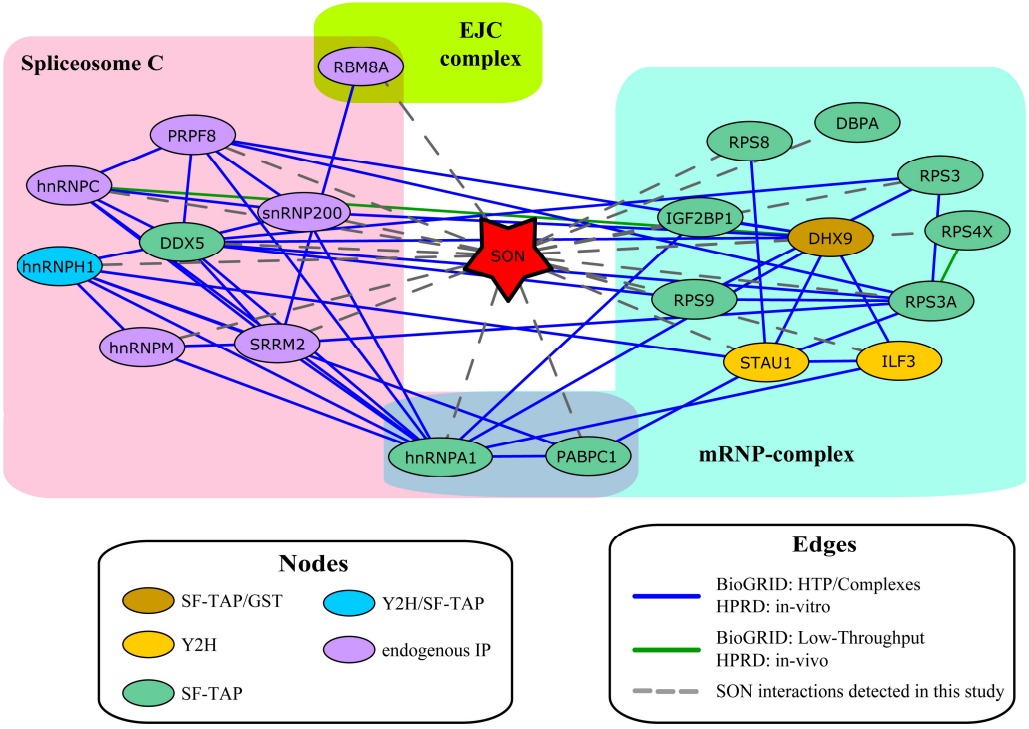

Figure 16: Interacting partners of SON are presenting a high interconnectivity. The interacting partners of SON and SONc proteins obtained from table 24 have been processed with the network generator Pathway Palette Upon data upload, Pathway Palette generates a protein-protein interaction (PPI) network based on information curated by BioGRID and HPRD databases. (Edges) Reported *in-vivo* interactions (BioGRID: Low-Throughput and HPRD (+mouse): *in-vivo*) are represented by green solid lines, *in-vitro* interactions (BioGRID: HTP/Complexes and HPRD (+mouse): *in-vitro*) are represented by solid blue lines and newly identified interactions are represented by dashed-grey lines. Interacting partners of SON and SONc (nodes) are color depicted according to the different techniques used to identify these interactors.

3.8 Dissection of the C-terminal portion of SON

As reported previously [70] SON has been correlated with efficient splicing and correct introns excision (figure 6) of target genes and in particular the C-terminal portion of SON has been hypothesized to be involved in this process (figure 17). However, only final effects have been reported, ignoring completely the mechanistic explanation within SON and its interacting partners involved in splicing.

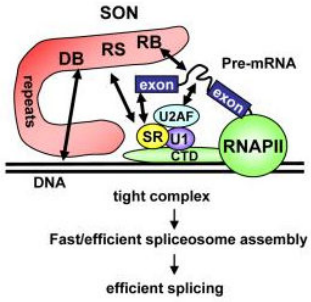

Figure 17: The C-terminal portion of the SON protein is responsible for mRNA splicing. As shown in the sketch (adapted from [70]) it has been hypothesized a direct role of the final portion of the SON protein in efficient splicing. In particular, it is suggested that the C-terminal portion of SON is responsible for splicing.

Since the C-terminal portion is consisting of two fundamental domains for RNA-recognition/-binding (G-patch domain) and RNA processing (double stranded RNA-binding domain) (figure 19), it has been questioned whether the binding sites of SONc responsible for the identified interactions could be determined. A preliminary consideration about this approach should be done: although the SONc dissection could describe previously shown interactions and the SONc domains responsible for binding, it has to be carefully considered that the protein domains are not self-standing entities, but are part of a structured protein.

3.8.1 Domain production

During the design of the C-terminal portion of SON dissection, the positions of the previously described significant domains (dsRNA-BD and G-Patch domains) have been precisely located in the SONc sequence. Afterwards, 3 sub-domains of the C-terminal portion of SON were produced presenting an overlap of one exon with the neighboring

domain. This peculiarity has been intentionally created in order to detect interactions that could be mediated by two adjacent exons and to facilitate correct structure formation.
The first domain (named SONc1) covers exons 5 to 8 (amino acids 2079-2234) and lacks any annotated domain; the second domain (SONc2) covers exons 8 to 10 (amino acids 2319-2346) and contains the G-Patch domain; the last domain produced (SONc3) covers exons 10 to 12 (amino acids 2261-2426) and contains the dsRNA-BD (figure18). Given that the domains could be not intended as self-standing entities and according to previous experiences [70], physiologically coherent combination of these domains were produced. Due to this SONc12 (combination of domains 1 and 2) and SONc23 (combination of domains 2 and 3) were produced as indicated in figure 18.

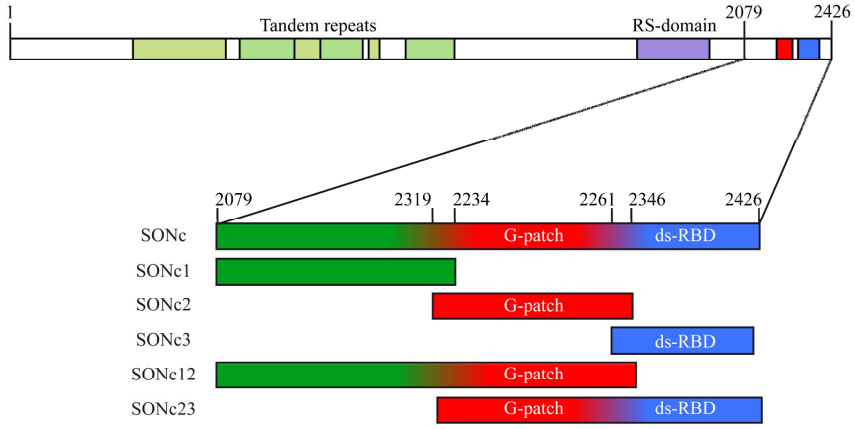

Figure 18: Domain dissection of the SONc sequence. SON´s peculiar structure, presenting in the N-terminus tandem repeat-rich sequences (green boxes), additionally an RS-domain (purple box), a G-patch domain (red box) and a double strand RNA-binding domain (dsRBD) (blue box). Numbers are indicating the amino acids used as N and C terminal ends of SONc). The SONc sequence has been divided into 3 sub-domains (SONc1, SONc2 and SONc3) and their combinations (SONc12 and SONc23) in order to study more in detail the minimal domain (and consequently the protein motif) involved in interactions detected with the SONc sequence. In the SONc sequence the amino acids used to produce specifically the SONc sub-domains have been indicated.

The different SONc domains were separately cloned into suitable vectors for the interaction studies with three different approaches (SF-AP, yeast two-hybrid and GST pull-down). Figure 19 shows the selections of SONc domains which were used in this study and the domain choice will be extensively discussed in the following chapters.

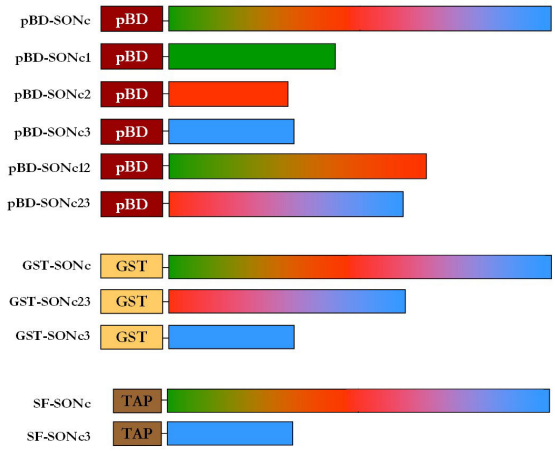

Figure 19: Representation of SONc and the SONc domains used for the domain-based interaction studies. A schematic representation of the vectors (and relative fused protein products) is shown here, maintaining the same domain-color code used for figure 18. For yeast two-hybrid studies (upper panel) SONc, the separate domains and their fusions (SONc12 and SONc23) were cloned into the pBD-GAL4_Cam/DEST vector, to produce BD-GAL4 fused proteins. For GST pull down (middle panel) only SONc, SONc3 and SONc23 were cloned into the pDEST -15 vector to produce GST-fused proteins. For SF-AP (lower panel) SONc and SONc3 were cloned into the pDEST-N-SF-AP vector to obtain N-terminal strep-FLAG tagged proteins.

3.8.2 Identification of an atypical nuclear localization signal (NLS) in the C-terminal sequence of the SON protein.

SONc, as described previously, was found to localize correctly in the nucleus indicating the presence of a nuclear localization signal (NLS) in its sequence (figure 12). As described in the planning of SONc´s interaction partner discovery (chapter 3.1) similar attention was given to the localization of the SONc´s domains. HEK293T cells were transfected with the GFP-fused SONc domains or GFP alone (control) and the fluorescent localization pattern was studied. For the single domains (SONc1, SONc2 and SONc3), it was observed that only GFP-SONc3 is able to locate correctly in the nucleus (figure 20: d and j), differently from GFP-SONc1 and GFP-SONc2 (figure 20: b and h for SONc1, c and i for SONc2). Analyzing the combined domains, GFP-SONc23 is able to localize in the nucleus (figure 20: f and l), while a diffuse cytoplasmic distribution was detected for GFP-SONc12 (figure 20: e and k). However, as observed for the GPF-SONc previously, GFP-SONc3 and GFP-SONc23 (and none of the other domain/domains tested) are not presenting a comparable distribution pattern to the endogenous SON protein (figure 20: v and x).

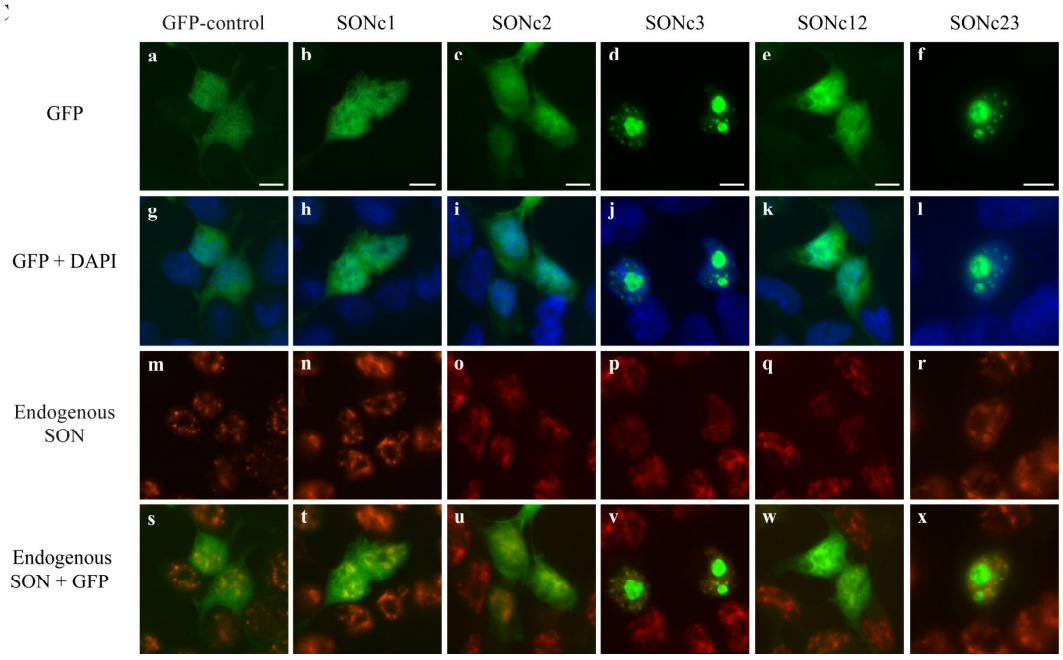

Figure 20: SONc domains distribution in HEK293T cells reveal a NLS in the SONc3 domain. GFP-fused domains have been produced and transfected in HEK293T cells, using GFP protein as control (a). Upon transfection of GFP-SONc1 (b), GFP-SONc2 (c) and GFP-SONc12 (e) domains a diffused cytoplasm staining was detected, resembling the diffused pattern of the control GFP protein. Upon transfection of the GFP-SONc3 (d) and GFP-SONc23 (f) domains a nuclear staining partially overlapping with nucleoli structure was obtained (j, l). Endogenous SON localization was also studied (s-x) in order to evaluate a potential co-localization with the GFP-fragments. These findings indicated the presence of a NLS sequence in the domain 3 that is necessary for a nuclear localization of the SONc protein but not sufficient for the fine localization compared to the endogenous pattern (v). Scale bar: 12.5 microns

These evidences are pointing to the presence of a NLS in the third domain. However the NLS sequence could not been detected with *in-silico* approaches, according to the classical NLS amino acid sequences (e.g. PKKKRKV, QRKRQK or KSKKKAQ) [114-115]. Therefore it could be concluded that the nuclear localization signal contained in the SONc3 domain, is an atypical NLS (aNLS). Even if domain 3 (alone or in combination with the neighboring domain) is able to locate the SONc protein in the nucleus, the nuclear distribution of SONc is still differing from the endogenous SON protein, suggesting that elements external to SONc are responsible for fine speckle localization. The presence of a NLS sequence in the first 140 amino acids of the full length sequence has been reported [54] and, most probably, these two sequences together are responsible for the physiological distribution observed with the endogenous SON protein. Further fine dissections of this domain could be performed to identify the minimal amino acid sequence of the aNLS

3.8.3 Quantitative analysis of the SON Strep-FLAG domain-based interactome

After collecting valuable data from the SONc-based interactome, it has been investigated whether a more comprehensive analysis of the portion of SONc is responsible for the binding with the previously identified interactors could be provided. In order to obtain not only a qualitative but also a quantitative description of the binding of the different domains, the SILAC labeling approach was performed as described for the SONc protein. However one limitation of the SILAC approach is the availability of only three different labels (see chapter 2.2.1.2). Due to this only 2 SONc baits could be used in one experiment, since the third labeling has been used for the SF-GFP control. According to the previous findings, the domain 3 was prioritized since it contains one of the RNA-related motifs and the newly discovered aNLS. Therefore a fusion protein SF-SONc3 has been produced and used in comparison with SF-SONc (as reference) and SF-GFP (as control). The sequences encoding for these three fusion proteins were transfected into HEK293T cells grown in differential SILAC media and their expression was tested (figure 21). In order to avoid protein identification caused by labeling effects, the SILAC media were switched for the constructs to obtain for each bait different labeling (table 23).

Bait	Experiment 1	Experiment 2	Experiment 3
SF-GFP	0/0	8/10	4/6
SF-SONc	4/6	0/0	8/10
SF-SONc3	8/10	4/6	0/0

Table 23: SILAC labeling strategy for the SF-AP domain experiments. As reported in the table, for each experiment the SILAC labeling was switched in order to minimize labeling-dependent artifacts. Labeling was performed with SILAC light media (Arg0, Lys0: **0/0**), medium media (Arg4, Lys6: **4/6**) heavy media (Arg8, Lys10: **8/10**).

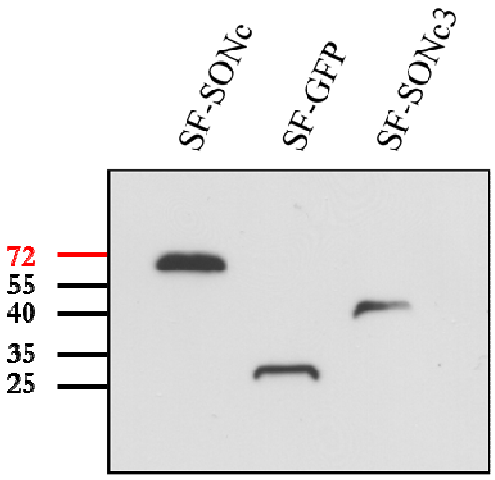

Figure 21: SONc and SONc3 fragments used for domain-based SF-AP interaction studies. In this figure the 3 baits used for the domain-based affinity purification are shown. The SF-SONc and the SF-GFP were previously produced for the first identifications of SONc interacting partner, while the SF-SONc3 was newly produced. In the figure, the domains have been detected using an anti-FLAG antibody in order to evaluate the correct size and expression levels. Each lane contains 10 micrograms of bait-transfected HEK923T cells lysate. For our experiments the amounts of lysates used for the affinity purification were adapted to the SF-fragments expression.

After performing the single-step affinity purification, as described before, the collected eluates were pooled together and subjected to mass spectrometry and MaxQuant software analysis. Obtained interacting partners were assumed as specific only if identified with at least 2 unique peptides, presenting an enrichment compared to the GFP control (ratio >2) and identification probability over 95% ($p<0.05$).

Ten potential interactors of SONc3 were identified (table 24A) and two independent trends for them were detected: 1) 4 proteins have been previously detected with SONc (RPL13A, RPL19, RPS25 and PSMA1), suggesting that the SONc3 is the minimal functional domain responsible for binding in the complete SONc sequence; 2) 6 proteins were newly identified suggesting that this binding is dependent on the portion of the SONc (SONc3) that, most probably, is not accessible in the whole SONc fragment. Considering the first group, the PSMA1 protein has been found significantly enriched in SONc3/SONc (table 24B) indicating that only domain 3 is enough for binding, while for RPL13A, RPL19 and RPS25 proteins, the domain 3 is also responsible for binding but not significantly more than the full length SONc.

Table 24A

UniProt ID	Gene Name	Identified protein	SONc3/GFP normalized	SONc3/GFP significance	SONc3/GFP peptides
Nuclear proteins					
Q8WZ42	TTN	Titin	35,69846	8,95E-63	11
Q9NVH2	INTS7	Integrator complex subunit 7	14,48221	5,68E-37	5
Not nuclear proteins					
Ribosomal proteins					
P40429	RPL13A	60S ribosomal protein L13a	3,189565	1,98E-09	5
P84098	RPL19	60S ribosomal protein L19	2,228012	9,26E-06	20
P62851	RPS25	40S ribosomal protein S25	2,126798	2,31E-05	11
Miscellaneous					
P25786	PSMA1	Proteasome subunit alpha type-1	2,983027	1,16E-08	4
Q9Y2K3	MYH15	Myosin-15	2,523055	6,56E-07	5
P81605	DCD	Dermcidin	2,422583	1,61E-06	6
P84090	ERH	Enhancer of rudimentary homolog	2,395574	2,05E-06	9
Q06830	PRDX1	Peroxiredoxin-1	2,111479	2,65E-05	23

Table 24B

UniProt ID	Gene Name	Identified protein	SONc3/SONc normalized	SONc3/SONc significance	SONc3/SONc peptides
Not nuclear proteins					
Ribosomal proteins					
P40429	RPL13A	60S ribosomal protein L13a	n.sig.	n.sig.	7
P84098	RPL19	60S ribosomal protein L19	n.sig.	n.sig.	22
P62851	RPS25	40S ribosomal protein S25	n.sig.	n.sig.	11
Miscellaneous					
P25786	PSMA1	Proteasome subunit alpha type-1	2,705232	0,002498	5

Tables 24A and 24B: Interactors list from SF-tag affinity purification sorted by domains of SONc responsible for binding. The proteins identified as interactors of SONc and SONc3 are listed. The total protein list is additionally divided according to cellular localization and sub complexes (ribosomal proteins). For this list only proteins having a ratio SONc3/GFP higher than 2 and an identification probability over 95% ($p<0.05$) have been considered as significantly enriched compared to the control. Additionally, this list has been compared to the one obtained with only SF-SONc (table 12) and as a result 4 common proteins were identified. For these proteins SONc3/SONc ratio has been calculated to determine whether domain 3 is solely responsible for binding. In one case we detected a significant enrichment (PSMA1) of SONc3 compared to SONc. Not significant enrichment (lower than 2) or significances ($p>0.05$) have been labeled with "n. sig.".

3.8.4 Y2H domain based interactome analysis

From the previously performed yeast two-hybrid screening there were 32 interactors identified for pBD-SONc. In order to narrow down the sequence, and consequently the domain responsible for binding, the separate and the combined domains of SONc (see figure 19) were also cloned in pBD vectors. These domains were transformed in yeast in combination with the vectors containing the preys previously identified (table 14) fused to

pAD-GAL4 tag. Prior to domain-based interaction experiments, in order to remove any possible false negative result, the interaction occurring between pBD-SONc and the 32 pAD-preys were reconfirmed. To evaluate the domain-based interactions, the six different combinations of the SONc-domains (SONc, SONc1, SONc2, SONc12, SONc3 and SONc23) were plated into a unique SD-WLHA plate. Different pattern of interactions have been identified and visualized. After performing the experiment, 22 interacting partners re-confirmed the interaction with pBD-SONc but no interaction patterns were detected for any SONc sub-domains. In this case, it can be concluded that the whole SONc sequence is required for these interactions and even a small domain deletion could abrogate completely the binding. For the other 10 remaining interactors it was possible to determine a specific domain (or more) responsible for binding (figure 22 and table 25).

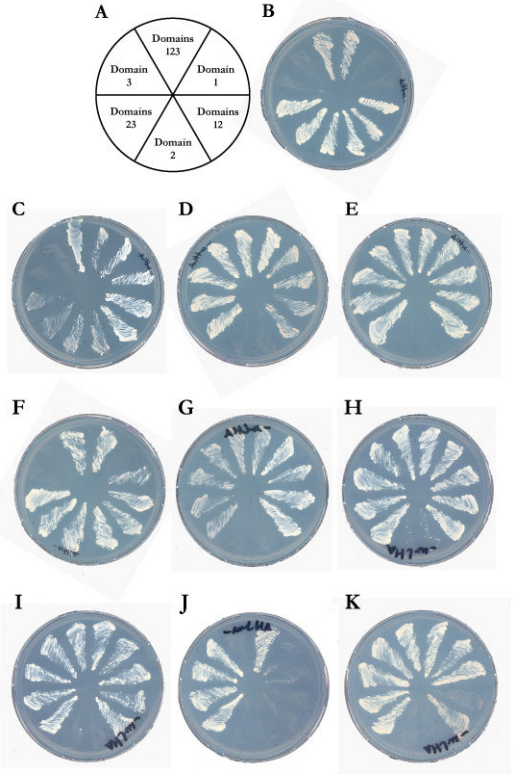

Figure 22: Identification of SONc domains responsible for binding of the detected interactors with the yeast two-hybrid technique. In this figure the resulting interaction studies performed in yeast cells combining the found interactors for SONc with SONc and its domain are shown. With this approach, the sequence responsible for interaction is narrowed down to a 100-200 amino acids sequence. Out of the 32 interactors detected for pBD-SONc, only 10 of them were providing positive results also for domains of SONc. The plates were divided into 6 parts and 2 independent colonies were used to prove each domain-based interaction. Figure legend: **A** domain division example; **B** ILF3; **C** CTSL1; **D** PRKRA; **E** SUMO1; **F** CTSC; **G** PIAS2; **H** TARBP2; **I** HNRNPH1; **J** STAU1; **K** CHD3

UniProt ID	Gene name	Identified protein	Domain(s) involved
Nuclear proteins			
Q15633	TARBP2	TAR (HIV-1) RNA binding protein 2	SONc1, SONc3
Q12906	ILF3	Interleukin enhancer binding factor 3	SONc2
Q569K4	ZNF385B	Zinc finger protein 385B	Not def.
P63165	SUMO1	SMT3 suppressor of mif two 3 homolog 1	SONc1, SONc3
P63279	UBE2I	Ubiquitin-conjugating enzyme E2I	Not def.
Q8NDW4	ZNF248	Zinc finger protein 248	Not def.
O75928	PIAS2	Protein inhibitor of activated STAT, 2	SONc1, SONc3
Q96PM9	ZNF385A	Zinc finger protein 385A	Not def.
Q9UBB5	MBD2	Methyl-CpG binding domain protein 2	Not def.
Q9H898	ZMAT4	Zinc finger, matrin type 4	Not def.
Q9HCU9	BRMS1	Breast cancer metastasis suppressor 1	Not def.
O95833	CLIC3	Chloride intracellular channel 3	Not def.
Q08117	AES	Amino-terminal enhancer of split	Not def.
Q12873	CHD3	Chromodomain helicase DNA binding protein 3	SONc1, SONc3
Ribonucleoprotein			
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H1	SONc1. SONc3
Not nuclear proteins			
O75569	PRKRA	Interferon-inducible double stranded RNA dependent activator	SONc1, SONc3
P14902	INDO	Indoleamine-pyrrole 2,3 dioxygenase	Not def.
P09912	IFI6	Interferon, alpha-inducible protein 6	Not def.
P02747	C1QC	Complement component 1, q subcomponent, C chain	Not def.
P07711	CTSL1	Cathepsin L1	SONc1, SONc2
Q7Z5U7	CTSC	Cathepsin C	SONc1, SONc2
O75843	AP1G2	Adaptor-related protein complex 1, gamma 2 subunit	Not def.
P07858	CTSB	Cathepsin B	Not def.
P14174	MIF	Macrophage migration inhibitory factor	Not def.
Q14152	EIF3A	Eukaryotic translation initiation factor 3, subunit A	Not def.
O95793	STAU1	Staufen, RNA binding protein, homolog 1	SONc3
Q86YQ8	CPNE8	Copine VIII	Not def.
Q8N1I0	DOCK4	Dedicator of cytokinesis 4	Not def.
Q96PE1	GPR124	G protein-coupled receptor 124	Not def.
Q96RW7	HMCN1	Hemicentin 1	Not def.
Q9NX74	DUS2L	Dihydrouridine synthase 2-like, SMM1 homolog	Not def.

Table 25: Interactors list from the yeast two-hybrid approach with description of domains responsible for binding. In this table the proteins identified previously with the yeast two hybrid technique and the domain of SONc responsible for their binding are listed. In the last column the domain/domains of SONc responsible for the binding are reported. "Not.def." identifies interactors that were detected only with SONc and no domain information was available.

3.8.5 GST domain based interactome

Preparing the experiments for the GST-pull down with the SONc domains the constructs containing domains and their combinations fused to the GST tag at the N-terminus were

produced. These sequences were expressed and purified in transformed BL21 *E. coli* as described for the GST-SONc approach (chapter 3.1). Surprisingly, the GST-SONc1 and GST-SONc12 protein products were presenting a high tendency to precipitate in macro protein agglomerates and none of the tested extraction buffers could provide a sufficient yield for interaction studies. To compensate these proteins lack, subsequent deletion mutants of the SONc sequence were used, in particular the GST-SONc23 and GST-SONc3 constructs (figure 23). This strategy is also corroborated by previous findings such as the third domain sequence structure and the role in binding shown in yeast two-hybrid (chapter 3.8.4).

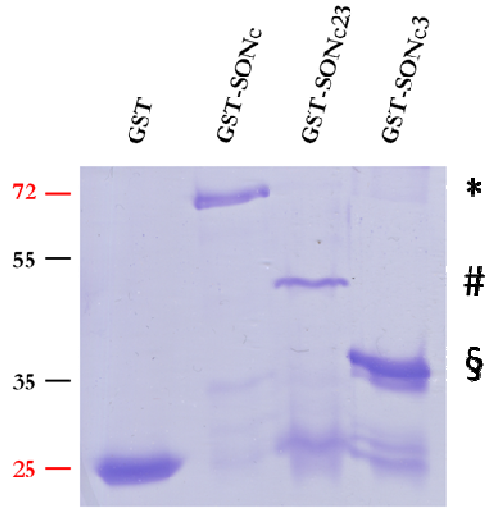

Figure 23: GST domains used for the domain-based approach. In this figure the SONc and its domains used for the GST pull-down after Coomassie staining are shown. The bands shown are resulting from the purification of the transfected bait proteins from a similar amount of bacterial lysate. GST-SONc (band indicated with *), GST-SONc23 (band indicated with #) and GST-SONc3 (band indicated with §) are shown and correspond to the predicted molecular weight. In order to perform correctly the interaction experiments, the relative amount of the bait of interest were quantified, ignoring the bands belonging to bacterial proteins or bait degradation. This additional quantification is a necessary step in order to obtain reliable and comparable result for each bait.

The experiments were performed using the two previously mentioned SONc domains (GST-SONc3 and GST-SONc23) and the GST protein as negative control. Eluates from the GST-affinity purification were subjected to mass spectrometry and the proteins were identified with Mascot analysis. Only proteins identified with at least 2 unique peptides with an identification probability over 95% ($p<0.05$) were assumed to be specific and significant binders. Additionally the proteins identified also with the GST control were removed from the interactor lists. Sixty-one proteins were identified, 14 of them have

already been identified previously with the GST-SONc (table 25). Among these re-confirmed proteins, a minimal domain responsible for binding was identified for 9 proteins (indicated with an asterisk in table 26). As observed previously with the SF-AP domain-based experiments (chapter 3.8.3), the isolated domains of SONc are able to interact with protein not previously detected with the full SONc. Analyzing the 47 newly identified proteins more in detail revealed that these can be grouped in 3 main classes:

- Proteins identified exclusively with GST-SONc3, indicate that domain 3 is the solely responsible for binding and the presence of the neighboring domains could interfere with the recognition efficiency. Belonging to this group are proteins such as SSB and hnRNPH1 previously identified with the SF-AP approach (chapter3.2.2, table 12).
- Proteins identified exclusively with GST-SONc23, suggesting that domain 3 of SONc is not required for binding or even deleterious for the interaction. The DDB1 protein, not identified with GST-SONc and GST-SONc3, indicates that this interaction takes place only through the domain 2.
- Proteins identified with both GST-SONc3 and GST-SONc23. Two different trends have been identified: the two SONc domains are able to bind the protein as GST-SONc does, indicating that domain 3 is the minimal required sequence for binding (e.g. DHX15); the second case, represented by RCN2, shows a protein that binds domain 3 as minimal sequence but is also able to interact with a more complex domain structure (GST-SONc23) but not with the full length GST-SONc.

Among the overall data obtained from the GST pull-down approach, noteworthy are two important proteins identified previously as interacting partners of endogenous SON: DHX15 and HNRNPM. Both proteins are able to bind to domain 3 of SONc, but they present different interaction behavior to the GST-SONc (HNRNPM does not bind the SONc construct). Additionally, for the HNRNPH1 protein the SONc3 domain has been confirmed as minimal responsible sequence, exactly as found previously with the yeast two-hybrid method (chapter 3.8.4).

All these evidences are furthermore confirming that different methods, even in a domain-based procedure, could be combined to obtain more comprehensive information about the interacting partners.

Uniprot ID	Gene name	Identified protein	GST-SONc23			GST-SONc3			Previously identified with GST-SONc
			#1	#2	#3	#1	#2	#3	
P18583	SON	SON DNA binding protein	12	9	10	9	6	8	
Interacting partners identified exclusively with GST-SONc23									
Nuclear proteins									
Q16531	DDB1	Damage-specific DNA binding protein 1		2	2				
Not nuclear proteins									
Q13576	IQGAP2	IQ motif containing GTPase activating protein 2		3	3			3	*
Interacting partners identified exclusively with GST-SONc3									
Nuclear proteins									
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA binding protein 3				5		2	
O15347	HMGB3	High-mobility group box 3				2		2	
P05455	SSB	Lupus La protein				4		6	
P12956	XRCC6	X-ray repair cross-complementing protein 6				10		7	
P13010	XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5				6		4	
P16401	HIST1H1B	Histone cluster 1, H1b				2		2	
Q01844	EWSR1	Ewing sarcoma breakpoint region 1 protein	2			3		2	
Q08211	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9	3			15		19	*
Q13435	SF3B2	Splicing factor 3b, subunit 2				3		2	
Q15393	SF3B3	Splicing factor 3b, subunit 3	2			6		3	
Q92499	DDX1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1		9			13	5	
Ribosomal proteins									
P05388	RPLP0	60S ribosomal protein, P0	3			6	2	5	
P61247	RPS3A	40S ribosomal protein S3A				5		2	
Q08ES8	RPL11	60S ribosomal protein L11	2			2	2	2	*
Ribonucleoproteins									
Q9UKM9	RALY	Autoantigen p542				3	2	2	
P22626	HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2/B1	3			4		3	
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H1	3			8	2	6	
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	3			5		2	
Q1KMD3	HNRNPUL2	Heterogeneous nuclear ribonucleoprotein U-like 2				9		3	
P14678	SNRPB	Small nuclear ribonucleoprotein polypeptides B	3			3	2	3	
Not nuclear proteins									
O75569	PRKRA	Protein kinase, interferon-inducible double stranded RNA dependent activator	2				2	3	
P00450	CP	Ceruloplasmin				5		7	
P04004	VTN	Vitronectin	3			4		3	

Uniprot ID	Gene name	Identified protein	GST-SONc23			GST-SONc3			Previously identified with GST-SONc
			#1	#2	#3	#1	#2	#3	
P17948	FLT1	Vascular endothelial growth factor receptor 1				5		3	
P18085	ARF4	ADP-ribosylation factor 4	3			3		2	*
P80723	BASP1	Brain abundant, membrane attached signal protein 1	2			2		2	
Q05682	CALD1	Caldesmon 1	3			3		2	
Q13219	PAPPA	Pregnancy-associated plasma protein A				2		2	
Q5TCU6	TLN1	Talin 1				5		2	
Q96CF6	HDLBP	High density lipoprotein binding protein	2			8		3	
Q9H0U4	RAB1B	Ras-related protein Rab-1B	2			2	2		*
Q9HCE1	MOV10	Mov10, Moloney leukemia virus 10				2		3	
Q9NYU1	UGCGL2	UDP-glucose ceramide glucosyltransferase-like 2				2		3	
Q9Y2J2	EPB41L3	Erythrocyte membrane protein band 4.1-like 3				6		2	
Q9Y6Q1	CAPN6	Calpain 6				5		2	*
Ribosomal proteins									
Q6IRZ0	RPL31	60S ribosomal protein L31				2		2	
Q6IBH6	RPL26	60S ribosomal protein L26				2		2	
Q6IPH7	RPL14	60S ribosomal protein L14				4		2	
Q07020	RPL18	60S ribosomal protein L18				3		2	
Q49AJ9	RPL3	60S ribosomal protein L3				3		2	
P62249	RPS16	40S ribosomal protein S16				4	2	3	
P62701	RPS4X	40S ribosomal protein S4, X-linked				4		3	
P62888	RPL30	60S ribosomal protein L30	4			5	4	6	*
P62241	RPS8	40S ribosomal protein S8				2		3	
P05387	RPLP2	60S ribosomal protein, P2				6	2	6	
P08708	RPS17	40S ribosomal protein S17				3		2	
P18124	RPL7	60S ribosomal protein L7				6		5	
P26373	RPL13	60S ribosomal protein L13				3		2	
P36578	RPL4	60S ribosomal protein L4				6		2	
P61353	RPL27	60S ribosomal protein L27				2		2	
Q02878	RPL6	60S ribosomal protein L6				7		4	
Interacting partners identified with GST-SONc23 and GST-SONc3									
Nuclear proteins									
O43143	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	31	26	24	7		2	*
Not nuclear proteins									
Q14257	RCN2	Reticulocalbin 2, EF-hand calcium binding domain	10	5	4	8	6	8	
Q8IZ29	TUBB2C	Tubulin, beta 2C	7	2		11		7	*

Table 26: GST-SONc23 and GST-SONc3 interactors list from the GST pull down approach. In order to increase the knowledge about the domains of SONc responsible for the binding and the identification of new interacting partners, we performed a GST pull down using 2 subsequent deletion mutants of GST-SONc: GST-SONc23 and GST-SONc3. Proteins identified also in the GST alone control were removed from the list as non-specific binders. In the columns GST-SONc23 and GST-SONc3 the peptides used for identification of the interacting partners in the three independent experiments (labeled with #1, #2 and #3) are listed. In the last column the proteins also identified as GST-SONc in the previous experiment are indicated with an asterisk (*) (see table 16). The interactors list is sorted according to the domain responsible for binding and according to the sub-cellular localization and sub-groups (ribosomal and ribonucleoproteins). In these experiments only proteins identified in at least 2 experiments with 2 peptides were considered as significant with an identification probability over 95% ($p < 0.05$).

3.8.6 Combination of different techniques after the SONc and SONc-domain-based approaches.

As listed in the aim of the study, the main interest of this study was to prove the possibility to combine different approaches, not compulsorily all of them, to obtain a more comprehensive overview about an unknown protein and its interacting partners. With the SONc and the endogenous SON proteins, 3 proteins have been identified which were detected with more than 2 methods (DHX9, DHX15 and HNRNPH1, see table 20). Surprisingly, those proteins were also detected with more methods as with the domain-based domain approaches. The combination of the SONc domains interaction datasets obtained with SF-AP, yeast two-hybrid and GST pull-down identified 13 proteins (table 27). Three of them (DHX15, ERH and HNRNPM) were previously identified with the endogenous SON immunoprecipitation Analyzing them more in detail, these SON interacting partners have been detected with the SONc3 domain, confirming the functional hypothesis of the C-terminal portion previously introduced [70]. The most valuable information obtained from these results is the possibility to identify minimal sequences responsible for binding with the protein of interest and its portion (SON and SONc for example).

Considering all the information obtained for these 13 proteins, it has been noticed that for some interactors the minimal sequence differs substantially according to the method used. A critical analysis of these discrepancies and their impact on the interaction datasets will be done in the discussion chapter.

Uniprot ID	Gene name	Identified protein	Identifying techniques			
			SF-affinity purification	Yeast two-hybrid	GST pull down	eIP
O43143	DHX15	DHX15	SONc		SONc SONc23 SONc3	#
Q08211	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9	SONc		SONc SONc3	
P84090	ERH	Enhancer of rudimentary homolog	SONc3			#
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H1	SONc	SONc SONc1 SONc3	SONc3	
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M			SONc3	#
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA binding protein 3	SONc		SONc3	
O75569	PRKRA	Protein kinase, interferon-inducible double stranded RNA dependent activator		SONc SONc1 SONc3	SONc3	
P26373	RPL13	60S ribosomal protein L13	SONc		SONc3	
P62249	RPS16	40S ribosomal protein S16	SONc		SONc3	
P61247	RPS3A	40S ribosomal protein S3A	SONc		SONc3	
P62701	RPS4X	40S ribosomal protein S4, X-linked	SONc		SONc3	
P62241	RPS8	40S ribosomal protein S8	SONc		SONc3	
P05455	SSB	Lupus La protein	SONc		SONc3	

Table 27: Combination of the different approaches identified 13 common interactors. In this table the interactors identified with two or more methods upon domain-based experiments combination are listed. In the columns labeled "identifying techniques" the techniques and domains responsible for specific protein identification are marked. Proteins identified with SON endogenous IP are marked with #.

3.9 Interaction and pathway analysis confirm the functional role of SON in the splicing machinery.

From the previously described results of SON´s interacting partner identification, SON has been clearly linked to the 2 major splicing sub-complexes, the spliceosome C and the mRNP complexes. More in detail, the overall results are pointing to SON as key molecular linker between those complexes. In a previously reported study [12] it has been proposed that an *in-silico* approach was able to determine the putative role of a protein of interest according to similarities with interacting networks formed by interacting partners. These protein modules are partially or totally conserved among proteins belonging to the same cellular machinery. In order to understand more about SON´s function, 2 proteins have been selected for the similarity study: DHX9 and DDX5, members of the mRNP complexes and of the spliceosome C, respectively. Protein-protein interactions occurring to

these proteins were searched in the BioGRID and HPRD interaction databases and depicted as graph with the Pathway Palette web-tool (figure 24).

Initially the interacting partners of DHX9 have been analyzed, obtaining a list of 106 specific binders. Among those, 10 proteins were newly identified in this study as SON interacting partners (ACTB, DDX5, DICER1, EWSR1, HNRNPA1, HNRNPA2B1, HNRNPC, ILF3, MBD2, NOLC1, PRPF8, SF3B3, SNRPB, SNRNP200, STAU1, UBE2I, XRCC5 and XRCC6) and 7 proteins were previously reported as SON interactors (NONO, NXF1, PRPF40, SF3B1, SFPQ, SUMO2 and UBC). Noticeably among the 4 interactors identified with pairwise techniques, 3 proteins were identified with the same approach as SON´s interactors (ILF3, MBD2 and UBE2I). These evidences are indicating a physical interaction between SON and proteins belonging to the DHX9-containing modules, demonstrating the reasonable functional relation between SON and the mRNP complex (table 28).

Subsequently the proteins interacting with DDX5 have been analyzed in the above mentioned databases, obtaining a list of 115 specific binders. Among these, 22 proteins have been identified as new interacting partner of SON in this study (DDX1, DHX9, RPL4, RPL6, RPS3A, RPL17, RPL26, RPS9, RPS3, DDX21, RPL7, CHD3, HNRNPH1, SRRM2, SUMO1, PRPF8, HNRNPA1, NOLC1, HNRNPA2B1, HNRNPC, SF3B3, SNRNP200) and 5 proteins have been previously reported in the proteomic databases (SF3B1, NONO, UBC, SFPQ, SUMO2) (table 28).

Lastly, the 2 lists obtained have been combined in order to consider exclusively the interactors common to DDX5 and DHX9 proteins. A total of 48 common proteins have been identified. Among them 7 proteins were identified in this study as new SON interacting partners (HNRNPA1, HNRNPA2B1, HNRNPC, NOLC1, PRPF8, SF3B3 and SNRNP200) and 5 proteins have been already previously reported as interacting partner of SON (NONO, SF3B1 SFPQ, SUMO2 and UBC) (table 28).

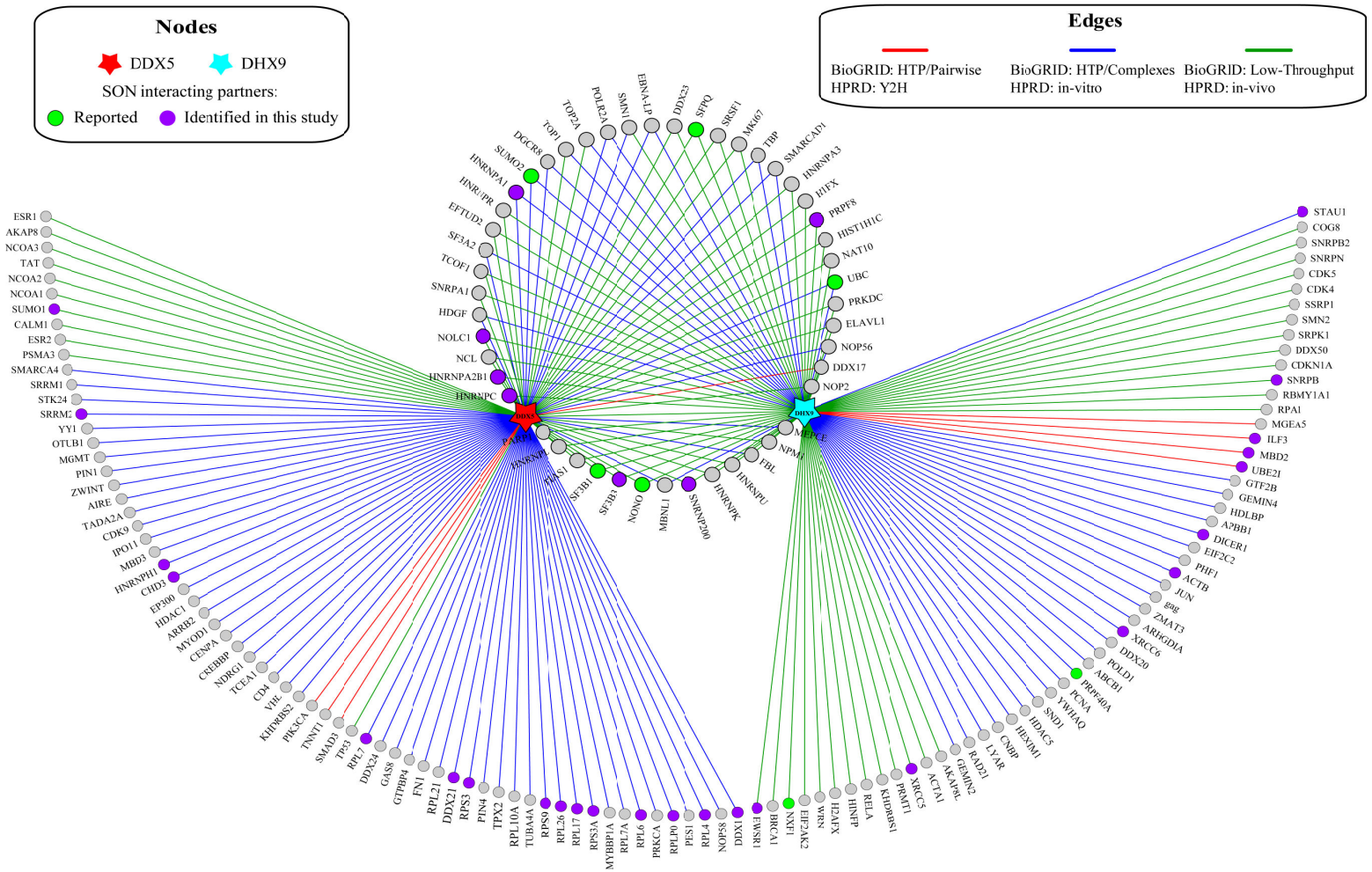

Figure 24: Interacting partners of DHX9 and DDX5 present a significant interconnectivity. In this figure the network of the overall interactions reported for DHX9 and DDX5 are depicted, generated with Pathway Palette based on information curated by BioGRID and HPRD databases. Reported *in-vivo* interactions (edges) (BioGRID: Low-Throughput and HPRD (+mouse): *in-vivo*) are represented by green solid lines, *in-vitro* interactions (BioGRID: HTP/Complexes and HPRD (+mouse): *in-vitro*) are represented by solid blue lines, binary interactions (BioGRID: Pairwise and HPRD Y2H) are represented by solid red lines. The nodes represent the proteins interacting with DHX9 (turquoise node) and DDX5 (red node) proteins. Proteins previously identified interacting with SON are depicted in green, while the newly identified SON´s interacting

Interacting partners of DHX9	SON´s interactors		% of total
	Previously reported	In this study	
106	7	11	16,03%

Interacting partners of DDX5	SON´s interactors		% of total
	Previously reported	In this study	
115	5	22	23,47

Common interacting partners of DDX5 and DHX9	SON´s interactors		% of total
	Previously reported	In this study	
48	5	7	27,03

Table 28: Significant numbers of proteins interacting with DHX9 and DDX5 have been found interacting with SON. The proteins interacting with DHX9 and DDX5 have been checked against interacting partners of SON reported in BioGRID and HPRD databases. Additionally proteins identified in this study as SON´s interacting partners have been also included in the search. The resulting findings are reported in the table. In particular for each protein of interest (DHX9 or DDX5) the number and the percentage of proteins found also interacting with SON has been calculated. Additionally the sub-group of interactors characterized by the simultaneously interaction with both DHX9 and DDX5 have been considered.

The analysis of the previously identified common interacting partners between SON, DHX9 and DDX5, provided valuable information to the significant interconnectivity of SON within these sub-complexes. However, the pathways involving SON has been cautiously considered according to its modularity with DHX9 and DDX5. An enrichment of the reported KEGG pathways was calculated for all the members of the protein complexes formed by DHX9, DDX5 or both through the DAVID web-tool (table 29). For the DHX9 protein complex, pathway analysis identified 17 proteins (out of 106 identified, ~16% of the total) reported in the spliceosome machinery (hsa03040). For the DDX5 protein complex, 13 proteins (out of 115 identified, ~11% of the total) have been reported belonging to the spliceosome machinery and 12 proteins (~10% of the total) belonging to the ribosome machinery (hsa03010).

Considering exclusively the common interactors of DHX9 and DDX5, only the spliceosome pathway has been found to be enriched with 14 members (30% of total common proteins). Out of these, 6 proteins (43% of spliceosome-related proteins) have been identified (newly or previously reported) as SON interacting partners: HNRNPA1, HNRNPC, PRPF8, SF3B1, SF3B3 and SNRNP200. All these findings are certainly pointing to an active role of SON in the splicing machinery.

KEGG Term	Enrichment Score	Count	%	P Value	Bonferroni	Benjamini	FDR
DHX9 interactors							
hsa03040:Spliceosome	22.852	17	15,89	1,83E-15	1,13E-13	1,13E-13	1,89E-12
DDX5 interactors							
hsa03040:Spliceosome	19,942	13	11,33	8.78E-11	7.2E-9	7.2E-9	9.38E-8
hsa03010:Ribosome	8,684	12	9,67	3.3E-10	2,71E-08	1,35E-08	3,53E-07
Total interactors DDX5 and DHX9							
hsa03040:Spliceosome	22,410	13	26	1.36E-15	1.99E-14	1.99E-14	9.21E-13
DHX9 and DDX5 common interactors							
hsa03040:Spliceosome	22.806	17	10,12	1,24E-11	1,20E-09	1,20E-09	1,37E-08
hsa03010:Ribosome	7.316	12	7,14	2,90E-08	2,81E-06	1,41E-06	3,20E-05

Table 29: Pathway enrichment analysis reveals two major cellular mechanisms involved in the DHX9 and DDX5 interactomes. Pathway KEGG enrichment analysis of the proteins interacting with DHX9 and DDX5 have been performed using the DAVID web-tool. Both interactomes are presenting two major pathways significantly enriched, the spliceosome and the ribosome. For each analysis the proteins belonging to a specific pathway (Count), the percentage of these proteins compared to the control (%) and the significance of the pathway identification (p value) are indicated. The Bonferroni calculation, the Benjamini index and the false discovery rate (FDR) values were calculated to provide significance to the pathway identification.

3.10 Effects of SON protein knock-down on the associated modules of the splicing machinery

As previously reported, the SON protein has been linked to the correct localization of splicing factors and their organization [50]. It has been noticed how the splicing factors, such SF2/ASF, SC35, U1-70K and MAGOH (this last one identified with the endogenous SON immunoprecipitation, table 20) present an altered nuclear distribution upon SON depletion (see figure 5). Given that SON has been involved in some splicing sub-machineries like the spliceosome C and the mRNP-/mRNP granule-complexes, the impact of SON depletion in these splicing complexes has been studied.

Considering previously reported silencing experiments [55, 70], HeLa cells have been selected as the most suitable system for SON silencing through small interference RNA (siRNA). Transfection efficiency and appropriate conditions have been tested to obtain a

significant depletion of the SON protein. Seventy-two hours post-transfection the protein levels of SON protein were completely abolished (figure 25A) and cells were not presenting significant changes in morphology (figure 25B).

3.10.1 SON depletion in HeLa cells shows no significant re-localization of DDX5 and DHX9

In order to evaluate the functional impact of SON depletion on the splicing machinery, two proteins were selected as representative for the two major splicing sub-machineries identified: DDX5, member of the spliceosome C, and DHX9, member of the mRNP-/mRNP granule-complex. The cellular distribution of these proteins was analyzed to evaluate whether SON silencing could cause a re-localization, similarly to previous reported results [50]. By performing immunohistochemistry experiments, no evident miss localization of any of the target protein was detected in SON depleted cells (figure 25B). However, these findings are not excluding any involvement of the SON protein in splicing machinery function or architecture. Indeed the cellular distributions of DDX5 and DHX9 are not an evidence for the loss of interconnectivity between the 3 proteins, and could indicate that some proteins could still provide a minimal but not functional organization of some splicing machinery sub-structure.

Although no evident changes have been detected in the cellular distribution of DDX5 and DHX9, subtle changes on the protein network involving these proteins could maybe not be detectable without a quantitative proteomic approach. In order to evaluate even minor, but still significant, changes to the protein complexes containing DDX5 or DHX9 upon SON depletion, quantitative SILAC-based experiments were performed. SILAC-labeled HeLa cells were transfected with the SON-specific siRNA (or control siRNA) and the cells lysates were used for immunoprecipitation with the DDX5 or DHX9 antibodies. The eluates obtained from these experiments were subjected to mass spectrometry and the proteins were identified as well as quantified as described previously (chapter 3.3).

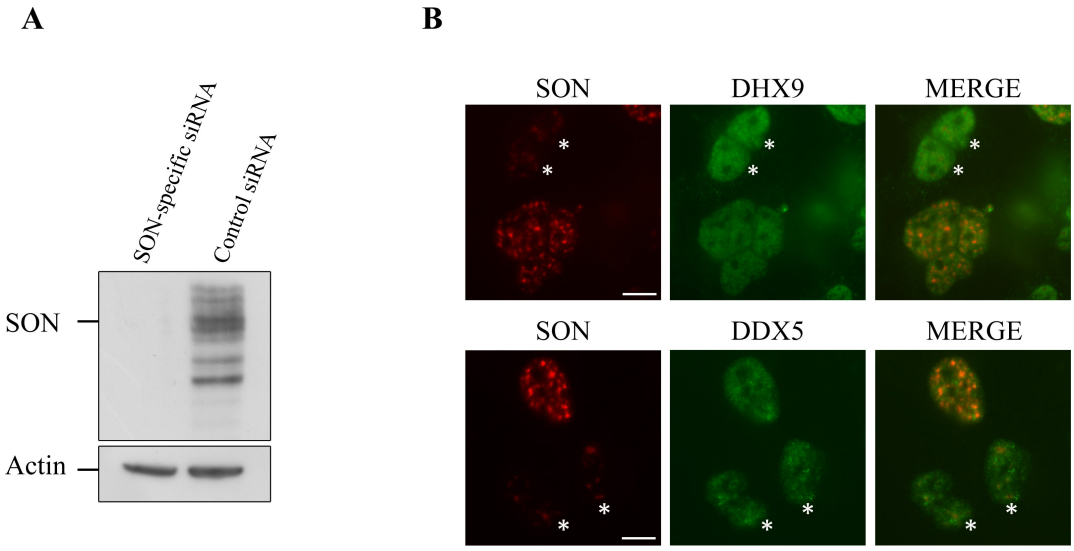

Figure 25: SON co-localizes with splicing factors in HeLa cells and their distribution is not affected after SON silencing. In order to study changes occurring in the splicing machinery upon SON depletion, SILAC labeled HeLa cells depleted of SON were produced through siRNA silencing. (**A**) 72 hours after the transfection of specific siRNA targeting the SON sequence, we observed a complete depletion of the SON protein level compared to the HeLa cells treated with an unspecific (control) siRNA. Furthermore these cellular lysates were used for DDX5 and DHX9 immunoprecipitation and mass spectrometry analysis. (**B**) Although changes in SON protein level led to a destabilization of the DDX5 and DHX9 protein complexes, we were not able to identify any significant change in the local distribution of the 2 target proteins. Remarkably DHX9 and DDX5 proteins, not identified in the endogenous IP as SON interacting partners, are showing high signal overlap with endogenous SON. SON silenced HeLa cells are indicated with white asterisks. Scale bar: 12.5 microns

3.10.2 SON depletion affects the DDX5 interaction pattern, involving proteins responsible for mRNA binding and mRNA transport.

By analyzing the interactions pattern of the DDX5 protein upon SON depletion, particularly proteins showing significant positive or negative enrichment have been considered.

The first group represents all those proteins showing a significant increased affinity for the DDX5 protein subsequent SON depletion. These proteins are presenting a significant positive enrichment compared to the control (represented in table 30 with the ratio "SON knock-down/control siRNA" higher than 2) and have been identified with at least 2 unique peptides ($p<0.05$). Nineteen proteins were showing all these requisites and their role in the splicing machinery and mRNA-related complexes was investigated (table 30). Three proteins have been previously reported to be involved in mRNA binding. The heterogeneous ribonucleoproteins A0 (HNRNPA0) is a member of ribonucleosomes and

has been linked with the post-transcriptional regulation of specific mRNA coding for cytokines [116]. The signal recognition particle 9 KDa (SRFP9) and the signal recognition particle 14 KDa protein (SRP14) have been described as binders of specific mRNA sequences and responsible of regulation and correct localization of target mRNAs. Three additional proteins have been involved in mRNA export and are belonging to the TREX mRNA export complex [117]. NXF1 and NXT1 have been described to form the NXF1/NXT1 heterodimer responsible for the mRNA active transport to the cytoplasm. ALYREF (also indicated as THOC4) is another protein involved in mRNA transport that associates to the NXF1/NXT1 complex in order to export specific subsets of mRNAs (e.g. the HSP70 mRNA) from the nucleus to the translation site.

The second group represents all those proteins showing a significant decreased affinity for the DDX5 protein subsequent to SON depletion. These proteins are presenting a significant negative enrichment compared to the control (represented in table 30 with the ratio "SON knock-down/control siRNA" lower than 0.5) and have been identified with at least 2 unique peptides ($p<0.05$). Twelve proteins were identified according to these parameters and splicing machinery involvement was investigated (table 30). Differently from the previous group, no proteins involved in mRNA binding or processing were identified. However, many de-regulated proteins were described as cytoskeleton and motor protein members. Three proteins have been linked to actin filaments/cytoskeleton components (ACTA1, ACTB and ACTBL2), 3 proteins are involved in actin filament binding (LMO7, LIMA1 and TPM3) and 4 proteins are related to actin-based motor components (MRLC3, MYL6, MYO1C and MYH9). All these proteins have been related to active transport of proteins and protein-mRNA macro complexes [70] and their decreased affinity with DDX5 could point to an impairment of the transport of mRNA-protein complexes from the splicing to the translation sites [118].

UniProt ID	Gene Name	Description	Normalized ratio SON KD/Control	SON KD/Control Significance	Peptides FOR	Peptides REV
DDX5 interactors negatively enriched upon SON knock-down						
Q53X45	MRLC3	Myosin regulatory light chain MRCL3 variant	0,2240842	0,00665828	39	15
P60660	MYL6	Myosin light polypeptide 6	0,2300303	0,00757382	31	22
Q9UHB6	LIMA1	LIM domain and actin-binding protein 1	0,233779	0,00819379	42	4
Q8WWI1	LMO7	LIM domain only protein 7	0,2524276	0,01180039	59	4
P62158	CALM1	Calmodulin 1	0,2589233	0,01327252	35	7
Q16643	DBN1	Debrin 1	0,2860904	0,02074078	28	2
P06753	TPM3	Tropomyosin alpha-3 chain	0,3101591	0,02924717	18	7
P68133	ACTA1	Actin, alpha skeletal muscle	0,3118551	0,02991669	13	7
P35579	MYH9	Cellular myosin heavy chain, type A	0,3297831	0,03757303	202	115
O00159	MYO1C	Myosin I beta	0,3378573	0,04137058	69	3
P63261	ACTB	Actin, cytoplasmic 2	0,3505531	0,0477852	5	2
Q562R1	ACTBL2	Beta-actin-like protein 2	0,3535362	0,04937131	5	3
DDX5 interactors positively enriched upon SON knock-down						
O00629	QIP1	Importin subunit alpha Q1	3,524487	0,03873591	6	5
O14524	TMEM194A	Transmembrane protein 194A	3,598276	0,03521665	2	6
Q12905	ILF2	Interleukin enhancer-binding factor 2	3,645619	0,03313989	15	12
P13639	EEF2	Elongation factor 2	3,768523	0,02833639	20	8
Q9UBU9	NXF1	Nuclear RNA export factor 1	3,854338	0,02542754	2	2
Q96SW2	CRBN	Protein cereblon	3,862729	0,02516077	2	2
P81605	DCD	Dermcidin	3,872076	0,02486716	3	2
P37108	SRP14	Signal recognition particle 14 kDa protein	4,022001	0,02062722	6	6
Q9UKK6	NXT1	NTF2-related export protein 1	4,031823	0,02037792	4	3
Q13151	HNRNPA0	Heterogeneous nuclear ribonucleoprotein A0	4,197742	0,01662109	14	12
P18754	RCC1	Regulator of chromosome condensation 1	4,527719	0,01118172	6	5
Q86V81	ALYREF	THO complex subunit 4	4,661725	0,00955044	9	8
P49458	SRP9	Signal recognition particle 9 kDa protein	4,741481	0,00870251	8	10
P01040	CSTA	Cystatin-A	4,971877	0,00667673	2	3
P06733	MPB1	C-myc promoter-binding protein	5,650331	0,00315146	8	2
P07737	PFN1	Profilin I	5,735948	0,00287513	7	4
P06702	CAGB	Calgranulin-B	7,766022	0,00038623	6	7
P31944	CASP14	Caspase-14	10,02479	5,62E-05	6	6
P05109	CAGA	Calgranulin-A	50,39156	1,69E-12	4	6

Table 30: Upon endogenous SON knock-down, DDX5 interacting partners show differential enrichment. Upon SON knock-down (SON-KD) in SILAC-labeled HeLa cells, immunoprecipitation of DDX5 was performed and eluates were subjected to quantitative mass spectrometry. HeLa cells transfected with control siRNA were used as negative control. In the table the proteins that have been detected as negatively enriched (ratio SON knock-down/control equal or lower than 0.5 and $p<0.05$) or positively enriched (ratio SON knock-down/control equal or higher than 2 and $p<0.05$) are listed. The negatively enriched proteins are the interacting members destabilized by SON KD, while the positively enriched proteins are more stable partners in the DDX5 complex upon SON KD. In the last two columns the pairs of peptides that have been used for quantification (FOR: forward; REV: reverse) are listed.

3.10.3 SON depletion affects the DHX9 interaction pattern, involving proteins responsible for mRNA production/maturation and mRNA transport.

In a similar way as done before for DDX5, the DHX9 protein complex was analyzed and divided according to its positive or negative enrichment.

The first group represented by proteins showing a significant increased affinity for DHX9 subsequent SON depletion, was formed by 21 proteins (table 31). Four of these proteins are components of histones complexes (H2BFL, H2AFC, H4FB and H3.3A), entities involved in the remodeling of chromatin and responsible of the positive/negative regulation of the gene transcription. One additional protein (HNRNPC) is a key component of the early spliceosome assembly [111] and was previously identified as specific endogenous SON interactor.

The second group, composed by proteins showing a significant decreased binding affinity to the DDX5, contains 19 proteins (table 31). Five proteins have been linked to binding activity to actin cytoskeleton (TMOD3, TPM4, TPM1, LMO7 and LIMA1), six proteins are characterized as motor proteins with actin affinity and three proteins are linked to the actin-based motor components (MYO1B, MYL6, MRLC3, MLC1SA, MYO1C and MYO5C) [118]. Surprisingly, five proteins identified with a decreased affinity binding with DDX5 have been also identified with the same parameters in the DHX9 experiments. Two of them are binding-associated with the actin cytoskeleton (LMO7 and LIMA1) and three proteins are linked with the actin-based motor components (MRLC3, MYL6 and MYO1C).

UniProt ID	Gene Name	Description	Normalized ratio SON KD/Control	SON KD/Control Significance	Peptides FOR	Peptides REV"
DHX9 interactors negatively enriched upon SON knock-down						
O43795	MYO1B	Unconventional myosin I-b				
Q8IVT2	C19orf21	Uncharacterized protein C19orf21	0,2370543	1,25E-13	45	16
Q9UHB6	LIMA1	LIM domain and actin-binding protein 1	0,3072722	9,95E-10	19	7
P60660	MYL6	Myosin light polypeptide 6	0,3087116	1,15E-09	64	14
Q53X45	MRLC3	Myosin regulatory light chain MRCL3 variant	0,3252051	5,68E-09	32	32
P14649	MYL6B	Myosin light chain 6B	0,3265517	6,43E-09	37	21
P62158	CALM1	Calmodulin 1	0,3287868	7,88E-09	2	2
Q8WWI1	LMO7	LIM domain only protein 7	0,3405898	2,22E-08	36	14
O00159	MYO1C	Myosin I beta	0,3507155	5,12E-08	79	15
Q16643	DBN1	Drebrin 1	0,356657	8,18E-08	82	17
P08754	GNAI3	Guanine nucleotide-binding protein G(k) subunit alpha	0,3949954	1,23E-06	34	13
Q9NYL9	TMOD3	Tropomodulin-3	0,4083095	2,80E-06	7	3
P67936	TPM4	Tropomyosin alpha-4 chain	0,4118046	3,45E-06	9	3
P09493	TPM1	Tropomyosin alpha-1 chain	0,4152802	4,22E-06	13	5
Q9UBI6	GNG12	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	0,4300427	9,65E-06	9	7

UniProt ID	Gene Name	Description	Normalized ratio SON KD/Control	SON KD/Control Significance	Peptides FOR	Peptides REV"
P80723	BASP1	Brain acid soluble protein 1	0,4320772	1,08E-05	8	4
Q9NQX4	MYO5C	Unconventional myosin-Vc	0,4374406	1,43E-05	23	14
P13987	CD59	CD59 glycoprotein	0,4401678	1,65E-05	13	2
P62879	GNB2	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	0,4423713	1,84E-05	7	2
			0,4436837	1,97E-05	13	2
DHX9 interactors positively enriched upon SON knock-down						
P62807	H2BFL	Histone H2B type 1-C/E/F/G/I				
P0C0S8	H2AFC	Histone H2A type 1	2,166839	0,04884531	13	7
P49327	FASN	Fatty acid synthase	2,305832	0,03262471	9	10
P07910	HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	2,408878	0,02420132	89	50
P62805	H4FB	Histone H4	2,628294	0,01285723	7	3
P15924	DSP	Desmoplakin	2,692935	0,0106843	13	11
P84243	H3.3A	Histone H3.3	2,802281	0,00782339	40	37
P05161	ISG15	Ubiquitin-like protein ISG15	2,812513	0,00759931	4	3
P14923	JUP	Junction plakoglobin	4,545185	7,60E-05	5	2
P29508	SERPINB3	Serpin B3	5,773681	4,26E-06	5	4
Q9HCM1	KIAA1551	Uncharacterized protein KIAA1551	5,965447	2,79E-06	2	9
P81605	DCD	Dermcidin	8,096756	3,69E-08	2	2
P47929	LGALS7	Galectin-7	8,706986	1,20E-08	5	5
Q01469	FABP5	Epidermal-type fatty acid-binding protein	14,25082	2,20E-12	2	3
P01040	CSTA	Cystatin-A	23,99252	3,82E-17	4	7
P31944	CASP14	Caspase-14	25,58951	8,62E-18	4	6
P05109	CAGA	Calgranulin-A	29,59817	2,68E-19	7	9
P06702	CAGB	Calgranulin-B	85,03853	3,40E-32	5	6
P11532	DMD	Dystrophin	106,4928	2,17E-35	6	6

Table 31: Upon endogenous SON knock-down, DHX9 interacting partners show differential enrichment. Upon SON knock-down (SON-KD) in SILAC-labeled HeLa cells, immunoprecipitation of DHX9 was performed and eluates were subjected to quantitative mass spectrometry. HeLa cells transfected with control siRNA were used as negative control. In the table the proteins that have been detected as negatively enriched (ratio SON knock-down/control equal or lower than 0.5 and $p<0.05$) or positively enriched (ratio SON knock-down/control equal or higher than 2 and $p<0.05$) are listed. The negatively enriched proteins are the interacting members destabilized by SON KD, while the positively enriched proteins are more stable partners in the DHX9 complex upon SON KD. In the last two columns the pairs of peptides that have been used for quantification (FOR: forward; REV: reverse) are listed.

These evidences are pointing to a specific role of the SON protein in the correct organization of the DDX5 and DHX9 protein networks. In particular, proteins involved in the mRNA process and transport present a lower affinity for such splicing complexes upon SON depletion. In the discussion paragraph the possible link between SON and mRNA transport process will be further analyzed.

4 DISCUSSION

4.1 Critical analysis of used techniques and their combination to determine protein networks.

In this study interaction analysis using different available techniques in order to provide a more comprehensive description of SON's interactome have been performed. During the planning of the experiments four methods according to their strength and the value of information they could provide have been selected, without ignoring the weakness and drawbacks that each method inherently presents.

The first technique used was the tag-based affinity purification based on the interaction occurring between a tagged protein and a resin presenting high affinity for the tag used. This approach is allowing the identification of binding partners of a protein of interest, or its mutations and deletions, in target mammalian cell systems. Additionally, the localization of the protein in the cell, and its sub-compartments, can be studied using an antibody- or fluorescence-based detection. However, this technique presents two majors drawbacks, all linked to the need of introducing and expressing an exogenous protein in the cells of interest. The first one includes the possibility to use only cell lines where transfection or transduction can be performed efficiently. Primary cell lines are difficult to transduce and the exogenous protein is expressed with lower efficiency, while tissues (as biopsies) cannot be studied for the elementary lack of the tagged protein of interest. The second drawback is incidental to the exogenous protein expression driven by strong promoters such the viral CMV-promoter. This expression system drives the protein production machineries towards the exogenous bait, diverting the protein production efforts from the endogenous and constitutive proteins. Additionally, in yeast experiments has been reported how exogenous proteins expression can cause ER stress causing an overall cell suffering. Since ER stress cannot be ignored or excluded from the experiment, only a similar stress can standardize the out-coming results. According to that, it was decided to use an exogenous standard protein (in this study the GFP protein) fused to the Strep-FLAG tag in order to normalize cell-stress.

The second technique used was the yeast two-hybrid screening, a valuable high throughput approach that can be performed without expensive machineries requirement. As described

extensively in the previous chapters, this system is based on the exogenous expression of the protein of interest in combination with a wide range of proteins obtained from a cDNA library. The main advantages are tightly connected with the yeast themselves: these cells are the most elementary eukaryotic system available and are easy to handle and transfect with the target DNAs. Additionally the choice of suitable cDNA libraries permits interaction discriminations according to the origin of the tissue of the library. Furthermore, the design of the yeast two-hybrid technique provides valuable information about the modality of interaction since only binary bindings between proteins are detectable. The main drawback of this technique lies in the detection of protein-protein interactions that have to occur in the nucleus of the cell. This mechanism is mediated by the fusion tags used (pAD-GAL4 and pBD-GAL4) containing both a nuclear localization signal that divert the exogenous proteins in the nucleus. This fact is forcing the interaction to take place in the nucleus even if the bait protein is a cytoplasm or a membrane protein, causing large false positive detections. Additionally, since the yeast are simple eukaryotic cells, some protein post-translational modifications are missing [119], introducing the chance of false negative results. However, recent genetic modifications to the yeast strains used are including also enzymes involved in acetylation and phosphorylation [120].

The third technique used in the study was the GST pull-down, a variation of the affinity-purification method described before. The main difference is the different cellular system producing the protein of interest: *E. coli* cells are the source of the exogenous protein used later for the experiments. The major strength of this technique is the possibility to combine the protein of interest, portions of it or mutants, with virtually every tissue of choice. However, the production of the protein in a bacterial system has two main disadvantages: 1) the bacterial cells are missing some important post-translational modification [121] causing some protein interactions, mediated by phosphorylation or other modification, not to occur as expected by physiological approaches; 2) bacterial cells are lacking important folding proteins, causing some protein folding to not happen correctly [122].

The last technique used in the study was the immunoprecipitation of the endogenous SON protein. This method is the most suitable to study protein-protein interaction in mammalian cell systems or in target tissues, since the investigation analyses the physiological protein in its natural environment. Additionally, this technique not only allows the detection of physiological interactions, but also can be used to differentiate

protein-interaction patterns occurring in different tissues or in disease samples. Although this is the method of choice for interaction studies, a major drawback is incidental to the antibody-based system. In order to purify efficiently the complexes containing the protein of interest, the antibody to be used has to present a high affinity for the target protein and a significant specificity for the epitope.

As hypothesized in the "Aim of the study" chapter, the possibility to combine efficiently the different approaches has been investigated. Combining the data obtained from the endogenous SON, the SONc and the SONc-domains experiments, reported in table 26, 13 proteins were identified with more than one method. In particular, with the SONc protein the HNRNPH1 protein was identified as SON´s interactor with the SF-AP, with the GST pull-down and, interestingly, with the yeast two-hybrid approach. These findings not only prove the compatibility of different approaches, but also the binary nature of this interaction is revealed.

4.1.1 Domain-based interaction experiments identified a minimal SONc sequence responsible for binding

The study of specific sub-domains of SONc provided important evidences of minimal sequences responsible for binding with the identified interacting partners. After combination of the overall results obtained from the SONc domain experiments in SF-AP, yeast two-hybrid and GST pull-down, 3 different trends for binding properties have been identified.

The first group includes proteins identified with the endogenous immunoprecipitation method and with the SONc fragment and its domains in two additional experiments. The DHX15 protein has been identified as significant interacting partner of the endogenous SON (table 18) and of the SONc fragment in the SF-AP and GST pull-down techniques (tables 12 and 16, respectively). More in particular, the domain based study proposed the SONc3 portion as minimal sequence detected in the GST pull-down. On the other hand, SF-AP experiments were not able to determine a SONc minimal sequence involved in the described binding. These differences could be explained considering the substantial differences between the approaches used since in the SF-AP approach the SONc3 is produced and modified in a mammalian environment, differently from the bacterial environment producing the bait protein for the GST pull-down.

The second group includes proteins identified with the SONc fragment with 2 independent methods and with the SONc sub-domains in only one of the two techniques. The DHX9 protein has been identified as SONc interacting partner using the SF-AP and the GST pull-down technique. As for the previous group, a minimal portion for binding was described for the GST pull-down but not with the SF-AP approach. These differences could be again explained considering the cellular system used to produce the SONc and its domains.

The third group includes proteins identified with the endogenous immunoprecipitation and exclusively with SONc-domain approaches. The ERH and HNRNPM proteins have been initially identified as specific interactors of the endogenous SON protein and, subsequently, with the domain approaches. These proteins present a minimal binding portion within SONc (SONc3) but not with the whole SONc fragment. These precious findings show the minimal domain responsible in the full-length SON for binding with such interacting partners, providing additional confirmation of the functional interaction of the far end of the SON sequence with proteins involved in cell cycle (ERH) and splicing (HNRNPM).

The fourth group includes proteins identified with at least one SONc-based method and with an additional technique exclusively with the SONc3 domain. Proteins such HNRNPH1 and RPS4X, both involved in splicing machineries have been identified with one or more SONc-approach: SF-AP and Y2H for HNRNPH1 and SF-AP for RPS4X. More detailed the SONc-domain based analysis identified both proteins as GST-SONc3 binders (see table 26).

It can be suggested that the above listed discrepancies found with the different approaches, could be determined by the different biological systems used and in particular all these differences could be linked to post-translational and structure dependent interactions and to the different accessibility of the domains of SONc compared to the whole SONc fragment.

As mentioned before different systems are producing different mature proteins or portions of it. SF-AP and GST pull-down approaches are requiring proteins with folding and post-translational modification very different from each other. Additionally the different cellular lysates used to probe our bait were different (mammalian cell line compared to human tissue), providing differences in interactomes occurring for the SON protein. Nevertheless, the occurrence of the interaction cannot be ignored and provides more insights about significant differences when considering different approaches and target tissues.

Furthermore, the domain production could expose binding sequences previously masked by the full protein, leading to interactions based on the protein sequence. Nevertheless, even in this case, the obtained information can be fruitfully used to determine the nature of the interaction occurring.

4.1.2 Practical applications and perspectives of methods combination.

As previously described, the combination of different methods open new insights on interactome studies, in particular explaining how interactions are taking place. In this thesis information about the interacting proteins and how this interaction takes place is provided. Additionally it could be distinguish between binary and not binary interactions, being able to determine the proteins, belonging to the SON complex, that have a direct and physical contact with SON protein. These findings can be used in future to determine how the SON protein complex is physically composed and how interacting proteins could be involved in SON-dependent diseases.

Furthermore, the domain-based interaction studies provided us information about the SON´s protein sequence responsible for the binding with the detected interacting members. These findings could be extended to other proteins described to be responsible in disease onset. Newly identified mutations could be combined with proteomic datasets in order to understand the underlying mechanism that is altered in the disease. We expect, in a nearly future, to be able to connect each possible protein modification of the SONc sequence with our interaction datasets, providing a better understanding of the mechanistic onset of a peculiar disease.

4.2 Proteomic data provide a mechanistic explanation of the role of the SON protein in splicing

The original aim of the study was to provide a more comprehensive description of SON´s interacting partners, in order to assign a putative role for our protein of interest. However, in the following years many evidences of SON´s involvement in the splicing processes have been given but without a clear mechanistic explanation. I took the chance to combine the new data obtained from the different proteomic approaches together with the previous findings. The SON protein interacting partners have been identified as members of two

majors splicing complexes of the splicing machinery, the spliceosome C and the mRNP/mRNP-granule complex. As described before, these two complexes have been associated with the correct maturation of mRNAs and their export to the translation active sites, the ribosomes. Additionally, these two splicing sub-complexes have been shown to be highly interconnected, as indicated in figure 16. After combination of the SONc and SONc domains interaction datasets has been shown that SON protein (figure 26) is tightly linked with both sub-machineries showing almost equal distribution of interactors.

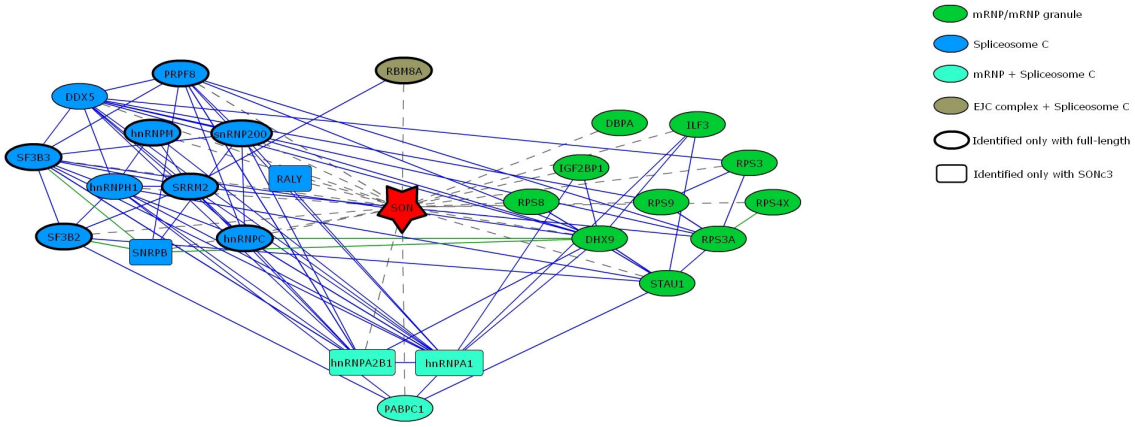

Figure 26: Interacting partners of SON, SONc and SONc-domains are presenting a high interconnectivity. The interacting partners of SON and SONc proteins have been processed with the network generator Pathway Palette. Upon data upload, Pathway Palette generates a protein-protein interaction (PPI) network based on information curated by BioGRID and HPRD databases. Reported *in-vivo* interactions (BioGRID: Low-Throughput and HPRD (+mouse): *in-vivo*) are represented by green solid lines, *in-vitro* interactions (BioGRID: HTP/Complexes and HPRD (+mouse): *in-vitro*) are represented by solid blue lines and newly identified interactions are represented by dashed-grey lines. Proteins identified exclusively with the full length SON protein have been marked with a thicker node border and the ones identified exclusively with the SONc3 domain are depicted with a rectangle shape.

In chapter 3.9 the significant partial overlap of the SON protein network with the DDX5 and DHX9 protein networks has been demonstrated. Considering the two networks independently the overlap of SON´s interactome is 16,03% for DHX9 and 23,47% for DDX5, indicating the reasonable functional similarities between proteins involved in the spliceosome C and mRNP/mRNP-granule complexes. According to Deszo [12] these findings are suggesting an active role of SON in these 2 complexes. Interestingly, the SON overlap with other protein networks is more significant when only the common protein of the DDX5 and DHX9 proteins were considered. These common interactors are representing only those proteins that are functionally linking the two complexes,

physiologically consecutive during mRNA maturation. SON interacting partners represent the 27% of these common interactors increasing the likelihood of a SON involvement in both complexes. In conclusion, the SON protein can be confidently assumed to act as a coordinating protein between these two complexes involved in the latest steps of mRNA maturation.

4.3 SON depletion causes subtle alteration in protein-protein interaction networks of DDX5 and DHX9

The SON protein, as previously reported, is responsible for an altered distribution of specific mRNA-splicing proteins. More in detail, these altered proteins are linked to the early stage of splicing like the recognition of the exon splicing enhancers (U1-70K) and the U1-/U2-snRNP splicing complexes (SF2/ASF and SC35) or, on the opposite, latest steps of splicing like the EJC complex (MAGOH). In this study it has been analyzed whether further splicing machineries could present the same reported alteration upon SON depletion. According to proteomic findings the attention was pointed on the spliceosome C and mRNP complexes and selecting critically two proteins for subsequent analysis.
Although SON depletion didn´t caused evident miss-localization of DDX5 and DHX9 proteins, it cannot be reasonably excluded that changes are effectively taking place but more subtly. However, applying proteomic tools, significant changes in the interaction networks of DDX5 and DHX9 became more evident. In particular, interesting analogies have been detected for interacting partners of the target proteins showing decreased affinity upon SON depletion. The majority of these proteins (10 for DDX5 and 11 for DHX9) have been connected with the actin cytoskeleton and molecular machineries involved in transport of protein complexes. Consistently with the previous findings of molecular connection between the spliceosome C and mRNP-complex, almost 50% of these transport-related members were in common between the DDX5 and the DHX9 proteins. Actin and myosin members, known to be essential for transport of macro protein complexes, have been also recently linked to efficient mRNA transport [118]. In particular the complexes binding to these motor proteins have been found to contain also mature mRNAs.

4.3.1 SON is essential for correct mRNA export and localization.

The above described evidences are pointing to the SON protein as a key member for mRNA loading form the spliceosome C and mRNP-complex to the actin-based export modules. Additionally, to these findings, an additional motor protein involved in mRNA transport was identified with the eIP approach (table 18), the dynein light chain 1. This protein has been initially described as a motor protein involved in active transport of proteins in specific cell structure; in particular the dynein has been linked to active retrograde transport of complexes in cilia [123-124].

More proves to the important role of SON on mRNA export and fate, are represented by the STAU1 and IGF2BP1 proteins identified as SON´s interacting partners. STAU1 is a member of a protein macro-complex involved in the fine regulation and nuclear export of mature mRNA to the ribosomes [106]. Analyzing more in detail the composition of these STAU1-containing complexes, identified in granules containing untranslated mRNAs [125], many different classes of proteins were detected. Among them motor proteins (e.g dyneins – also identified as SON´s interactor), nuclear proteins involved in mRNA processing (e.g. DHX9 – SONc´s interactor) and ribosomal proteins (e.g. RPLP0 – SONc3 interactor) were isolated [126]. The IGF2BP1 protein is an important player on mRNA stabilization and translation control of target genes. Comparing the list of IGF2BP1-associated proteins reported with the SON´s interactors identified in this study, 7 SON (and SONc) interacting partners have been identified as members of the IGF2BP1-containing complex: IGF2BP1 itself, IGF2BP3, PABPC1, DBPA, ILF3, DHX9 and HNRNPH1. Given that IGF2BP1 has been correlated with mRNA regulation and expression of the c-myc oncogene, it appears clear that there is a connection between SON´s role on cell proliferation and mRNA control.

5 Conclusions

In this thesis I tried to provide a molecular link between the newly reported roles of the SON protein in cell survival and mRNA production, and the cellular machineries in which SON could be a key member. As planned in the thesis objectives, I was successfully able to identify new SON´s interacting partners applying different methods, revealing molecular machineries and pathways in which SON could be involved, in particular the splicing machinery. Furthermore, the complexity and the interconnections between different sub-modules of the splicing macro-complex have been fruitfully analyzed, showing the tight connections existing between SON and the spliceosome C and mRNP complexes. According to these results, I could narrow the reported generic participation of SON in the splicing machinery to the two specific sub-complexes previously mentioned, leading to an increase of the existing knowledge about SON´s interactors and a deeper mechanistic understanding. However a comprehensive description of the splicing machinery alterations upon SON depletion, that produce the previously reported effects, was not successfully achieved. Even though SON´s absence is not re-localizing members of the spliceosome C and mRNP complexes, as reported for other splicing proteins, has been here proposed a more subtle function for SON. More in detail, I hypothesized in this study a broader scaffold function for SON protein considering the significant affinity changes for motor and cytoskeleton proteins occurring to the DDX5 and DHX9 proteomes upon SON depletion. Overall SON protein is an important but not necessary link between different modules of the splicing machinery, but still a fundamental key protein for correct mRNA production.

6 References

[1] George EL, Georges-Labouesse EN, Patel-King RS, Rayburn H, Hynes RO. Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. Development. 1993;119:1079-91.

[2] Susaki E, Nakayama KI. Functional similarities and uniqueness of p27 and p57: insight from a knock-in mouse model. Cell cycle. 2009;8:2497-501.

[3] Peters RH, van Doorninck JH, French PJ, Ratcliff R, Evans MJ, Colledge WH, et al. Cystic fibrosis transmembrane conductance regulator mediates the cyclic adenosine monophosphate-induced fluid secretion but not the inhibition of resorption in mouse gallbladder epithelium. Hepatology. 1997;25:270-7.

[4] Thompson S, Clarke AR, Pow AM, Hooper ML, Melton DW. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell. 1989;56:313-21.

[5] Muller U. Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. Mechanisms of development. 1999;82:3-21.

[6] Deininger M, Buchdunger E, Druker BJ. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood. 2005;105:2640-53.

[7] Emens LA. Trastuzumab: targeted therapy for the management of HER-2/neu-overexpressing metastatic breast cancer. American journal of therapeutics. 2005;12:243-53.

[8] Chakravarti B, Mallik B, Chakravarti DN. Proteomics and systems biology: application in drug discovery and development. Methods in molecular biology. 2010;662:3-28.

[9] Heller MJ. DNA microarray technology: devices, systems, and applications. Annual review of biomedical engineering. 2002;4:129-53.

[10] Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, et al. Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Molecular & cellular proteomics : MCP. 2002;1:323-33.

[11] Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999;402:C47-52.

[12] Dezso Z, Oltvai ZN, Barabasi AL. Bioinformatics analysis of experimentally determined protein complexes in the yeast Saccharomyces cerevisiae. Genome research. 2003;13:2450-4.

[13] Wang PI, Marcotte EM. It's the machine that matters: Predicting gene function and phenotype from protein networks. Journal of proteomics. 2010;73:2277-89.

[14] Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989;340:245-6.

[15] Letteboer SJ, Roepman R. Versatile screening for binary protein-protein interactions by yeast two-hybrid mating. Methods in molecular biology. 2008;484:145-59.

[16] Suter B, Kittanakom S, Stagljar I. Two-hybrid technologies in proteomics research. Current opinion in biotechnology. 2008;19:316-23.

[17] Chien CT, Bartel PL, Sternglanz R, Fields S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proceedings of the National Academy of Sciences of the United States of America. 1991;88:9578-82.

[18] Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, et al. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature. 2000;403:623-7.

[19] Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:4569-74.

[20] Schwikowski B, Uetz P, Fields S. A network of protein-protein interactions in yeast. Nature biotechnology. 2000;18:1257-61.

[21] Brizzard BL, Chubet RG, Vizard DL. Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. BioTechniques. 1994;16:730-5.

[22] Voss S, Skerra A. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. Protein engineering. 1997;10:975-82.

[23] Hochuli E, Dobeli H, Schacher A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. Journal of chromatography. 1987;411:177-84.

[24] Kolodziej PA, Young RA. Epitope tagging and protein surveillance. Methods in enzymology. 1991;194:508-19.

[25] Smith DB, Johnson KS. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene. 1988;67:31-40.

[26] di Guan C, Li P, Riggs PD, Inouye H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene. 1988;67:21-30.

[27] Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. Comparison of affinity tags for protein purification. Protein expression and purification. 2005;41:98-105.

[28] Brymora A, Valova VA, Robinson PJ. Protein-protein interactions identified by pull-down experiments and mass spectrometry. Current protocols in cell biology / editorial board, Juan S Bonifacino [et al]. 2004;Chapter 17:Unit 17 5.

[29] Magnaghi-Jaulin L, Masutani H, Robin P, Lipinski M, Harel-Bellan A. SRE elements are binding sites for the fusion protein EWS-FLI-1. Nucleic acids research. 1996;24:1052-8.

[30] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nature biotechnology. 1999;17:1030-2.

[31] Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nature cell biology. 2004;6:97-105.

[32] Gloeckner CJ, Boldt K, Schumacher A, Roepman R, Ueffing M. A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. Proteomics. 2007;7:4228-34.

[33] Gloeckner CJ, Boldt K, Ueffing M. Strep/FLAG tandem affinity purification (SF-TAP) to study protein interactions. Current protocols in protein science / editorial board, John E Coligan [et al]. 2009;Chapter 19:Unit19 20.

[34] Kaake RM, Wang X, Huang L. Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. Molecular & cellular proteomics : MCP. 2010;9:1650-65.

[35] Boxem M, Maliga Z, Klitgord N, Li N, Lemmens I, Mana M, et al. A protein domain-based interactome network for C. elegans early embryogenesis. Cell. 2008;134:534-45.

[36] Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & cellular proteomics : MCP. 2002;1:376-86.

[37] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature biotechnology. 1999;17:994-9.

[38] Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics. 2005;5:4-15.

[39] Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert FM, Vandermoere F, et al. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. The Journal of cell biology. 2008;183:223-39.

[40] Ideker T, Sharan R. Protein networks in disease. Genome research. 2008;18:644-52.

[41] Zhong Q, Simonis N, Li QR, Charloteaux B, Heuze F, Klitgord N, et al. Edgetic perturbation models of human inherited disorders. Molecular systems biology. 2009;5:321.

[42] Charloteaux B, Zhong Q, Dreze M, Cusick ME, Hill DE, Vidal M. Protein-protein interactions and networks: forward and reverse edgetics. Methods in molecular biology. 2011;759:197-213.

[43] Saitoh N, Spahr CS, Patterson SD, Bubulya P, Neuwald AF, Spector DL. Proteomic analysis of interchromatin granule clusters. Molecular biology of the cell. 2004;15:3876-90.

[44] Takata H, Nishijima H, Ogura S, Sakaguchi T, Bubulya PA, Mochizuki T, et al. Proteome analysis of human nuclear insoluble fractions. Genes to cells : devoted to molecular & cellular mechanisms. 2009;14:975-90.

[45] Sun CT, Lo WY, Wang IH, Lo YH, Shiou SR, Lai CK, et al. Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. The Journal of biological chemistry. 2001;276:24059-67.

[46] Birney E, Kumar S, Krainer AR. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic acids research. 1993;21:5803-16.

[47] Aravind L, Koonin EV. G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. Trends in biochemical sciences. 1999;24:342-4.

[48] Saunders LR, Barber GN. The dsRNA binding protein family: critical roles, diverse cellular functions. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2003;17:961-83.

[49] Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009;136:701-18.

[50] Sharma A, Takata H, Shibahara K, Bubulya A, Bubulya PA. Son is essential for nuclear speckle organization and cell cycle progression. Molecular biology of the cell. 2010;21:650-63.

[51] Martens JH, Stunnenberg HG. The molecular signature of oncofusion proteins in acute myeloid leukemia. FEBS letters. 2010;584:2662-9.

[52] Yan M, Kanbe E, Peterson LF, Boyapati A, Miao Y, Wang Y, et al. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. Nature medicine. 2006;12:945-9.

[53] Yan M, Burel SA, Peterson LF, Kanbe E, Iwasaki H, Boyapati A, et al. Deletion of an AML1-ETO C-terminal NcoR/SMRT-interacting region strongly induces leukemia development. Proceedings of the National Academy of Sciences of the United States of America. 2004;101:17186-91.

[54] Ahn EY, Yan M, Malakhova OA, Lo MC, Boyapati A, Ommen HB, et al. Disruption of the NHR4 domain structure in AML1-ETO abrogates SON binding and promotes leukemogenesis. Proceedings of the National Academy of Sciences of the United States of America. 2008;105:17103-8.

[55] Huen MS, Sy SM, Leung KM, Ching YP, Tipoe GL, Man C, et al. SON is a spliceosome-associated factor required for mitotic progression. Cell cycle. 2010;9:2679-85.

[56] Kruhlak MJ, Lever MA, Fischle W, Verdin E, Bazett-Jones DP, Hendzel MJ. Reduced mobility of the alternate splicing factor (ASF) through the nucleoplasm and steady state speckle compartments. The Journal of cell biology. 2000;150:41-51.

[57] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860-921.

[58] Tarn WY, Steitz JA. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell. 1996;84:801-11.

[59] Sharp PA. The discovery of split genes and RNA splicing. Trends in biochemical sciences. 2005;30:279-81.

[60] Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. Nature reviews Genetics. 2004;5:389-96.

[61] Burns CG, Ohi R, Mehta S, O'Toole ET, Winey M, Clark TA, et al. Removal of a single alpha-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in Saccharomyces cerevisiae. Molecular and cellular biology. 2002;22:801-15.

[62] Venables JP. Aberrant and alternative splicing in cancer. Cancer research. 2004;64:7647-54.

[63] Ward AJ, Cooper TA. The pathobiology of splicing. The Journal of pathology. 2010;220:152-63.

[64] Srebrow A, Kornblihtt AR. The connection between splicing and cancer. Journal of cell science. 2006;119:2635-41.

[65] Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, et al. p53 isoforms can regulate p53 transcriptional activity. Genes & development. 2005;19:2122-37.

[66] Mazoyer S, Puget N, Perrin-Vidoz L, Lynch HT, Serova-Sinilnikova OM, Lenoir GM. A BRCA1 nonsense mutation causes exon skipping. American journal of human genetics. 1998;62:713-5.

[67] Chen LL, Sabripour M, Wu EF, Prieto VG, Fuller GN, Frazier ML. A mutation-created novel intra-exonic pre-mRNA splice site causes constitutive activation of KIT in human gastrointestinal stromal tumors. Oncogene. 2005;24:4271-80.

[68] Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nature structural & molecular biology. 2007;14:185-93.

[69] Ge K, DuHadaway J, Du W, Herlyn M, Rodeck U, Prendergast GC. Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. Proceedings of the National Academy of Sciences of the United States of America. 1999;96:9689-94.

[70] Ahn EY, DeKelver RC, Lo MC, Nguyen TA, Matsuura S, Boyapati A, et al. SON controls cell-cycle progression by coordinated regulation of RNA splicing. Molecular cell. 2011;42:185-98.

[71] Sharma A, Markey M, Torres-Munoz K, Varia S, Kadakia M, Bubulya A, et al. Son maintains accurate splicing for a subset of human pre-mRNAs. Journal of cell science. 2011;124:4286-98.

[72] Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nature reviews Genetics. 2010;11:345-55.

[73] Raymond C, Tom R, Perret S, Moussouami P, L'Abbe D, St-Laurent G, et al. A simplified polyethylenimine-mediated transfection process for large-scale and high-throughput applications. Methods. 2011;55:44-51.

[74] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry. 1976;72:248-54.

[75] Peterson GL. Determination of total protein. Methods in enzymology. 1983;91:95-119.

[76] Frangioni JV, Neel BG. Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. Analytical biochemistry. 1993;210:179-87.

[77] Askenazi M, Li S, Singh S, Marto JA. Pathway Palette: a rich internet application for peptide-, protein- and network-oriented analysis of MS data. Proteomics. 2010;10:1880-5.

[78] Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic acids research. 2006;34:D535-9.

[79] Mishra GR, Suresh M, Kumaran K, Kannabiran N, Suresh S, Bala P, et al. Human protein reference database--2006 update. Nucleic acids research. 2006;34:D411-4.

[80] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28:27-30.

[81] Herbert BR, Harry JL, Packer NH, Gooley AA, Pedersen SK, Williams KL. What place for polyacrylamide in proteomics? Trends in biotechnology. 2001;19:S3-9.

[82] Zhou G, Li H, DeCamp D, Chen S, Shu H, Gong Y, et al. 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. Molecular & cellular proteomics : MCP. 2002;1:117-24.

[83] Zhang G, Neubert TA. Automated comparative proteomics based on multiplex tandem mass spectrometry and stable isotope labeling. Molecular & cellular proteomics : MCP. 2006;5:401-11.
[84] Brun V, Masselon C, Garin J, Dupuis A. Isotope dilution strategies for absolute quantitative proteomics. Journal of proteomics. 2009;72:740-9.
[85] Everley PA, Krijgsveld J, Zetter BR, Gygi SP. Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. Molecular & cellular proteomics : MCP. 2004;3:729-35.
[86] Mattanovich D, Gasser B, Hohenblum H, Sauer M. Stress in recombinant protein producing yeasts. Journal of biotechnology. 2004;113:121-35.
[87] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology. 2008;26:1367-72.
[88] Fuller-Pace FV. DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic acids research. 2006;34:4206-15.
[89] Zonta E, Bittencourt D, Samaan S, Germann S, Dutertre M, Auboeuf D. The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. Nucleic acids research. 2013;41:554-64.
[90] Weidensdorfer D, Stohr N, Baude A, Lederer M, Kohn M, Schierhorn A, et al. Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. Rna. 2009;15:104-15.
[91] Jonson L, Vikesaa J, Krogh A, Nielsen LK, Hansen T, Borup R, et al. Molecular composition of IMP1 ribonucleoprotein granules. Molecular & cellular proteomics : MCP. 2007;6:798-811.
[92] Chauvin TR, Herndon MK, Nilson JH. Cold-shock-domain protein A (CSDA) contributes posttranscriptionally to gonadotropin-releasing hormone-regulated expression of Egr1 and indirectly to Lhb. Biology of reproduction. 2012;86:53.
[93] Kimura M, Ishida K, Kashiwabara S, Baba T. Characterization of two cytoplasmic poly(A)-binding proteins, PABPC1 and PABPC2, in mouse spermatogenic cells. Biology of reproduction. 2009;80:545-54.
[94] Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R. Isolation of an active step I spliceosome and composition of its RNP core. Nature. 2008;452:846-50.
[95] Lee SB, Kwon IS, Park J, Lee KH, Ahn Y, Lee C, et al. Ribosomal protein S3, a new substrate of Akt, serves as a signal mediator between neuronal apoptosis and DNA repair. The Journal of biological chemistry. 2010;285:29457-68.
[96] Russell L, Naora H, Naora H. Down-regulated RPS3a/nbl expression during retinoid-induced differentiation of HL-60 cells: a close association with diminished susceptibility to actinomycin D-stimulated apoptosis. Cell structure and function. 2000;25:103-13.

[97] Garand C, Guay D, Sereduk C, Chow D, Tsofack SP, Langlois M, et al. An integrative approach to identify YB-1-interacting proteins required for cisplatin resistance in MCF7 and MDA-MB-231 breast cancer cells. Cancer science. 2011;102:1410-7.

[98] Hao Y, Kong X, Ruan Y, Gan H, Chen H, Zhang C, et al. CDK11p46 and RPS8 associate with each other and suppress translation in a synergistic manner. Biochemical and biophysical research communications. 2011;407:169-74.

[99] Lindstrom MS, Nister M. Silencing of ribosomal protein S9 elicits a multitude of cellular responses inhibiting the growth of cancer cells subsequent to p53 activation. PloS one. 2010;5:e9578.

[100] Song KY, Choi HS, Law PY, Wei LN, Loh HH. Post-transcriptional regulation of mu-opioid receptor: role of the RNA-binding proteins heterogeneous nuclear ribonucleoprotein H1 and F. Cellular and molecular life sciences : CMLS. 2012;69:599-610.

[101] Stark M, Bram EE, Akerman M, Mandel-Gutfreund Y, Assaraf YG. Heterogeneous nuclear ribonucleoprotein H1/H2-dependent unsplicing of thymidine phosphorylase results in anticancer drug resistance. The Journal of biological chemistry. 2011;286:3741-54.

[102] Liu J, Huang B, Xiao Y, Xiong HM, Li J, Feng DQ, et al. Aberrant expression of splicing factors in newly diagnosed acute myeloid leukemia. Onkologie. 2012;35:335-40.

[103] Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell. 1998;92:315-26.

[104] Ritchie DB, Schellenberg MJ, MacMillan AM. Spliceosome structure: piece by piece. Biochimica et biophysica acta. 2009;1789:624-33.

[105] Lemm I, Ross J. Regulation of c-myc mRNA decay by translational pausing in a coding region instability determinant. Molecular and cellular biology. 2002;22:3959-69.

[106] Kim YK, Furic L, Parisien M, Major F, DesGroseillers L, Maquat LE. Staufen1 regulates diverse classes of mammalian transcripts. The EMBO journal. 2007;26:2670-81.

[107] Kuwano Y, Kim HH, Abdelmohsen K, Pullmann R, Jr., Martindale JL, Yang X, et al. MKP-1 mRNA stabilization and translational control by RNA-binding proteins HuR and NF90. Molecular and cellular biology. 2008;28:4562-75.

[108] Lau CK, Diem MD, Dreyfuss G, Van Duyne GD. Structure of the Y14-Magoh core of the exon junction complex. Current biology : CB. 2003;13:933-41.

[109] Tanackovic G, Ransijn A, Thibault P, Abou Elela S, Klinck R, Berson EL, et al. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. Human molecular genetics. 2011;20:2116-30.

[110] Laggerbauer B, Achsel T, Luhrmann R. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proceedings of the National Academy of Sciences of the United States of America. 1998;95:4188-92.

[111] Huang M, Rech JE, Northington SJ, Flicker PF, Mayeda A, Krainer AR, et al. The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. Molecular and cellular biology. 1994;14:518-33.

[112] Hovhannisyan RH, Carstens RP. Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively spliced exons. The Journal of biological chemistry. 2007;282:36265-74.

[113] Grainger RJ, Barrass JD, Jacquier A, Rain JC, Beggs JD. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. Rna. 2009;15:2161-73.

[114] Jans DA, Xiao CY, Lam MH. Nuclear targeting signal recognition: a key control point in nuclear transport? BioEssays : news and reviews in molecular, cellular and developmental biology. 2000;22:532-44.

[115] Kutay U, Guttinger S. Leucine-rich nuclear-export signals: born to be weak. Trends in cell biology. 2005;15:121-4.

[116] Rousseau S, Morrice N, Peggie M, Campbell DG, Gaestel M, Cohen P. Inhibition of SAPK2a/p38 prevents hnRNP A0 phosphorylation by MAPKAP-K2 and its interaction with cytokine mRNAs. The EMBO journal. 2002;21:6505-14.

[117] Viphakone N, Hautbergue GM, Walsh M, Chang CT, Holland A, Folco EG, et al. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nature communications. 2012;3:1006.

[118] Jansen RP, Niessing D. Assembly of mRNA-protein complexes for directional mRNA transport in eukaryotes--an overview. Current protein & peptide science. 2012;13:284-93.

[119] Guo D, Hazbun TR, Xu XJ, Ng SL, Fields S, Kuo MH. A tethered catalysis, two-hybrid system to identify protein-protein interactions requiring post-translational modifications. Nature biotechnology. 2004;22:888-92.

[120] Bao L, Redondo C, Findlay JB, Walker JH, Ponnambalam S. Deciphering soluble and membrane protein function using yeast systems (Review). Molecular membrane biology. 2009;26:127-35.

[121] Dieckelmann M, Roddam LF, Jennings MP. Purification of post-translationally modified proteins from bacteria: homologous expression and purification of histidine-tagged pilin from Neisseria meningitidis. Protein expression and purification. 2003;30:69-77.

[122] Haacke A, Fendrich G, Ramage P, Geiser M. Chaperone over-expression in Escherichia coli: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. Protein expression and purification. 2009;64:185-93.

[123] Boldt K, Mans DA, Won J, van Reeuwijk J, Vogt A, Kinkl N, et al. Disruption of intraflagellar protein transport in photoreceptor cilia causes Leber congenital amaurosis in humans and mice. The Journal of clinical investigation. 2011;121:2169-80.

[124] Gagnon JA, Kreiling JA, Powrie EA, Wood TR, Mowry KL. Directional transport is mediated by a Dynein-dependent step in an RNA localization pathway. PLoS biology. 2013;11:e1001551.

[125] Villace P, Marion RM, Ortin J. The composition of Staufen-containing RNA granules from human cells indicates their role in the regulated transport and translation of messenger RNAs. Nucleic acids research. 2004;32:2411-20.

[126] Lebeau G, DesGroseillers L, Sossin W, Lacaille JC. mRNA binding protein staufen 1-dependent regulation of pyramidal cell spine morphology via NMDA receptor-mediated synaptic plasticity. Molecular brain. 2011;4:22.

7 Annex

Figure index

Table index

8 Publication and oral presentations

Publications

Kucharska J, Del Río P, Arango-Gonzalez B, Gorza M, Feuchtinger A, Hauck SM, Ueffing M. Cyr61 activates retinal cells and prolongs photoreceptor survival in rd1 mouse model of retinitis pigmentosa. J Neurochem. 2014 Mar 4

Iuso A, Sibon OC, Gorza M, Heim K, Organisti C, Meitinger T, Prokisch H. Impairment of Drosophila orthologs of the human orphan protein C19orf12 induces bang sensitivity and neurodegeneration. PLoS One. 2014 Feb 21.

Texier Y, Toedt G, Gorza M, Mans DA, van Reeuwijk J, Bolz S, Horn N, Willer J, Katsanis N, Roepmann R, Gibson TJ, Ueffing M, Boldt K. EPASIS: Elution profile analysis of SDS-induced sub-complexes by quantitative mass spectrometry. Mol Cell Proteomics. 2014 Feb 21

Haack TB; Gorza M; Danhauser K; Mayr JA; Haberberger B; Wieland T; Kremer L; Strecker V; Graf E; Memari Y; Ahting U; Kopajtich R; Wortmann SB; Rodenburg RJ; Kotzaeridou U; Hoffmann GF; Sperl W; Wittig I; Wilichowski E; Schottmann G; Schuelke M; Plecko B; Stephani U; Strom TM; Meitinger T; Prokisch H; Freisinger P. Phenotypic spectrum of eleven patients and five novel MTFMT mutations identified by exome sequencing and candidate gene screening. Mol Genet Metab. 2014 Mar;111(3):342-52

Gai X, Ghezzi D, Johnson MA, Biagosch CA, Shamseldin HE, Haack TB, Reyes A, Tsukikawa M, Sheldon CA, Srinivasan S, Gorza M, Kremer LS, Wieland T, Strom TM, Polyak E, Place E, Consugar M, Ostrovsky J, Vidoni S, Robinson AJ, Wong LJ, Sondheimer N, Salih MA, Al-Jishi E, Raab CP, Bean C, Furlan F, Parini R, Lamperti C, Mayr JA, Konstantopoulou V, Huemer M, Pierce EA, Meitinger T, Freisinger P, Sperl W, Prokisch H, Alkuraya FS, Falk MJ, Zeviani M. Mutations in FBXL4, encoding a

mitochondrial protein, cause early-onset mitochondrial encephalomyopathy. Am J Hum Genet. 2013 Sep 5;93(3):482-95.

Haack TB, Haberberger B, Frisch EM, Wieland T, Iuso A, Gorza M, Strecker V, Graf E, Mayr JA, Herberg U, Hennermann JB, Klopstock T, Kuhn KA, Ahting U, Sperl W, Wilichowski E, Hoffmann GF, Tesarova M, Hansikova H, Zeman J, Plecko B, Zeviani M, Wittig I, Strom TM, Schuelke M, Freisinger P, Meitinger T, Prokisch H. "Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing." J Med Genet. 2012 Apr;49(4):277-83.

Zucchelli S, Vilotti S, Calligaris R, Lavina ZS, Biagioli M, Foti R, De Maso L, Pinto M, Gorza M, Speretta E, Casseler C, Tell G, Del Sal G, Gustincich S. "Aggresome-forming TTRAP mediates pro-apoptotic properties of Parkinson's disease-associated DJ-1 missense mutations." *Cell Death Differ*. 2009 Mar;16(3):428-38

Oral presentations

Young researcher Vision camp; Leibertingen, Germany: "Identification of interaction networks relevant for retinal degeneration", June 2010

4th Annual meeting RETNET; Budapest, Hungary: "SON-DNA binding protein characterization", June 2007

Medical Biomolecular meeting; Trieste, Italy: "Discovery of a novel interactor of DJ-1: TTAP", September 2005

9 Acknowledgements

In these long years I had the pleasure and luck to work and establish great relationships with many colleagues. They contributed not only to succeed with my thesis but most importantly, they helped me with my scientific formation and with new skills development, each one in a special way.

I would like to deeply thank Marius for the great chance he gave me to work in a high committed lab together with great people. In these years he strongly supported me with the thesis and he pushed me firmly to understand my limits and to overcome them. He didn´t limit himself to the mere supervision but his criticism and suggestions gave me the precious chance to improve myself far beyond my expectations.

I would like to thank Prof. Jerzy Adamski for the great interest he showed for my thesis and the important critics to my work. I am thankful for the advices about my thesis and the complicated bureaucratic procedures connected with the university.

I would like to thank Karsten for the huge help he provided me in the last years of my work. His contribution to my formation has been fundamental. He introduced me to the complex and fascinating world of mass spectronomy, spending most of the time teaching me all the basis and principle of this technology. He taught me how to keep focused on the topic, bringing me to the final form of the thesis and the relative paper. Most important of all, thanks to Karsten I discovered that people can be friend without affecting the quality of their work.

I would like to thank Elöd, a really eccentric scientist. From him I learned that are existing many different ways to look at a problem and do exist even more possible solutions. In the beginning of my PhD work he introduced me to the funny world of yeast two-hybrid. He spend a lot of his time teaching me many new techniques and, for these reasons, I am really grateful to him.

For the great help provided me during the writing of this work and the relative paper, I am indebted to three special people, really committed to the success of my PhD work. I am really grateful to Andrea for the outstanding teaching of protein complexes depiction and analysis, and for the important corrections of the paper. I am really thankful to Andreas (GV) for the extensive corrections on the thesis and for the great friendship. The time

spent with him as bench-neighbors have been amazing. Heartfelt thanks to Norbert for the support with the annoying scientific questions I had. He has been a great example as a highly-skilled person but still relaxed enough to let an easy talk to take place.

From the PROT laboratory in Munich, I would like to thank many people that in these years had always a moment to speak with me and have helped me, at least at psychological level to remain of sound mind. Thanks a lot to Alex for the great time spent teaching me the various practical issues about mass spectrometry and for the outstanding help provided for the german sections of this thesis. Thanks a lot to Joanna for the nice friendship and the rare opportunity to establish collaboration in the lab. Thanks a lot to Marcel for being patient with me during the last year of the thesis writing and accepting my messy desk. Hertfelt thanks to Sandra, Silke and Hakan for the great help they provided me during the setting up of the first mass spectronomy experiments, for the great patience demonstrating in those confused starting periods and for the great scientific contribution they gave to my work. I had really appreciated their frankness, even if sometimes the reality could be harsh. Many thanks also to Saskia and Steffi for the support provided in the long years in the lab.

From the lab in Tübingen I would like to thank Yves and Felix for the great friendship they gave me. Both of them have been really important for experiment set-up and for nice scientific and not scientific discussions. I would like to thank them for the great time spent out of the lab.

Special thanks to the IHG laboratory that in the last 16 months of my thesis gave me the great chance to work for them and, in the meantime, to allow me to finalize the thesis and the paper. Heartfelt thanks to Prof. Meitinger and to Holger that truly believed in me and offered me an unexpected chance to work for them. Holger lead me to the new world of mitochondria and allowed me to obtain the new job. Great thanks to Tobias and Arcangela, outstanding people, which took me by hand in the mitochondria and drosophila world. Their contributions to my further skill development have been extremely precious, like the great friendship they gave to me. Thanks a lot to Robert and Marieta for the funny and serious moments in the lab.

I would like to thank in particular some special people that endlessly supported me in these long years. They are the most important people of my life and without them would have been impossible for me to reach this milestone. I thank my father, my mother and my

brother, for the great chance they gave me, for the affectionate endless support and for the great time we had together in Italy and here in Munich.

My last thanks, the most important of all, are for my beautiful and patient Jennifer. In these three years she was able to keep my motivation high enough to finish this thesis, being also able to provide me superlative support. Since 50% of my internal strength came from her, she has been the biggest contributor to this work.

10 Curriculum vitae

Personal Information

Name	Gorza, Matteo
Address	Oberhofer platz 5, 80807 München
Email	matteogorza@hotmail.com
Nationality	Italian
Date of birth	07/04/1981
Place of birth	Udine (Italy)

Education

January 2008 – July 2014	PhD thesis At the department of protein science at the Helmholtz Zerntrum München,German Research Center for Environmental Health, Munich, Germany. Supervisor Prof. Dr. Marius Ueffing Title: "Molecular dissection of SON's interactome reveals the molecular link to splicing"
February 2005 – April 2006	Master degree thesis at the International School for Advanced Studies (SISSA), Trieste, Italy at the Molecular Neurobiology department. Supervisor: Prof. S. Gustincich Title: "Molecular characterization of a DJ-1 interactor: TTRAP"
September 2003 – April 2006	Studies in medical biotechnology (MSc) At the Medicine and Surgery faculty, University of Trieste, Italy.

	Qualification: Master degree. Final score: 1.0
February 2003 – July 2003	Bachelor degree thesis at the Institute for Scientific Hospitalization and Cure (I.R.C.C.S.) Burlo Garofolo, Trieste, Italy, at the Human Genetics department Supervisor: Prof. S. Crovella Title: "Genetics of antimicrobial peptides: hBD-4 in non-human primates"
September 2000 – July 2003	Studies in Biotechnology (BSc) At the Medicine and Surgery faculty, University of Trieste, Italy. Qualification: Bachelor degree.
September 1995 – July 2000	High school diploma. At high school focused on Sciences "Liceo Scientifico Copernico", Udine, Italy

Working experience

Since October 2013	Part-time researcher At the Gene Center and Department of Biochemistry LMU München, Munich, Germany Dr. Fabiana Perocchi
May 2012 – September 2013	Part-time researcher On neurodegenerative disorder (NBIA). At the Institute of human genetics, Helmholtz Zentrum München, Munich, Germany

	Dr. Holger Prokisch
January 2008 – April 2012	PhD student At the Department of Protein Science, Helmholtz Zentrum München, Munich, Germany Prof. Dr. Marius Ueffing
August 2006 - December 2007	Early stage researcher Marie Curie European Retinal Research Training Network (FP6) at the Department of Protein Science, Helmholtz Zentrum München, Munich, Germany Prof. Dr. Marius Ueffing
April 2006 – August 2006	Post-diploma training in animal behavior and perfusion At the Neuroscience, Physiology and Pathology department, University of Trieste, Italy Prof. Giampiero Leanza

Training and supervision

January 2012 – April 2012	Training of master students from institute of Human Genetics, Helmholtz Zentrum Muenchen. Responsible: Dr. Prokisch. Subject: protein-related techniques. Outcome: Publication
March 2010 – July 2010	Training of PhD student from UCL, London, UK. Responsible: Dr. Körtvely. Subject: protein-related techniques and mass spectrometry

Language skills

Italian and Friulan: Mother tongue

English: Fluent in writing and speaking

German: Intermediate in writing and speaking

French: Basic skills