



From sequencing to function: A triad of RNA-guided processes in the nucleus - medium-sized ncRNAs, nuclear speckles and Rbfox1-mediated splicing

Dissertation

Zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

Vorgelegt der

Naturwissenschaftlichen Fakultät I – Biowissenschaften
der Martin-Luther-Universität Halle-Wittenberg

Von

Peter Zorn

1. Gutachter/in: Prof. Dr. Stefan Hüttelmaier
2. Gutachter/in: Prof. Dr. Christian R. Eckmann
3. Gutachter/in: Prof. Dr. Michaela Müller-McNicoll

Datum der öffentlichen Verteidigung: 29.05.2024

“The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day.”

- Albert Einstein

Table of Content

1. Introduction	1
1.1 Non-coding RNAs: important regulators of cellular functions	1
1.2 Non-coding RNA synthesis by RNA polymerase III	2
1.3 Non-coding RNAs synthesized by RNA polymerase III promoter types I,II	3
1.4 Non-coding RNAs synthesized by RNA polymerase III promoter types III	4
1.5 The medium sized non-coding RNA 4.5S_H	5
1.6 Membraneless organelles direct cellular complexes	7
1.7 Membraneless organelles in the nucleus: nuclear speckles	8
1.8 Membraneless organelles in the nucleus: paraspeckles	10
1.9 The long non-coding RNA Malat1	11
1.10 RNA-binding proteins: intricate modulators of cellular processes in RNPs	13
1.11 The RNA binding protein and splicing regulator SRSF1	14
1.12 The disordered RS-domain of SRSF1	15
1.13 RNA splicing mediated by the complex and highly dynamic spliceosome	17
1.14 The relevance of alternative splicing in higher eukaryotes	20
1.15 The RNA binding protein Rbfox1 mediates alternative splicing	21
1.16 Aim of the study	22
2. Materials and Methods	24
2.1 Materials	24
2.1.1 Chemicals, enzymes and reagents	24
2.1.2 Bacteria strains	24
2.1.3 Cell lines	24
2.1.4 LNA and siRNA sequences	25
2.1.5 Plasmids	25
2.1.6 Oligonucleotides and probes	26
2.1.7 Primary and secondary antibodies	29
2.1.8 Standard Kits and Assays	30
2.1.10 Devices	32
2.2 Methods	32
2.2.1 Cell biological methods	32
2.2.1.1 Cell cultivation	32
2.2.1.2 Cell differentiation	33
2.2.1.3 LNA and siRNA transfection	33
2.2.1.4 Plasmid transfection	34
2.2.2 Molecular biology methods	34

2.2.2.1 Cloning.....	34
2.2.2.2 RNA Isolation	36
2.2.2.3 Reverse transcription and quantitative real-time PCR (qRT-PCR)	36
2.2.2.4 Library preparation and msRNAseq.....	37
2.2.2.5 Northern blot analysis.....	37
2.2.2.6 <i>In vitro</i> transcription.....	38
2.2.3 Protein-biochemical methods	39
2.2.3.1 Protein extraction, SDS-PAGE and Western blot	39
2.2.3.2 Protein pulldowns	39
2.2.3.2 Protein co-immunoprecipitation / RNA immunoprecipitation (Co-IP / RIP).....	40
2.2.3.4 Quantitative proteomics by LC MS/MS.....	40
2.2.4 Microscopy.....	41
2.2.4.1 Bright field microscopy	41
2.2.4.2 Fluorescence microscopy with CLSM technique	41
2.2.4.3 Live-cell imaging via Incucyte S3	42
2.2.5 Computational analyses	43
2.2.5.1 Bioinformatic analyses of the RNA-Sequencing reads.....	43
2.2.5.2 Analysis of putative Rbfox1 targets	43
2.2.5.3 Functional annotation clustering.....	43
2.2.5.4 Cell counting of multinucleated cells	43
2.2.5.5 Publicly available datasets, databases and visulaization	44
3. Results	45
3.1 Modification of RNA-Sequencing for medium-sized ncRNA analyses.....	45
3.1.1 Adapting the library generation protocol for msRNA-Sequencing analysis ..	45
3.1.2 Identification of RNA-Polymerase III driven msRNA transcripts	47
3.1.3 Characterization of the medium-sized non-coding RNA POLAR	49
3.2 The medium sized non-coding RNA 4.5S _H	52
3.2.1 The 4.5S _H RNA is a nuclear POLIII transcript.....	52
3.2.2 The nuclear 4.5S _H RNA is mainly associated with splicing factors	55
3.2.3 SRSF1 and its RS-Domain recruit SRSF-family members to the 4.5S _H RNA..	58
3.2.4 Malat1 and 4.5S _H RNA expression is linked in mouse cells.....	61
3.2.5 Malat1 and 4.5S _H RNA contribute to splicing speckle integrity	63
3.3 Identification of novel Rbfox1 splicing targets.....	67
3.3.1 Rbfox1 regulates alternative splicing of focal adhesion genes.....	67
3.3.2 Alternative splicing of mVcl and ePxn is mediated by Rbfox1	71
3.3.3 Alternative splicing of ePxn is mediated by Rbfox1	73

3.3.3 Expression of Rbfox1 facilitates cardiomyoblast morphology and multinucleation	75
4. Discussion	79
4.1 The identification of unknown msRNAs by next generation sequencing	79
4.2 The intriguing relation of 4.5S_H RNA and SRSF1	82
4.3 Cooperation for nuclear speckle integrity - 4.5S_H and Malat1	84
4.4 Alternative focal adhesions and Rbfox1 – factors for multinucleation	88
5. Summary	92
6. References	95
7. Appendix	120
7.1 List of publications	120
7.2 List of abbreviations	121
7.3 List of Figures	124
7.4 List of Tables	125
Eidesstattliche Erklärung	I
Danksagung	II
Curriculum Vitae	IV

1. Introduction

1.1 Non-coding RNAs: important regulators of cellular functions

Gene expression is one of the fundamental processes in life playing a crucial role in development and therefore is relevant for cellular function throughout all organisms. The central dogma of biology in eukaryotes represents the synthesis of mRNA from the DNA in the nucleus and translocation to the cytoplasm for translation of this sequence into proteins. However, in humans the amount of protein coding genes is astonishingly low with just 2% of the genome encoding for translated mRNAs. Still, a large fraction of the DNA is transcribed into RNAs harboring fully functional and regulatory non-coding RNAs (ncRNAs) (Mattick and Makunin 2006). These *trans*-acting molecules have emerged as key regulators in all kingdoms of life, including prokaryotes (Storz et al. 2011), archaea (Babski et al. 2014) and eukaryotes (Grosshans and Filipowicz 2008). While most of the genome in bacteria contains protein coding genes, higher organisms contain a low, but stable number of genes translating to proteins (Hayashi et al. 2001). In fact, eukaryotes display a rising number of non-coding RNAs generated from intergenic and intronic sequences with increasing cellular complexity hinting towards a transcriptional shift from protein coding genes towards non-coding RNAs (Mattick and Makunin 2006; Frith et al. 2005). Most well-characterized ncRNAs and their respective families are distinguishable by their molecular role including RNA splicing (snRNAs), RNA modification (snoRNAs), translational (rRNAs, tRNAs) and transcriptional (7SK RNA) regulation (Fu 2014). The ncRNAs are generally transcribed by all three RNA polymerases with RNA polymerase I mainly transcribing rRNAs (except for 5S) and the remainder to be synthesized by RNA polymerase II and III. In addition to RNA polymerase driven transcription, snoRNAs are generated through RNA processing of the host gene (Calvo Sánchez and Köhn 2021). Due to their limited size and correlating with their function, most of these RNAs encompass structural features to enable interaction with RNA-binding proteins (RBPs) and facilitate their molecular role as ribonucleoproteins (RNPs) (e.g. snRNAs and snoRNAs).

In contrast to the small non-coding RNAs mentioned above, non-coding RNAs with a sequence length of more than 200 nucleotides are considered long non-coding RNAs (lncRNAs). These are often transcribed by RNA polymerase II as sense or antisense transcripts in relation to coding genes and processed through removal of introns by RNA splicing. Additionally, lncRNAs might contain typical mRNA 5'-end modifications like a cap structure and 3'-end polyadenylation (Wu et al. 2017). Although a high amount of predicted lncRNAs to be transcribed have not been functionally characterized, the importance of lncRNA mediated cellular processes has been revealed for several RNAs recently (e.g. XIST and Malat1) (Statello et al. 2021). LncRNAs serve a role in a variety of biological functions ranging from gene expression control in proximity to their own transcription loci (*cis*-acting) as well as

complex cellular processes spatially separated from their place of biosynthesis (*trans*-acting). These often require the RNA as a structural component much like the small ncRNAs and function is then facilitated by interaction with RNA-binding proteins or with nucleic acids directly (Kopp and Mendell 2018). The association of RNA-binding proteins is primarily mediated by sequence-specific contacts to RNA-elements. These motifs frequently lead to RNA-association of RBPs inside the nucleus and thereby nuclear retention of the lncRNA can be observed, which often correlates with their function (Statello et al. 2021). Over the past decades, many non-coding RNAs have been identified and characterized. However, the recent discovery of circular RNAs (circRNAs) and RNA interference (RNAi) as well as the ongoing identification of new lncRNAs indicate a broad and more diverse regulatory role for ncRNAs than previously anticipated (Sakshi et al. 2021; Statello et al. 2021; Mattick and Makunin 2006).

1.2 Non-coding RNA synthesis by RNA polymerase III

The transcription of RNA from a DNA template is facilitated by three different RNA polymerases in eukaryotes (RNA Polymerase I-III). Each polymerase is required for synthesis of a different subset of RNAs in cells. The RNA polymerase I is exclusively associated with genes encoding ribosomal RNA (rRNA). In contrast, RNA polymerase II is responsible for transcription of protein coding mRNAs, small nuclear RNAs involved in RNA splicing, lncRNAs and a proportion of microRNAs (miRNAs) (Kuehner et al. 2011). RNA polymerase III (POLIII) has been identified to solely synthesize a variety of non-coding RNAs including 5S ribosomal RNA, U6 snRNA, Y RNAs, tRNAs, 7SK and 7SL RNA. Additionally, POLIII has been suggested to transcribe multiple other uncharacterized ncRNAs based on ChIP-seq experiments (Park et al. 2017).

As mentioned before, RNA polymerase II transcripts are extensively modified involving splicing and 3'-end polyadenylation (mRNAs, lncRNAs) as well as a 5'-end capping (mRNAs, snRNAs). In contrast, transcripts generated by RNA polymerase III are modified to a much lower extent (Borodulina and Kramerov 2008). A broad feature of POLIII-transcripts is the oligo-U terminator at the 3'-end of the sequence commonly recognized by the La-protein, enhancing nascent RNA stability and presumably further RNA processing (Wolin and Cedervall 2002).

RNA polymerase III genes show a distinct set of promoters that are classified in three different categories (type I, II and III promoters). In contrast to RNA polymerase II transcription which can be activated by a huge variety of transcription factors (TFs), RNA polymerase III transcription can be initiated by assembly of specific POLIII TFs depending on the promoter type (TFIIIA, TFIIIB and TFIIIC). In particular, TFIIIB is required for all three transcription types (see Figure 1). This transcription complex consists of BDP1 (B double prime 1), TBP (TATA-box binding protein) and BRF1 (B-related factor 1) for type I/II and BRF2 for type III promoters,

respectively. Type II promoters feature specific internal A- and B-box elements in the gene sequence. These boxes are recognized by TFIIIC thereby recruiting TFIIIB to direct the protein complex of RNA-polymerase III to the transcription start site (TSS) of the target gene. In addition to the A-box, type I promoters contain the intermediate element (IE) and a C-box which are specifically bound by TFIIIA. For type III transcription initiation, the SNAPc (small nuclear RNA activating protein complex) associates with the proximal sequence element upstream of the transcription start site to recruit TFIIIB to a downstream TATA box. This complex directs the RNA polymerase III to the TSS to guide synthesis of the gene (Turowski and Tollervey 2016).

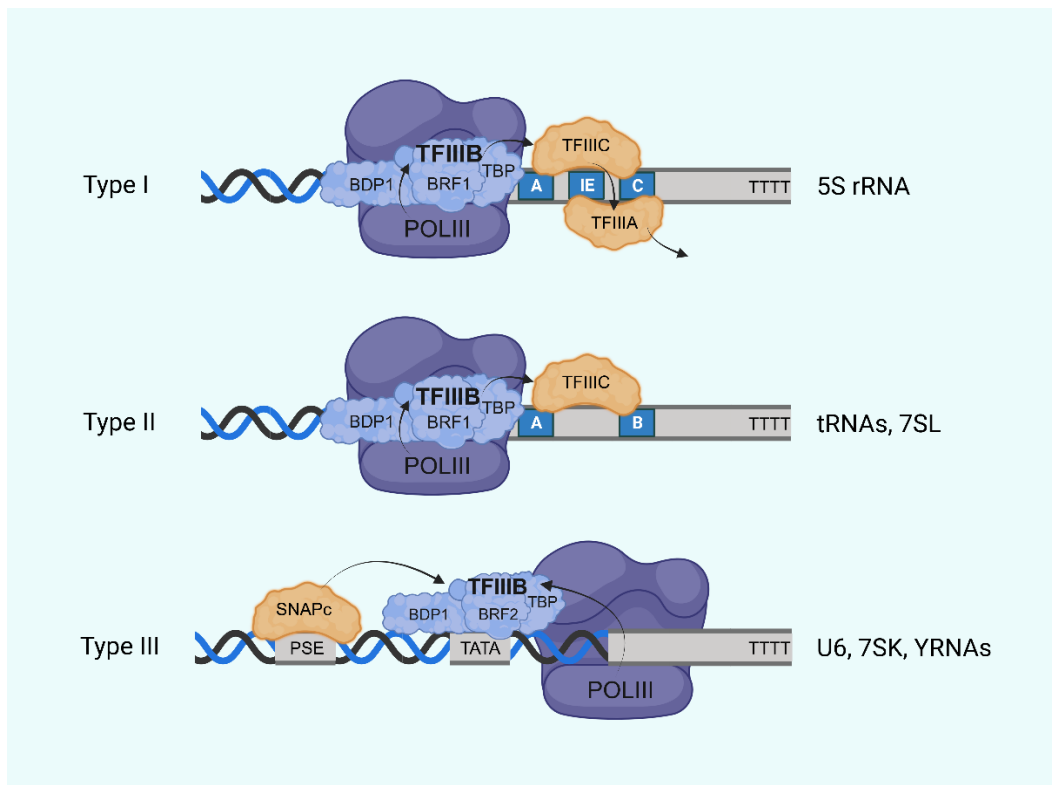


Figure 1. Schematic illustration of RNA Polymerase III transcription with promoter types I, II and III. Depiction represents DNA sequences including promoter elements and transcription factor complexes. The genes are represented in gray with oligo-T tail. Active POLIII transcription requires a set of transcription factors TFIIIA, TFIIIC, SNAPc and TFIIIB. The TFIIIC complex consists of three subunits BDP1, TBP, BRF1 for type I and II and BRF2 for type III transcripts, respectively. Transcripts for the individual promoter types are indicated on the right. *Cis*-regulatory elements are depicted in blue boxes for internal sequences (A – A-box, B – B-box, C – C-box, IE – intermediate element) and in gray boxes for upstream promoter types (PSE – proximal sequence element, TATA – TATA-box). Figure adapted from (Park et al. 2017)

1.3 Non-coding RNAs synthesized by RNA polymerase III promoter types I,II

The type I promoter facilitates transcription of the small ribosomal RNA 5S which has been shown to be essential for ribosome assembly and function as part of the large ribosomal subunit (Sun and Caetano-Anollés 2009). The precursor RNA transcribed by POLIII is bound by the chaperone protein La at the 3'-oligo-U terminator sequence. Further processing of the

RNA involves 3'-exonucleolytic cleavage to generate a 119-nucleotide mature 5S RNA with the La protein dissociating in the process as previously shown in the *Drosophila* system (Gerstberger et al. 2017). Additionally, the RNA associates with the ribosomal proteins RPL5 and RPL11 resulting in the formation of a 5S RNP which is later integrated into the 60S ribosome required for mRNA translation.

Transcripts with promoter type II among others comprise genes for tRNAs. Eukaryotic pre-tRNAs are recognized by the La protein which protects the premature tRNAs from 3'-exonuclease-mediated turnover and also prevents misfolding (Maraia et al. 2017). Maturation of the tRNA occurs after 5'- and 3'-end processing by RNase P (5'-cleavage) and RNase Z (3'-cleavage). Subsequently, tRNA nucleotidyltransferases add a CCA-oligomer to the 3'-end which is required for loading of the tRNA with amino acids and protein synthesis at ribosomes (Berg and Brandl 2021). Another important RNA generated through RNA polymerase III – type II promoters is the 7SL RNA. Upon transcription 7SL RNA is modified at the 3'-end by removal of the oligo-U stretch and instead adenosines are added to generate the mature transcripts (Sinha et al. 1998). The 300-nucleotide long 7SL RNA serves as a scaffold for association of proteins constituting the signal recognition particle (SRP). This RNP recognizes an N-terminal signal sequence of nascent polypeptides that are determined to be delivered to the ER for protein membrane integration or subsequent protein secretion (Akopian et al. 2013). Noteworthy, it is suggested that the 7SL RNA sequence is the predecessor of short interspersed nuclear elements (SINEs), commonly known as Alu elements in primates or B1 in rodents. These retrotransposons have integrated in mammalian genomes, representing about 10% of the total human genome (Lander et al. 2001). Though originating from a type II repeat element containing A- and B-box for transcription, many SINEs evolved into relevant non-coding RNAs through sequence insertions and deletions, contributing to complex cellular functions (e.g. BC200 RNA) (Kramerov and Vassetzky 2005).

1.4 Non-coding RNAs synthesized by RNA polymerase III promoter types III

As mentioned above, most small nuclear RNAs (snRNAs) are synthesized by RNA-polymerase II, however 7SK and U6 snRNA are type III POLIII-transcripts. Despite transcription through RNA polymerase III, the gene architecture resembles RNA polymerase II genes with proximal and distal instead of internal regulatory elements. The 7SK RNA is also trimmed, 3'-end oligoadenylated upon maturation and also modified on the 5'-end by addition of a 5'-monomethyl cap by the methyl phosphate capping enzyme (MePCE) (Sinha et al. 1998; Yang et al. 2019). This modification is required for the interaction of 7SK with the positive transcription elongation factor b (P-TEFb) in the nucleus. RNA polymerase II initially transcribes up to 100 nucleotides at the TSS before promoter-proximal pausing. Transcription can resume if P-TEFb is phosphorylated by Cdk9. The 7SK RNP (comprised of 7SK RNA,

MePCE, LARP7 and HEXIM1) inhibits the kinase activity of Cdk9 effectively stalling RNA polymerase II transcription (Nguyen et al. 2001; Yang et al. 2001).

U6 snRNA, like other snRNAs (U1, U2, U4 and U5) is an essential component of the major spliceosome involved in the RNA processing known as splicing. It forms a snRNA-protein complex referred to as the U4/U6.U5 tri-snRNP which will be elaborated in further detail in a later chapter (*see RNA splicing mediated by the complex and highly dynamic spliceosome*) (Wan et al. 2016). The U6 snRNA undergoes several processing steps including 5'-monomethyl cap addition by MePCE and 3'-oligouridylation by TUT1 which subsequently is trimmed by USB1 resulting in a mature transcript of 100 nucleotides in length (Nomura et al. 2018; Trippe et al. 2006; Shimba and Reddy 1994).

In addition to U6 and 7SK ncRNAs, RNA polymerase III also synthesizes Y RNAs as type III promoter transcripts. These RNAs are conserved among vertebrate species and feature a characteristic stem loop structure with a short segment (~20-30 nts) at the 5'- and 3'-end. These sequences form the double stranded region required for association with the Ro60 protein (Köhn et al. 2013). The La protein recognizes the POLIII typical oligo-(U) tract at the 3'-end of the Y RNA influencing RNA stability, subcellular localization and protein interactions (Simons et al. 1996). Y RNA processing involves 3'-oligo(U) trimming with subsequent dissociation of La and triphosphate removal by DUSP11 (Burke et al. 2016). The human genome encodes four different Y RNAs with varying length in the loop structure and different subcellular localization (hY1, hY3, hY4 and hY5). However, only Y1 and Y3 RNAs could be identified in mouse (Köhn et al. 2013). Numerous interactions with proteins indicate Y RNAs to be involved in a wide range of cellular processes including DNA replication, translation (interaction with nELAVL and HuD) as well as packaging into extracellular vesicles and virions together with APOBEC3G (Scheckel et al. 2016; Christov et al. 2006; Driedonks and Nolte-'t Hoen 2018).

As outlined above, the majority of the RNAs synthesized by RNA-Polymerase III are considered to play an important role in manifold cellular pathways and have been identified as core components of RNP complexes mediating a broad range of essential biological processes. The identification and characterization of novel POLIII transcripts might expand the understanding of complex molecular mechanisms and emphasize the importance of RNAs as crucial factors for cell fate (Zhou and van Bortle 2023).

1.5 The medium sized non-coding RNA 4.5S_H

The 4.5S_H RNA is a 92-94 nucleotide long non-coding RNA that is exclusively expressed in certain rodent lineages with internal elements resembling A- and B-boxes indicating an RNA polymerase III type II gene (Harada and Kato 1980). In line with characteristics of POLIII transcripts, the 4.5S_H RNA displays a high expression in cells and

also features the typical oligo-U terminator commonly bound by the La protein for stabilization of the RNA (Leinwand et al. 1982). Additionally, the transcribed RNA displays sequence homology to the short interspersed nuclear elements (SINEs). These transcripts typically consist of a 'head' region that originated from tRNAs or in case of the B1-SINE from the 7SL RNA, harboring the A- and B-box for POLIII transcription. The nucleotide sequence of the 'body' is 50-200 nucleotides long and unique for the SINE-family. SINEs might contain 3'-end sequences of corresponding long interspersed nuclear elements (LINEs) as they are required for reverse transcription. However, the remaining 'body'-region is comprised of sequences that are of unknown origin and function. The 'tail'-region commonly harbors simple sequence repeats like A-rich sequences or T-stretches which function as terminators for POLIII-transcription (reviewed in (Kramerov and Vassetzky 2005)). The sequence of the 4.5S_H RNA is similar to the B1-SINE element, but features an additional 20 nucleotide insertion and lacks the poly-A tract generally associated with retrotransposition (Harada and Kato 1980; Labuda and Zietkiewicz 1994; Kramerov and Vassetzky 2005). Despite this, the mouse genome encompasses a high copy number of 4.5S_H RNA genes (700-800 copies) arranged in tandem repeats. These gene sequences are variable in the flanking genomic regions but are highly conserved for the actual transcript sequence indicating a functional purpose of the 4.5S_H RNA (Gogolevskaya et al. 2005). Initial characterization of the RNA suggested an interaction with poly-(A)+ cytoplasmic RNA based on cross-linking analysis (Schoeniger and Jelinek 1986). Due to this observation, it was originally assumed that the 4.5S_H RNA is localized in the cytoplasm. However, the first functional study described an association of the RNA with Nucleolin assayed by immunoprecipitation analysis, implying a more nuclear localization of the RNA. In addition to the interaction with Nucleolin and the La protein it was also revealed that the 4.5S_H RNA interacts with several other unidentified proteins in a mobility shift assay, highlighting a potential function of the RNA in RNP-complexes (Hirose and Harada 2008). In addition, it has been reported that 4.5S_H RNA interacts with transcripts containing antisense sequences of the B1-SINE based on luciferase assays in Neuro2A cells. The same study also suggested the nuclear retention of these transcripts by the described interaction with the 4.5S_H RNA, as asB1 RNAs accumulated in the cytoplasm upon 4.5S_H RNA knockdown (Ishida et al. 2015). Although the authors indicated an RNA-RNA mediated function for the 4.5S_H, it was also shown that the RNA is co-localized with serine/arginine-rich splicing factor 1 (SRSF1) inside nuclear speckles hinting towards a potential function of 4.5S_H in RNA splicing.

1.6 Membraneless organelles direct cellular complexes

In eukaryotes, cellular processes are well-organized in a variety of compartments some of which contain a lipid bilayer membrane (e.g. mitochondria, endoplasmic reticulum and Golgi apparatus) and the so called membraneless organelles (MLOs). Since the first discovery of nuclear bodies by Ramón y Cajal including nuclear speckles or splicing speckles, various other MLOs have been identified in the nucleus and cytoplasm of mammalian cells (Figure 2). These structures are biomolecular condensates and usually contain specific protein and nucleic acid compositions individual for each type of MLO (Spector 2006). Because of this, many types of MLOs can be detected as bright foci in immunofluorescence microscopy when stained with antibodies against the respective marker proteins.

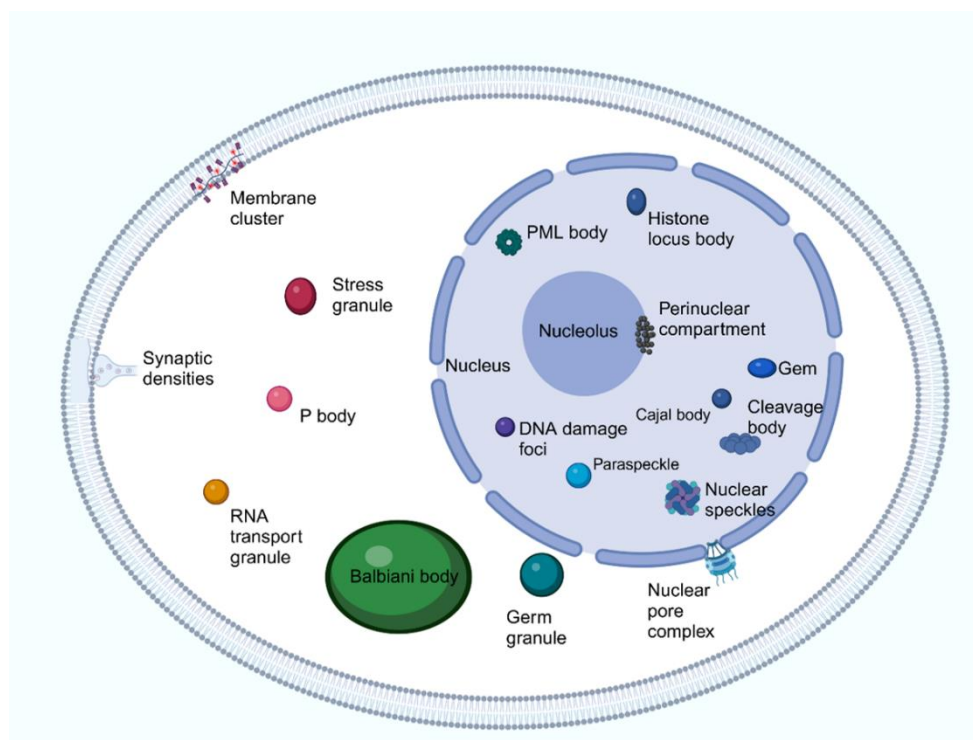


Figure 2. Membraneless organelles (MLOs) formed in eukaryotic cells. Illustration of known condensates found in the cytoplasm, nucleus and membrane of the cell. Certain organelles are cell type specific (Balbiani body and germ granule in germ cells; synaptic densities and RNA transport granules in neurons) while others require specific conditions to form (e.g., stress granules, DNA damage foci). Condensation of the depicted foci is often achieved by liquid-liquid phase separation (LLPS) of proteins detected in these MLOs. Figure adapted from (Banani et al. 2017).

The majority of proteins present in these organelles are connected to regulatory processes in the cell and often display RNA-binding capabilities (Mao et al. 2011). Additionally, the formation of some specific membraneless organelles can only be observed under cellular stress conditions as a protective mechanism (e.g. stress granules, nuclear stress bodies) (Riggs et al. 2020).

Despite their defined composition, the formation of MLOs is highly dynamic as they display liquid-like properties similar to liquid droplets and occur as a consequence of liquid-

liquid phase separation (LLPS). This process is dependent on the concentration of macromolecules in an aqueous solution as well as the characteristics of the biomolecule including length, hydrophobicity and charge distribution (Banani et al. 2017). These properties are relevant for a multivalent interaction of the proteins with each another, including intrinsically disordered proteins or nucleic acids like RNA often co-localizing inside the MLOs. While not exclusively present in condensates, intrinsically disordered proteins often feature low-complexity domains with a high percentage of charged or aromatic residues contributing to the phase separation process. However, folded protein domains, especially of RNA-binding proteins could also induce phase separation in combination with RNA (Shin and Brangwynne 2017). The prevalence of RNA-binding proteins suggests RNA to be an integral part of MLOs as they are either regulated in these condensates or contribute to exerting the function of the cellular bodies.

1.7 Membraneless organelles in the nucleus: nuclear speckles

The nucleus is a confined space with a large variety of different cellular processes involving countless proteins and nucleic acids that are required to exert their function on specific targets in a crowded environment. Subcellular organization and compartmentalization into MLOs enables clustering of proteins and RNAs relevant for a distinct process in close proximity to ensure rapid reactions or sequestering these factors to prevent unwanted interactions (Shin and Brangwynne 2017). The nucleus is highly organized and contains a variety of subcompartments or nuclear bodies including paraspeckles and nuclear speckles.

Nuclear speckles contain several snRNAs in associated RNPs (snRNPs), other non-coding RNAs, splicing factors and RNA polymerase II transcripts among others. These MLOs are detected in the nucleus and have previously been described as interchromatin granule clusters (ICGs) as they were observed in the interchromatin space by microscopy analyses (Spector and Lamond 2011). Recent findings by Fei et al. suggest nuclear speckles to be a multi-layered subcompartment with ICGs as the central 'core' region and an outside 'shell' (Figure 3, Fei et al. 2017). According to this publication the 'core' region is comprised of the SON DNA and RNA-binding protein (SON) and the serine/arginine repetitive matrix 2 (SRRM2) protein which have been shown to be relevant for nuclear speckle condensation (Ilik et al. 2020). The 'shell' is supposedly composed of nascent mRNA transcripts and snRNAs (U2 snRNA) as well as the long non-coding RNA Malat1 (metastasis associated lung adenocarcinoma transcript 1) and 7SK RNA. Additionally, proteomic analysis of the composition of ICGs revealed an enrichment of proteins involved in multiple cellular processes including epigenetic regulation, transcriptional modulation, 3'-end processing, splicing, mRNA modification and export (Galganski et al. 2017).

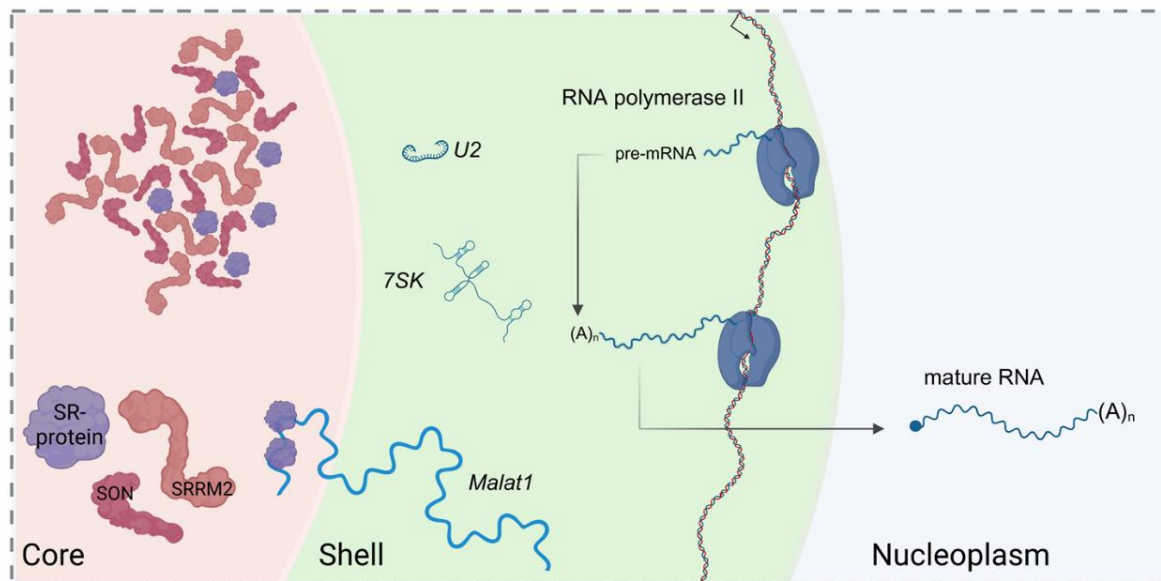


Figure 3. Nuclear speckle composition according to the ,core' and ,shell' hypothesis. Schematic representation of the organization of a nuclear speckle with a protein-rich core and an RNA-rich shell. The RS-domain containing proteins SRRM2 and SON have been identified as essential components for nuclear speckle formation. The liquid-liquid phase separation is supposedly mediated by the largely disordered RS-domain also involving other SR-proteins (Greig et al. 2020). The long non-coding RNA Malat1 is localized in the ,shell' together with U2 and 7SK RNAs. Additionally, nuclear speckles were identified in close proximity to active genomic loci with pre-mRNA and transcription factors detected in the shell. Figure adapted and modified from (Hirose et al. 2023)

A major class of factors enriched in splicing speckles are RS-domain containing proteins including the aforementioned SON and SRRM2 as well as the proteins of the serine/arginine-rich splicing factor family (SRSFs). Furthermore, corresponding kinases (CLK1, SRPK1) and phosphatases responsible for the phosphorylation state of these factors were also found in nuclear speckles (Spector and Lamond 2011).

The functional purpose of nuclear speckles has been debated for longer than the last two decades, however it remains controversially discussed. There are two general concepts regarding the biological role of nuclear speckles. The first mode of action is based on the enrichment of splicing factors and splicing related proteins inside nuclear speckles. In addition, no active transcription or DNA content was detected in ICGs which could implicate splicing speckles to be a reservoir for splicing and RNA processing factors. In line with this hypothesis, inhibition of transcription or splicing leads to enlarged splicing speckles, arguably buffering the concentration of splicing factors and RNA processing proteins in the nucleoplasm (Chen and Belmont 2019). The second model implicates nuclear speckles as 'gene expression hubs' aiding in transcriptional and co-transcriptional processes. This includes RNA splicing through cycling of factors between speckles and genes enabled by the liquid-like properties of MLOs (Fei et al. 2017). Supporting this hypothesis, the co-localization of splicing speckles in the periphery of active genes and gene-rich chromosomes is observed with increased frequency (Rieder et al. 2014; Shopland et al. 2003). Additionally, the presence of transcription elongation

factors (e.g. Elongin) as well as epigenetic regulators (e.g. TRIM28) in nuclear speckles indicates a role of this MLO connected to transcription (Galganski et al. 2017). Moreover, certain already spliced mRNAs (e.g. COL1A1) accumulate at the periphery or the 'shell' of nuclear speckles. This is even more pronounced when mRNA export was inhibited by depletion of the TREX complex (Wang et al. 2018; Hirose et al. 2023).

1.8 Membraneless organelles in the nucleus: paraspeckles

Paraspeckles can be observed in varying numbers as spherical bodies inside the nucleus of the majority of *in vitro* cell culture models with the exception of embryonic stem cells. These MLOs were originally identified by electron microscopy and named inter-chromatin granule-associated zones and were later described as nuclear bodies adjacent to ICGs or nuclear speckles (Visa et al. 1993). These subcompartments are enriched in the RNA-binding proteins NONO (non-POU domain containing octamer binding) and PSPC1 (Paraspeckle component 1) (Fox et al. 2002). Furthermore, the long non-coding RNA Neat1 (nuclear paraspeckle assembly transcript 1) is an essential component for the formation of paraspeckles, as depletion experiments resulted in the absence of these MLOs altogether. CRISPR-Cas mediated mutagenesis of the relevant transcript variant Neat1_2 emphasized the central region of the transcript as a major association site for RBP interaction, including NONO and SFPQ (splicing factor, glutamine-rich) (Yamazaki et al. 2019). In this context, paraspeckles also have been described to consist of a similar 'core' and 'shell' organization. The central part of Neat1_2 is localized in the 'core' region and both RNA ends (5'- and 3'-end) loop out into the 'shell'. The 'core' fraction is represented by association of NONO and SFPQ with the Neat1_2 central region and cooperative binding of FUS and RBM14 subsequently contributing to phase separation (Figure 4). The extended RNA 5'- and 3'-ends present in the 'shell' region supposedly sequester additional RBPs in accordance with the postulated function of paraspeckles (West et al. 2016; Yamazaki et al. 2018). In analogy to nuclear speckles, it has been suggested, that paraspeckles act as cellular sponges to sequester proteins and RNAs. It has been shown that SFPQ-dependent transcriptional targets were deregulated upon enhanced localization of SFPQ protein into paraspeckles. Moreover, certain mRNAs that contain a stemloop structure in the 3'UTR are constrained to paraspeckles to inhibit their nuclear export and translation (Hirose et al. 2014; Hirose et al. 2023). However, precise molecular mechanisms for the physiological role of paraspeckles remain to be elucidated.

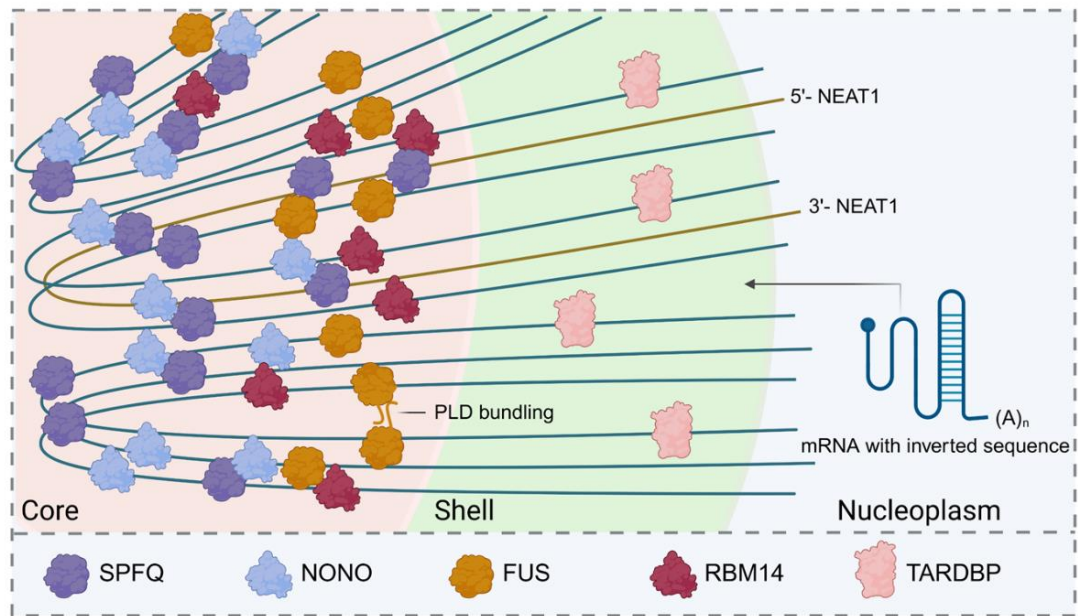


Figure 4 Paraspeckle composition according to the 'core' and 'shell' hypothesis. Schematic representation of the organization of a paraspeckle with the central Neat1 RNA sequence in the 'core' and both ends reaching into the 'shell'. The Neat1 RNA is essential for paraspeckle condensation but also requires NONO and SPFQ protein for MLO formation. Initial LLPS is mediated by recruitment of SPFQ and NONO followed by association of FUS and RBM14. The bundling of the NEAT1 RNA is supposedly achieved by multivalent interactions (e.g PLD – prion-like domains) of FUS consequently resulting in phase separation (Yamazaki et al. 2018). The NEAT1 RNA regions reaching into the shell sequester additional RBPs (e.g TARDBP) as well as mRNAs with loop structures in the 3'-region. Figure adapted and modified from (Hirose et al. 2023).

1.9 The long non-coding RNA Malat1

Despite the high amount of lncRNAs annotated in humans (17952 lncRNA genes) and mouse (13197 lncRNA genes) the majority displays low cellular expression, is restricted to certain cell types or developmental stages. (Djebali et al. 2012; Frankish et al. 2019). However, the lncRNA Malat1 is one of the few well studied lncRNAs due to its rather ubiquitous expression, high abundance and its link to pathologies like cancer (Gutschner et al. 2013; Hu et al. 2016). Initially, Malat1 was identified as Nuclear Enriched Abundant Transcript 2 (NEAT2) in tumor samples of non-small cell lung cancer in microarray analyses and is a highly expressed gene in metastases of this cancer type (Ji et al. 2003). The conservation of lncRNAs across species is rather poor if just the nucleotide level is considered. In addition to simple nucleotide conservation, the secondary structure and folding has to be taken into account for lncRNAs. Moreover, the same non-coding RNA can provide distinct functions in different organisms. The conservation of the chromosomal region from where the lncRNA is transcribed may also be important as this could affect neighboring genes in *cis* (Diederichs 2014). The gene locus of Malat1 is located at chromosome 11q13.1 in human and 19qA in mouse, which represents a multi-gene region with a large syntenic evolutionary conservation. Surprisingly,

the Malat1 transcripts also display an astonishing sequence conservation of more than 50% of the whole transcript and 80% at the 3'-end (Johnsson et al. 2014; Basu et al. 2013). In humans, the primary Malat1 transcript is processed at the 3'-end by RNase P and RNase Z which results in a 61 nucleotide long tRNA-like mascRNA (Malat1-associated small cytoplasmic RNA). The remainder of the 3'-end of the processed Malat1 transcript contains a genomically-encoded adenosine-stretch in contrast to conventional polyadenylation. This A-sequence has been shown to hybridize with an upstream poly-U sequence to form a triple helical structure enhancing Malat1 RNA stability and potentially enabling RBP-binding (Wilusz et al. 2012). Additionally, the Malat1 locus also transcribes an antisense transcript (TALAM1), that has been characterized to enhance 3'-end maturation of the sense transcript and contribute to Malat1 RNA expression by acting in a feed-forward positive regulatory loop (Zong et al. 2016).

The mature Malat1 transcript displays a predominantly nuclear localization with an enrichment in nuclear speckles, especially in the periphery or 'shell' as described previously (Hutchinson et al. 2007; Fei et al. 2017). The observed localization is mediated by several splicing factors also present in nuclear speckles such as RNPS1, SRRM1, and IBP160 that are required for proper Malat1 nuclear distribution (Miyagawa et al. 2012). Additionally, Malat1 RNA has been shown to interact with several proteins of the SRSF-family, including SRSF1, SRSF3 and SRSF5 which supposedly interact with recognition sites at the 5'-end of Malat1 (Tripathi et al. 2010). The same study also implied that Malat1 regulates the phosphorylation status of these SR-proteins resulting in altered localization and ultimately affect alternative pre-mRNA splicing. Likewise, upregulation of Malat1 RNA enhanced SRSF1 mRNA expression and protein amount, altering SRSF1 mediated alternative splicing towards pro-oncogenic isoforms (Malakar et al. 2017). It was shown by Bernard et al. that Malat1 recruits SRSF1 to actively transcribed loci of transgenic LacI and knockdown of Malat1 resulted in decreased co-localization of SRSF1 at the transcription site (Bernard et al. 2010). Additionally, Malat1 RNA has been cross-linked *in vivo* to chromatin of actively transcribed genes and was shown to interact with several transcription factors and transcription activators like LTBP3, FOXO1, PC2 and HMGA2 (Yang et al. 2011; Li et al. 2014; Zhou et al. 2018). The molecular role of Malat1 in pre-mRNA splicing is defined by its subcellular localization in nuclear speckles and the association with splicing factors (Bernard et al. 2010). In accordance with this, modulation of Malat1 expression has been shown to regulate alternative splicing of pre-mRNAs in various cancer types (Malakar et al. 2017; Zhang et al. 2012; Arun et al. 2016). Another study found RNA-RNA interaction of Malat1 with the U1 snRNA and has proposed association of the Malat1 with nascent pre-mRNAs. This interaction is supposedly mediated through interaction with SR-proteins enabling RNA splicing of these transcripts (Engreitz et al. 2014).

In view of the ambiguous functions of Malat1 RNA it appears to be a critical component for RNA splicing. However, despite the high conservation of key sequence elements in mouse (Nakagawa et al. 2012) Malat1 it appears to be less relevant compared to human Malat1. While Malat1 is not essential for the formation of splicing speckles in human HeLa cells, knockdown of the RNA results in abnormal localization of splicing factors including SRSF1 (Tripathi et al. 2010). In contrast, deletion of Malat1 in mouse models did not alter the localization of splicing factors SRSF1 or SRSF2 in various organs analyzed by FISH/immunofluorescence analyses (Nakagawa et al. 2012). Consistently, mouse Malat1 has been reported to be dispensable for mouse development and physiology, as knockout mouse models did not report any modulation of SR-protein expression, phosphorylation or aberrant splicing defects. Although gene expression remained largely unaffected, neighboring genes of Malat1 in brain tissues were apparently upregulated by 1.5-2 fold implying a potential regulatory role of Malat1 in *cis* (Zhang et al. 2012). Nonetheless, the discrepancy of Malat1 RNA is hinting towards a functional redundancy in mouse compared to human cells (Nakagawa et al. 2012).

1.10 RNA-binding proteins: intricate modulators of cellular processes in RNPs

Many cellular processes are facilitated by RNA-protein interactions mediated by RNA-binding proteins forming dynamic ribonucleoprotein complexes (RNPs). It is widely accepted that this interaction is facilitated by recognition of short nucleotide sequences or structural folds in the RNA by globular RNA-binding domains (RBDs) including RNA recognition motifs (RRMs), hnRNP K homology domains (KH-domains), zinc fingers and the DEAD box helicase domains among others (Lunde et al. 2007). However, up to one third of identified RBPs do not confer with the classical organization of structured RBDs only. They also comprise disordered amino acid sequences including R/S- and RG/G-repeats contributing to RNA-binding or even enable RNA-binding directly (Kiledjian and Dreyfuss 1992; Valcárcel et al. 1996). The RNA binding capability and supposedly function however is depending on the type of RBP, the number of RBDs and the association with other RBPs on one RNA molecule (Lunde et al. 2007). Considering the highly variable types of RNA molecules based on sequence and secondary structure, RNA-binding specificity is required to be very flexible. Essential cellular processes of mRNA metabolism require RNA-binding proteins to recognize virtually all mRNAs, which is often mediated by conserved sequence motifs present in the majority of transcripts or interaction with the phosphate-backbone of the RNA. These core factors facilitate important basic functions necessary for mRNA maturation including mRNA processing in the form of RNA splicing, polyadenylation, capping and mRNA export. In contrast to this, RBPs with specific RNA targets recognize defined sequences or associate to secondary structures in their target transcripts with a higher affinity compared to other RNAs. Consequently, these RBPs target only subclasses of mRNAs and regulate their fate by interaction with more global

RBPs involved in RNA processing (Gehring et al. 2017). One example involves proteins of the SR-family recognizing specific exonic splicing enhancer sites (ESEs) of transcripts in *cis* and recruiting *trans*-acting regulatory factors. Thereby SR-proteins can guide these *trans*-acting factors as well as the splicing machinery to adjacent splice sites effectively modulating alternative splicing of the transcript (Wahl et al. 2009).

1.11 The RNA binding protein and splicing regulator SRSF1

The serine/arginine-rich splicing factor 1 (SRSF1) is the archetypical RNA-binding protein among the serine/arginine-rich family consisting of 12 proteins in *homo sapiens*. The modular structure of the SRSF-proteins is comprised of one or two RRM domains and a C-terminal RS-domain of varying length with multiple Arg-Ser dipeptide repeats (Manley and Krainer 2010). The SRSF1 protein was first discovered among the family members due to its regulation of alternative splicing of the SV40 early pre-mRNA and promoting spliceosome assembly and constitutive splicing in S100 HeLa extracts (Ge and Manley 1990; Krainer et al. 1990a, 1990b). Although originally identified as a splicing factor, recent discoveries have implied additional roles of SRSF1 in transcriptional and translational control, nuclear export and nonsense-mediated decay. Due to this involvement in crucial cellular functions, SRSF1 has been shown to be an essential gene since SRSF1 null mice display lethal phenotypes (Xu et al. 2005). Additionally, enhancing the expression of SRSF1 has led to oncogenic transformation of cells originating from mouse and human (Das and Krainer 2014).

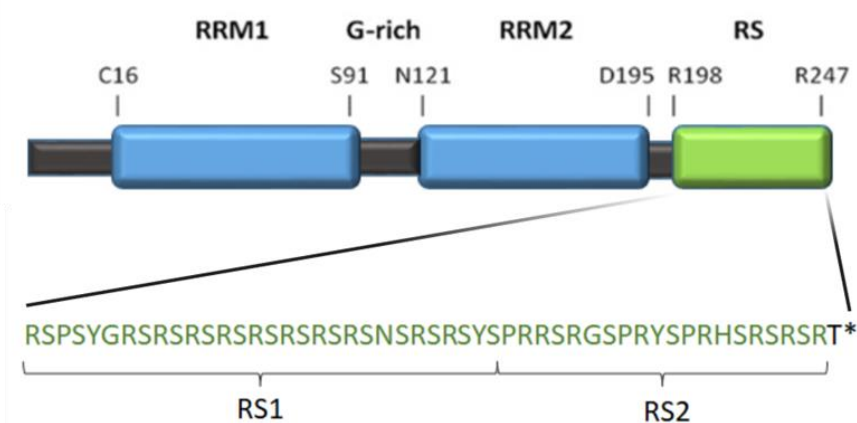


Figure 5. The domain organization of SRSF1. The SRSF1 protein consists of two separate RRM domains followed by a C-terminal RS-domain. The RRM domains impose a synergistic effect on RNA binding with RRM2 mediating binding specificity towards GA-rich RNA sequences. The RS-domain is largely disordered and undergoes several phosphorylation steps mediated by SRPK1 for RS1 and CLK1 phosphorylating proline-rich RS-repeats in RS2. The phosphorylation is required for subcellular localization into nuclear speckles. Additional modifications can also be observed e.g. methylation of R93, R97 and R103 in the G-rich linker region.

The SRSF1 protein contains two N-terminal RRM-domains separated by a short linker and the C-terminal RS-domain with 15 dipeptide repeats (Figure 5). The double RRM structure imposes a synergistic effect on RNA binding and splicing. However, the pseudo-RRM2 has

been shown to mediate substrate specificity and is responsible for association with GA-rich ESE sequences as demonstrated in NMR analysis (Cléry et al. 2013; Cáceres et al. 1997). The association of SRSF1 with the ESE sequences leads to recruitment of the U2 snRNP (U2AF35/65) to the 3'-splice site and the U1 snRNP (U1-70k) to the intronic 5'-splice site enhancing early spliceosome assembly (Wu and Maniatis 1993; Staknis and Reed 1994). SRSF1 interaction with the U1-70k protein of the U1 snRNP has been suggested to be mediated by the RRM domains rather than the RS-domain, which previously was described to be important for protein-protein interaction (Wu and Maniatis 1993; Cho et al. 2011). Additionally, one study found that SR-proteins are required for maturation of *in vitro* assembled pre-spliceosome by recruiting the U4/U6.U5 tri snRNP (Rosciogno and Garcia-Blanco 1995). In case of alternative splicing events, SRSF1 has been shown to be involved in facilitating exon inclusion as well as exon skipping depending on ESE binding site position. The promotion of exon inclusion is occurring on preferential binding to the ESEs present in the alternatively spliced exon whereas association with constitutive flanking exons was shown to enhance exon skipping (Pandit et al. 2013). This mode of regulation is antagonized by hnRNP A/B family members in a concentration dependent manner. Increased association of SRSF1 promotes proximal 5'-splice site selection *in vitro* and *in vivo* whereas hnRNP A/B proteins tend to enhance distal 5'-splice site choice (Cáceres et al. 1994; Expert-Bezançon et al. 2004).

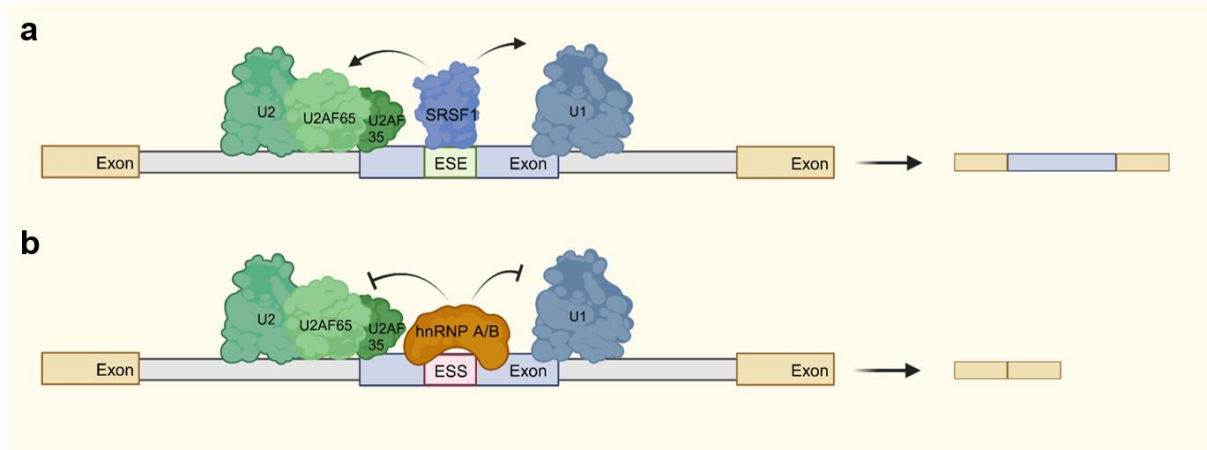


Figure 6. SRSF1 and hnRNP A/B mediated recruitment of U1 and U2 snRNP. **a** SRSF1 binds to exonic splicing enhancer (ESE) sites actively recruiting U1 snRNP and U2AF35/65 followed by U2 snRNP to the 3'- and 5'-splice site effectively mediating exon inclusion. **b** hnRNP A/B recognizes exonic splicing silencer (ESS) sequences in the exon inhibiting association of U1 and U2 snRNP consecutively promoting exon skipping. Flanking exons are marked in yellow and regulated exon in blue with introns represented by gray bars.

1.12 The disordered RS-domain of SRSF1

The C-terminal RS-domain comprised of several Arg-Ser-dipeptide repeats is extensively phosphorylated altering SRSF1 functions and subcellular localization. The RS-domain of SRSF1 can be divided into RS1 and RS2 domains, with the former domain containing Arg-Ser dipeptide repeats and the latter additional Pro-Ser repeats in (Figure 5).

Primary phosphorylation of the N-terminal (RS1) part is occurring in the cytoplasm mediated by SR-protein kinase 1 (SRPK1) of at least two out of eight serine residues generating a hypophosphorylated SRSF1 (Aubol and Adams 2011). The nuclear import is then facilitated by interaction of the hypophosphorylated SRSF1 and the SR-specific transportin-SR protein (Lai et al. 2000). Upon nuclear entry SRSF1 accumulates inside nuclear speckles and can further be phosphorylated by Clk/Sty kinase (CLK1) at RS2 (see Figure 5). This hyperphosphorylated SRSF1 has been implicated to exit nuclear speckles and is potentially recruited to active sites of transcription to promote spliceosome assembly (Paz et al. 2021). Originally, the RS-domain was thought to be involved in constitutive splicing by mediating protein-protein interactions with U1 and U2 snRNPs bridging 5'- and 3'-splice site for splicing activity (Wu and Maniatis 1993). In addition, phosphorylation of the RS-domain has been shown to be relevant for spliceosome assembly in *in vitro* pre-spliceosomal complexes and also enhances splicing activity (Cao et al. 1997). However, more recent studies implicated the RS-domain to be dispensable for RNA-binding and SRSF1-mediated splicing of constitutive and alternative exons. Here the exception are few mRNA transcripts containing weak splice sites that require U2AF35 association with AG dinucleotides at the 3'-splice site (Merendino et al. 1999; Hastings and Krainer 2001).

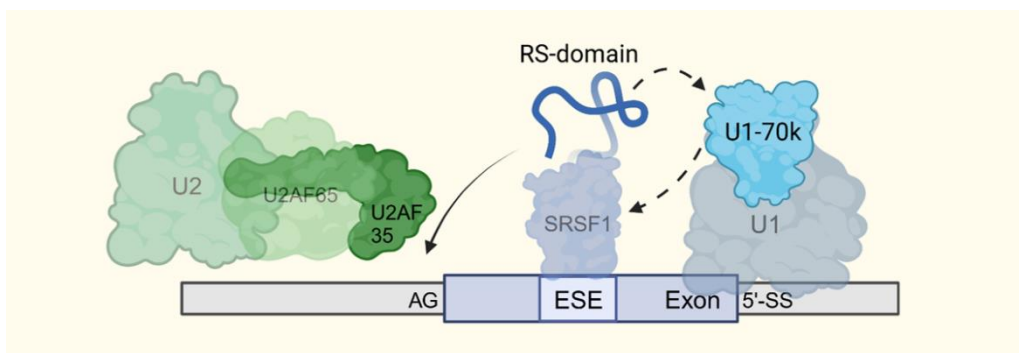


Figure 7. The RS-domain of SRSF1 directs U2AF35 and U1-70k. The RS-domain mediates recruitment of U2AF35 to the AG of poorly conserved 3'-splice sites but is dispensable for RNA binding. Additionally, the RS-domain interacts with the U1-70k protein assisting in complex formation of U1-70k and SRSF1 through the RRMs. The regulated exon is marked in blue with intronic regions represented in gray. 5'SS – 5'-splice site, ESE – exonic splicing enhancer site.

Further investigations on the importance of the RS-domain for SRSF1 mediated splicing also included mutation studies. Reduction of the RS-domain to 10 RS-repeats revealed sufficient functionality for *in vitro* splicing assays and also RD and RE mutants complemented HeLa S100 extracts (Cazalla et al. 2002). In contrast, the replacement of the RS-repeats into KS and RT led to impairment of splicing function (Cáceres and Krainer 1993). However, these analyses included phosphomimics but did not use actual phosphorylated serines in the RS-repeats. Another study investigated the importance of phosphorylation of the RS-domain for association of the U1 snRNP component U1-70k and found phosphorylation to be relevant for interaction of U1-70k and the RRM1 of SRSF1 (Cho et al. 2011). Regardless of the importance for SRSF1 mediated splicing, the RS-domain is crucial for correct nuclear localization of SRSF1 inside

nuclear speckles. It has been shown that SRSF1 Δ RS, KS and RT mutants display aberrant localization in the cytoplasm and nucleoplasm whereas shortened 10 RS-repeat and RD and RE variants locate inside nuclear speckles (Cazalla et al. 2002).

However, the RNA-binding capabilities of SRSF1 are not limited to mRNA transcripts but also involve association with the long non-coding RNA Malat1 that serves as a structural anchor in nuclear speckles and recruits SRSF1 to active transcription sites (see *The long non-coding RNA Malat1*). The interaction between SRSF1 and Malat1, especially in the human cell system represents an important axis of SRSF1 mediated constitutive and alternative splicing. The Malat1 RNA modulates SRSF1 expression and phosphorylation required for proper SRSF1 function (Tripathi et al. 2010).

1.13 RNA splicing mediated by the complex and highly dynamic spliceosome

Many RNP complexes that harbor catalytic activity are featuring a stable composition of proteins and RNAs on pre-assembled active sites. The ribosome as a ribonucleoprotein complex relevant for translation of mRNAs is one example of an RNP machinery that largely remains in the same constitution except for interchanging elongation factors. Additionally, the decoding center and the peptidyl transferase are presented in preformed complexes referring to the small and large ribosomal subunits (Steitz 2008). Further analyses revealed the same functionality for the subunits independently of isolated subcomplexes or the whole ribosome indicating a rather stable composition of the individual RNP complexes throughout the entire process of peptide elongation (Ogle et al. 2001; Selmer et al. 2006; Wohlgemuth et al. 2006).

In contrast to this, the spliceosome represents a highly dynamic RNP with remarkable structural and compositional changes during the splicing process. The pre-mRNA is not only the substrate during the process, but is required for initiating the complex assembly. This is followed by a stepwise sequestration of snRNPs at splice sites into inactive conformations to be released at appropriate time points to continue the splicing process (Wahl et al. 2009). These snRNPs consist of U1, U2, U4/U6 and U5 snRNAs with the respective associated proteins and represent the main components of the major spliceosome responsible for intron removals of precursor RNAs. The splicing reaction of the pre-mRNA requires short consensus motifs including the 3'-splice site (3'-SS), a variable branch point sequence (BPS) encompassing a specific adenosine and a lowly conserved 5'-splice site (5'-SS) (Figure 8).

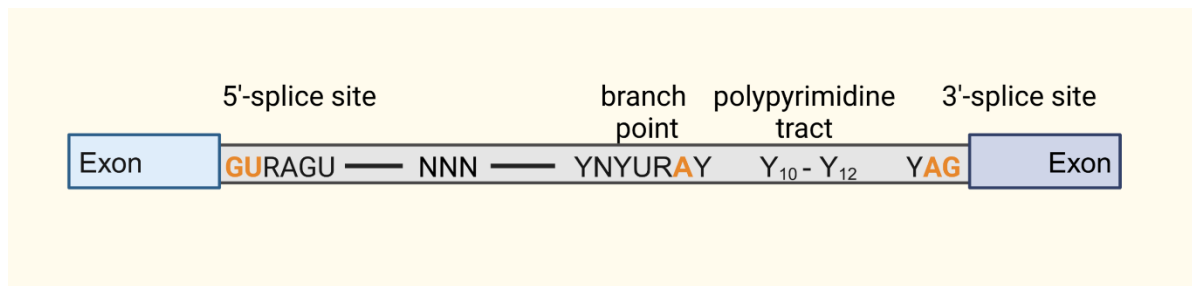


Figure 8. Conserved sequence elements required for splicing recognition in metazoans. The blue boxes represent consecutive exons connected by an intron (gray box) containing conserved sequence elements. The 5'- and 3'- splice site contain a GU and AG dinucleotide, respectively. The branchpoint sequence (BPS) harbors the adenosine required for branching followed by the polypyrimidine tract. N – any nucleotide, Y – pyrimidines, R – purines. Adapted from Wahl et al 2009.

The chemical reaction underlying the intron removal is based on a nucleophilic substitution (S_N2). In short, the 2'-hydroxyl group of the BP adenosine attacks the phosphodiester of the 5'-SS forming a 3'-exon-intron lariat and a free 3'-hydroxyl group on the 5'-exon. Subsequently, the 3'-hydroxyl group of the 5'-exon engages the phosphodiester of the 3'-SS resulting in exon ligation and removal of the intron lariat (Wilkinson et al. 2020). Due to the low conservation of branch point sequences and 5'-splice sites, introns in nascent transcripts rarely adopt a three dimensional fold. This can be explained by missing secondary or tertiary structural information encoded in the sequence and thus requires the assistance of *trans*-acting splicing factors for efficient intron removal (Sheth et al. 2006). The splicing complex formation starts with the U1 snRNP and requires base pairing of the 5'-end of the U1 snRNA with the 5'-SS in the intron, which is considered a weak RNA-RNA interaction (Figure 9). In higher eukaryotes the association of the U1 snRNP with the 5'-SS is enhanced by the interaction of SR-proteins with the U1-70k protein as explained previously (see Figure 6). Subsequently, the U2 snRNP consisting of U2 auxiliary factor 35 and 65 (U2AF35/65) recognizes the 3'-SS and the polypyrimidine tract (PPT) downstream of the BPS, resulting in the formation of the spliceosomal E complex. Following the formation of the E complex U2 snRNA associates with the pre-mRNA at the BPS in an ATP-dependent manner constituting the spliceosomal A complex. Notably, binding of the RS-domain containing protein U2AF65 to the PPT leads to interaction of the RS-domain with the BPS, enhancing U2 snRNA mediated RNA-RNA interaction (Valcárcel et al. 1996). Additionally, U2 snRNA interaction with the pre-RNA is supported by heteromeric splicing factor complexes SF3a and SF3b as components of the mature U2 snRNP (Gozani et al. 1996).

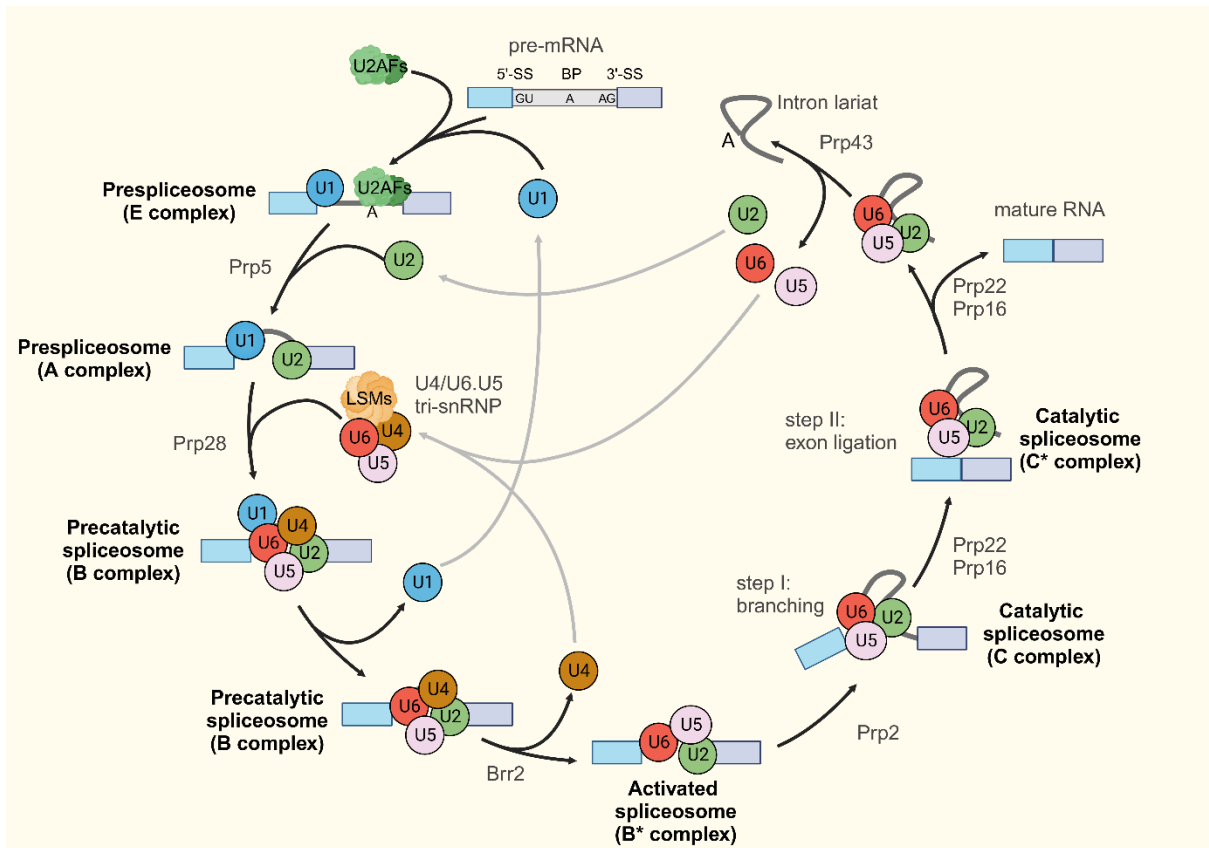


Figure 9. Cross-intron assembly and disassembly of the major spliceosome. Stepwise progression of the interactions with the conserved intronic sequence elements and spliceosomal snRNPs (colored circles) resulting in intron removal and exon ligation (blue shaded boxes). The illustration is simplified and only features resolved spliceosomal complexes in mammals as well as conserved DExD/H-type RNA-dependent ATPases and Helicases (Prp5, Prp28, Brr2, Prp2, Prp16, Prp22, Prp43). The enzymes catalyze specific steps of the rearrangement of the spliceosome and remodeling of the RNPs. U2AFs – U2AF35/65, LSMs – LSM ring consisting of LSM proteins 2-8, 5'-SS – 5'-splice site, 3'-SS – 3'-splice site, BP – branch point, A – adenosine. This model was adapted from Wahl et al. 2009 and modified with published data from Wilkinson et al 2020.

Consecutively to A complex formation, recruitment of the preassembled U4/U6.U5 tri-snRNP is facilitated resulting in the catalytically inactive B complex. The U6 snRNA will eventually form the catalytic center of the spliceosome and is required to be inactive until association with the pre-mRNA occurs in order to prevent premature splicing activity. For this reason, U6 snRNA is extensively paired with the U4 snRNA and the processed oligo-U 3'-end originating from RNA polymerase III synthesis interacts with a preformed heptameric LSm ring (LSm 2-8) (S eraphin 1995; Montemayor et al. 2018). The DEAD-box helicase Prp28 transfers the 5'-SS from U1 snRNA to U6 snRNA followed by separation of U4 snRNA from U6 by RNA helicase Brr2 resulting in U1 and U4 snRNA release. The coordinated reactions of Prp28 and Brr2 are mediated by the C-terminal part of Prp8 (Maeder et al. 2009) leading up to the catalytically active B* complex with the adenosine of the branch point in close proximity to the 5'-SS and subsequent restructuring of the spliceosome to form the C complex. Additional steps following the removal of factors involved in lariat formation and recruitment of exon ligation factors including Prp22 eventually lead to the release of spliced mRNA and lariat intron degradation

(Wilkinson et al. 2020). It has to be noted that, in addition to the aforementioned splicing by the major spliceosome, a low amount of mRNAs are also spliced by the so called minor spliceosome differing in sequence requirements and protein composition (Turunen et al. 2013).

1.14 The relevance of alternative splicing in higher eukaryotes

Recognition of 5'- and 3'-splice sites mediated by U1 and U2 snRNPs is crucial for the splicing process and determines the processed mRNA sequence and ultimately the protein translated by the ribosome. In higher eukaryotes, splice site conservation among the variety of transcripts is degenerate and requires *cis*-acting splicing factors for efficient spliceosome recruitment. The binding sites for these factors are often located in flanking regulatory sequences in the pre-mRNA termed exonic and intronic splicing enhancers (ESEs and ISEs) or silencers (ESSs and ISSs) which impose a positive or negative effect on splice site selection (Singh and Valcárcel 2005). The cumulative binding of enhancers and silencers determines the exon structure of spliced mRNAs giving rise to a variety of alternatively spliced transcripts originating from one gene (Figure 10).

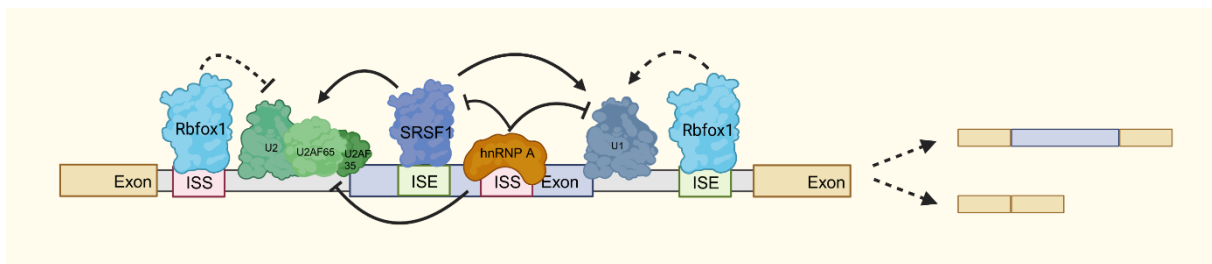


Figure 10. Regulation of alternative splicing by *trans*-acting proteins. Simplified illustration of exon definition in the final transcript mediated by alternative splicing regulators (SRSF1, hnRNP A1 and Rbfox1). Constitutive exons are marked in yellow and alternative exon in blue with the intronic sequence displayed by the gray bar. Positive regulatory sequence elements are depicted in green and inhibitory elements in red. Positive regulation is reflected by arrows and inhibition by bars. Rbfox1 mediated regulation of snRNPs has not been resolved yet.

Alternative splicing prevalent in higher eukaryotes enables an additional layer of splicing regulation by altering expression of specific enhancer and silencer proteins. This results in a different splicing outcome and thereby protein isoforms contributing to cell differentiation, tissue specialization and homeostasis as well as organ development (Wang et al. 2008). The importance of alternative splicing and generation of tissue-specific protein isoforms has been shown extensively in correlation with human diseases e.g. mutation in *cis*-acting elements in the pre-mRNAs (including 5'- and 3'-SS and enhancer/silencer elements) and *trans*-acting splicing factors (Scotti and Swanson 2016). The upregulation of the splicing factor and exonic enhancer SRSF1 has been demonstrated to promote aberrant splicing of proteins involved in apoptosis (BIN1, BCL2L11, MCL1) resulting in anti-apoptotic isoforms (Karni et al. 2007; Anczuków et al. 2012). Additionally, SRSF1 was proposed to induce a short isoform of the RPS6KB1 (S6 kinase 1) involved in PI3K/AKT/mTOR signaling as well as alternatively spliced

MKMK2 (MAPK/ERK pathway), resulting in increased oncogenic potential and SRSF1-mediated transformation of cells (Karni et al. 2007; Ben-Hur et al. 2013).

1.15 The RNA binding protein Rbfox1 mediates alternative splicing

Alternative splicing augments the range of RNA and proteins expressed in certain tissues and extends functionality in a tissue specific context. Especially in neuronal tissue, splicing factors such as PTBP1 and NOVA2 were suggested to be involved in crucial splicing events for neuronal differentiation and neuronal migration (Zhang et al. 2010; Bouts et al. 2007). Another important group of splicing regulators are members of the Rbfox-family (comprised of Rbfox1, Rbfox2 and Rbfox3), which have been associated with critical alternative splicing regulation of neuronal (Rbfox1, Rbfox3), heart and muscle tissue (Rbfox1, Rbfox2) in terms of development and functionality (Conboy 2017). The RNA-binding protein Rbfox1 binds to sequence elements in intronic regions of alternatively spliced transcripts and has been shown to recognize a specific (U)GCAUG motif (Jin et al. 2003; Lambert et al. 2014). Alternative splicing regulation by Rbfox1 has been associated with distinct binding patterns of Rbfox1 in relative to the alternatively spliced exon. Recognition of the motif in a downstream intron enhances exon inclusion while binding to the upstream intronic region results in exon skipping (see Figure 10) (Zhang et al. 2008). Additionally, the Rbfox1 transcript also displays a broad distribution of alternatively spliced exons with tissue-specific variants present in muscle and neuronal protein isoforms (Nakahata and Kawamoto 2005). Altered exon structure also encompasses an auto-regulatory splicing event leading to exclusion of a 93 nucleotide exon in the RRM domain rendering Rbfox1 unable to bind to RNA. Even subcellular localization of Rbfox1 is shifted by inclusion of a short exon (53 nt) resulting in a frameshift and subsequent cytoplasmic distribution of the protein (Kuroyanagi 2009).

Studies investigating the molecular function of Rbfox1 found significant reduction of Rbfox1 expression in mouse models suffering from heart failure and human hearts associated with dilated cardiomyopathy (Gao et al. 2016). It was suggested that Rbfox1 regulates alternative splicing of mutually exclusive exons ($\alpha1/\alpha2$) of the MEF2 transcription factor family (Mef2a, Mef2c, Mef2d), resulting in a shift of transcriptional activity and a 'fetal-like' expression pattern in the heart. Furthermore, knockout of Rbfox1 in mouse hearts displayed similar phenotypes with increased $\alpha1$ exon inclusion in Mef2c/Mef2d and development of pathological cardiac hypertrophy, which could be reverted by re-expression of Rbfox1. Genome-wide analysis of alternative splicing events in zebrafish combined with Rbfox1 knockout also indicated altered splicing events causing progressive cardiac contractile dysfunction and heart failure due to loss of Rbfox1 mediated splicing (Frese et al. 2015). These findings were also supported by RNA annealing, selection and ligation sequencing (RASL-Seq) of alternatively spliced exons in pathological hypertrophy mouse models (generated by transverse aortic

constriction – TAC). This study identified the majority of splicing changes were also present in Rbfox1 knockout mice, indicating a protective role of Rbfox1 mediated alternative splicing towards heart pathologies (Gao et al. 2016).

1.16 Aim of the study

In recent years, numerous lncRNAs emerged as potential regulators of cellular pathways and have been investigated in the context of cancer and other diseases. Furthermore, newly discovered classes of ncRNAs including miRNAs and circRNAs display key regulatory functions in cellular gene expression, contributing to the diverse transcriptome in eukaryotes. The transcripts generated by RNA Polymerase III represent major medium-sized ncRNAs involved in a multitude of processes from transcription (7SK RNA), translation (5S RNA, tRNAs) to spliceosome activity (U6 snRNA). Without any doubt the de-regulation of these ncRNAs may impact biological functions. However, due to modifications of these ncRNAs differential gene expression analysis using RNA-Sequencing methods remain challenging. Furthermore, characterization of unknown ncRNAs first requires identification of these transcripts which proves to be difficult as many ncRNAs are transcribed from multiple loci resulting in multi-mapped reads, commonly discarded in the bioinformatic pipeline. Once identified, the ncRNAs need to be characterized in terms of a potential biological function. The first objective of this thesis was to improve the library generation for medium-sized RNAs by size selection and removal of modifications obstructing adapter ligation. Subsequently, enhancing the pipeline downstream of the RNA-Sequencing to recognize transcripts mapping to several genomic loci was an additional research topic. The dataset obtained for human HEK-293 cells was intended to be subjected for further analyses to evaluate the quality of the method including detection of several medium-sized RNAs and potentially unknown RNAs. The results from the RNA-seq of the mouse B16-F10 cells were designated as a starting point for the characterization of the previously identified 4.5S_H RNA. This RNA is a rodent specific RNA, which has been suggested to be highly expressed and to interact with a variety of RNA-binding proteins, thus implying a physiological function. The second objective of this thesis is the characterization of this RNA in terms of associated biomolecules and to evaluate a potential biological role in the B16-F10 cell model.

In addition to gene regulation by non-coding RNAs, *trans*-acting splicing regulators also contribute to the variety of protein isoforms in eukaryotic cells. The generation of alternatively spliced isoforms enables tissue-specific expression of proteins required for cell differentiation and homeostasis as well as adaptability to environmental changes. The de-regulation of alternative splicing factors has been associated with a variety of diseases including cancer and cardio vascular diseases. The Rbfox1 protein modulates transcript variants in muscle and has been suggested to be important for maintaining proper heart function. Differential isoform

analyses upon Rbfox1 knockdown has revealed a global impact on isoform expression in the heart context and emphasizes the relevance of Rbfox1 as a major splicing regulator. The third objective of this thesis encompasses the identification of novel Rbfox1 targets in the heart-muscle context and to further characterize the impact of Rbfox1 on muscle cell differentiation.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals, enzymes and reagents

The chemicals used in this thesis have been purchased from Sigma Aldrich, Roth, PeqLab and Thermo Fisher Scientific if not stated otherwise. Tissue culture dishes were acquired from TPP and serological pipettes from Corning. Cultivation Medium (DMEM) and supplements including GlutaMAX, Opti-MEM and Trypsin as well as PBS for cell culture were purchased from Thermo Fisher Scientific. The FBS has been bought from Thermo Fisher Scientific and PAN-Biotech. Dynabeads conjugated with streptavidin or protein A for pulldowns and immunoprecipitation, respectively were also purchased from Thermo Fisher Scientific. Enzymes and reaction buffers required for PCR reactions, cloning and *in vitro* transcription were purchased from New England Biolabs and Promega. The Master Mixes for qRT-PCR were manufactured by High Qu GmbH and Promega. The protein and Nucleotide ladder used for agarose- and SDS-polyacrylamid gels were purchased from Thermo Fisher Scientific.

2.1.2 Bacteria strains

The bacterial strain used for transformation of cloned plasmids in this thesis was *Escherichia coli* TOP10 (genotype: F-mcrA $\Delta(mrr-hsdRMS-mcrBC)$ $\Phi80lacZ\Delta M15 \Delta lacX74 recA1 deoR araD139 \Delta(ara-leu)7697 galU galK rpsL$ (StrR) *endA1 nupG*). The bacteria for plasmid preparations were cultivated in LB liquid culture medium (Luria Bertani – 1 % (w/v) Trypton, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl) supplemented with antibiotics (30 μ g/mL Kanamycin or 150 μ g/mL Ampicillin) required for selection of bacteria. Initial transformation of bacteria during the cloning procedure were spread on LB plates containing 1.5% (w/v) agar for colony growth (including 100 mg/mL Ampicillin or 50 mg/mL Kanamycin) and subsequent analysis by PCR for positive transformants.

2.1.3 Cell lines

Cell lines not listed in the following table have been acquired as pellets or RNA samples from the Köhn and the Hüttelmaier Lab and were not cultivated for the duration of this thesis.

Table 1. Cultivated cell lines

cell line	origin	distributor	publication
HEK293T	human embryonic kidney	ATCC	(DuBridg e et al. 1987)
B16-F10	mouse melanoma	ATCC	(Fidler 1975)
NIH: OVCAR-3	human ovary	CLS	(Hamilton et al. 1983)
MC57G	mouse fibrosarcoma	ATCC	(Trinchieri et al. 1976)

CHO-K1	hamster ovary	ATCC	(Hu et al. 1999)
H9C2	rat heart	ATCC	(Kimes and Brandt 1976)
C2C12	mouse skeletal muscle	ATCC	(Yaffe and Saxel 1977)

2.1.4 LNA and siRNA sequences

Locked Nucleic acids (Table 2) used in this thesis were purchased from Integrated DNA Technologies Inc. The siRNAs (Table 3) were ordered from Eurofins Genomics GmbH.

Table 2. Oligonucleotides for siRNA-mediated inhibition of gene expression

siRNA	labeled	sequence (5' - 3')
siRbfox1 pool	siRb1/siRbfox1	CAGUUACGGACGAGUUUUAU GGUCGCACCGUGUACAACA GUUAUGCUGCGUACCGCUA GUGGUCGCACCGUGUACAA CCUACGGCGUUGGUGCCAU GGUUAUGCUGCGUACCGCU GCCGCUGCCUACAGUGACA CCACCACACACUUGCUGCA
Cel-miR-239b	siC	UUUGUACUACACAAAAGUACUG
siRNA La protein #1	siLA#1	GGAAAUCAGUGAAGAUAAA
siRNA La protein #2	siLA#2	GGACAAGUUUCUAAAGGAA
siRNA BDP1#1	siBDP1#1	GGUCAAGCCUGUAGGUAAA
siRNA BDP1#2	siBDP1#2	GGGCCAAUGUAGAGAAGAA
siSRSF1 pool	siS1	CTGGATAACACTAAGTTTA CGTGGAGTTTGTACGGAAA

Table 3. Modified oligonucleotides for LNA mediated inhibition of gene expression

LNA	labeled	sequence (5' - 3')
Y4	Y4	GACACTAATGTTAATA
4.5S _H #2	4.5S _H	CCTCCTTCAGCAAATC
Malat1 #1	Malat1	GAGTTACTTGCCAACT

2.1.5 Plasmids

Commercially available vectors (Table 4) were used for transfection of cells and as cloning vectors to generate eukaryotic expression plasmids (Table 5).

Table 4. Commercial plasmids

plasmid	resistance	company	order numb.
pCR TM -Blunt II - TOPO TM	Kanamycin	Thermofisher Scientific	450245

pcDNA3.1zeo_Flag	Ampicillin	Köhn Lab	Derived from Thermo Fisher V79020
pcDNA3.1zeo_SBP_Flag	Ampicillin	Köhn Lab	Derived from Thermo Fisher V79020
pcDNA3.1zeo_SBP_Flag_RFP	Ampicillin	Köhn Lab	Derived from Thermo Fisher V79020
pEGFP-C1	Kanamycin	Addgene	ClonTech - Discontinued
pEGFP-C2	Kanamycin	Addgene	ClonTech- Discontinued
pGEM-T	Ampicillin	Promega	A3600

Table 5. Plasmids used for eukaryotic expression

plasmid name	empty plasmid	restriction sites used
SRSF1_CDS	pcDNA3.1zeo SBP_Flag	EcoRI – XhoI
SRSF1 Δ RS_CDS	pcDNA3.1zeo SBP_Flag	EcoRI – XhoI
SFR-Rbfox1	pcDNA3.1zeo_SBP_Flag_RFP	EcoRI – XhoI
Flag-Rbfox1	pcDNA3.1zeo_Flag	EcoRI – XhoI
Vcl MG wt	pEGFP-C1	BglII-NotI / PsPOMI-BamHI
Vcl MG BS1	pEGFP-C1	BspEI – XhoI
Vcl MG BS2	pEGFP-C1	PsPOMI – SacII
Vcl MG BS3	pEGFP-C1	SacII – PFIMI
Vcl MG BS4	pEGFP-C1	PFLMI – BamHI
Vcl MG BS5	pEGFP-C1	PFLMI – BamHI
Pxn MG wt	pEGFP-C2	EcoRI – HindIII / HindIII – XhoI
Pxn MG BS1	pEGFP-C2	BspEI – EcoRV
Pxn MG BS2	pEGFP-C2	BbsI – BstZ17
Pxn MG BS3	pEGFP-C2	BbsI – BstZ17
Pxn MG BS4	pEGFP-C2	BbsI – BstZ17
4.5S _H (gene)	pGEM-T	TA - cloning
4.5S _H mut (gene)	pGEM-T	TA - cloning
4.5S _I (gene)	pGEM-T	TA - cloning
hY4 (gene)	pGEM-T	TA - cloning

2.1.6 Oligonucleotides and probes

The oligonucleotides for cloning (Table 6) and amplification in qRT-PCR analysis (Table 7) were purchased from Eurofins Genomics GmbH and Microsynth AG. Oligos for the *in vitro* transcription and fluorescent Northern blot probes (Table 8 and 9) were acquired from Eurofins Genomics GmbH.

Table 6. Oligonucleotides used for insert generation and cloning

insert	sense-primer (5' – 3')	antisense-primer (5' – 3')
SRSF1	GGGAATTCATGTCGGGAGGTGGT GTGATTCGTGGCCCC	GGCTCGAGTTATGTACGAGAGCG AGATCTGCTATGACG
SRSF Δ RS	GGGAATTCATGTCGGGAGGTGGT GTGATTCGTGGCCCC	GGCTCGAGTTATTTAACCCGGAT GTAGGCAGTTTCT
Rbfox1	GGGAATTCATGTTGGCGTCGCAA GGAGTCCTTCTGCAT	GGCTCGAGTTAATATGGAGCAA ACGGTTGTATCCCC
Vcl MG p1	GGGGATCCCTAACTGATGAGCTG GCTCCTCCTAAGCCA	GGGAATTCTTTGGAAAGTGACCC CCATATCCTTGTATT
Vcl MG p2	GGGAATTCATCCCCCCCCCTTTT GTGGCCAATACTGGA	GGGCGGCCGCTTACTGTAAGAGA TTGGTTCTAATCCGCTTATC
Pxn MG p1	GGCTCGAGAATCTGCCCGGCGAT GGGCTTCTTGCAGGCTCC	TACGCATTAAAGCTTACTCATTG TACTGT
Pxn MG p2	ACAGTACAAATGAGTAAGCTTTAA TGCCTA	GGCTCGAGAATCTGCCCGGCGAT GGGCTTCTTGCAGGCTCC
4.5S _H	GGAGATCTATCGATAGGAGCCGG TGCTCGTCGCTCGTCGCTCGT	GGGGATCCAAAATGTGTAGCCCA GGCTGGCCTCGAACT
4.5S _H mut	GGAGATCTATCGATAGGAGCCGG TGCTCGTCGCTCGTCGCTCGT	GGGGATCCAAAATGTGTAGCCCA GGCTGGCCTCGAACTCGTGATTC AGCAAAT
4.5S _I	GGAGATCTATCGATTAAGTTCAAC ACGAAGACAAATTCACGGCC	GGGGATCCAAAAGGTGGCTGGA GAGACAGCCGTGGGTG
hY4	GGGGATCCGGCTGGTCCGATGG TAGTGG	GGGAATTCAAAAAGCCAGTCAA TTTAGCAGT

Table 7. Oligonucleotides for amplification in RT-qPCR analyses

gene	sense-primer (5' – 3')	antisense-primer (5' – 3')
<i>Homo sapiens</i>		
PPIA	GTCAACCCCACCGTGTCTT	CTGCTGTCTTTGGGACCTTGT
EEF2	GGAGTCGGGAGAGCATATCA	GGGTCAGATTTCTTGATGGG
BDP1	CTGATCTGCTTCCATCTCCAAGT	GTGGTTGCCTCCCCTTGAG
POLAR	ACGGCAGTATGGTTAAGTGGG	CCAAACCCAGACAACCTCAGGT
PAMIR	TAGCCTGTGGGCCTAAGCTG	ATCCAAGGCCACACAGTTAGTGA
COIR	AACTGTCTAGTGCCGTGGCT	GGAATCATTCCCTAAGGTGGGA
U1	ATACTTACCTGGCAGGGGAGATA CC	CAGGGGAAAGCGCGAACGCAGT CC
MALAT1 5'	GCGCTAACGATTTGGTGGTG	GCGCTAACGATTTGGTGGTG

<i>Mus musculus</i>		
Malat1	GGGGGAATGGGGGCAAATA	TCCCTGTTAAGACCACTCCCA
Malat1 5'	CCATACGGATGTGGTGGTGA	AGGAATGTTACCGCACCGCA
Eef2	TCATCGACTCTCCAGGCCAT	TGTTTCATCATCAGGGCGGGC
Actb	GCCCTAGGCACCAGGGTGTG	ACAGGGTGCTCCTCAGGGGC
Ppia	AGGATTCATGTGCCAGGGTG	CCAAAGACCACATGCTTGCC
Cfl1	GTGAAGAAACGCAAGAAGGCG	GTAGGGGTCGTCCACAGTCT
Actg1	TGGAAGAAGAAATCGCCGCA	TGAGTCTTTCTGGCCCATGC
Eif5a	AATGGTTTTGTGGTGTCAAAGG	ATCAGCTGGAAGTCATTCCGTT
preEef2	GAGGGGTTGTGCTTGGTGAT	GGATCAAAGGCCACAATGCG
prePpia	CGTGTTCTTCGACATCACGGC	AAATGGCCCCTCGTCAGGAAAT
preCfl1	CGTGTTCTTCGACATCACGGC	AAATGGCCCCTCGTCAGGAAAT
preActg1	AAATCGCCGCACTCGTCATT	GGAGGCCACTCCTCTCAACT
preRplp2	TACTAGACAGCGTGGGCATCG	GGCCTCCTCCATCCATCCTTTT
Rbfox1	GTGGTTATGCTGCGTACCG	GGAGCAAGTGTGTGGTGGTA
meta Vcl Ex19-20	GAACCTGAGCTGCTGTTAATGC	TGGCGATATCCTTGGCACAC
Vcl Ex18-20	TGGTCTAGCAAGGGCAATGAC	GGCGATATCCTTGGCACACT
ePxn Ex6-7	AACCAGGGAGAGATGAGCAGT	AGCCACAGAAGGTAAAAGGACC
Pxn Ex6-11	AACCAGGGAGAGATGAGCAGT	TGCTCCCAGTTTTCCCCTGG
<i>Rattus norvegicus</i>		
Rbfox1	GTGGTTATGCTGCGTACCG	GGAGCAAGTGTGTGGTGGTA
meta Vcl Ex19-20	GAACCTGAGCTGCTGTTAGTGC	TGGCGATATCCTTGGCACAC
Vcl Ex18-20	TGGTCTAGCAAGGGCAATGAC	TGGCGATATCCTTGGCACAC
ePxn Ex6-7	AACCAGGGAGAGATGAGCAGT	CAATGGGCTTCTTGCAGGCTC
Pxn Ex6-11	AACCAGGGAGAGATGAGCAGT	TGCTCCCGTTTTCCCCTGG
Actb	GCCCTAGGCACCAGGGTGTG	ACAGGGTGCTCCTCAGGGGC
Eef2	GCACTGTAGGCTTTGGCTCT	GCTGACTTGCTGAACTTGCC
Myod1	AACTGCTCTGATGGCATGATGG	CCCTGTTCTGCATCGCTTGA
Myog	TGGGCGTGTAAGGTGTGTAAG	AGTTGCATTCACTGGGTACCACC
Tnnt2	GAGGACACCAAACCCAAGCC	CCCGACGCTTTTTCGATCCT
GfP_s	GGGATCACTCTCGGCATGGA	

Table 8. Oligonucleotides for *in vitro* transcription

insert	sense-primer (5' – 3')	antisense-primer (5' – 3')
T7-4.5S _H	GTAATACGACTCACTATAGCCGG TAGTGGTGGCGCACGCCGG	AAAATGTGTAGCCCAGGCTGGC CTC
T7-4.5S _H mut	GTAATACGACTCACTATAGCCGG TAGTGGTGGCGCACGCCGG	AAAATGTGTAGCCCAGGCTGGC CTC
T7-4.5S _I	GTAATACGACTCACTATAGGCTG GAGAGATGGCTCAGCCGTTAAA	AAAAGGTGGCTGGAGAGACAGC CGTG
hY4	GTAATACGACTCACTATAGGCTG GTCCGA	AAAAAGCCAGTCAAATTTAGCA GT

Table 9. Fluorescently-labeled anti-sense probes for Northern blot analysis

probe	labeled	sequence (5' - 3')
POLAR-FISH	Cy3	CTTACCCACTTAACCATACTGCCGTTTATC
4.5S _H -FISH	Cy3	TTCAGCAAATCCTACCGGCGT
4.5S _H	Atto680	TTCAGCAAATCCTACCGGCGT
U6	DY782	CAGGGGCCATGCTAATCTTCTCTGT
5S	DY782	AAGTACTAACCAGGCCCGAC
U2	Atto680	GGTACTGCAATACCAGGTCGATGCGTGGAG

2.1.7 Primary and secondary antibodies

Table 10. primary antibodies for Western blot detection and immunofluorescence

antibody	species	company	order numb.
La	rabbit	Bethyl	A303-902A
VCL	mouse	Sigma-Aldrich	V9131
PSF	rabbit	Bethyl	A301-322A
NONO	mouse	Upstate	05-950
RBM14	rabbit	Bethyl	A300-845A
EIF3A	rabbit	Bethyl	A302-002A
PSP1	rabbit	Bethyl	A303-205A
SC35	mouse	Sigma-Aldrich	S4045
FBL	rabbit	Abcam	ab5821
PRPF8	rabbit	Bethyl	A303-922A
SmB/B'/N	mouse	Santa Cruz	sc-130670
GAPDH	mouse	Proteintech	60004-1-Ig
Phospho-SR (1H4)	mouse	Sigma-Aldrich	MABE50
DHX15	mouse	Santa Cruz	sc-271686

SF3B1	rabbit	Santa Cruz	sc-102102
RBFOX1	mouse	Merck	MABE985
PXN	mouse	Uptstate	5H11
RPL7	rabbit	Bethyl	A300-741A
RFP	mouse	Thermo	MA5-15257

Table 11. Fluorescently-conjugated secondary antibodies

Antibody	comany	order numb.
IRDye680RD donkey anti-mouse	LI-COR	926-68072
IRDye680RD donkey anti-rabbit	LI-COR	926-68073
IRDye800CW donkey anti-mouse	LI-COR	926-32212
IRDye800CW donkey anti-rabbit	LI-COR	926-32213
Antibody	comany	order numb.
Alexa Fluor® 488 AffiniPure F(ab') ₂ Fragment Donkey Anti-Rabbit IgG (H+L)	DiaNova	711-546-152
Alexa Fluor® 488 AffiniPure F(ab') ₂ Fragment Donkey Anti-Mouse IgG (H+L)	DiaNova	711-546-150
Cy™3 AffiniPure F(ab') ₂ Fragment Donkey Anti-Mouse IgG (H+L)	DiaNova	715-166-151
Cy™3 AffiniPure F(ab') ₂ Fragment Donkey Anti-Mouse IgG (H+L)	DiaNova	711-166-152
Cy™5 AffiniPure Donkey Anti-Mouse IgG (H+L)	DiaNova	715-175-150

2.1.8 Standard Kits and Assays

All kits and assays mentioned below have been used according to manufacturer's protocols and instructions. If methods were changed it is specifically indicated in the method section.

Table 12. Commercial Kits and Assays

Kit	comany	order numb.
DC Protein Assay	Bio-Rad	5000111
Phusion® High-Fidelity PCR Kit	New England Biolabs	E0553
OneTaq® 2X Master Mix with Standard Buffer	New England Biolabs	M0482
Q5® High-Fidelity 2X Master Mix	New England Biolabs	M0492
Zero Blunt™ TOPO™ PCR Cloning Kit	Thermo Fisher Scientific	451245
Agarose Gel Extraction Kit - Column Kit	Jena Bioscience	PP-202L
Plasmid Mini-Prep Kit - Column Kit	Jena Bioscience	PP-204L

PureLink™ HiPure Plasmid-Midiprep-Kit	Thermo Fisher Scientific	K210004
PCR Cloning Kit	New England Biolabs	E1202
PureLink™ Genomic DNA-Minikit	Thermo Fisher Scientific	K182002

2.1.9 Standard Buffers

Table 13. Chemical composition of buffers

buffer solution	components
PBS (phosphate buffered saline)	137 mM NaCl 2.7 mM KCl 10 mM Na ₂ HPO ₄ 2 mM KH ₂ PO ₄
PBS-T (PBS – Tween)	1 x PBS 1% Tween-20
2x SSC (Saline sodium citrate)	300 mM NaCl 30 mM Na ₃ C ₆ H ₅ O ₇
Low stringency wash buffer	2x SSC 0.1% SDS
TAE (Tris/Acetate/EDTA)	40 mM Tris 20 mM Glacial acetic acid 1 mM EDTA
NuPAGE (transfer buffer)	50 mM Tris 40 mM Glycerin 0.04 % SDS 10 % Methanol
5x TBE (Tris/Borate/EDTA)	450 mM Tris 450 mM Boric acid 10 mM EDTA
Ponceau-S	0.1 % Ponceau-S 5 % acetic acid
TRIZOL	0.8 mM Guanidine thiocyanate 0.4 mM Ammonium thiocyanate 0.1 mM Sodium acetate (pH 4.5) 5 % Glycerin 48 % Roti®Aqua Phenol
Total lysis buffer	50 mM Tris (pH 7.4) 50 mM NaCl 1 % SDS 2 mM MgCl ₂
Binding buffer	10 mM HEPES-KOH pH 7.2 5 mM MgCl ₂ 150 mM KCl 0.5% NP-40
BW buffer	5 mm Tris-HCl pH 7.5 0.5 mm EDTA 1 m NaCl
Fractionation buffer	10 mm HEPES-KOH pH 7.2 5 mm MgCl ₂ 150 mm KCl 2 µg/µL Digitonin

2x FASB	20 mM EDTA 0.025 % SDS 0.025 % Bromophenol blue 25 mM Formamid
---------	---

2.1.10 Devices

Table 14. Manufacturer's list of used devices

type	devices
Imaging	Nikon TE-2000E Leica SP5X Leica SP8X Sartorius Incucyte S3® Nikon TS-100
Infrared scanner	LiCOR Odyssey Infrared Scanner
SDS-PAGE	Life Technologies NuPAGE® Bis-Tris Electrophoresis System
Western/Northern Blot	XCell II™ Mini-Cell Blot Module (Thermo Fisher)
Magnetic rack	Invitrogen™ DynaMag™-2 Magnet
qRT-PCR	Roche Light Cycler 480 II / Bio-rad iCycler iQ
Spectroscopy	Tecan Spark® multimode microplate reader GloMax® 96 well Microplate Luminometer
UV Crosslinker	Biostep® Crosslinker 254nm
Thermocycler	Eppendorf Mastercycler Nexus II (96well)
Centrifuges	Heraeus Biofuge Stratos Eppendorf Centrifuge 5424 R Eppendorf MiniSpin
Agarose gel electrophoresis	Peqlab PerfectBlue Mini L
Cell counting chamber	Neubauer (0.1 mm / 0.0025 mm ²)

2.2 Methods

2.2.1 Cell biological methods

2.2.1.1 Cell cultivation

The cells used in this thesis have been cultivated in standard conditions (37 °C and 5% CO₂) with DMEM supplemented with 5 mM GlutaMAX™ and 10% FBS for all cell lines (HEK293, OVCAR-3, C2C12, B16-F10, CHO-K1, MC57G and U2-OS) except H9C2 cells which have been supplemented with 20% FBS. For passaging and seeding during transfection, cells were washed with PBS and subsequently detached from the plastic dish using 0.05% Trypsin and 0.4 mM EDTA. Cells were resuspended in fresh DMEM with FBS and plated on a

new cell culture dish. The H9C2 cells were cultivated at 60-70% confluency to maintain differentiation potential and prevent differentiation during the cultivation process.

2.2.1.2 Cell differentiation

The differentiation protocol for H9C2 cells has been adapted from Ménard et al. Differentiation of the H9C2 cells was achieved by seeding $1.0 - 1.5 \times 10^5$ H9C2 cells per well in a 6-well-plate at 60-70% confluency in DMEM containing 20% FBS and 5 mM GlutaMAX prior to differentiation induction to allow attachment of the cells. The next day, cultivation medium was removed and cells were washed with PBS followed by replacement with low-serum DMEM (1% FBS) for muscle-like differentiation and low-serum DMEM supplemented with 2 μ M retinoic acid (Sigma Aldrich) for cardiac-like differentiation. The differentiation medium was replaced every other day and H9C2 cells were treated with 2 μ M retinoic acid on a daily basis over the course of 7 days.

2.2.1.3 LNA and siRNA transfection

For knockdown experiments using LNAs or siRNAs cells were seeded into appropriate cell culture dishes and consequently transfected. The LNA mediated knockdowns in B16-F10 ($3.0 - 4.0 \times 10^5$ cells per 6-well) and U2-OS cells ($2.0 - 3.5 \times 10^5$ cells) were performed with 5 μ L DharmaFECT 1 transfection reagent (provided by Horizon Discovery Ltd.) and 100 pmol of the respective LNA. Transfection solutions were prepared and applied according to manufacturer's protocol. The transfected cells were monitored and harvested 24 h post transfection for Northern blot and qPCR analysis or subjected to immunofluorescence staining. The knockdown of SRSF1 in B16-F10 cells was performed with a two siRNA pool and followed the procedure of LNA transfection with DharmaFECT1 and 100 pmol total siRNA. These cells were harvested 48h post transfection followed by analysis on Western blot. Transfection of siRNA in HEK293 (5.0×10^5 cells) conducted with 7 μ L Lipofectamine RNAiMAX (Thermo Fisher Scientific) and 100 pmol of a two siRNA pool against BDP1 and two single siRNAs targeting the La protein. Transfection solutions were prepared in OptiMEM and applied to the cells according to manufacturer's protocol. Experiments with siRNAs against the La protein were harvested 72 h post transfection and BDP1 transfected cells were harvested after 36 h due to harsh proliferative effects of the cells. The H9C2 cells were originally transfected with Viromer Blue (formerly from Lipocalyx) and later using DharmaFECT1 transfection reagent as described previously. For the course of the knockdown and differentiation experiments, H9C2 cells were transfected with a 100 pmol of a pool of siRNAs targeting Rbfox1 prior to differentiation induction and harvested at the final day of the differentiation.

2.2.1.4 Plasmid transfection

Plasmid transfections in this thesis have been performed with the use of DharmaFECT kb (Horizon Discoveries Ltd), Viromer Red and Yellow (formerly from Lipocalyx) according to manufacturer's protocol. Production of Viromer was discontinued and DharmaFECT was used as a replacement. For DharmaFECT kb transfections OptiMEM was used as serum-free medium for preparation of the transfection solution. B16-F10 cells were seeded at a density of 2.5×10^5 cell per 6-well for SBP-tagged SRSF1 and SRSF1 Δ RS overexpression using 7 μ L DharmaFECT kb and 4 μ g of plasmid in 200 μ L OptiMEM per transfection. The cells were lysed 48 h post transfection and the cell lysate subjected to pulldown assays and subsequent analyses on Western and Northern blot. For the overexpression of Rbfox1 containing plasmids 1.5×10^5 H9C2 cells per 6-well were seeded and transfected with Viromer Yellow or DharmaFECT kb according to protocol. Transfection medium was replaced with fresh DMEM and 20% FBS the following day and cells were harvested for analyses 24 h later. The immunostainings required re-plating of 20 - 30.000 previously transfected H9C2 cells onto coverslips in 24-well plates 24 h post transfection which were subjected to the procedure of immunostaining on the next day. The minigene analyses were performed in C2C12 cells as a representative rodent cell line with high transfection efficiency and cell number. C2C12 were seeded at a density of 3.0×10^5 cell per 6-well and transfected with Viromer Red or DharmaFECT kb with co-transfection of 2 μ g plasmid each encoding Flag-Rbfox1 and the respective minigene constructs of Pxn and Vcl.

2.2.2 Molecular biology methods

2.2.2.1 Cloning

The cloning procedure for the overexpression plasmids containing Rbfox1, SRSF1 full length and SRSF1 Δ RS followed the same steps starting with amplification of the target transcript from cDNA with Q5® High-Fidelity 2X Master Mix and the respective oligonucleotides with flanking restriction sites (Table 6) in a PCR reaction. The amplified DNAs were loaded onto a 2% agarose gel containing ethidium bromide and separation was performed by gel electrophoresis in 1 x TAE buffer with subsequent detection of double stranded DNA by UV light. Amplicons matching the expected nucleotide length were excised as a gel piece. The DNA was extracted and purified with the Agarose Gel Extraction Kit followed by elution with ddH₂O. Eluted PCR products were subjected to ligation into linearized blunt end vector pCR™-Blunt II - TOPO™ according to manufacturer's protocol. Next, the solution containing ligated plasmid was incubated with chemically competent *E. coli* (TOP10) for 20 minutes at 4 °C followed by transformation of the bacteria through heat shock at 42 °C for 50 seconds and a 4-minute resting phase on ice. Subsequently, the transformed bacteria

were cultivated in 800 μ L LB-medium for 1 h at 37 °C and spread on LB-agar plates containing 50 mg/mL Kanamycin for overnight growth. The clones on the plate were tested for successful insertion in a PCR reaction and positive clones were cultivated in liquid LB-medium with 30 mg/mL Kanamycin at 37 °C overnight. The plasmid DNA was isolated with the Plasmid Mini-Prep Kit from Jena Bioscience according to the manufacturer's protocol and correct sequence insertions were analyzed by Sanger-Sequencing at Eurofins Genomics GmbH. The final cloning step required enzymatic digestion of 2 μ g of the empty expression plasmids (Table 4) and the pCR™-Blunt II - TOPO™ containing the valid gene sequence with 0.5 μ L per restriction enzyme (New England Biolabs) in the corresponding buffer and separation on agarose gel electrophoresis. The gel pieces containing linearized expression plasmid and gene sequence were excised and DNA was isolated and eluted in 30 μ L ddH₂O followed by a ligation reaction containing 8 μ L insert, 2 μ L linearized vector, 1 μ L T4 DNA ligase (New England Biolabs), 3 μ L corresponding 10 x ligation reaction buffer in a total volume of 30 μ L. The ligation reactions were transformed into bacteria as described above after 2 h at room temperature. Plating, PCR testing, liquid cultivation was carried out as described previously with liquid LB and agar plates containing Kanamycin or Ampicillin for positive clone selection (see 2.1.2).

The wild type minigene constructs for the Metavinculin exon and the ePxn cassette were generated by sequential insertion of two independently produced amplicons from mouse heart genomic DNA isolated with the PureLink™ Genomic DNA-Minikit according to manufacturer's protocol (Thermo Fisher Scientific). The cloning process was executed as described above with the oligonucleotides and plasmids from Table 4 and Table 6. Notably, amplification of the gene sequences was achieved using the Phusion® High-Fidelity PCR Kit with 0.5 μ g genomic DNA, 2% DMSO, GC-Buffer and 0.5 μ L Phusion DNA Polymerase (2U/ μ L) in 20 μ L reaction volume. The Vcl minigene part 1 contained BamHI-EcoRI and part 2 EcoRI-NotI restriction sites subcloned sequentially into the pcDNA3.1 SFB_Flag_RFP expression vector. The complete minigene sequence of the Vcl minigene was excised using BamHI-NotI restriction enzymes and ligated into the pEGFP-C1 cut with BglII-PspOMI. For the Pxn minigene the individual parts were inversely subcloned into the pEGFP-C1 using EcoRI-HindIII and HindIII-XhoI restriction sites and the final minigene sequence was inserted into the pEGFP-C2 utilizing EcoRI-XhoI enzymes. For the mutated minigenes, sequences were ordered from GenScript Biotech resembling regions in the wild type with point mutations in putative Rbfox1 binding sites. The mutated sequences were excised with the indicated restriction enzymes (Table 5) and subcloned into the pEGFP-C1 Vcl wild type minigene and pEGFP-C2 Pxn wild type minigene, respectively.

2.2.2.2 RNA Isolation

Total RNA was isolated by TRIZOL (Table 13) extraction. In general, 1 mL of TRIZOL was added to a single confluent 6-well and the solution was transferred to a reaction tube. The phase separation was accomplished by addition of 200 μ L chloroform and centrifugation at 13.000 rpm for 30 minutes at 4 °C. The aqueous phase was transferred to a new reaction tube and RNA was precipitated with an equal volume of isopropanol and centrifugation at 13.000 rpm for 1 hour at 4 °C. The RNA pellet was washed in 1 mL 80% ethanol, centrifuged and dried at 65 °C, followed by careful resuspension of the RNA in 20 μ L ddH₂O. Determination of RNA quality and concentration were realized by measuring absorbance at wavelengths of 260 nm and 280 nm with the Tecan Spark® multimode microplate reader.

2.2.2.3 Reverse transcription and quantitative real-time PCR (qRT-PCR)

Reverse transcription for cDNA generation was conducted by random hexamer priming on 2 μ g of isolated RNA. The master mix for the reaction contained 4 μ L 5x M-MLV reverse transcriptase buffer (Promega), 1 μ L dNTPs (10mM) and 0.5 μ L M-MLV reverse transcriptase (Promega, Cat# M1701, 200U/ μ L) that was added to the RNA in ddH₂O with a total reaction volume of 20 μ L. The reaction was incubated in a thermocycler with reverse transcription at 42 °C for 2 h and subsequent inactivation of the transcriptase at 75 °C for 15 min. The RNA abundance and expression changes due to transfection of the various cell lines has been determined by qRT-PCR with SYBRGreen fluorescent dye on a LightCycler 480 II (Roche) on a 384-well plate or the Biorad iCycler iQ on a 96-well plate. Each well contained equal volumes of ORA qPCR Green ROX L Mix (HighQu) with 0.2 μ M of each oligonucleotide and 1:10 diluted cDNA resulting in a total reaction volume of 5 μ L. The oligonucleotides used for the qPCR listed in Table 7 were designed with NCBI (National Center for Biotechnology Information) Primer-BLAST and set for optimal annealing temperature at 60 °C. The reaction parameters for the PCR involved initial activation of the polymerase at 95 °C for 5 minutes followed by cDNA denaturation at 95 °C for 10 secs, annealing of the oligos at 60 °C for 10 secs and elongation at 72 °C for 15 secs in 50 cycles. The purity of amplified PCR products was determined by a melting curve gradient from 55 °C – 95 °C at the end of the PCR reaction. Newly synthesized double-stranded DNA contained intercalated fluorescent SYBRgreen dye which was quantified. The second derivative maximum was used to determine the cycle of quantification (Ct). Calculations, based on Δ Ct- or the $\Delta\Delta$ Ct- method revealed RNA expression changes relative to one or two housekeeping genes (EEF2, ACTB, PPIA) and control conditions (Livak and Schmittgen 2001).

$\Delta\Delta\text{Ct}$ - analysis:

$$\text{relative mRNA level} = 2^{-(\text{CtX}-\text{CtN})_{\text{population}} - (\text{CtX}-\text{CtN})_{\text{control population}})}$$

X – target transcript *N* – housekeeping transcript

2.2.2.4 Library preparation and msRNAseq

Library preparation and subsequent msRNAseq has been published in (Zorn et al. 2021). The RNA was isolated from HEK293 and B16-F10 cells using TRIZOL extraction (2.2.2.2). Adapter ligation to the RNA species required the removal of 5'-pyrophosphates which was facilitated by processing of 15 µg of RNA with RNA 5' Pyrophosphohydrolase (RppH, New England Biolabs) according to manufacturer's protocol for 1 h at 37 °C. The processed RNA was loaded on a TBE-urea gel (22% bottom / 7% top acryl amide) in addition to an RNA marker for size identification purposed and separated using gel electrophoresis. Recognition of the RNA species was realized by ethidium bromide staining and RNAs with sizes between 50 to 300 nucleotides were excised as gel pieces and centrifuged through perforated tubes resulting in gel fragmentation. The gel pieces were mixed with extraction buffer (20 mM Tris 7.4, 250 mM Na-acetate, 1 mM EDTA, 0,25% SDS) and incubated at 65 °C for 1 h, followed by spin filtered centrifugation to separate the small gel pieces and extracted RNA solution. Isolation of the RNA was achieved by isopropanol precipitation, washing with ethanol and dissolving in ddH₂O as described for RNA isolation above. The quality and integrity of the RNA was determined by loading on a TBE-urea gel and electrophoresis separation followed by ethidium bromide staining to identify RNA species as individual bands. The library preparation was facilitated according to the Illumina TruSeq protocol and performed by Vertis biotechnology AG (Freising, Germany) in addition to the RNA-Sequencing on a NextSeq 500 system producing single reads from the 5'-end and 150 bp read length.

2.2.2.5 Northern blot analysis

For direct analysis of RNA abundances and analysis of altered expression of medium-sized non-coding RNAs a Northern blot was performed. The RNA was isolated from the cells as described above with concentration determination and RNA quality control based on absorbance at 260 and 280 nm. Subsequently, 4 µg of total RNA were diluted with ddH₂O to 10 µL and mixed with equal amount of 2x FASB (Table 13) for loading and separation on a 15% denaturing TBE-Urea-gel followed by blotting onto nylon membranes (Roche). These membranes were crosslinked with 260mJ of UV light and pre-hybridized using PerfectHyb Plus buffer (Sigma-Aldrich) for 10 minutes at room temperature. The probes (Table 9) were diluted to 100 ng/µL in PerfectHyb Plus buffer and hybridized at 30 °C for 2 hours. Detection of the fluorescent signal was conducted with the Odyssey Scanner (LI-COR).

2.2.2.6 *In vitro* transcription

The DNA templates for 4.5S_H, 4.5S_H mut, 4.5S_I and hY4 were amplified from the respective pGEM-T vector constructs with the specific amplification oligos containing the T7 RNA polymerase sequence (Table 8) and the 2x One-Taq Mix (New England Biolabs). The PCR product was separated on a 2% agarose-gel in TAE-buffer and DNA was extracted from the gel with Agarose Gel Extraction Kit (Jena Bioscience). Transcription was carried out with the total volume of eluted DNA plus 20 µL transcription optimized 5x buffer, 10 µL DTT (100mM), 1 mM rNTPs (ATP, GTP, CTP), 0.3 mM unlabeled UTP, 1µL RNasin® ribonuclease inhibitor and 4 µL T7 RNA polymerase (all products from Promega). Biotinylated UTP was supplemented to the reaction with a final concentration of 100 µM (Jena Bioscience) in 100 µL reaction volume and transcription was realized at 37 °C for 4 h. The DNA template was digested after the reaction by addition of 5 µL RQ1-DNAse (Promega) and RNA was purified by TRIZOL precipitation (2.2.2.2).

2.2.2.7 Biotinylated RNA pulldown

For RNA pulldowns 15 µL of Dynabeads™ MyOne-Streptavidin-T1 (Thermo Fisher Scientific) per pulldown were transferred to a reaction tube, washed with equal volume of BW buffer (Table 13) three times utilizing a magnetic rack and incubated with 1 µg of *in vitro* transcribed RNA or no RNA for the bead control for 20 min at RT. In the meantime, cells were lysed with binding buffer (Table 13), incubated for 10 min at RT followed by centrifugation at 13.000 rpm for 5 min. The supernatant was transferred to a new reaction tube and 20 µL lysate extracted for the input sample, while 200 µL of the cell lysate were resuspended with the RNA-coupled beads and incubated for 30 min at RT. Subsequently, pulldown samples were washed three times with 200 µL binding buffer and separated on a magnetic rack to remove unspecific proteins and RNA. For analysis of associated proteins, the beads were resuspended in 40 µL 1 x NuPage™ LDS sample buffer (Thermo Fisher Scientific) and boiled at 95 °C for 5 minutes to elute proteins, which were separated from the beads on a magnetic rack and subjected to SDS-PAGE and Western blot analysis.

2.2.2.8 Cell fractionation

For cell fractionation 5×10^6 cells were harvested and lysed in 200 µL fractionation buffer (Table 13) for 7 min at RT followed by centrifugation for 3 min at 4000 rpm. The cytoplasmic fraction in the supernatant was carefully transferred to a separate tube and TRIZOL was added up to a total volume of 1 mL for RNA isolation. The remaining pellet containing the nuclear fraction was thoroughly resuspended in fractionation buffer to extract nuclear RNAs, transferred to a new tube and subjected to RNA isolation using TRIZOL. The isolated RNAs in each fraction were subjected to either on Northern blot or qRT-PCR analyses.

2.2.3 Protein-biochemical methods

2.2.3.1 Protein extraction, SDS-PAGE and Western blot

For the analysis of protein abundance in transfected and control cells, proteins were extracted using the total lysis buffer (Table 13) supplemented with 1 μL of Benzonase for genomic DNA digestion (250 U/ μL , Jena Bioscience). The lysate was resuspended after 10 minutes of incubation and centrifuged at 13.000 rpm for 5 minutes to pellet cell debris followed by transfer of the protein-rich supernatant to a new tube. Protein concentration was assessed with the DC Protein Assay measuring absorbance at 700 nm with the GloMax[®] 96 well Microplate Luminometer. Protein samples were set to equal amounts and mixed with 4 x NuPage[™] LDS sample buffer containing 50 mM DTT and boiled at 95 °C for 5 min resulting in complete denaturation. The samples were loaded into a NuPAGE[®] Novex[®] 4-12% Bis-Tris protein gel and separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE) with the corresponding NuPAGE[®] MOPS SDS running buffer. Immobilized proteins in the gel were transferred to a nitrocellulose membrane (Amersham Bioscience) with a wet blot system using NuPAGE transfer buffer with 10 % methanol (Table 13). If necessary, efficiency of the transfer was evaluated by Ponceau S staining followed by 30 min of blocking the membrane with either 5% (w/v) skimmed milk or 5% BSA (w/v) in PBS, depending on the requirements of the primary antibody and if detection involved phosphorylated residues. Primary and fluorescently labelled secondary antibody (Table 10 and 11) solutions were prepared in PBS-T with the respective blocking agent. Incubations of the primary antibody were carried out for 1 h at RT or overnight at 4° C and 1 h at RT for the secondary antibody. Fluorescent signals were subsequently detected at 680 nm or 800 nm with the Odyssey Infrared Imaging System (LI-COR) and quantification of the protein amount based on intensities via Image Studio[™] software occurred (LI-COR). Relative protein amounts were determined based on loading control and control conditions.

2.2.3.2 Protein pulldowns

For protein pulldowns streptavidin-binding peptide-tagged (SBP-) proteins were overexpressed in cells and harvested followed by lysis of the cell pellet with binding buffer (Table 13) for 10 min at RT. Next, 15 μL of Dynabeads[™] MyOne-Streptavidin-T1 per pulldown were washed twice with 200 μL binding buffer for equilibration utilizing a magnetic rack. Lysed cell suspension was centrifuged at 13.000 rpm for 5 min and the supernatant transferred to a new tube. For the inputs 20 μL of the lysate were transferred to a new tube and 7 μL 4 x NuPage[™] LDS sample buffer (containing 50 mM DTT) or 980 μL TRIZOL for analysis of proteins on Western blot and RNAs on Northern blot, respectively. The washed beads were mixed with 200 μL cell lysate containing SBP-tagged overexpressed proteins or empty vector

control followed by 30 min incubation at RT. Removal of unspecific binders was facilitated by washing with 200 μ L binding buffer twice and transfer of the pulldown samples in a new tube for the third washing step using a magnetic rack. For the elution of associated proteins, beads were resuspended in 40 μ L 1 x NuPage™ LDS sample buffer (Thermo Fisher Scientific) and boiled at 95 °C for 5 minutes, while elution of RNAs was facilitated by incubation of 40 μ L binding buffer containing 1% SDS at 65 °C for 5 minutes. Beads were removed from the eluate with the magnetic rack and samples were subjected to Western blot and Northern blot analysis, respectively.

2.2.3.2 Protein co-immunoprecipitation / RNA immunoprecipitation (Co-IP / RIP)

The co-immunoprecipitation was performed with lysates originating from mouse MC57G and hamster CHO-K1 cells. The cells were harvested, pelleted and lysed with Binding buffer (Table 13) for 10 minutes at RT as described for pulldown experiments. In the meantime, 10 μ L Protein-G coupled magnetic beads (Dynabeads™ Protein G for immunoprecipitation, Thermo Fisher Scientific) were washed twice with 200 μ L binding buffer and the supernatant was removed using a magnetic rack. Next, 2 μ L SmB antibody (Table 10) were coupled to the washed beads for 15 minutes at RT while rotating the tube. The inputs (20 μ L) for either RNA or protein were taken from the lysates and mixed with 980 μ L TRIZOL or 7 μ L 4 x NuPage™ LDS sample buffer (containing 50 mM DTT), respectively. The antibody-coupled beads were resuspended with 200 μ L lysate and incubated for 2 h at 4 °C. Subsequently, samples were washed three times with 200 μ L binding buffer and eluted in Binding buffer with 1% SDS for RNA analysis and 1 x NuPage™ LDS sample buffer for protein analysis. Co-precipitated proteins were analyzed on Western blot and precipitated RNAs were determined by Northern blot.

2.2.3.4 Quantitative proteomics by LC MS/MS

For the analysis of associated proteins with the 4.5S_H RNA, proteins were enriched with an RNA pulldown using biotinylated 4.5S_H RNA compared to bead control. The cell lysate was extracted from 5×10^6 B16-F10 cells with binding buffer (10 mM HEPES-KOH pH 7.2, 5 mM MgCl₂, 150 mM KCl 1% DDM) suitable for mass spectrometry analysis followed by RNA pulldown as described above. Pulldown samples were left in binding buffer after the third washing step and sample preparation and subsequent LC MS/MS analysis were executed by Dr. Christian Ihling (AG Sinz, Institute for Pharmacy, Martin-Luther University) according to the following protocol provided by Dr. Ihling:

Pulldowns were prepared for liquid chromatography/tandem mass spectrometry (LC-MS/MS) following a modified FASP (filter-aided sample preparation) protocol (Wiśniewski JR, Zougman A, Nagaraj N, Mann M. *Nat Methods*. 2009 May;6(5):359-62. doi:

10.1038/nmeth.1322. Epub 2009 Apr 19. PMID: 19377485). Briefly, protein samples were incubated and washed with 8 M urea in 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 10 mM TCEP, pH 8.5. Washing steps were performed using 0.5 mL centrifugal filter units (30-kDa cutoff, Sigma Aldrich, Darmstadt, Germany) and two centrifugation steps (14,000× g, 10 min). Afterwards, alkylation of proteins was performed with 50 mM iodoacetamide in 8 M urea, 50 mM HEPES, pH 8.5 at room temperature (20 min in the dark). Samples were washed twice with 50 mM HEPES, pH 8.5 (18,000× g, 2 x 10 min) before they were incubated with 1 µg trypsin in 50 mM HEPES, pH 8.5 at 37 °C overnight. After enzymatic digestion, samples were acidified with trifluoro acetic acid (TFA) at 0.5 % (v/v) final concentration. Proteolytic peptide mixtures were analyzed by LC/MS/MS on an UltiMate 3000 RSLC nano-HPLC system coupled to an Orbitrap Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). Peptides were separated on reversed phase C18 columns (trapping column: Acclaim PepMap 100, 300 µm × 5 mm, 5µm, 100 Å (Thermo Fisher Scientific); separation column: µPAC™ 75 cm C18 (Pharmafluidics). After desalting the samples on the trapping column, peptides were eluted and separated using a linear gradient from 3% to 35% B (solvent A: 0.1% (v/v) formic acid in water, solvent B: 0.08% (v/v) formic acid in acetonitrile) with a constant flow rate of 300 nl/min over 180 min. Data were acquired in data-dependent MS/MS mode with stepped higher-energy collision-induced dissociation (HCD) and normalized collision energies of 27%, 30%, and 33%. Each high-resolution full scan (m/z 375 to 1799, $R = 140,000$ at m/z 200) in the orbitrap was followed by high-resolution product ion scans ($R = 17,500$) of the 10 most intense signals in the full-scan mass spectrum (isolation window 2 Th); the target value of the automated gain control was set to 3,000,000 (MS) and 200,000 (MS/MS), maximum accumulation times were set to 100 ms (MS) and 250 ms (MS/MS). Precursor ions with charge states $<3+$ and $>8+$ or were excluded from fragmentation. Dynamic exclusion was enabled (duration 60 seconds, window 3 ppm).

2.2.4 Microscopy

2.2.4.1 Bright field microscopy

Bright field imaging served to analyze cell growth in normal tissue culture routine and to visualize cell morphology using a Nikon TE-2000-E phase contrast microscopy with a 4x, 10x or 20x magnification objective.

2.2.4.2 Fluorescence microscopy with CLSM technique

The localization of RNAs and proteins was analyzed with confocal laser scanning microscopy (CLSM) based on fluorescence *in situ* hybridization (FISH) and immunofluorescence staining. For imaging, cells were seeded onto glass cover slips which

were treated with 40 % (v/v) ethanol and 60 % (v/v) hydrochloric acid prior to the experiment. Standard immunofluorescence staining was carried out by fixation of the cells using 4% formaldehyde solution in PBS for 20 min at RT and subsequent permeabilization of the membranes with 0.5 % (w/v) Triton-X-100 in PBS for 4 min. Non-specific binding of the antibodies was blocked with 5 % BSA (w/v) in PBS for 45 min RT. The primary (1:250 dilution) and fluorescently-labelled secondary antibody solutions (1:500) were prepared in 2 % BSA in PBS (Table 10 and 13) and subsequently incubated for 1 h at RT. For visualization of the nucleus cells were stained with DAPI (4',6-diaminidino-2-phenylindole; 1:4000) in PBS for 3 min. The coverslips were washed with water and dehydrated in 99 % ethanol consecutively three times and placed onto an object slide mounted with ProLong™ Diamond Antifade (Thermo Fisher, Cat# P36970) to preserve fluorescent signals.

Preparation of the cells for FISH required fixation with 4% formaldehyde solution in PBS for 5-10 min at RT. The cells were permeabilized with 1 mL ice-cold ethanol at 4 °C for 1 h and washed with PBS twice before cells were preincubated with 250 µL PerfectHyb™ Plus Buffer for 5 min at RT. The fluorescently-labeled probes (Table 9) were diluted with PerfectHyb™ Plus buffer to a final concentration of 2 nmol/µL and incubated with the coverslips for 4 h and RT in the dark to preserve fluorescence. Unbound probe was removed by washing twice with 2 x SSC buffer followed by immunofluorescence staining of proteins for co-localization according to the procedure described above.

Images were acquired with the Leica SP5X or SP8X confocal microscopes using a 63x oil objective. The fluorescent dyes were excited with a white light laser and emission was detected with HyD detectors. Nucleus staining was imaged with DAPI excitation using a diode laser at 405 nm wavelength. Images for the LNA-treated B16-F10 and U2-OS have been compiled in a maximum intensity projection featuring multiple confocal planes of the nuclei to have a better representation of the splicing speckles. Analysis of the speckle parameters in the images was conducted with ImageJ using the NuclearParticleDetector2D plugin of the MiToBo package on standard settings. The eccentricity of H9C2 cells overexpressing Rbfox1 was determined manually based on the cell roundness parameter in ImageJ.

2.2.4.3 Live-cell imaging via Incucyte S3

To detect multinucleated H9C2 cells under differentiation conditions, bright field images were taken with the Incucyte S3 (Sartorius) at the endpoint of differentiation using a 10x magnification object. Images were taken to capture the whole well with 25 images in a 5 x 5 grid for each condition.

2.2.5 Computational analyses

2.2.5.1 Bioinformatic analyses of the RNA-Sequencing reads

Downstream analyses of the sequencing reads of the above mentioned RNA-Sequencing was handled by Dr. Danny Misiak (Martin-Luther University, Halle) and executed as described in (Zorn et al. 2021) under the materials and methods “bioinformatic analyses”.

2.2.5.2 Analysis of putative Rbfox1 targets

The analysis was performed with intron sequences of the whole mouse genome using the biomaRt tool from the Ensembl database. These sequences were analyzed for Rbfox1 binding motifs (TGCATG and GCATG) with the desktop application BLAST+ from the National Library of Medicine (NIH). The potential binding sites were compiled in a list with each binding site on the nucleotide sequence in the respective intron. Each binding site received was summed up to calculate the total amount of binding sites in all introns of each specific gene for further analysis. The heart specific transcripts containing putative Rbfox1 binding sites were accessed through comparison of gene expression from the whole mouse genome with an RNA-Sequencing dataset from mouse heart (E-MTAB-3725). For further analysis and to increase reliability of the target list, an artificial threshold of 10 transcripts per 1 million was implemented. Moreover, only genes with a minimum of 20 putative Rbfox1 binding sites were considered. The regulation by Rbfox1 could still occur with less than 20 binding motifs, but these genes were not included for the initial investigation. In addition, non-coding RNAs have been reported to contain Rbfox1 binding sites, but were also discarded for the analysis in this thesis.

2.2.5.3 Functional annotation clustering

The clustering of putative Rbfox1 target genes and proteins received from the LC-MS/MS analysis were uploaded to the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al. 2009). Proteins associated and enriched with the 4.5S_H RNA were converted to a gene list which was subjected to a Gene Ontology classification. The list of putative Rbfox1 targets received from the target prediction based on Rbfox1 intronic binding sites was analyzed using the functional annotation tool to generate functional clusters.

2.2.5.4 Cell counting of multinucleated cells

The H9C2 cells were transfected with a siRNA pool against Rbfox1 or non-targeting siControl (Table 2) and subjected to cell differentiation as described above (2.2.1.2). To determine multinucleated cells, H9C2 were fixated and permeabilized as described for immunofluorescence (2.2.4.2) and images were taken with the Incucyte S3 (Sartorius). The

average cell number per image was calculated from manually counting total cell numbers based on visible nuclei for one randomly picked image out of 25 images per well for each replicate and transfection condition. Next, 5 out of the remaining 24 images were chosen using a random number generator and multinucleated cells were counted manually. The fraction of multinucleated cells was calculated for each image based on the average total cell number that has been determined before.

2.2.5.5 Publicly available datasets, databases and visualization

The datasets were extracted from the GEO databank (Gtf3c1 and Rpc155 ChIP GSE75191; mouse Srsf1-CLIP-seq GSE141349; human SRSF1-CLIP GSE130867; MI and TAC RNA-Seq GSE96561) and BDP1, BRF1 and POLR3G CHIP-seqs were accessed from the ENCODE databank. The RNA-sequencing dataset for the EGFR-KO mice was kindly provided by Dr. Barbara Schreier (Martin-Luther-University, Halle). The tRNA sequences were retrieved from two tRNA databases (GtRNAdb 2.0 (Chan and Lowe 2016) and tRNAdb 2009 (Jühling et al. 2009)). The molecular weight of the alternatively spliced variants Metavinculin and ePxn and the kinase recognition motifs have been extracted from the human protein reference database HPRD (Keshava Prasad et al. 2009). For the folding prediction of the ePxn isoform AlphaFold Protein Structure Database (Jumper et al. 2021) was used and the model presented in this thesis was created with PyMOL. The secondary structure prediction of the 4.5S_H RNA was visualized with the desktop application of the VARNA Visualization Applet for RNA for lowest free energy of the 4.5S_H RNA sequence (Darty et al. 2009). Illustrations for the figures in the introduction and the minigene schematics were created with BioRender (biorender.com) and are applicable for publication with the following licences: *SR25XLDMEL*, *DE25XLDMP*, *ST25XLDMZ5*, *FZ25XLDN9I*, *NB25XLDNOG*, *QM25XLDO0X*, *GE25XLDOAW*, *IG25XLDOSY*, *VE25XLDPAN*, *XI25XLDPM7*, *TK25XLDPWR*, *BE25XLDQ9D*.

3. Results

3.1 Modification of RNA-Sequencing for medium-sized ncRNA analyses

3.1.1 Adapting the library generation protocol for msRNA-Sequencing analysis

The current approaches to identify and characterize ncRNAs by RNA-Sequencing are primarily directed at two types of ncRNA species: long ncRNAs, such as lincRNAs, which are over 200-300 nucleotides (nts) in length, and short ncRNAs, like miRNAs, which are below 30 nts in size. However, the established library preparation protocols often do not cover medium-sized ncRNAs (msRNAs), which show intermediate lengths between these two thresholds. RNA-Sequencing libraries for small ncRNAs are often generated without special RNA treatments since the removal of 5'-end modifications such as cap structures or triphosphates commonly found in msRNAs are generally not required. On the other hand, lincRNA protocols usually require RNA fragmentation based on the length of the molecule to cover the whole RNA. However, this approach also imposes limitations on the coverage of msRNAs, which are similarly fragmented and may not meet the size requirements of subsequent library preparation steps or result in read peaks that do not accurately represent the full-length transcript. Additionally, msRNAs are frequently encoded by multiple loci hindering annotation analyses of RNA-Sequencing data. Several established protocols regarding short and long ncRNA-Sequencing are putting emphasis on RNA fragmentation, size selection and 5'/3'-end modifications (McGinn and Czech 2014; Podnar et al. 2014). However, the sequencing analysis of msRNAs required a new approach, with a modified library preparation protocol and a novel computational pipeline depicted in Figure 11.

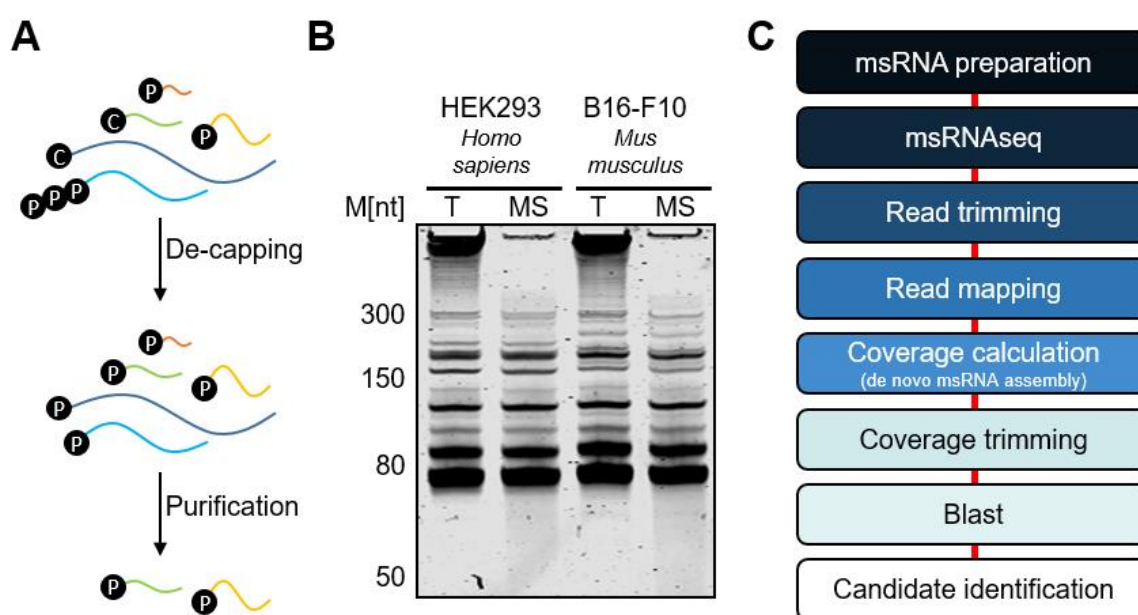


Figure 11. The msRNA-Sequencing pipeline. A Schematic representation of msRNA purification for the sequencing analysis. Isolated total RNA was de-capped and a size selection for RNA in the target

range of 50-300 nts was conducted. **B** Representative 15% TBE-Urea-Gel for msRNA purification in HEK293 (human) and B16-F10 (mouse) cells T – total RNA; MS – purified msRNAs. **C** Depiction of the bioinformatic pipeline for the msRNA-Sequencing analysis. (Published data from Zorn et al. 2021)

In order to analyze msRNA levels, total RNA was isolated from human HEK293 and mouse B16-F10 cells and subsequently treated with RNA 5'-Pyrophosphohydrolase (RppH) (Figure 1A). This step was necessary since RNA Polymerase II and RNA Polymerase III transcripts usually contain unwanted cap-structures or triphosphates at their 5'-end. In addition to the 3'-OH group ligation of RNA-linkers for the purpose of RNA-sequencing also requires a monophosphate at the 5'-end for a proper linker ligation at both ends of the RNA molecule. Since mature tRNAs are rapidly loaded with amino acids to generate the aminoacyl-tRNA, the 3'-OH group of this msRNA class is not accessible for linker ligation. The detection of tRNAs was therefore mainly restricted to pre-tRNAs (not fully processed) and a small fraction of unloaded mature tRNAs.

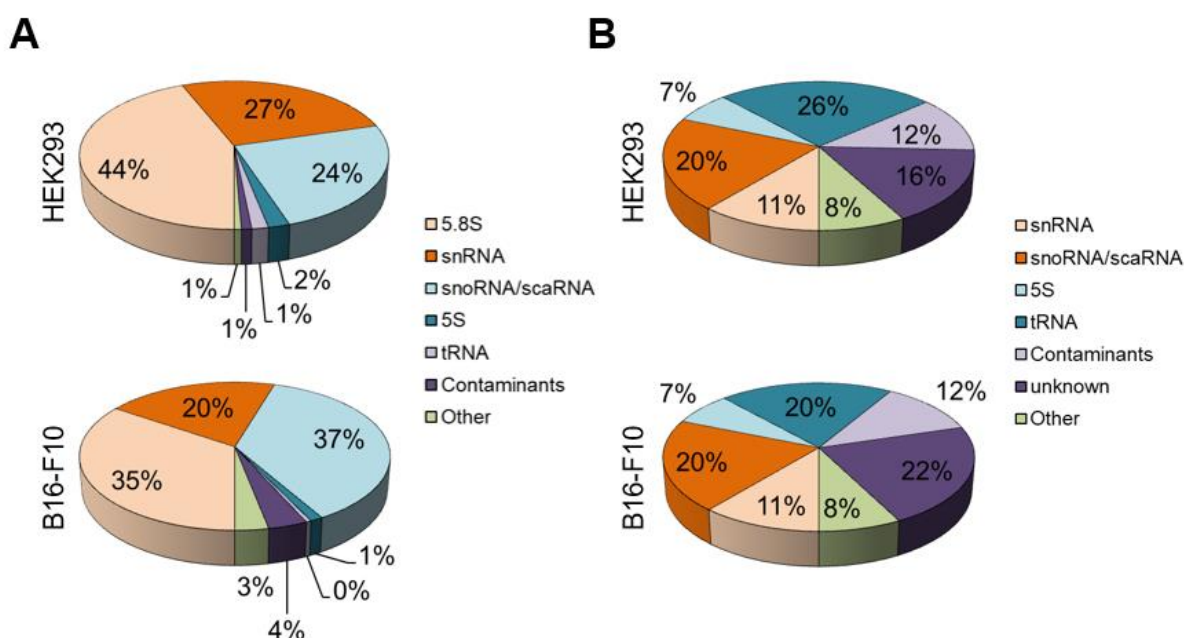


Figure 12. Classification of the coverage regions from the msRNA-Sequencing. **A** The indicated fraction of each msRNA family was calculated based on their total amount of reads obtained in the analysis. **B** The indicated fraction of each msRNA family based on their number of genes that the reads were mapped to. The contaminants are mainly comprised of mitochondrial RNAs, histone mRNAs and large ribosomal RNAs (Published data from Zorn et al. 2021)

For the evaluation of the msRNA-Seq and the bioinformatic pipeline from Figure 11C, the reads of RNAs and the blast results were categorized into their respective RNA species based on the total read amount (Figure 12A). The 5.8S RNA was the most abundant RNA that could be sequenced in the msRNA-Seq with 44% in HEK293 cells and 35% in B16-F10 cells. This result was expected, since established RNA-Sequencing protocols remove rRNA routinely before the analysis. However, for the msRNA-Seq the depletion of rRNA was deliberately avoided for an unbiased approach of the analysis. Additional enriched types of msRNAs with a high read count include snRNAs (27% and 20%) and snoRNAs (24% and 37%) in HEK293 and B16-F10

cells, respectively. The small fraction of contaminants in the pie chart refers to larger rRNAs, mitochondrial RNAs and histone mRNAs. This result was anticipated due to the fact that histone mRNAs represent the smallest category of mRNA molecules, with their lengths typically starting around ~300 nucleotides and are thereby included in the cutoff for the msRNA-Seq. As stated previously, msRNAs are frequently encoded by several genomic loci. Hence, the genes of the major classes of msRNAs (snRNAs, snoRNAs, 5S rRNAs and tRNAs) were mapped to multiple genomic regions in human and mouse genomes (Figure 12B) emphasizing that the same transcript can be produced from various genes. Nonetheless, a large fraction of genes could not be attributed to unannotated msRNA genes, indicating the identification of previously undescribed msRNAs (HEK293: 16%, B16-F10 22%).

3.1.2 Identification of RNA-Polymerase III driven msRNA transcripts

A significant amount of the msRNA transcripts are synthesized by RNA-Polymerase III, which utilizes three different promoter types for transcription of target genes (Type I, II and III, see also Introduction). The type II promoters are mainly responsible for tRNA synthesis and contain conserved A- and B-box sequences to initiate transcription by TFIIC and ultimately POLIII (Schramm and Hernandez 2002). Additionally, candidate sequences with type II promoters can be crosschecked with tRNA databases to filter out known tRNA sequences (Chan and Lowe 2016; Jühling et al. 2009).

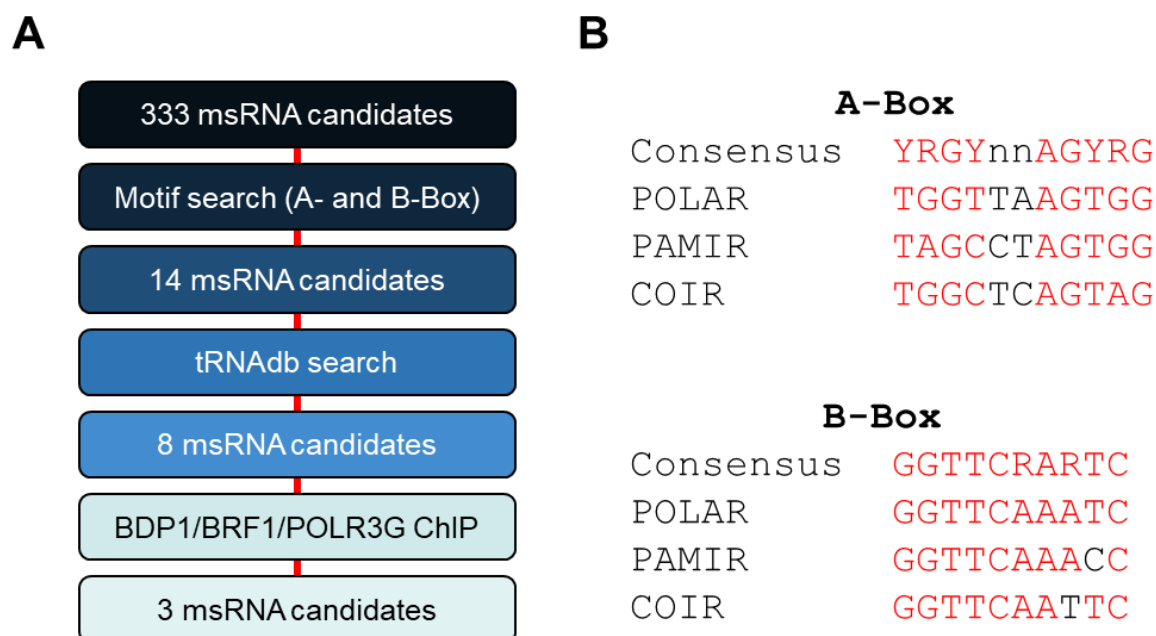


Figure 13. Identification of msRNAs transcribed by RNA-Polymerase III **A** The schematic representation of the analysis pipeline for identification of novel POLIII-transcribed msRNAs. **B** Consensus sequences of A- and B-boxes recognized by RNA-Polymerase III compared to the three putative msRNA candidates POLAR, PAMIR and COIR. (Published data from Zorn et al. 2021)

Hence, the 333 putative candidate msRNAs identified in the msRNA-Seq from the HEK293 cells (Figure 12B) were screened for motifs resembling A- and B-boxes using the FIMO-

package (Cuellar-Partida et al. 2012) (Figure 13A). As a result, 14 msRNAs exhibiting A- and B-boxes could be identified, which infer a type II transcription by RNA-Polymerase III. Furthermore, the sequences of the 14 candidate genes were compared to a tRNA database and it emerged that six msRNA sequences were matching to reported tRNAs. The remaining msRNA candidate genes were investigated for BDP1/BRF1/POLR3G ChIP peaks using publicly available ChIP datasets (derived from ENCODE). The association of these transcription factors would indicate an RNA-Polymerase III-dependent transcription of the msRNA genes. Three out of the eight remaining msRNA candidates displayed peaks BDP1/BRF1/POLR3G in close proximity to their genomic loci, suggesting the presence of specific transcription factor recognition motifs within these genes. Moreover, these three candidate genes contain sequences matching the consensus sequence of A- and B-boxes (Figure 13B). These three msRNAs were used for further characterization and named after their genomic location PAMR1 interspersed RNA (PAMIR), C1QTNF6 interspersed RNA (COIR) and POLR3E antisense RNA (POLAR). First, the POLIII-dependent transcription of the RNAs was validated by transient knockdown of transcription factor BDP1 in HEK293 cells (Figure 14A).

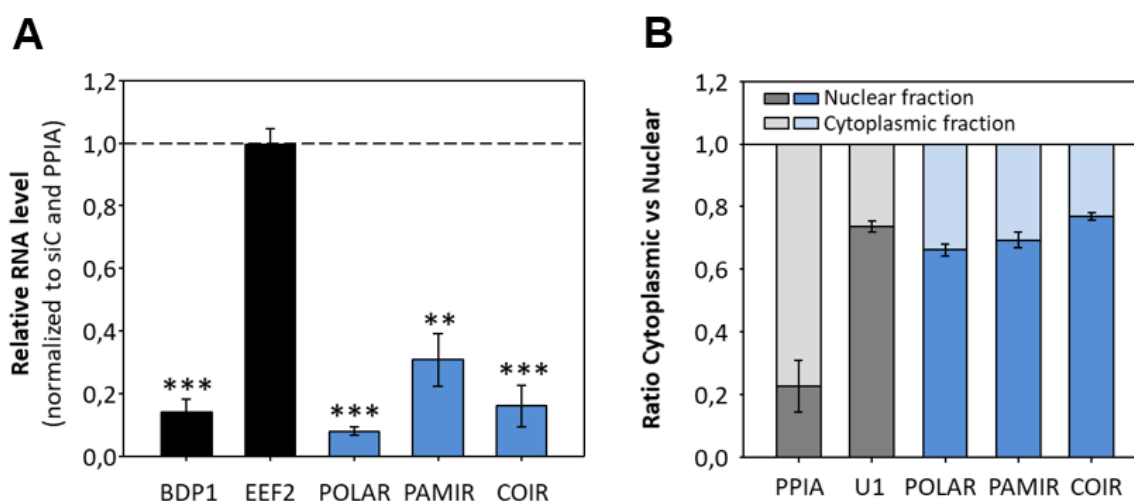


Figure 14. Initial characterization of the ncRNAs identified by the msRNA-Sequencing. **A** qRT-PCR-analysis of total RNA extracted from transient siRNA-mediated depletion of BDP1 in HEK293 cells. RNA amounts of POLAR, PAMIR and COIR were significantly reduced following the BDP1 knockdown. The EEF2 mRNA served as a negative control. **B** Cellular distribution ratios of the three candidate genes based on nuclear-cytoplasmic fractionation was assessed. The U1 snRNA served as a nuclear and the mRNA PPIA as a cytoplasmic control. (Published data from Zorn et al. 2021)

The analysis of POLAR, PAMIR and COIR in a qRT-PCR analysis revealed decreased RNA amounts for all three RNAs after the siRNA mediated reduction of BDP1. In contrast to this, the RNA-Polymerase II transcript EE2 did not show altered RNA expression upon BDP1 knockdown. Furthermore, the subcellular localization of the three RNAs was analyzed by nuclear-cytoplasmic fractionation of HEK293 cells. Total RNA was isolated from each fraction and analyzed by qRT-PCR (Figure 14B). The majority of the mRNA PPIA was found in the

cytoplasmic fraction, while snRNA U1 was mainly enriched in the nuclear fraction. The msRNAs POLAR, PAMIR and COIR showed a similar distribution as the U1 RNA, indicating a more nuclear localization. The initial characterization of the three msRNAs identified by the msRNA-Seq confirmed transcription by RNA-Polymerase III based on knockdown experiments involving transcription factor BDP1. To further validate, that identified msRNAs from the msRNA-Seq are indeed RNAs with cellular function, additional experiments were carried out for the POLAR RNA.

3.1.3 Characterization of the medium-sized non-coding RNA POLAR

As already mentioned previously, the genomic locus for the POLAR RNA was mapped to the first intron of the human POLR3E gene as indicated in Figure 15A (upper panel) in antisense orientation, which led to the name POL3RE-antisense RNA (POLAR). The ChIP-Seq dataset of POLIII transcription factors BDP1, BRF1 and RNA-Polymerase III subunit POLR3G from K562 cells indicated transcription of the POLAR gene by RNA-Polymerase III (Figure 15A bottom panel).

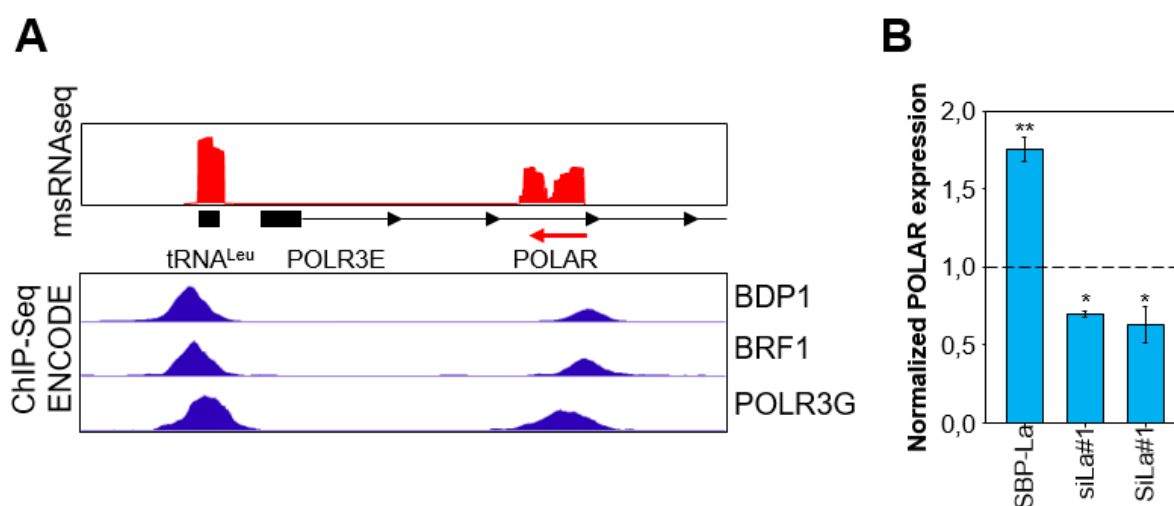


Figure 15. Initial characterization of the POLAR msRNA. **A** The genomic locus of the POLR3E (chr16: 22,297,409-22,335,096) with read coverage for the tRNA^{Leu} and the newly identified POLR3E-antisense RNA (POLAR) from the msRNA-Seq are depicted in the top panel. The ChIP-Seq analyses in K562 cells from the ENCODE database for three RNA-Polymerase III co-factors at the same genomic locus is depicted in the bottom panel. **B** qRT-PCR analysis of isolated total RNA of HEK293 cells with transient overexpression of SBP-tagged La protein or transient siRNA-mediated knockdown of La using two independent siRNAs (siLa#1, siLa#2). The RNA expression of POLAR was determined in all conditions normalized to empty vector transfection or control siRNA (Published data from Zorn et al. 2021)

The tRNA^{Leu} gene locus upstream of the POLR3E gene is a known POLIII-driven transcript and served as a positive control in regard to the ChIP-Seq peaks of the RNA-Polymerase III factors. For both loci, peaks of the transcription factors BDP1 and BRF1 could be detected immediately upstream of the genes. Additionally, POLR3G signal could be seen throughout the whole transcript resembling active POLIII transcription. Another prominent feature of nascent POLIII transcripts is an oligo-uridine tract at the 3'-end, typically recognized by the La

protein (Maraia et al. 2017). The La protein is a critical factor for POLIII-transcript homeostasis and turnover including protection from RNA degradation and assisting in RNA processing. For this reason, POLAR RNA levels were measured after transient overexpression of SBP-tagged La protein and single siRNA-mediated transient knockdown (siLa#1, siLa#2) in HEK293 cells (Figure 15B). The abundance of the POLAR RNA was significantly increased with La overexpression and was found to be significantly decreased with reduced amounts of La. Furthermore, the POLAR RNA was significantly enriched in La-pulldowns, providing additional evidence for La association with the msRNA POLAR (data not shown).

Non-coding RNAs that are associated with specific cellular functions are often organized in membraneless organelles (MLOs) which contain a specific composition of RNAs and proteins (Banani et al. 2017, see Introduction). To further characterize POLAR in regard to its potential molecular function, cellular localization analyzes were carried out using a combination of fluorescence *in situ* hybridization (FISH) and immunostaining (Figure 16).

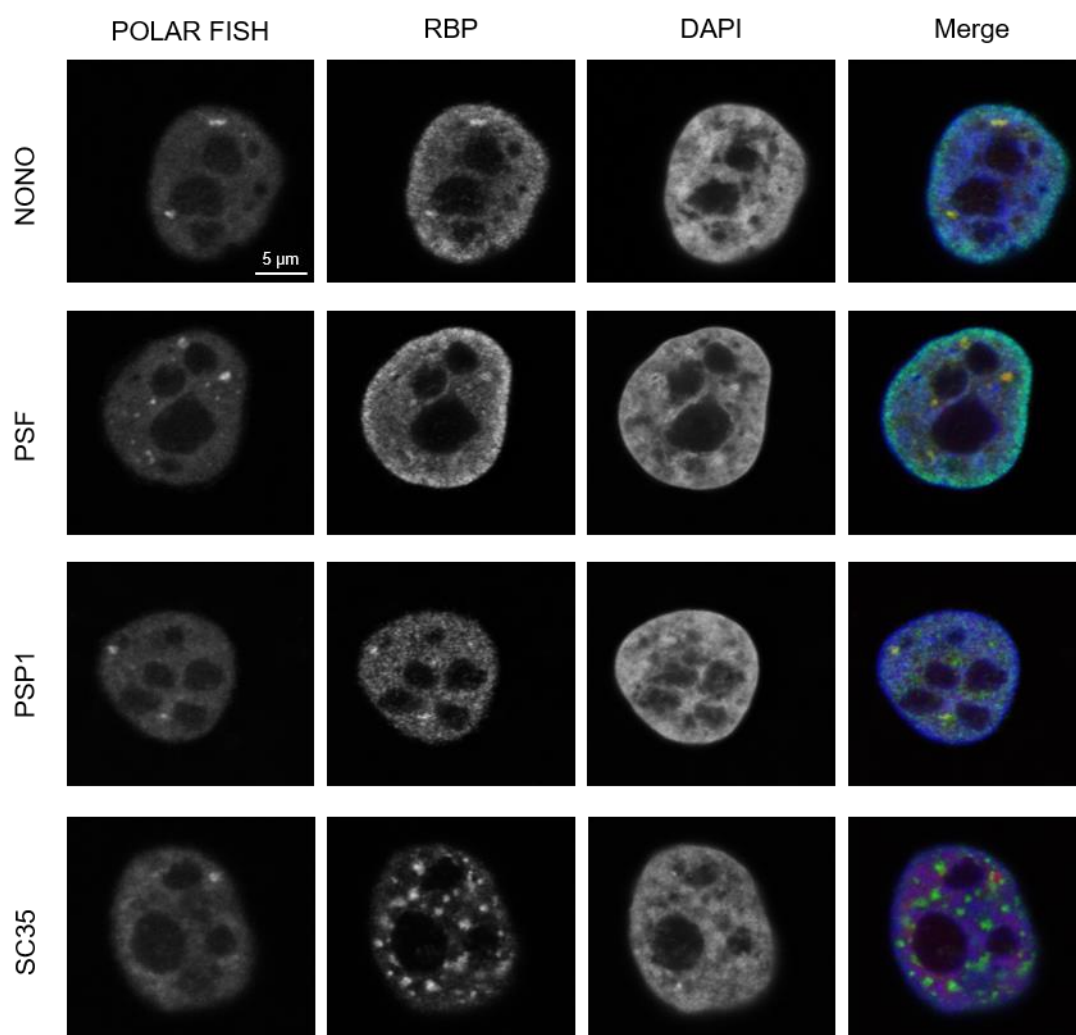


Figure 16. Analysis of cellular localization of POLAR in OVCAR-3 cells. Fluorescent *in situ* hybridization (FISH) followed by immunostainings with paraspeckle-associated proteins (NONO, PSF and PSP1) and the splicing speckle marker SC-35. Localization of POLAR is depicted in the first and

the respective RBP in the second column. DAPI staining was used to label the nucleus (third column). (Published data from Zorn et al. 2021)

The antibodies for the immunostaining targeted proteins localized in paraspeckles (PSP1, PSF and NONO) and splicing speckles (SC-35). Due to suboptimal properties of HEK293 cells for microscopic analyzes, these experiments were conducted in OVCAR-3 cells with comparable expression of endogenous POLAR. Consistent with the results from the nuclear-cytoplasmic fractionation (Figure 14B) the FISH displays a mainly nuclear localization for the RNA. However, the RNA appears to be concentrated in distinct areas rather than distributed homogenously in the nucleus, similarly to splicing speckles (Galganski et al. 2017). Nonetheless, co-staining with splicing speckle marker SC-35 did not reveal co-localization of POLAR in splicing speckles. In addition, the foci containing the RNA were fewer in numbers and varying in different cells, properties which could be attributed to paraspeckles in the interchromatin space of human cells (Fox et al. 2002). This could be confirmed when staining with antibodies against marker proteins PSF, PSP1 and NONO which clearly co-localize with POLAR, indicating the RNA to be localized in paraspeckles. The association of POLAR with paraspeckle proteins could also be confirmed in an RNA pulldown assay (Figure 17). For this experiment, biotinylated *in vitro* transcribed POLAR RNA was used with OVCAR-3 cell lysate. The Y4 RNA served as a control, since the RNA is also transcribed by RNA-Polymerase III in humans but localizes to the cytoplasm (Köhn et al. 2015).

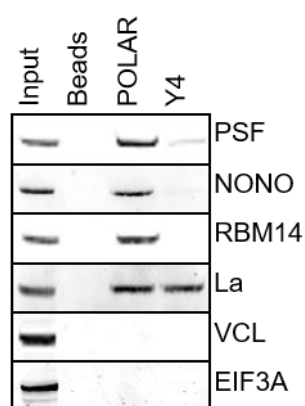


Figure 17. Western blot analysis of an RNA pulldown using OVCAR-3 cell lysate with biotinylated *in vitro* transcribed POLAR RNA. Beads only and the Y4 RNA served as a negative control for association with paraspeckle proteins PSF, NONO and RBM14. VCL and EIF3A were used as a specificity control for binding to POLAR and La served as a positive control for binding to POLIII transcripts. (Published data from Zorn et al. 2021)

The paraspeckle proteins PSF, NONO and RBM14 could be pulled down with POLAR RNA, but not with Y4 RNA. The La protein, recognizing the oligo-U stretch in POLIII transcripts, was enriched in both RNA pulldowns, while VCL and EIF3A served as negative controls and did not associate with either RNA.

Taken together it could be shown that the modification of the RNA-Sequencing protocol is a useful method for the identification of medium sized RNAs. In addition to the sequencing

and mapping of known msRNAs, three previously unannotated RNAs – POLAR, PAMIR and COIR could be identified in HEK293 cells. Initial characterization revealed RNA-Polymerase III dependent transcription based on ChIP datasets and consensus A- and B-boxes in the RNAs. Further investigation of the POLAR RNA indicated association of the genomic locus with RNA-Polymerase III co-factors BDP1 and BRF1 as well as the POLIII subunit POLR3G. Moreover, the POLAR RNA associated with the La protein and showed expression changes upon modulation of La, further implying a POLIII characteristic oligo-U stretch at the 3'-end. Based on co-localization studies POLAR appears to be concentrated in paraspeckles and is associated with specific paraspeckle markers PSF, NONO and RBM14 suggesting a potential function of POLAR in paraspeckles.

3.2 The medium sized non-coding RNA 4.5S_H

3.2.1 The 4.5S_H RNA is a nuclear POLIII transcript

The msRNA-Seq not only provided data for the identification of potentially novel RNAs but also serves as a valid starting point for studying other msRNAs. After the analysis of the msRNA-Sequencing data from the human HEK293 cells the dataset of the B16-F10 was investigated. The medium-sized non-coding RNA 4.5S_H has been described previously as a RNA-Polymerase III transcript based on the existence of consensus sequences resembling A- and B-boxes. This RNA is highly abundant and arranged in tandem repeats with a large copy number of the gene (700-800 copies per genome, see Introduction). The genomic regions flanking the gene sequence are weakly conserved but the transcript sequence itself displays a high conservation in all genetic loci (Gogolevskaya et al. 2005). This indicates a functional purpose of the 4.5S_H RNA which is poorly understood and has been investigated in this thesis.

The synthesis of 4.5S_H RNA has been suggested based on sequences resembling A- and B-boxes, indicating RNA-polymerase III transcription. For clarification, the msRNA-Seq data has been combined with ChIP-Seq datasets of RNA polymerase III components Gtf3c1 (POLIII initiation factor) and Rpc155 (POLIII catalytic subunit) in a similar manner as for POLAR (Figure 15).

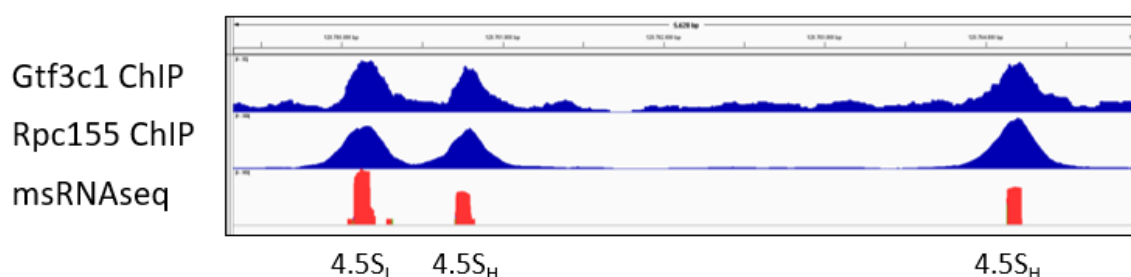


Figure 18. Genomic view of one msRNA-region of the mouse genome (chr6:128,759,326-128,764,958). We used publicly available ChIP-Seq datasets of the RNA polymerase III components

Gtf3c1 and Rpc155 (GSE75191) and combined this with the msRNA-Seq dataset. Peaks highlighted in red indicate msRNA expression (from 5' to 3': one 4.5S_I and two 4.5S_H RNA genes). Peaks in blue indicate ChIP-signals for Gtf3c1 and Rpc155, respectively. Note that all msRNA peaks overlap with these ChIP-signals indicating POLIII-driven RNA synthesis.

The msRNA-Seq data (Figure 18 bottom panel) revealed robust expression of two representative genetic loci transcribing 4.5S_H and a third locus encoding another medium-sized non-coding RNA - 4.5S_I. The ChIP-seq data of the POLIII-factors (Figure 18 upper panels) displays strong association with the respective loci for all three RNAs indicating a RNA-polymerase III-driven transcription of the 4.5S_H and 4.5S_I RNA.

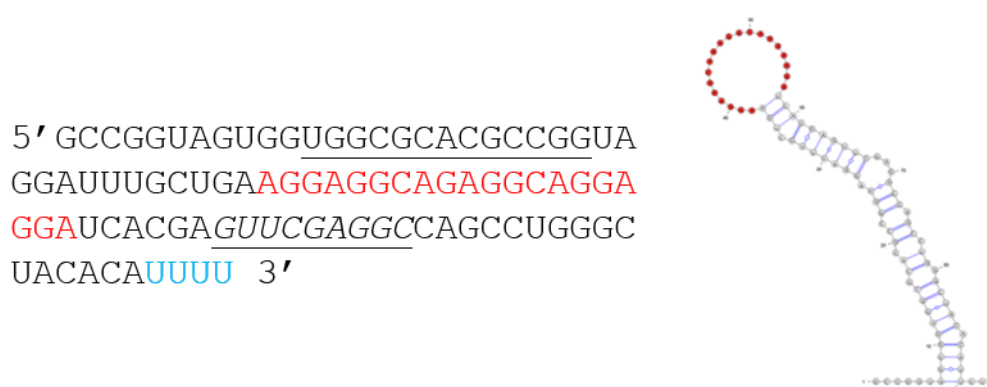


Figure 19. The 4.5S_H RNA sequence and predicted RNA fold. The RNA sequence is depicted and putative A- and B-boxes are underlined. The sequence highlighted in red is indicating the proposed RNA loop region. The Oligo-U terminator sequence is colored in blue.

Additionally, the 4.5S_H RNA sequence depicted in Figure 19 features an oligo-U terminator characteristic for POLIII transcripts that enables the association of the La protein. Interestingly, the RNA also contains an A- and G-rich region located in the center of the sequence. Upon RNA fold prediction using VARNA Visualization Applet (Darty et al. 2009) it became evident that 4.5S_H RNA might fold into a stem-loop structure. The sequence of the central loop consists of the purine-rich bases including three GAGG-repeats.

It has been suggested that the 4.5S_H RNA gene originated from the B1-SINE retrotransposon element exclusively present in rodents and the sequence still features limited homology (Gogolevskaya et al. 2005; Labuda and Zietkiewicz 1994; Vassetzky et al. 2003). The 4.5S_H RNA however, could only be detected in six rodent families that are part of the myodonta clade (Gogolevskaya et al. 2005). To determine the expression of 4.5S_H RNA in various mammalian cell lines a Northern blot analysis was performed (Figure 20A). In line with published data 4.5S_H RNA is absent in cell lines derived from human (HEK293 and HeLa) and can also not be detected in 104C1 cells from guinea pig (Harada et al. 1979; Gogolevskaya et al. 2005). Expectedly, 4.5S_H RNA is highly expressed in cells originating from mouse (MEF, RAW264, MC57G, P19, NIH3T3), rat (AR42J) and hamster (CHO-K1) as these belong to the myodonta clade mentioned above. Published data suggested 4.5S_H RNA to be localized in the

cytoplasm due to its potential interaction with cytoplasmic poly(A)⁺ mRNA (Jelinek and Leinwand 1978).

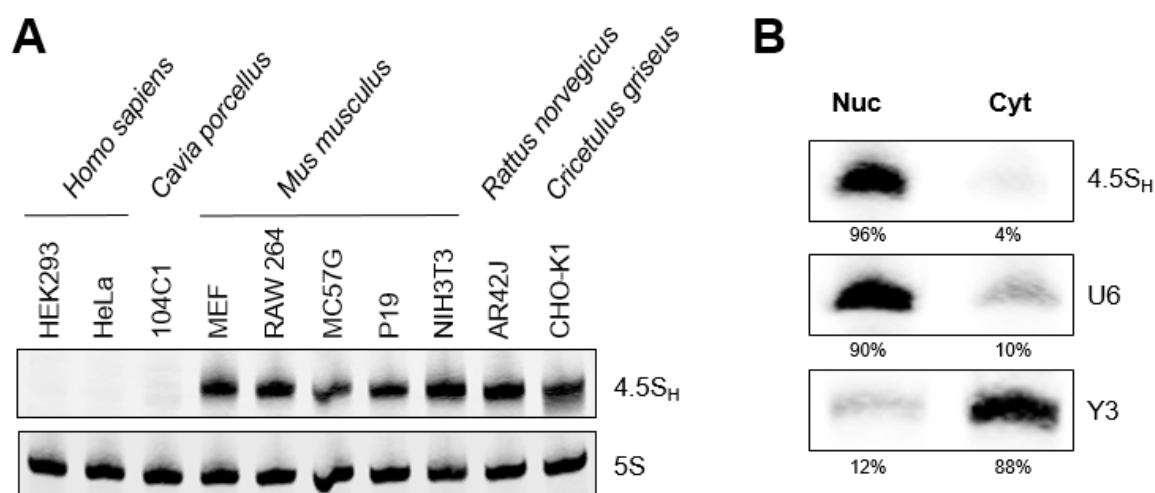


Figure 20. The 4.5S_H RNA expression and subcellular distribution. **A** Northern blot analysis of 4.5S_H RNA expression in various cell lines from indicated species. 5S rRNA served as loading control. **B** Northern blot analysis of fractionated B16-F10 cells into nuclear (Nuc) and cytoplasmic RNA pools (Cyt) to evaluate the cellular distribution of 4.5S_H RNA. U6 and Y3 RNA served as controls for proper nuclear and cytoplasmic fractionation, respectively.

This subcellular localization was questioned by other groups that performed nuclear-cytoplasmic fractionation and identified 4.5S_H RNA to be a nuclear RNA with varying ratios (Ishida et al. 2015; Tatosyan et al. 2017). To confirm these findings, we conducted fractionation of B16-F10 mouse melanoma cells and examined the nuclear and cytoplasmic fractions using Northern Blot analyses (Figure 20B). The success of the fractionation was controlled using U6 snRNA probes as a control for a nuclear RNA and Y3 RNA as a control for the cytoplasmic fraction. The signal for the 4.5S_H RNA clearly indicates an almost exclusive subcellular localization with 96% versus 4% in the nucleus after fractionation. Additionally, it has been previously described that 4.5S_H RNA interacts with nuclear RNA-binding proteins including Nucleolin (Hirose and Harada 2008). In order to confirm these results, cellular localization and nuclear distribution was validated with FISH experiments in combination with immunostaining of nuclear factors in B16-F10 cells (Figure 21).

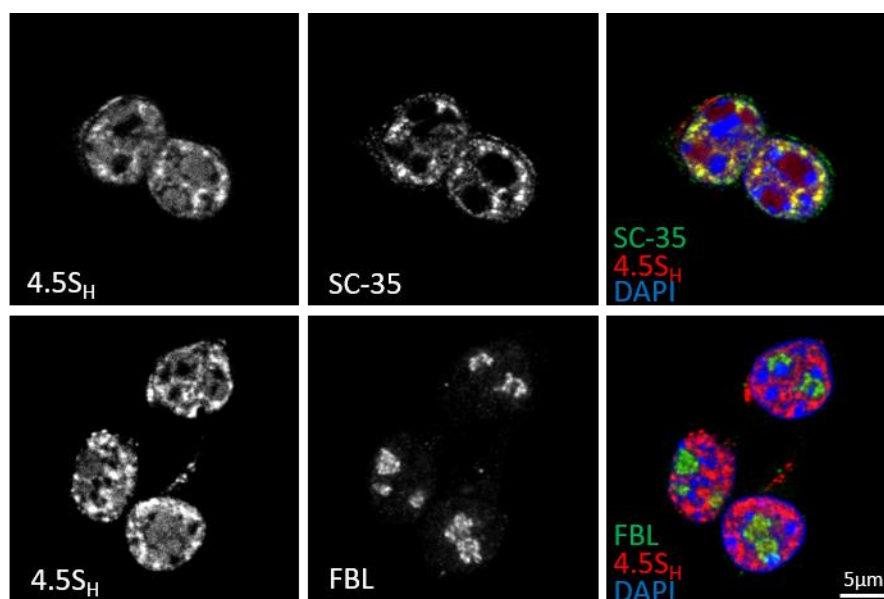


Figure 21. Fluorescence *in situ* hybridization (FISH) of 4.5S_H RNA combined with immunostaining. The image displays certain markers of nuclear MLOs in B16-F10 cells. The 4.5S_H RNA is highlighted in red and immunostaining of either SC-35 (splicing speckles) or FBL (nucleoli) is shown in green.

Fluorescent *in situ* hybridization revealed nuclear localization of 4.5S_H RNA in B16-F10 in concentrated foci resembling membraneless organelles. Immunostaining with marker proteins for nucleoli (FBL) and splicing speckles (SC-35) exposed co-localization of 4.5S_H with SC-35 but not FBL, hinting towards the localization of 4.5S_H RNA in splicing speckles.

3.2.2 The nuclear 4.5S_H RNA is mainly associated with splicing factors

It has been previously reported that 4.5S_H interacts with several nuclear proteins of which only Nucleolin could be identified as specific binder (Hirose and Harada 2008). In order to further characterize the non-coding RNA 4.5S_H, the identification of interacting proteins can provide insight in regard to its potential function. For this reason, RNA pulldowns with biotinylated *in vitro* transcribed 4.5S_H and cell lysates from B16-F10 were carried out. The eluted fraction of proteins from the pulldowns was then subjected to mass spectrometry analysis (LC MS/MS) for an unbiased approach to identify RNA-associated proteins. The enriched proteins with 4.5S_H over bead control are depicted in Figure 22. Notably, especially proteins involved in splicing (blue dots) were strongly enriched within the 4.5S_H RNA eluates. A prominent fraction of these included proteins that contain serine/arginine-rich (RS) domains (indicated by the red dots) and belong to the family of serine/arginine-rich splicing factors (SRSFs). These RNA-binding proteins usually contain one or more RNA-recognition motifs (RMM) and an extended disordered RS-domain of varying length (Haynes and Iakoucheva 2006; Twyffels et al. 2011; Xiang et al. 2013, see Introduction).

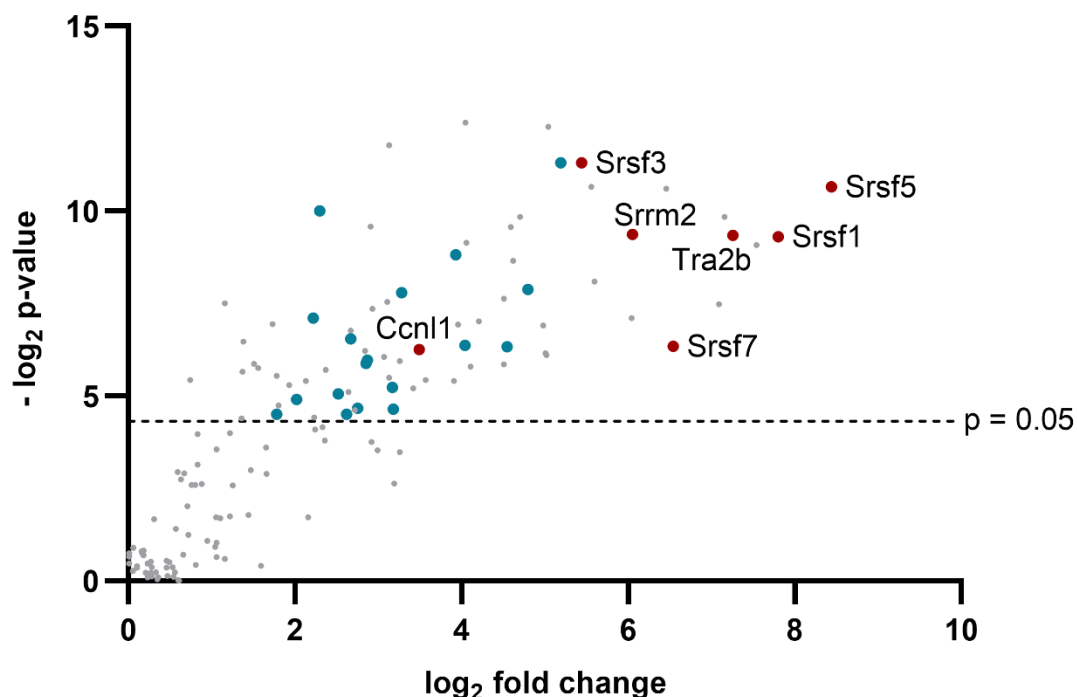


Figure 22. Proteins identified by pulldown and LC MS/MS analyses to be associated with 4.5S_H RNA. The fold change is based on identified peptides in the respective samples and enrichment of proteins from B16-F10 cells has been calculated for 4.5S_H RNA vs bead control. Proteins associated with splicing are highlighted in blue of which RS-domain containing proteins are highlighted in red.

In addition, functional clustering of significantly enriched proteins (above $p = 0.05$) using the DAVID bioinformatics database (Huang et al. 2009) exhibited mRNA processing, RNA splicing and mRNA splicing via spliceosome as the most significant Gene Ontology terms (Figure 23A).

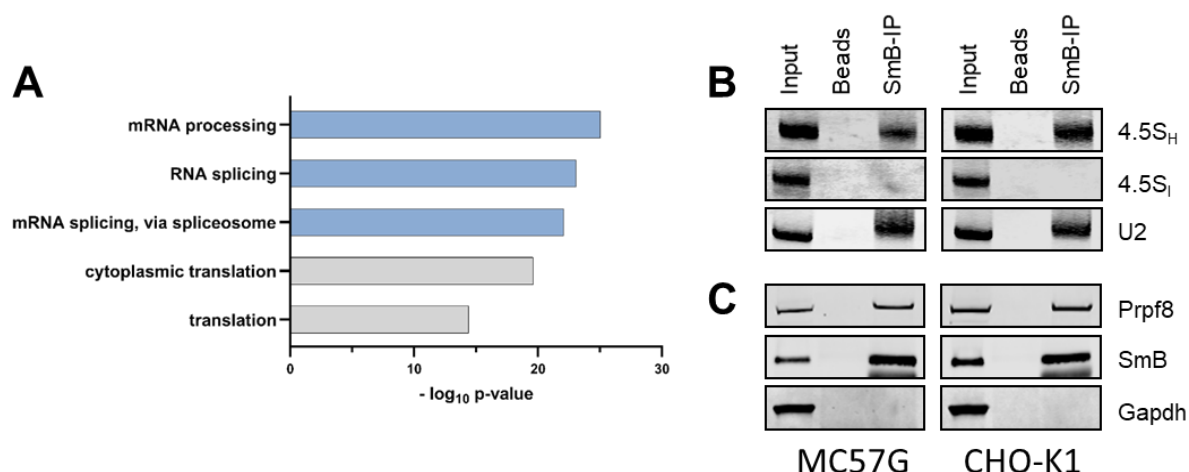


Figure 23. Functional clustering and association with spliceosomal proteins. **A** Functional clustering analysis for enriched proteins with 4.5S_H RNA vs bead control performed with the DAVID bioinformatic database. **B** Northern blot analysis of associated RNAs with spliceosomal component SmB after immunoprecipitation from total cell lysate of MC57G and CHO-K1. Fluorescent probes against 4.5S_H and 4.5S_I were used for detection and snRNA U2 served as a control. **C** Western blot analysis of the immunoprecipitation from (B). Antibodies against spliceosomal protein Prpf8 and Gapdh were used as positive and negative control, respectively.

Next, the association of spliceosomal proteins with 4.5S_H RNA discovered in the LC MS/MS-analysis was investigated by immunoprecipitation (IP) of small nuclear ribonucleoprotein polypeptide B (SmB) in more detail (Gray et al. 1999). Since SmB is known to be a core component of the spliceosome (Bacrot et al. 2015) this aimed at a more global picture of spliceosomal complexes by investigating interactions of this protein. Endogenous SmB was immunoprecipitated and the detection of bound RNAs was performed by Northern blot (Figure 23C). The 4.5S_H RNA was clearly detected in the input and the SmB-IP but was absent in the bead control similar to the U2 snRNA serving as a positive control for RNAs involved in splicing. These findings were consistent for both, the mouse MC57G and the hamster CHO-K1 cell lysates, indicating an involvement of 4.5S_H RNA in RNA splicing. The function of 4.5S_I RNA is evidently unrelated to splicing since it was not specifically precipitated with SmB in either of the cell lines. The successful IP of SmB was validated by Western blot analyses (Figure 23C) and co-precipitation of the spliceosomal factor Prpf8 (Wickramasinghe et al. 2015). Based on the previous findings, 4.5S_H RNA potentially interacts with splicing regulators, including several SRSF-family members (SR-proteins). To examine the interaction site between the 4.5S_H RNA and these SR-proteins a mutant form of the RNA was generated. This version of the RNA still contains the A- and B-box as well as the oligo-U terminator, but lacks the predicted single stranded loop region (see Figure 19). Then, wild type and mutant 4.5S_H RNA were *in vitro* transcribed using biotinylated UTP and an RNA pulldown was performed with B16-F10 cell lysate. The associated proteins were subsequently analyzed on a Western blot for each of the RNAs (Figure 24A).

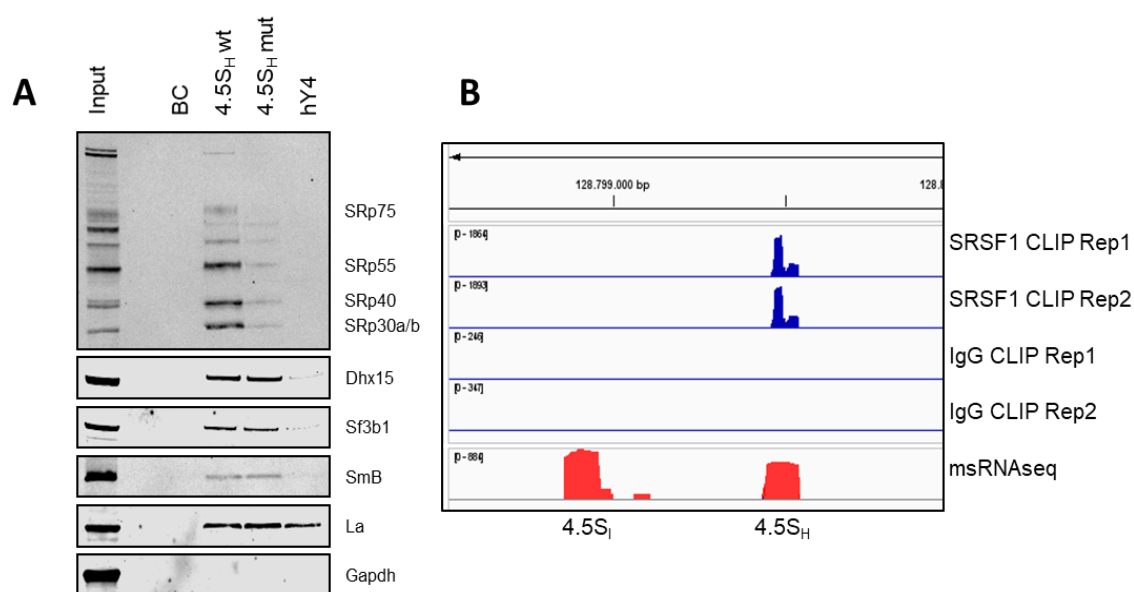


Figure 24. The 4.5S_H RNA is associated with SR-proteins and SRSF1. **A** Western Blot analysis of pull-down samples from B16-F10 mouse cells with 4.5S_H wt, 4.5S_H mut and hY4 RNAs as baits. hY4 RNA served as a specificity control. **B** Genome view of the mouse locus at (chr6:128,798,523-128,800,000). Publicly available CLIP datasets of the splicing factor SRSF1 (GSE141349) were combined with the msRNA-Seq datasets. Peaks highlighted in red indicate msRNA expression (from 5'

to 3': one 4.5S_I and 4.5S_H RNA gene). Peaks in blue indicate CLIP-signals for two independent SRSF1 replicates, IgG CLIP replicates were included as a specificity control.

First, the SRSF-proteins identified by the mass spectrometry analysis were validated by the use of an antibody recognizing phosphorylated serines specifically in RS-domains. According to the previous findings, several SR-proteins could be detected as seen by the band pattern in the top panel (SRp30a/b, SRp40, SRp55 and SRp75). Nonetheless, the association with the RNA is severely diminished when the RNA sequence does not contain the predicted single stranded loop as seen for 4.5S_H mut. This finding might suggest that this loop region serves as a potential interaction surface for the SR-proteins. Confirming our finding that 4.5S_H RNA is present after immunoprecipitation of splicing protein SmB in MC57G and CHO-K1, the RNA pulldown also shows association of SmB in B16-F10. Additionally, other splicing factors like U2 snRNP component Sf3b1 (Bertram et al. 2017) and spliceosomal ATP-dependent helicase Dhx15 (Martin et al. 2002) were pulled down effectively by 4.5S_H. In contrast to the SR-proteins, these factors were pulled down with either RNA regardless of the loop sequence. The human Y4 RNA was used as a negative control since it also folds into a predicted stem-loop structure but its function is most likely unrelated to splicing (Köhn et al. 2013). Accordingly, hY4 did not show significant association with splicing factors like 4.5S_H RNA, hinting towards specific binding partner detection. The La protein recognizes the oligo-U terminator of all three RNAs and was used as a general control for the RNA pulldowns.

Among the identified proteins interacting with the 4.5S_H RNA only the SR-proteins displayed a binding-dependency with the loop sequence. Interestingly, the GA-rich region resembles an RNA sequence previously described to be recognized by SRSF1 (Wang et al. 2011). To validate a potential SRSF1 - 4.5S_H interaction, publicly available mouse SRSF1 CLIP datasets were analyzed (GSE141349). The genomic regions with reported SRSF1 CLIP sites were compared to the msRNA-Seq expression (Figure 24B). The depicted genomic locus of the mouse contains a 4.5S_H RNA as well as a 4.5S_I RNA locus. However, SRSF1 CLIP peaks could only be detected at sites where 4.5S_H RNA transcription occurs.

These findings suggest a substantial interaction of 4.5S_H with splicing factors especially SR-proteins. Additionally, the SRSF1 protein could potentially directly associate with the RNA via the AG-rich single stranded loop region.

3.2.3 SRSF1 and its RS-Domain recruit SRSF-family members to the 4.5S_H RNA

The domain structure of the SRSF1 protein is organized in two ordered RRMs and an extended arginine-serine-repeat sequence with 50 amino acids in length. It has been suggested for the RS-domain to be relevant for correct localization of SRSF1 into splicing speckles as well as contributing to the splicing function of the protein in *in vitro* splicing assays (Cazalla et al. 2002; Shen et al. 2004, see Introduction). As part of this thesis, the interaction

of the RS-domain with other SR-proteins as well as the contribution of the repeat to recognition and binding of the 4.5S_H RNA in a cellular context was investigated. For this reason, SRSF1 full length protein and SRSF1 Δ RS (residues 1-193) were overexpressed in fusion with a streptavidin-binding peptide tag in B16-F10 cells. After lysis of the cells, a protein pulldown was performed with subsequent analysis of SRSF1 associated proteins and RNA on Western and Northern blot, respectively (Figure 25A and 25B).

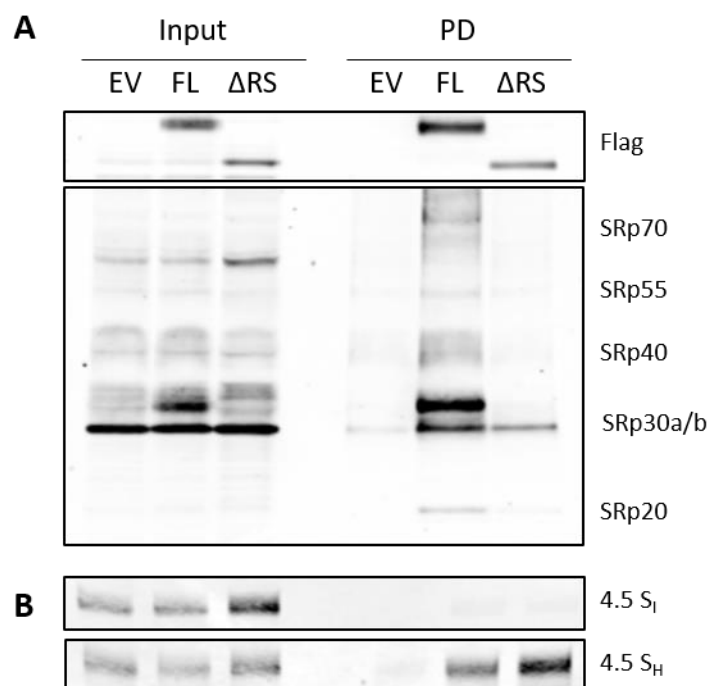


Figure 25. SRSF1 pulldown for associated RS-domain containing proteins and RNAs. A Western Blot analysis of a streptavidin pulldown with overexpressed SBP-tagged full length SRSF1 (FL), without RS-domain (Δ RS, residues 1-193) and empty vector as a control (EV) in B16-F10 cells. Western blot shows phosphorylated RS-repeats from several SRSF-proteins of indicated size and overexpressed proteins with the Flag antibody. **B** Northern blot analysis of associated RNAs from (A) to investigate RNA binding capabilities with full length SRSF1 and without RS-domain. Fluorescently labeled probes were used to detect 4.5S_H and 4.5S_I RNA, respectively.

As seen in Figure 25A, overexpression and pulldown of the full length SRSF1 protein (FL) leads to enrichment of several other RS-domain containing proteins. Notably, SRSF7 (SRp20), SRSF5 (SRp40), SRSF6 (SRp55), SRSF2 (SRp30b) and endogenous SRSF1 (SRp30a) can be pulled down with full length SRSF1, but are not significantly enriched in the pulldowns of SRSF1 Δ RS. It has been suggested, that the RS-domain is also required for facilitating protein-protein interactions between serine/arginine rich proteins (Wu and Maniatis 1993). Contrary to these findings, the SRSF1 Δ RS protein appears to associate with endogenous full length SRSF1, independently of its RS-domain. Despite limited protein-protein interactions, the Northern blot analysis reveals no significant change in RNA-binding capability for the SRSF1 Δ RS protein when compared to the input signal (Figure 25B). Based on these findings it is possible that association of other RS-domain containing proteins is not solely relying on an intact RS-domain and might rather depend on associated RNAs. This also implies that

association of endogenous SRSF1 in the SRSF1 Δ RS protein pull-down is mediated by RNA bound to the deletion variant. In order to verify, if the other SRSF-proteins detected in the pull-down interact with bound RNA or mainly via SBP-SRSF1 an RNA pull-down with 4.5S_H and B16-F10 cell lysate previously transfected with siRNAs against SRSF1 (siS1) and control siRNA (siC) was performed (Figure 26A).

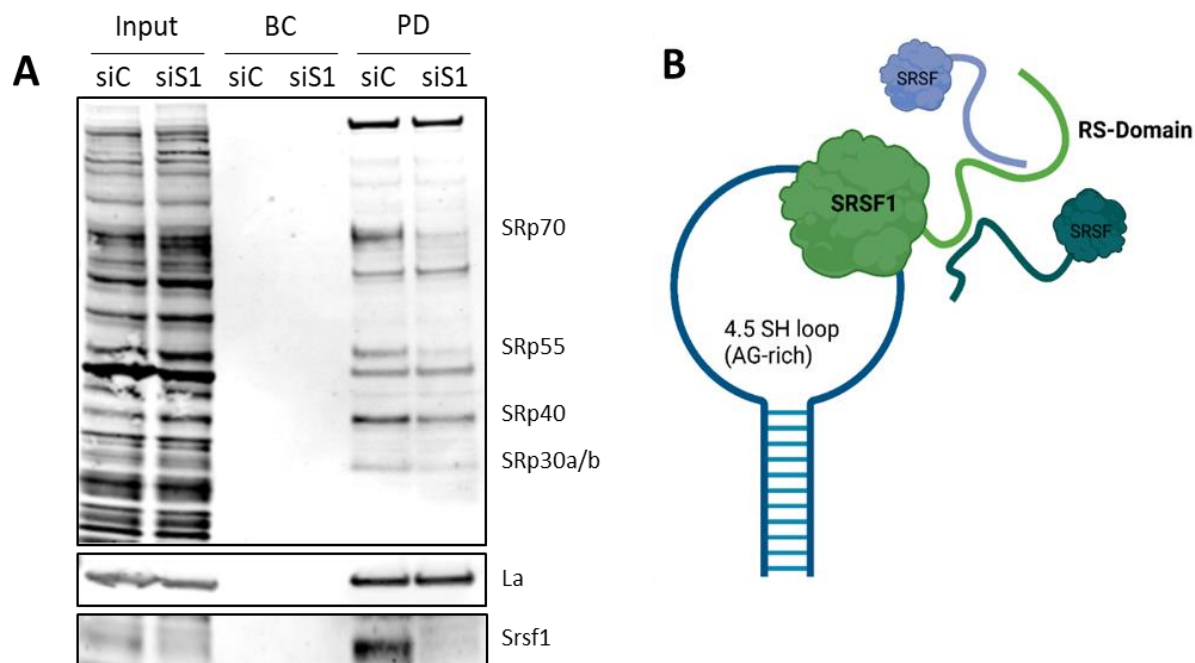


Figure 26. SRSF1 recruits SRSF-family members to the 4.5S_H RNA. **A** Western blot of an RNA pull-down with biotinylated *in vitro* transcribed 4.5S_H RNA and B16-F10 cell lysate. B16-F10 cells were transfected with siRNA against SRSF1 (siS1) and control siRNA (siC). Antibody against the La-protein indicates equal efficiency for the pull-down of the individual samples. **B** Binding model of SRSF1 to 4.5S_H RNA and association of other SR-proteins via the RS-domain. The binding of the 4.5S_H RNA is independent of the RS-domain of SRSF1.

As expected, association of several SR-proteins was clearly detected, including SRSF1 in the pull-down under control conditions (siC) in analogy to experiments performed in Figure 24A. However, the siRNA mediated knockdown of SRSF1 negatively affected the amount of SR-proteins present in the pull-down. According to the signal of the La protein which served as a control for equal protein and 4.5S_H RNA amounts in the pull-down condition, it can be inferred that the decrease in the other SR-proteins is a result of reduced interaction with the SRSF1 protein. This suggests that among the SR-proteins it is indeed SRSF1 that mainly associates with the 4.5S_H RNA and the interactions with the other family members is mediated by the RS-domain. The results were summarized in a schematic model of SRSF1 and 4.5S_H RNA binding and RS-domain dependent association of SRSF-proteins (Figure 26B). In summary, the 4.5S_H RNA recruits multiple SR-proteins and potentially serves as a scaffold for several proteins on one RNA molecule based on RS-RS domain interaction.

Besides the interaction with proteins, 4.5S_H RNA has been suggested to associate with RNA that contains an antisense sequence of the B1-SINE element and facilitate nuclear retention of the RNAs. This is due to the high similarity of B1-SINEs and the 4.5S_H RNA sequence (Ishida et al. 2015; Labuda and Zietkiewicz 1994). Among the RNAs identified in a BLAST analysis to contain an antisense B1 SINE element is the nuclear long non-coding RNA Malat1. To analyze a potential interaction between the 4.5S_H RNA and Malat1, RNA pulldowns with 4.5S_H and 4.5S_I RNA were performed and Malat1 levels were subsequently determined by qRT-PCR.

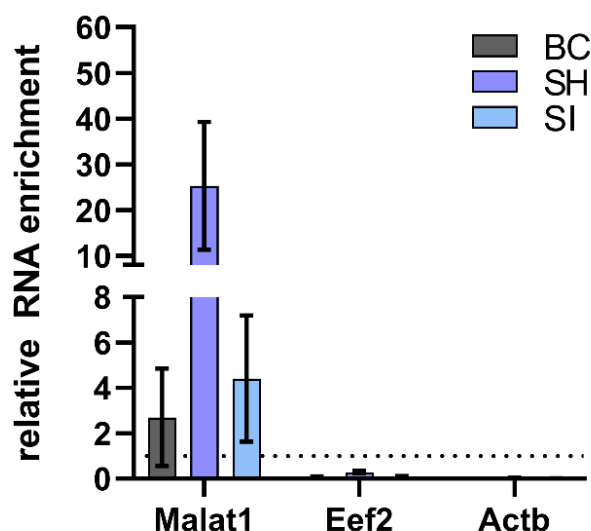


Figure 27. Quantitative real-time PCR of RNAs associated with 4.5S_H and 4.5S_I RNA. The pulldown was performed with total cell lysate from B16-F10 cells incubated with beads only (BC) and beads conjugated with biotinylated 4.5S_H and 4.5S_I RNA. Enrichment in the pulldown fractions has been determined compared to input RNA for each condition (n=3)

As shown in Figure 27, Malat1 RNA could be detected in all pulldown fractions (BC, 4.5S_I and 4.5S_H) with enrichment of the RNA compared to the input. However, the 4.5S_H RNA displayed the highest enrichment of Malat1 RNA with more than 20-fold in the pulldown fraction, indicating a rather specific interaction between the two RNAs. Similar to 4.5S_H, the Malat1 RNA has been reported to localize to splicing speckles and is bound by SRSF1 in several cancer entities (Gutschner et al. 2013; Hu et al. 2016; Malakar et al. 2017). Hence, Malat1 could be associated indirectly with 4.5S_H RNA via SRSF1 molecules attached to RS-domains according to our binding model (Figure 26B).

3.2.4 Malat1 and 4.5S_H RNA expression is linked in mouse cells

It has been shown, that human Malat1 participates in pre-mRNA splicing through the interaction with splicing factors like SRSF1 and potentially modulating the phosphorylation status of SR-proteins (Tripathi et al. 2010). Additionally, Malat1 levels are related to cancer cell

progression and metastasis (Arun et al. 2016). In contrast, expression changes of mouse Malat1 in transgenic mice did not result in observable phenotypes nor did it show an effect on mouse development (Zhang et al. 2012, see Introduction). The RNA sequence of the Malat1 transcript is largely conserved in mammals, but differs in transcript length in human (8779 nt) compared to mouse (6983 nt) with a 5'-extension of the human RNA (Hutchinson et al. 2007). In coherence with SRSF1 binding to Malat1, the two transcripts were screened for distinct SRSF1 binding sites in addition to their difference in length. For this purpose, publicly available SRSF1 HITS-CLIP datasets were analyzed in the 5'-end of the transcript (Figure 28A).

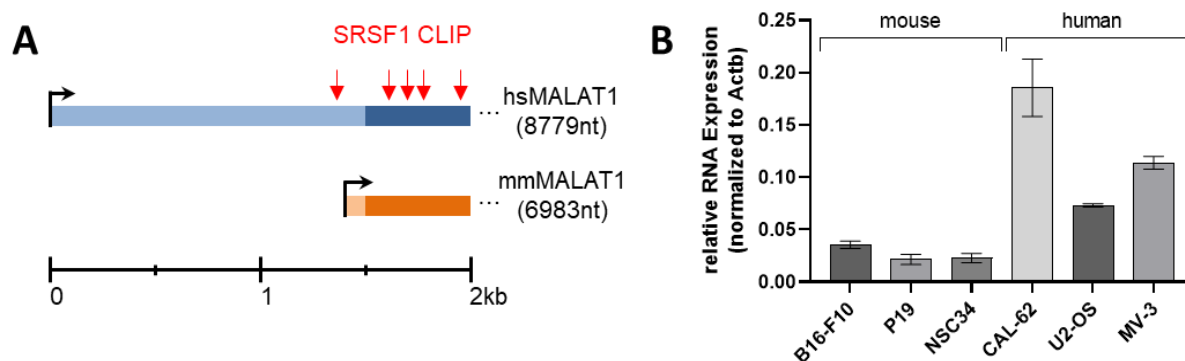


Figure 28. The Malat1 RNA in human and mouse. **A** The Malat1 transcript in human (hsMALAT1) and mouse (mmMALAT1) with reported SRSF1 HITS-CLIP sites indicated by red arrows at the 5'-end of the transcript. The hsMalat1 utilizes an alternative transcription start site (TSS) compared to the mouse transcript. **B** qRT-PCR analysis for Malat1 levels from isolated total RNA of the indicated cell lines.

However, the human Malat1 transcript contains only one additional endogenous SRSF1 CLIP site in the 5'-extension as analyzed in H1299 cells (GSE130867, (Fish et al. 2019)). Comparing the SRSF1 binding sites in human to CLIP datasets (GSE130867) revealed the majority of the interaction sites to be present in the conserved region o

f the Malat1 transcript. This suggests that there might be just minor differences in SRSF1 binding capabilities in terms of the 5'-extension based on the preserved binding sites in both transcripts. It was hypothesized that expression levels of human Malat1 are enhanced compared to the mouse transcript due to the alternative transcription start site, thereby providing additional SRSF1 binding sites overall. In order to verify this, Malat1 RNA expression was assessed in qRT-PCR analyses for three mouse (B16-F10, P19, NSC34) and human cancer cell lines (CAL-62, U2-OS, MV-3), respectively (Figure 28B). B16-F10 cells showed the highest expression of the RNA among the cell lines originated from mouse, while Malat1 was most abundant in the human CAL-62 cells. In line with the previous assumption, Malat1 RNA is on average expressed more abundantly by approximately four-fold in the analyzed human versus mouse cancer cells.

Based on the similarities of Malat1 RNA and 4.5S_H RNA in terms of localization and SRSF1 binding, it was questioned if the 4.5S_H RNA is related to Malat1-dependent functions

in rodents in respect to the differences in expression. For this purpose, a knockdown of 4.5S_H and Malat1 RNA levels using locked nucleic acids (LNAs) was performed in B16-F10 cells. Subsequently, expression changes of Malat1 and 4.5S_H were analyzed by Northern blot and qRT-PCR (Figure 29).

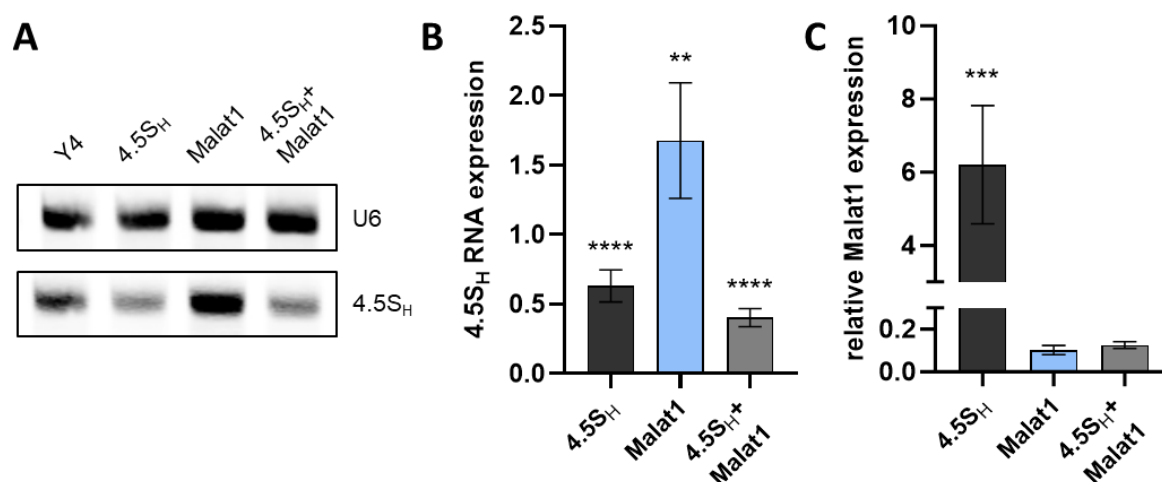


Figure 29. LNA mediated knockdown of 4.5S_H and Malat1 in B16-F10. **A** Northern Blot analysis of RNA from B16-F10 cells transiently transfected with the LNAs against the indicated transcripts. Fluorescently labeled probes were used against 4.5S_H and U6 snRNA as a control. The Y4 LNA is non-targeting in mouse and serves as a transfection control. **B** Quantification of 4.5S_H levels from (A) normalized to U6 expression and Y4 LNA treated cells. **C** qRT-PCR analysis of isolated total RNA determining relative Malat1 levels upon transient knockdown with LNAs from (A). Significance was calculated with a two-way Anova compared to Y4 control; ** p < 0.01, *** p < 0.001, **** - p < 0.0001

Control knockdowns were carried out transfecting an LNA against human Y4 RNA which is not present in the murine background. Notably, LNA-mediated knockdown of 4.5S_H RNA only results in a decrease of the RNA to 60% compared to the amounts in control knockdowns as quantified from the Northern blot in all experiments (Figure 29A and 29B). Interestingly, reduction of Malat1 RNA leads to a significant increase of 4.5S_H RNA (up to 1.5-fold). Similarly, decreasing 4.5S_H RNA in B16-F10 elevated Malat1 RNA levels up to 6-fold compared to Y4 LNA transfected cells, indicating a putative compensatory regulation between these two RNAs (Figure 29C). It could also be determined that 4.5S_H RNA levels were reduced more drastically by the combination of both LNAs as opposed to single LNA transfection (Figure 29B) which also hints towards transcript level compensation.

3.2.5 Malat1 and 4.5S_H RNA contribute to splicing speckle integrity

The Malat1 RNA has been shown to modulate localization of nuclear speckle components such as splicing factors (e.g. SRSF1, SF3a60 and U2AF-65) upon knockdown in HeLa cells (Tripathi et al. 2010). In contrast to this, an altered localization of splicing factors was not observed in isolated MEFs of Malat1 knock-out mice (Nakagawa et al. 2012). The

previous findings suggest similar cellular properties of Malat1 and 4.5S_H RNA in mouse cells hence a compensatory effect for 4.5S_H could potentially rescue Malat1 phenotypes observed in human cells. Along this line, splicing speckle behavior based on localization of splicing speckle marker SC-35 was investigated upon Malat1 and 4.5S_H RNA knockdown in a comparable analysis in human U2-OS and mouse B16-F10 cells (Figure 30 and 31)

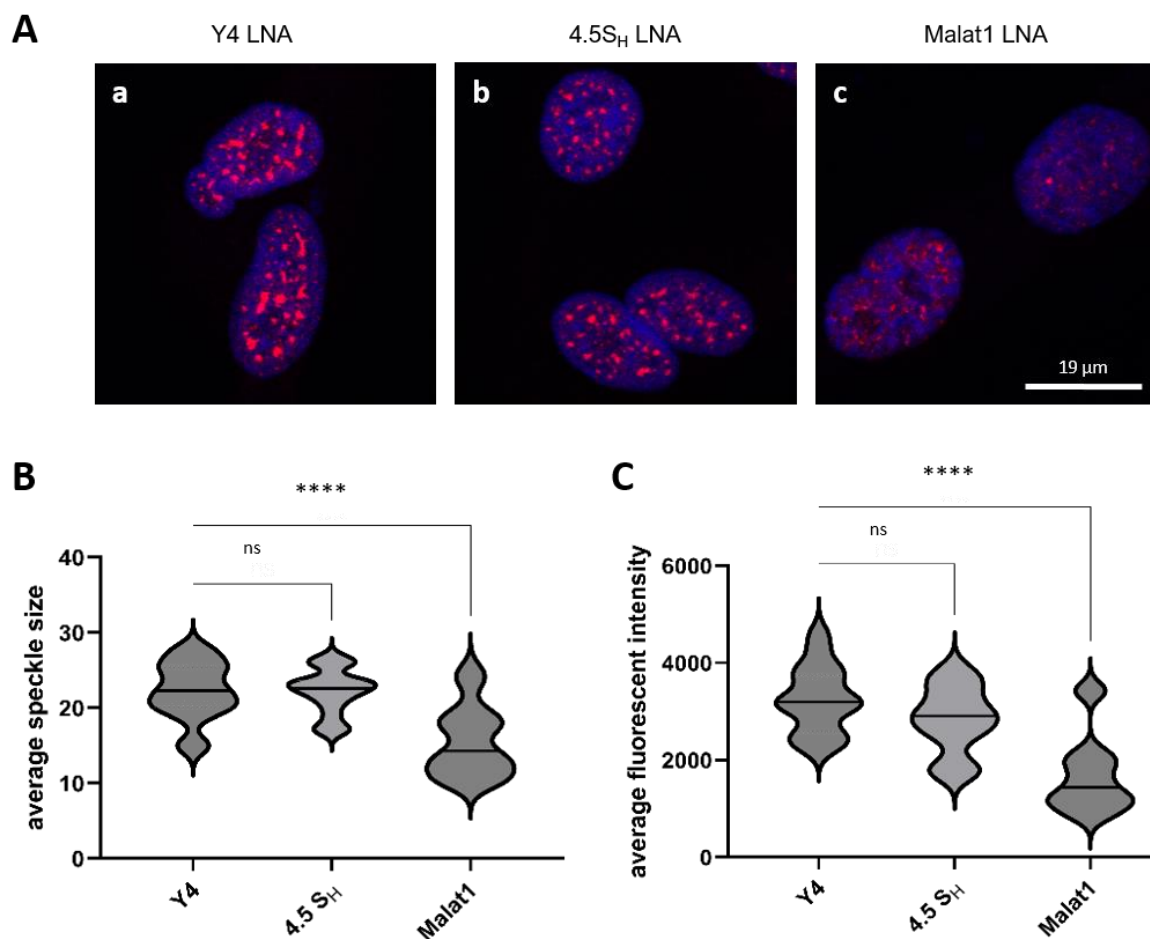


Figure 30. Splicing speckle integrity in human U2-OS cells upon ncRNA modulation. A Immunofluorescence staining of splicing speckles in U2-OS cells upon knockdown of Y4 (a), 4.5S_H (b) and Malat1 (c) using specific LNAs. Splicing speckle marker SC-35 is highlighted in red and DAPI in blue. **B** Average splicing speckle size and fluorescence intensity (**C**) of nuclear particles detected in human U2-OS cells 24 hours post transfection (Y4 N = 18, 4.5S_H N = 19, Malat1 N = 23). Significance was calculated with a two-way Anova compared to Y4 control; **** - p < 0.0001

Immunostainings have been carried out with an antibody against splicing protein SC-35 in three sets of U2-OS cells transfected with LNAs against Y4 (a), 4.5S_H RNA (b) and Malat1 (c). Microscopy images displayed in Figure 30A have been recorded as maximum stack projection to capture splicing speckles in several focal planes. In order to determine splicing speckle integrity, two parameters were analyzed using the 2D nuclear particle detector from the MiToBo toolkit based on fluorescence signal and average fluorescence intensity (Figure 30B and C). The knockdown of the cytoplasmic Y4 RNA (a) and transfection of the non-targeting

4.5S_H LNA (b) showed no abnormal localization of SC-35 outside of the expected splicing speckles. In contrast, LNA transfection against Malat1 RNA displayed reduced condensation of SC-35 (c). This discrepancy can also be assessed in significantly decreased average speckle size and fluorescence intensity compared to Y4 or 4.5S_H LNA transfection, suggesting a role for Malat1 in splicing speckle integrity in human U2-OS cells. Next, splicing speckles morphology was analyzed by immunofluorescence (Figure 31A) in mouse B16-F10 transfected with non-targeting LNA against Y4 RNA as control (a) and LNA mediated knockdown of 4.5S_H (b), Malat1 (c) and the combination of both (d).

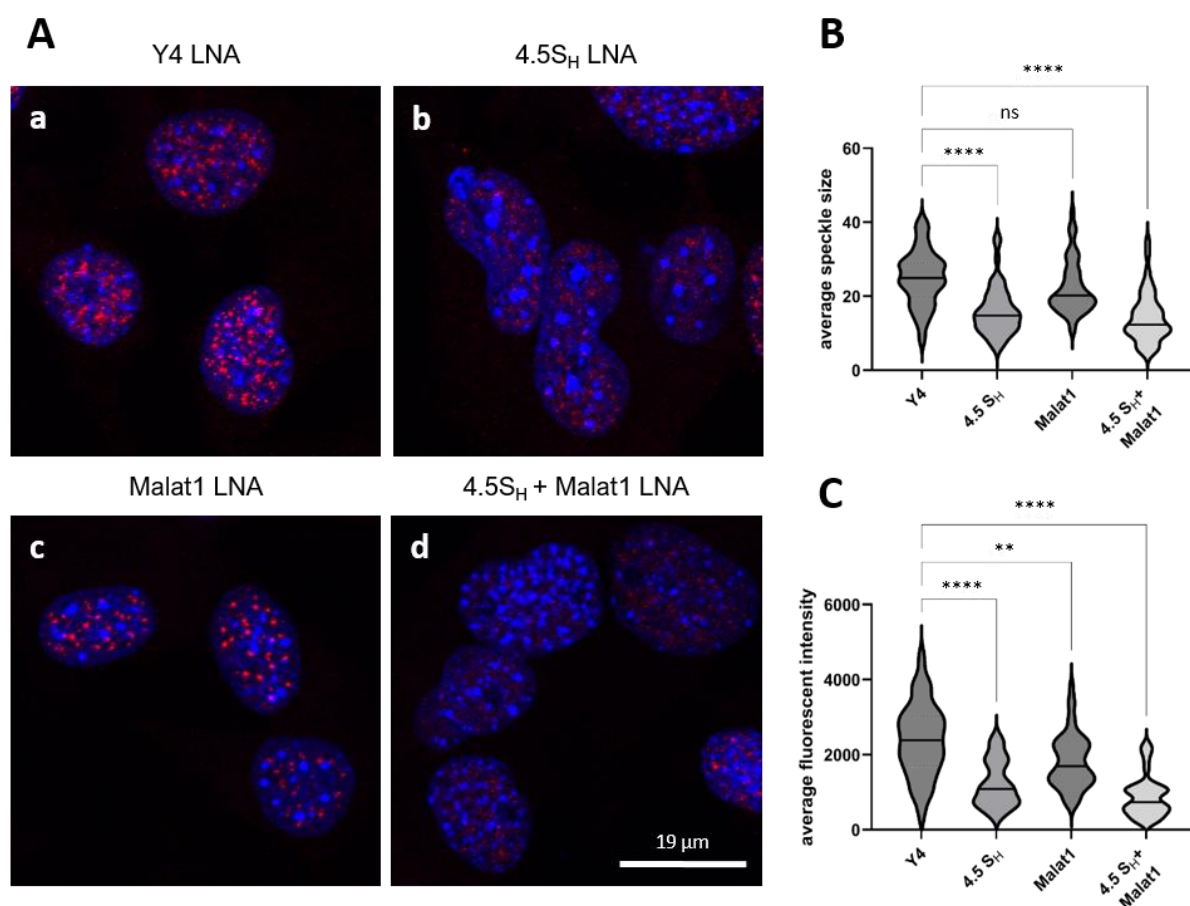


Figure 31. Splicing speckle integrity in mouse B16-F10 cells upon ncRNA modulation. A

Immunofluorescence staining of splicing speckles in B16-F10 cells upon knockdown of Y4 (a), 4.5S_H (b), Malat1 (c) and 4.5S_H + Malat1 using specific LNAs. Splicing speckle marker SC-35 is highlighted in red and DAPI in blue. **B** Average splicing speckle size and fluorescence intensity (**C**) of nuclear particles detected in mouse B16-F10 cells 24 hours post transfection (Y4 N = 33, 4.5S_H N = 46, Malat1 N = 51, 4.5S_H + Malat1 N = 44). Significance was calculated with a two-way Anova compared to Y4 control; ** - p < 0.01 **** - p < 0.0001

In accordance with the proposed hypothesis, depletion of 4.5S_H RNA led to a visible dispersion of SC-35 protein in the nucleus with a significant reduction of speckle size and also fluorescence intensity (Figure 31B and 31C) in line with the observations made for Malat1 knockdowns in U2-OS cells (Figure 30A). It has been reported, that Malat1 knockout did not lead to perturbation of splicing speckles (Nakagawa et al. 2012). Nonetheless, it appears that

transfection of B16-F10 with LNA targeting Malat1 resulted in more condensed nuclear speckles with detectable effect on fluorescence intensity, but no significant changes in average splicing speckle size. As shown in Figure 29C, decreasing Malat1 RNA levels elevated 4.5S_H RNA expression which might suggest that this RNA is required for speckle integrity in mouse and effects observed in U2-OS could be partially compensated by 4.5S_H RNA. Furthermore, combined knockdown of 4.5S_H RNA and Malat1 revealed a more drastic effect on splicing speckles as they appeared almost completely absent compared to the image of cells transfected with Y4 control LNA. Likewise, average speckle size and average fluorescence intensity was reduced even further compared to cells transfected with 4.5S_H LNA alone. In summary, these results indicate a crucial role for 4.5S_H RNA in B16-F10 cells to maintain integrity of splicing speckles in cooperation with Malat1 RNA.

The depletion of Malat1 in human cells has been shown to be associated with decreased recruitment of SR-proteins to active transcription sites and thereby affecting pre-mRNA splicing (Bernard et al. 2010). However, depletion of Malat1 RNA in mouse resulted in a decreased expression of Neat1 RNA, but did not result in changed localization of SRSF1 and SRSF2. Moreover, alterations of pre-mRNA splicing have not been investigated in this context (Nakagawa et al. 2012). To further characterize the effect of altered splicing speckle integrity observed in B16-F10 mouse cells, accumulation of pre-mRNA was investigated based on LNA-mediated knockdowns of 4.5S_H and Malat1 RNA (Figure 32).

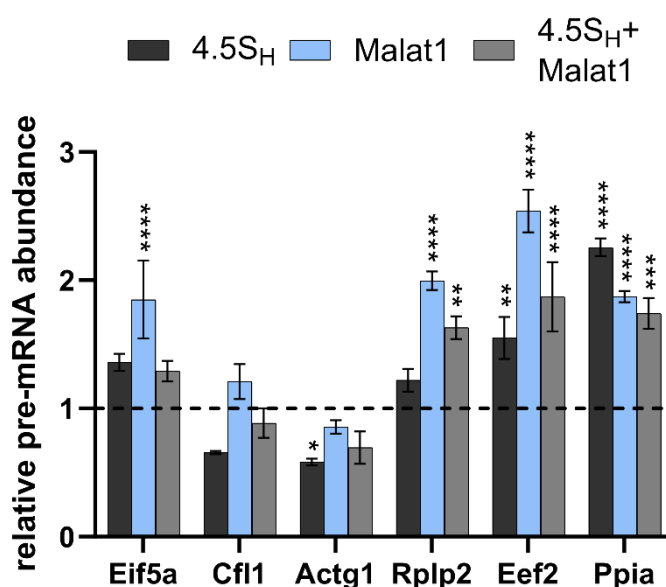


Figure 32. Analysis of pre-mRNA splicing upon modulation of 4.5S_H and Malat1. The quantity of pre-mRNA and spliced transcript was determined with qRT-PCR analysis and normalized to the ratio of pre-mRNA to spliced transcript in Y4 LNA transfected control cells (B16-F10). Error bars indicate SEM. Significance was calculated with a two-way Anova compared to Y4 control; * - $p < 0.05$ ** - $p < 0.01$ **** - $p < 0.0001$

The process of splicing and removal of intronic sequences by the spliceosome is a highly efficient process leading to rapid decay of spliced out introns and thereby low abundance of

measurable intron-containing pre-mRNAs. Due to this, highly abundant transcripts have been chosen here to be able to detect pre-mRNAs and associated based on protein abundance reported on PAXdb - protein abundance database (Wang et al. 2015). The pre-mRNA amount was determined by qRT-PCR with the reverse oligo binding inside the first intron sequence of the transcript. In addition to modulation of splicing speckle integrity (Figure 31), knockdown of 4.5S_H and Malat1 RNA also resulted in an overall increase in pre-mRNA transcripts in B16-F10 cells with exception of cofilin (Cfl1) and actin gamma 1 (Actg1). The depletion of Malat1 RNA led to a significant increase of pre-mRNAs for eukaryotic translation initiation factor 5a (Eif5a), ribosomal protein lateral stalk subunit P2 (Rplp2), eukaryotic translation elongation factor 2 (Eef2) and peptidylprolyl isomerase A (Ppia). Furthermore, knockdown of 4.5S_H RNA also increased pre-mRNA amount of these transcripts indicating an effect on constitutive splicing with reduction of Malat1 and 4.5S_H RNA. However, combined knockdown of Malat1 and 4.5S_H RNA did not result in an increased accumulation of pre-mRNAs when compared to Malat1 knockdown alone, indicating that the splicing process is less affected. Nonetheless, combined transfection of LNAs targeting 4.5S_H and Malat1 resulted in significantly more pre-mRNA transcripts for Rplp2 and Eef2 in contrast to B16-F10 cells with 4.5S_H knockdown. Notably, Ppia pre-mRNA displayed the highest abundance in cells with 4.5S_H knockdown, while Cfl1 and Actg1 pre-mRNAs appeared to be unaffected in terms of increased abundance due to potential modulation of the splicing process as seen for the other transcripts. This analysis revealed an important aspect of mouse Malat1 and 4.5S_H to modulate splicing of pre-mRNAs to a certain extent. In addition to altered splicing speckle integrity this highlights the relevance of Malat1 but also the 4.5S_H RNA for normal splicing function in mouse cells.

3.3 Identification of novel Rbfox1 splicing targets

3.3.1 Rbfox1 regulates alternative splicing of focal adhesion genes

The results presented in this thesis focused on the identification and initial characterization of medium-sized non-coding RNAs as part of an optimization of RNA-sequencing protocols. Additionally, the msRNA 4.5S_H RNA was characterized and has been identified to be relevant for splicing speckle integrity and interaction with splicing proteins of the SR-family, especially SRSF1. In respect to its splicing function, SRSF1 is involved in constitutive splicing as well as alternative splicing and has been described as an exonic splicing enhancer (Das and Krainer 2014). RNA splicing is an essential cellular process that leads to the generation of protein variants from one single gene through post-transcriptional processing. The formation of alternatively spliced transcripts relies on inclusion or exclusion of exons in which *trans*-acting splicing proteins are involved. As a result, alternative splicing proceeds to generate a huge variety of RNAs encoding for proteins to facilitate complex

functions in highly-specialized tissues like brain, heart and skeletal muscle (Baralle and Giudice 2017; Pedrotti et al. 2015; Weyn-Vanhentenryck et al. 2018). A relevant group of muscle and neuronal specific alternative splicing factors is the Rbfox-family (consisting of Rbfox1, Rbfox2 and Rbfox3) which contribute to tissue development and homeostasis through tissue-specific alternative splicing (Conboy 2017, see Introduction).

In the following part of the thesis, Rbfox1 was investigated as an alternative splicing mediator in order to determine regulated genes under the premise of heart diseases. To elucidate Rbfox1 contribution for the development of heart pathologies the expression of Rbfox1 was analyzed in publicly available RNA-sSequencing datasets (Figure 33). These datasets include mouse models with induced cardiac dysfunction by transgenic knockout of EGFR (EGFR-KO) and mechanical experimental procedures (myocardial infarction – MI, transverse aortic constriction – TAC, GSE96561) (Binas et al. 2020; Schreier et al. 2013).

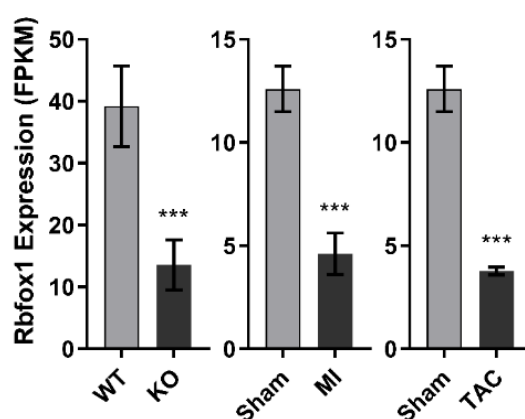


Figure 33. RNA Expression of Rbfox1 in RNA-Seq datasets. Expression is analysed from EGFR-KO, myocardial infarction (MI) and transverse aortic constriction (TAC) mouse models. Sham animals were subjected to surgery, but not treated otherwise.

In line with previously published results including other *in vivo* models (Frese et al. 2015; Park et al. 2011), Rbfox1 was significantly reduced in the analyzed mouse models with impaired heart function. Assuming that alternatively spliced transcripts regulated by Rbfox1 were affected under these conditions, the isoforms might contribute to compromised heart function. These putative Rbfox1 targets were investigated in further detail. In contrast to SRSF1 which is a splicing protein associated in exonic regions of the RNA, Rbfox1 has been reported to bind in intronic sequences of target transcripts. Additionally, it has been proposed that Rbfox1 specifically recognizes GCATG motifs and multiple occurrences within one intron might have synergistic effects on Rbfox1 mediated splicing (Conboy 2017). For the purpose of identifying putative Rbfox1 targets an *in silico* analysis was performed. The canonical Rbfox1 binding motif was screened in intronic regions of the whole mouse genome and compared to transcripts actually expressed in the mouse heart (Figure 34A). Notably, a large proportion of

these genes contain putative Rbfox1 binding sites in multiple introns with some transcripts containing more than 200 throughout all introns. It is worth noting, that the binding sites were not normalized to transcript length and number of introns. However, the target gene with the most abundant putative Rbfox1 binding sites has been reported for the Rbfox1 gene itself, which is in line with previously published data that suggest an auto-regulatory splicing mechanism for Rbfox1 (Damianov and Black 2010).

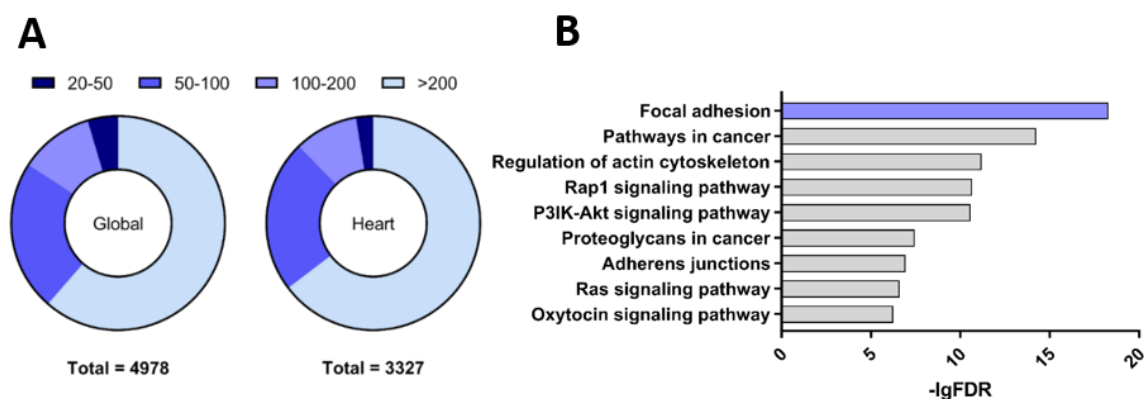


Figure 34. Computational analysis of putative Rbfox1 target genes. **A** *in silico* prediction of genes that contain more than 20 putative intronic Rbfox1 binding sites (GCATG) either in the whole mouse or in mouse hearts. Genes with more than 20 Rbfox1 binding sites were classified in four categories with different color coding. The total number of genes containing more than 20 Rbfox1 binding sites is indicated below. **B** DAVID functional clustering of identified heart transcripts from (A)

Next, identified putative Rbfox1 targets in mouse hearts (from Figure 34A) were sorted in functional groups using the DAVID functional annotation clustering tool (Huang et al. 2009). This step was necessary to reveal Rbfox1-regulated biological pathways with potential impact on heart functionality. According to the analysis, the genes in the focal adhesion cluster correlated the highest with the occurrence of intronic Rbfox1 binding sites in the transcripts (Figure 34B). Among the total number of reported genes in the focal adhesion cluster 31% were reported to localize in the actual focal adhesion sites directly. These included e.g. Paxillin (Pxn) and Vinculin (Vcl) as potentially regulated alternatively spliced transcripts through Rbfox1. Due to the uncharacterized interaction of Rbfox1 with these mRNA transcripts, the next experiments focused on investigating the splicing regulation in more detail. For this work, rat myoblasts H9C2 have been used as cell model as they can be differentiated to cells resembling developing skeletal muscle (SM) and cardiac muscle (CM) based on their phenotype and expression profile (Figure 35A). Western blot analysis of differentiated H9C2 cells revealed increased amounts of Rbfox1 protein compared to undifferentiated control cells (C). Notably, Rbfox1 expression is barely detectable in undifferentiated H9C2 cells. The analysis of Pxn and Vcl under these conditions showed atypical bands beside the detected proteins at their expected molecular weight. Based on the increased molecular weight compared to Vinculin, the additional band might represent the alternatively spliced meta

Vinculin isoform (mVcl) that has been reported previously (Siliciano and Craig 1982; Kanoldt et al. 2020). In contrast, the molecular weight for the detected isoform with the Paxillin antibody could not be referred to any characterized Paxillin isoforms. Searching transcript databases (Ensembl and NCBI) an alternative Pxn variant was reported matching the molecular weight estimated from the detected band on the Western blot that was termed ePxn (extended Paxillin) in this thesis.

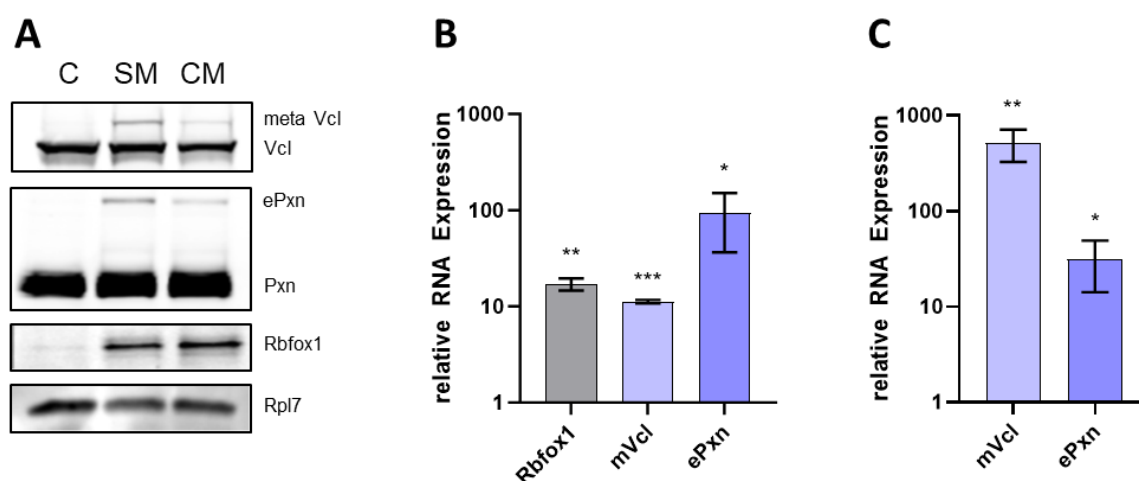


Figure 35. Rbfox1 controls alternative splicing of focal adhesion genes Vinculin and Paxillin. A Western Blot analysis from differentiated H9C2 cells with indicated antibodies. C – undifferentiated cells, SM – skeletal muscle cells and CM – cardiomyocyte precursor cells. **B** qRT-PCR was performed of CM versus C samples using specific oligonucleotides for indicated mRNAs. **C** qRT-PCR analysis of mVcl and ePxn after 48 h transient overexpression of Rbfox1 in H9C2 cells (n=3, significance was calculated with multiple unpaired t-test *- $p < 0,05$; ** - $p < 0,01$; ***- $p < 0,001$)

This variant contains potentially four additional exons (extensive exon 7 and three smaller exons – 8,9 and 10) which almost doubled the molecular weight of Pxn (60.8 kDa) to an extended 116.3 kDa of the uncharacterized Paxillin isoform that was named ePxn. Since it was necessary to validate that these isoforms were generated through alternative splicing of the transcript, the specific alternatively expressed exons were analyzed in a qRT-PCR comparing CM versus Control (Figure 35B). According to these results, transcripts containing the mVcl exon and exon 7 of ePxn were clearly upregulated upon differentiation of H9C2 cells into cardiomyocyte precursor cells. The upregulation of the alternative splice variants was accompanied by an increase in Rbfox1 RNA. In order to verify if the enhanced expression of mVcl and ePxn are connected to Rbfox1 expression, H9C2 cells were transfected with a plasmid encoding Rbfox1 CDS (Figure 35C). Consistent with the previous findings, exogenous Rbfox1 upregulation indeed led to a significant increase in exon inclusion of mVcl and ePxn transcripts thereby implying an Rbfox1-mediated splicing effect on these two target genes in differentiated cardiac myoblasts.

3.3.2 Alternative splicing of mVcl and ePxn is mediated by Rbfox1

The Metavinculin transcript contains an additional exon 19 that is alternatively spliced in a tissue-specific context in cardiac, skeletal and smooth muscle cells (Feramisco et al. 1982; Belkin et al. 1988; Kim et al. 2016). The exon 19 encodes for additional 68 amino acids in the C-terminal part of the vinculin protein modulating the function in the focal adhesion complexes (Tolbert et al. 2013; Oldfield and Dunker 2014). The Vinculin gene was reported in the initial screening of putative Rbfox1 target genes and further analyses revealed several Rbfox1 binding motif sequences flanking the alternatively spliced exon 19. In order to investigate Rbfox1 mediated splicing of this exon through these potential Rbfox1 binding sites, a series of minigenes were generated (Figure 36).

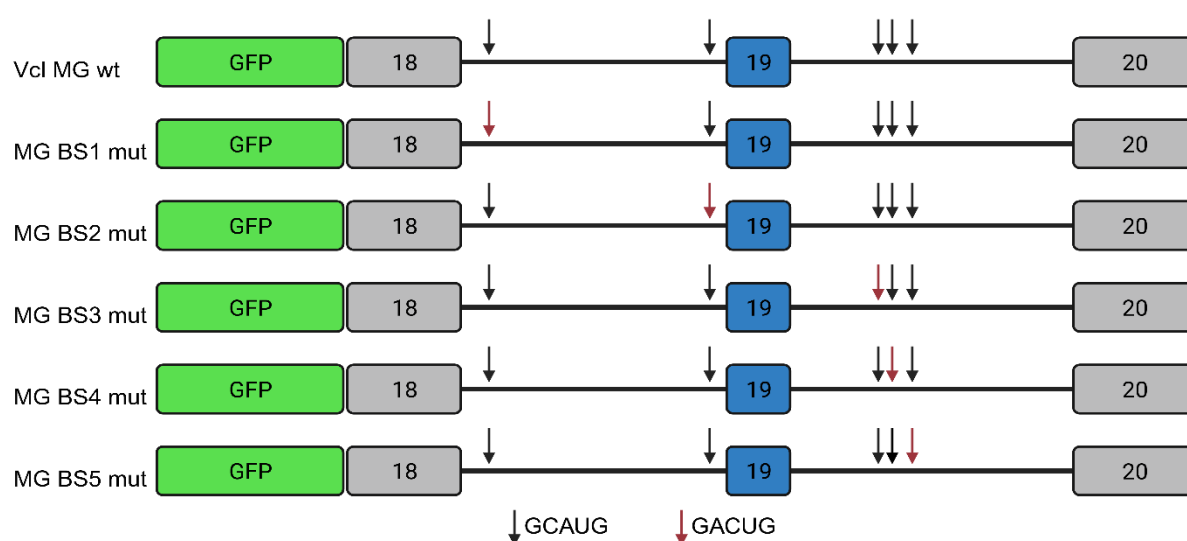


Figure 36. Vcl Minigene constructs in pEGFP-C1 vectors containing Vcl exon 18-20. The alternatively spliced exon 19 is depicted in blue and consecutive exons in gray. The arrows indicate putative Rbfox1 binding sites based on the consensus motif GCAUG with 2 binding sites upstream and 3 downstream of exon 19. Binding site mutations indicated by red arrows introduced a base switch to GACUG for each putative Rbfox1 motif in a separate minigene construct.

These constructs are comprised of exons 18 to 20 from the vinculin gene with the corresponding introns featuring two putative Rbfox1 binding sites upstream and three downstream of exon 19. Notably, only binding site 1 (BS1) contains the full UGCAUG sequence reported to provide the highest affinity for Rbfox1 binding (Conboy 2017). However, the GCAUG sequence has been shown to be sufficient for Rbfox1 mediated alternative splicing (Jin et al. 2003). In order to test the contribution of each putative Rbfox1 binding site on exon 19 inclusion, the binding sites were mutated individually resulting in five additional constructs with four complete GCAUG motifs each and one replaced by the GACUG sequence. Due to the low expression of Rbfox1 in undifferentiated cells, the minigene constructs were co-transfected with an additional plasmid either encoding Rbfox1 protein or the empty vector control. Subsequently, the transcript abundances of the isoforms originating from the

minigenes were determined by qRT-PCR analyses (Figure 37). Upon transfection of the wild type Vcl MG construct together with Rbfox1 in C2C12 cells, a substantial increase in exon 19 containing transcripts could be detected when compared to the empty vector conditions (Figure 37A). Consistent with the observed increase in endogenous mVcl levels upon Rbfox1 overexpression (Figure 35C), this result indicates the alternative splicing of this isoform to be promoted by Rbfox1 protein. In contrast, mutation of the individual binding sites resulted in an overall significant reduction of transcripts containing the alternatively spliced exon 19 (Figure 37B). Thus, each putative Rbfox1 binding site contributes to the generation of the mVcl

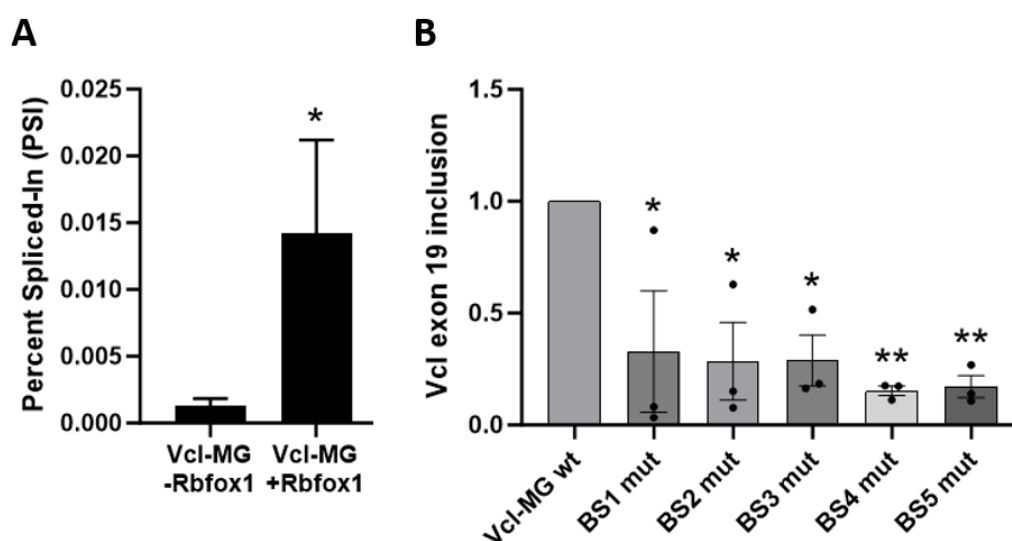


Figure 37. qRT-PCR analyses determining alternatively spliced transcripts from the mVcl minigene constructs. A PSI value of exon 19 containing transcripts compared to transcripts without exon 19 resulting from the MG wt minigene after co-transfection with EV and Rbfox1. **B** Fraction of transcripts including exon 19 as determined in (A) with normalization to the MG wt minigene for each mutated minigene construct. PSI values used for the calculation were determined in Rbfox1 overexpressing C2C12 cells. (n=3 significance was calculated with a one-sided student's t-test for figure A and multiple comparison one-way ANOVA for B. * - $p < 0,05$; ** - $p < 0,01$. Error bars display SEM

transcript, as any mutation inhibits exon 19 inclusion compared to the wild type. The mutation of binding site 4 and 5 appear to be the most crucial for Rbfox1 mediated splicing with an average reduction to 0.15 (BS4 mut) and 0.17 (BS5 mut) of the wild type minigene (MG wt). These results are in line with previous reports claiming distal Rbfox1 motifs (>500 nucleotides distance to any exon) can facilitate important alternative splicing reactions (Lovci et al. 2013). Additionally, it has been implied that Rbfox1 promotes exon inclusion with Rbfox1 binding sites positioned downstream of the alternatively spliced exon (Sun et al. 2012). However, the point mutations introduced to binding site 1 and 2 also decreased mVcl levels from the minigene (BS1 mut - 0.32; BS2 mut - 0.28) thereby decreasing exon inclusion in the same binding site mutant 3 downstream of exon 19 (BS3 mut - 0.28). Taken together these findings suggest mVcl exon 19 inclusion to be mediated by Rbfox1 through the consensus binding sites flanking the alternatively spliced exon. Furthermore, each individual binding site is potentially required

for enhanced inclusion, indicating that Rbfox1 recognizes and associates with each binding site upon mVcl splicing.

3.3.3 Alternative splicing of ePxn is mediated by Rbfox1

Paxillin is an important adaptor protein in focal adhesion complexes that contains highly flexible regions which are considered intrinsically disordered (Brown and Turner 2004; Neerathilingam et al. 2016). The alternatively spliced exons included in the ePxn variant provide a substantial amino acid extension of the protein, which could impact the function and folding of the protein. In order to visualize these changes the structure of the isoform was investigated using the AlphaFold Protein Structure Database (Jumper et al. 2021). The prediction of the folding behavior of ePxn revealed the additional exons to be largely disordered (Figure 38A, highlighted in red).

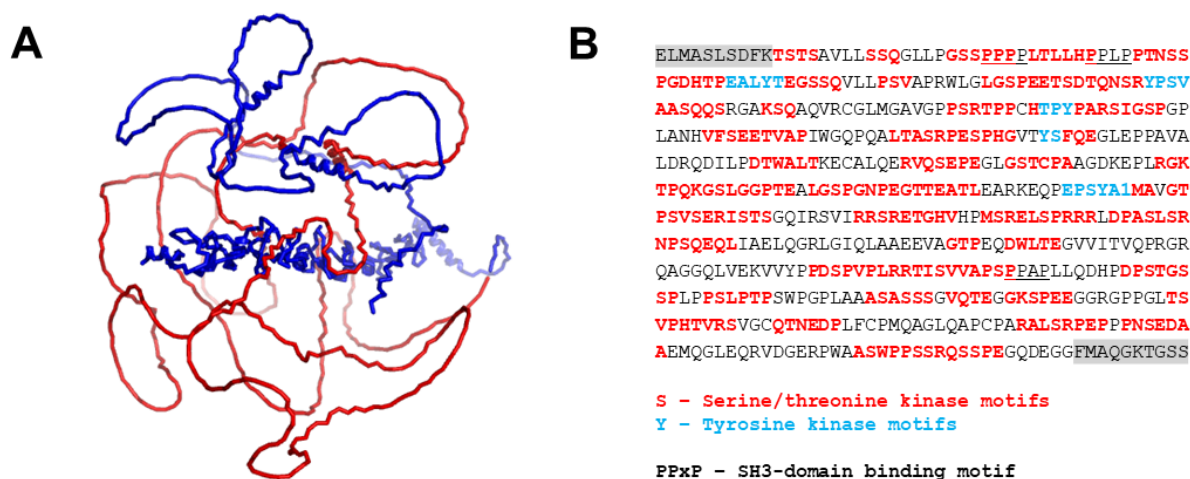


Figure 38. The ePxn isoform provides a large and highly disordered amino acid extension. A The structure of ePxn insertion in red and constitutive exons of Pxn visualized in PyMol (alphafold models: F5GZ78 for normal Pxn, A0A1B0GTU4 for ePxn variant). **B** Analysis of the amino acid sequence of the insertion in regard to kinase motifs with flanking exons 6 and 11 in gray.

Moreover, the additional exons translate into amino acids which show potential regulatory features as the sequence is comprised of multiple predicted serine/threonine and tyrosine kinase motifs in combination with SH3-domain binding motifs of SRC family kinases (Ingley 2008). Given the tissue specific expression of ePxn (high expression in skeletal and heart muscle, data not shown) in contrast to the ubiquitous Pxn isoform, these predictions imply a high regulatory capacity of this transcript in focal adhesion complexes potentially modulating their function. The upregulation of this isoform coincides with increasing Rbfox1 levels during the differentiation process of H9C2 cells and is induced upon exogenous Rbfox1 overexpression (Figure 35B and 35C). To further investigate the regulation of this alternatively spliced isoform by Rbfox1, minigenes were generated in a similar fashion to the mVcl minigenes (Figure 37). In contrast to the mVcl minigene, the ePxn transcript contains an exon

cassette composed of four exons of varying length with the exon 7 sequence to be longer than the other exons combined (831 nucleotides). Additionally, four out of five putative Rbfox1 binding sites were identified in the flanking introns of exon 7 and one GCAUG downstream of exon 10 (Figure 39). For technical reasons, the minigene analyses were focused on exon 7 as a representative exon of the whole cassette.

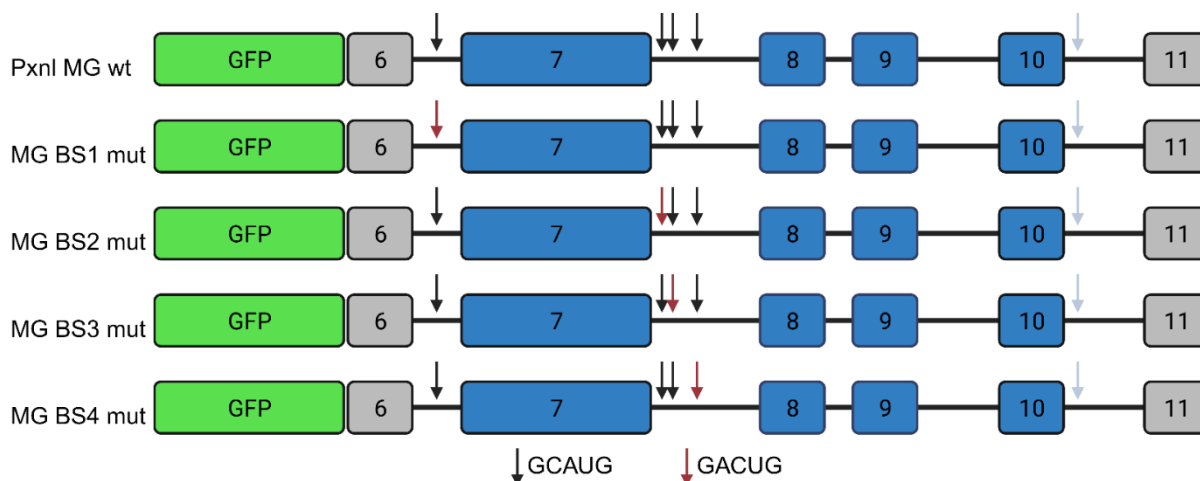


Figure 39. Pxn Minigene constructs in pEGFP-C1 vectors containing Pxn exon 6 to 11. The alternatively spliced cassette (exon 7-10) in blue and constitutive exons in gray. The arrows indicate putative Rbfox1 binding sites based on the consensus motif GCAUG with one binding site upstream of exon 7 and three downstream in addition to a single GCAUG downstream of exon 10. Binding site mutations indicated by red arrows introduced a base switch to GACUG for each putative Rbfox1 motif in a separate minigene construct.

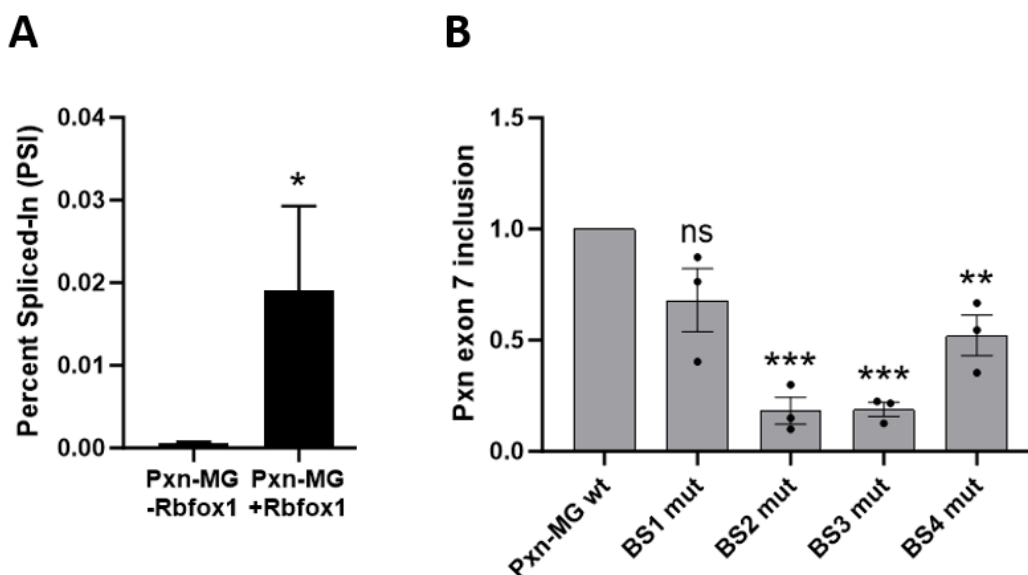


Figure 40. qRT-PCR analyses determining alternatively spliced transcripts from the ePxn minigene constructs. **A** PSI value of exon 7 containing transcripts compared to transcripts without exon 7 resulting from the MG wt minigene after co-transfection with EV and Rbfox1. **B** Percentage of transcripts including exon 7 as determined in **(A)** with normalization to the MG wt minigene for each mutated minigene construct. PSI values used for the calculation were determined in Rbfox1 overexpressing C2C12 cells. (n=3 significance was calculated with a one-sided student's t-test for figure

A and multiple comparison one-way ANOVA for B. * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. Error bars display SEM

Consistent with the analyses of the mVcl minigenes, the ePxn constructs were co-transfected with Rbfox1 into C2C12 cells to ensure proper Rbfox1 expression. Subsequently, the ratio of the splice variants with modulated exon 7 inclusion were determined by qRT-PCR. In line with the observations made for exon 19 inclusion in the mVcl minigene, the overexpression of Rbfox1 resulted in a significant increase in exon 7 usage (Figure 40A) hinting towards a possible regulation of this isoform by the Rbfox1 protein. Furthermore, the mutation of the putative Rbfox1 binding sites in the minigenes led to a noticeable reduction of transcript levels which include exon 7 (Figure 40B). These findings imply a shared contribution to the exon 7 inclusion for all the Rbfox1 binding sites, however, the effect on the alternative splicing is most prominent upon mutation of binding site 2 and 3 (BS2 mut – 0.18 and BS3 mut – 0.19) with a mild reduction for BS1 mut (0.68) and BS4 mut (0.52). Hence, the contribution of each binding site is not equally distributed which might also be dependent on the position of Rbfox1 association. The downstream binding sites (BS2 and BS3) are in close proximity which has been reported to enhance alternative splicing by Rbfox1 (Das et al. 2007) and thereby could affect the exon inclusion with the mutation of any of the two binding sites.

3.3.3 Expression of Rbfox1 facilitates cardiomyoblast morphology and multinucleation

Focal adhesions are a large complex in cells that have been well described as essential factors for maintaining cardiomyocyte function and divergent assembly of individual components often lead to heart failure (Samarel 2014). The focal adhesions mediate cell-matrix interactions in cells and are crucial for heart valve development and cardiac contractility as it has been shown in heart models of zebrafish (Hirth et al. 2016; Gunawan et al. 2019). Additionally, focal adhesions together with the actin cytoskeleton are important factors for transition of cells to a specific myogenic differentiation program called "primary myoblast fusion". During this process, the membranes of differentiated myoblasts and myotubes come in close proximity and the actin cytoskeleton undergoes restructuring at the contact sites (Rochlin et al. 2010; Abmayr et al. 2003). Since focal adhesion proteins function as adaptors between the extracellular matrix (ECM) and the actin cytoskeleton at the plasma membrane, it is plausible that they are essential in this process (Samarel 2014; Quach et al. 2009).

As the expression of the Rbfox1 protein and RNA is upregulated upon differentiation (Figure 35A and 35B), the role of Rbfox1 for the differentiation process itself was investigated in H9C2 cells. Given that Rbfox1 is assumed to coordinate alternative splicing of focal adhesion proteins, it was investigated if focal adhesions are structurally affected by exogenous Rbfox1 overexpression in undifferentiated H9C2 cells. RFP-tagged Rbfox1 was transfected and focal adhesion protein vinculin as well as cytoskeletal marker actin were co-immunostained to see

potential modulation of the structural components of the cells compared to empty vector control.

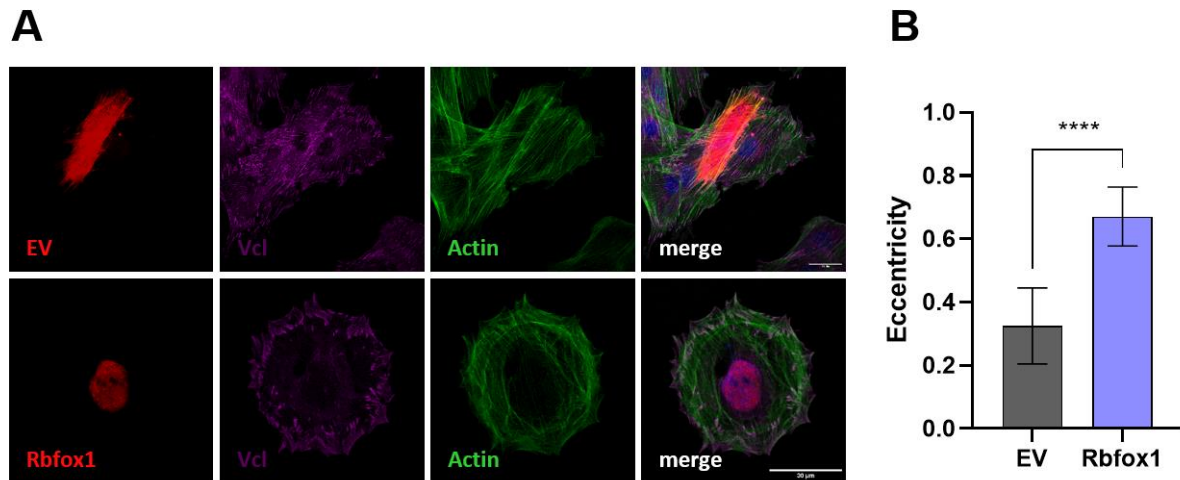


Figure 41. Rbfox1 overexpression alters the cell shape of H9C2 cells. **A** Immunofluorescence of RFP-overexpressing H9C2 cells (upper row) and overexpressing RFP-tagged Rbfox1 (lower row) in red co-stained with endogenous Vcl for focal adhesion complexes in magenta, endogenous Actin in green and nuclei staining with DAPI in blue. EV – Empty vector. Scale in the images is 30 μ m. **B** Cellular eccentricity of H9C2 cells quantified from (A) using the ImageJ software (N=15).

As seen in the immunofluorescence staining in Figure 41A Rbfox1 overexpression lead to clearly visible changes in the organization of focal adhesions represented by vinculin staining as well as rearrangement of the actin cytoskeleton 48 hours post transfection. Focal adhesions appear more concentrated and are localized in close proximity to the plasma membrane in comparison to control transfected cells in which they were spread more homogenously throughout the whole cell. In line with this result, the actin cytoskeleton in control cells consists of long fibers that are distributed from end to end while the bundles in Rbfox1 transfected cells display shorter filaments that are organized in a ring-shaped manner. Due to this observation, cellular eccentricity or ‘cell roundness’ was analyzed manually using ImageJ software measurements. Indeed, the Rbfox1 overexpressing H9C2 cells display a significant increase in eccentricity compared to the empty vector control as suggested by the microscopy images.

Apparent morphological changes in cell shape and organization of actin and focal adhesions could be related to Rbfox1 in undifferentiated cells. Since these processes are also relevant for the previously described myoblast fusion, altered morphology of H9C2 cells under differentiation conditions and upon Rbfox1 depletion were analyzed. In general, H9C2 cells can be differentiated from mono-nucleated myoblasts to multinucleated myotubes by stimulation with retinoic acid and serum-deprivation (Ménard et al. 1999). Consistently, both skeletal and cardiac muscle exhibit a large proportion of multinucleated cells, even though this more pronounced in skeletal muscle tissue (Landim-Vieira et al. 2020). The normal progression of H9C2 differentiation displays a higher amount of multinucleated cells than

undifferentiated H9C2. However, Rbfox1 knockdown prior to the differentiation resulted in altered differentiation advancement with fewer cells containing multiple nuclei (Figure 42).

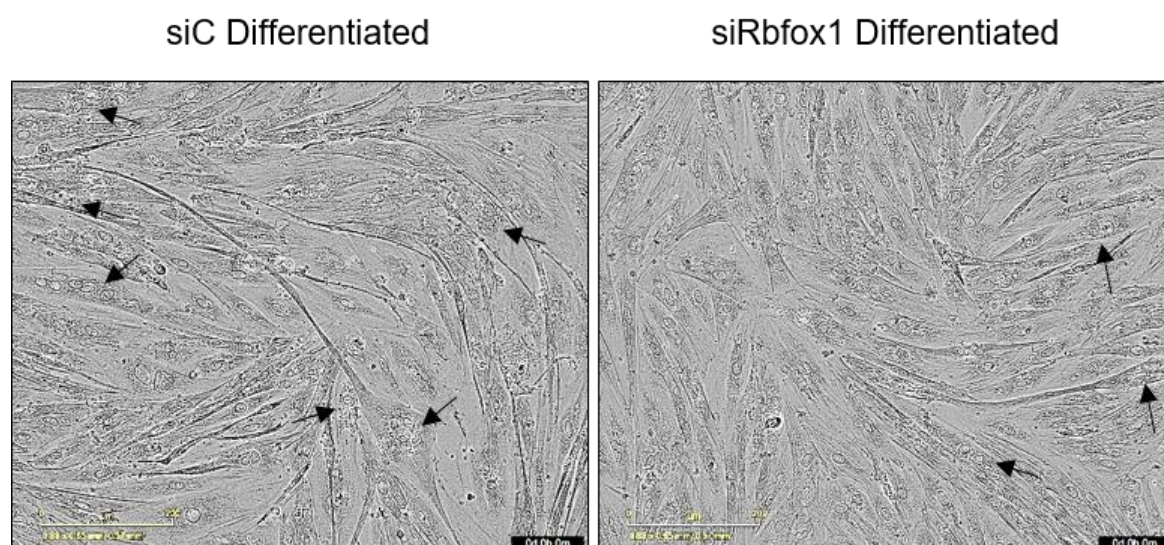


Figure 42. Rbfox1 mediated alternative splicing is required for cardiac cell fusion and differentiation progression. H9C2 cells have been transfected with siRNAs against Rbfox1 (right) and control (left) and cultivated in low-serum conditions with retinoic acid for 6 days. Arrows indicated multinucleated cells in representative images taken of the differentiation at the end point.

The differentiation of H9C2 cells transfected with control siRNA display normal differentiation behavior with alignment of elongated myocytes and multinucleation when compared to undifferentiated cells (data not shown). However, transfection of a siRNA pool targeting Rbfox1 resulted in a divergent morphology of the cells. Alignment and elongation were also visible but to a lower degree when compare to differentiated control cells (siC). Additionally, multinucleated cells were less frequently observed when Rbfox1 levels were decreased. Due to this observation, multinucleated cells were quantified in multiple images and normalized to mono-nucleated cells for both conditions (Figure 43A).

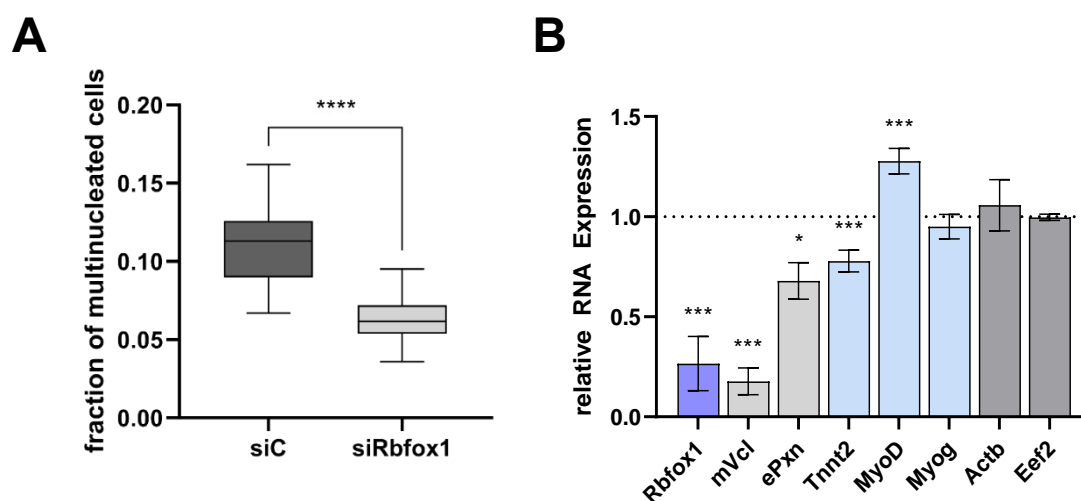


Figure 43. Depletion of Rbfox1 results in reduced amount of multinucleated cells and alters myogenic markers. **A** Quantification of the fraction of multinucleated cells from Figure 42 indicates the reduced amount of multinucleated cells upon Rbfox1 depletion. **B** RNA expression analysis of Rbfox1 splicing targets (mVcl and ePxn), myogenic differentiation markers (Tnnt2, MyoD and Myog) and housekeeper genes (Actb and Eef2) in differentiated H9C2 cells transfected with siRbfox1 vs control (n=3, significance was calculated with one-sided students t-test *- p<0,05; ** - p<0,01; ***- p<0,001; ****- p<0,0001)

In line with the noticeable morphological differences it could be determined that the Rbfox1 depletion led to a reduced number of multinucleated cells. As Rbfox1 could affect the fusion itself as well as altering the progression of differentiation, several myogenic markers were analyzed in a qRT-PCR under knockdown conditions (Figure 43B). The myogenic factors Tnnt2 and MyoD showed deregulation while Myog was not modulated upon Rbfox1 knockdown. Notably, mVcl and ePxn mRNA were significantly downregulated under these conditions. According to these results, it is assumed that Rbfox1 impacts the H9C2 differentiation program and is involved in multinucleation of myocytes through modulation of alternative splicing.

4. Discussion

4.1 The identification of unknown msRNAs by next generation sequencing

The method of next generation sequencing (NGS) enables a variety of analyses which can barely be done by other methods. However, NGS requires specific RNA preparation protocols, library generation and bioinformatic processing of the reads based on the desired outcome of the analysis and the availability of reference genomes as well as the RNA species to be investigated (Conesa et al. 2016). The majority of published RNA-sequencing analyses identify differential gene expression with a routine workflow that has not substantially changed in the recent years and includes RNA extraction followed by mRNA enrichment or ribosomal RNA depletion to increase read output for mRNA transcripts. The library preparation usually involves RNA fragmentation and cDNA synthesis which can be also strand specific, but standard analyses often utilizes oligo-(dT)-primed cDNA generation from mRNAs. The final step of library generation requires an adaptor-ligation reaction enabling PCR amplification for the sequencing analyses. The sequencing of these short-read cDNAs with a length of around 200 bp (due to fragmentation and bead based library purification) is then carried out on high-throughput devices (commonly Illumina) to generate a significant read depth of 10-30 million reads. The final procedure requires computational measures for the determination of read quality, alignment to published genomes or assembly of reads into a transcriptome, normalization of samples and statistical modelling of significant differences in gene expression for individual genes between different cohorts. The analysis of differential gene expression by RNA-Sequencing has become a routine procedure and often serves as an initial experiment for approaching new scientific projects related to RNA biology (Stark et al. 2019). However, divergence of the standard procedure and especially analyses of other RNA species besides mRNAs require adaptation of RNA and library preparation as well as modulation of the computational analyses pipeline. One major issue for the sequencing of ncRNAs is the heterogeneity of 5'-modifications as well as the 5'-end RNA structure and the lack of polyadenylation that serve as adaptor ligation sites for subsequent cDNA generation and PCR amplification during the sequencing process. Due to these properties of non-coding RNAs, the 5-end adapter-ligation preferentially targets specific species of RNAs while neglecting suboptimal reactants with modifications, 5'-secondary structure formation or triphosphates, which results in a sequencing bias (Dard-Dascot et al. 2018).

The msRNA-Sequencing method described in this thesis serves as an adaption for analysis of non-coding RNAs with a size range of 50-300 nucleotides generated by size selection with gel electrophoresis on a TBE-urea gel and excision of the nucleotide range followed by RNA extraction and RNA quality control (Figure 11B). Similar approaches have been established for early miRNA-Sequencing analyses from cell batches with a size selection

of 20-32 nucleotides representing miRNA species (Hafner et al. 2008). To minimize the sequencing bias due to 5'-modifications and triphosphates of the RNAs in the selected range preventing proper adapter-ligation, 5'-ends were de-capped and processed with RppH to generate a monophosphate more suitable for ligation (Figure 11A). Notably, the efficiency of the RppH treatment was not determined during the experiment, thereby inducing a potential sequencing bias by incomplete removal of modifications. Depending on the pipeline for mapping reads to transcripts, the allocation of certain reads can affect expression calculations, which implement an additional bias to gene expression analyses (Yang et al. 2015). This has proven especially difficult for reads mapping to multiple genes, which impose a minor problem for mRNA-Sequencing analyses, but due to the transcription of non-coding RNAs from several loci (e.g. YRNAs, snoRNAs and 4.5S_H RNA) handling of multi-mapped reads can bias the outcome of medium-sized and small non-coding RNA-Sequencing data sets. The established procedure for most computational pipelines is the exclusion of reads which cannot be uniquely mapped to the genome. This obstacle was largely overcome with the presented analyses pipeline (Figure 11C), which includes an unbiased alignment to the top five best matched gene loci and selection of the best fit as representative gene locus for every read. Additionally, the identification of msRNAs was coupled to a coverage threshold of 50 reads with a minimum length of 40 nucleotides at a genomic loci ensuring sufficient RNA expression to be considered a valid msRNA transcript region. As these adaptations represent a new approach to non-coding RNA-Sequencing, the identified msRNA sequences were extracted and blasted to a self-assembled msRNA database (including reported msRNAs from NCBI, ENSEMBL, GtRNAdb, snOPY and miRBase) for validation. The msRNA transcripts acquired from the msRNA-seq could be classified according to known msRNA genes (Figure 12) and sequences identified could be mapped to several genetic loci in human and mouse genomes due to inclusion of reads aligning to multiple regions, which emphasizes the analytical capacity of the introduced msRNA-Seq protocol. In respect to this, the method described ensured that msRNA sequences are not lost during the bioinformatic pipeline, however genes with exactly the same sequence could not be distinguished. Furthermore, the size selection for RNAs to be analyzed minimized potential contaminations with only 1% in HEK293 and 4% in B16-F10 cells, due to exclusion of larger RNAs. The main source of contaminants originates from small histone mRNAs (~300 nucleotides), which have been included due to the upper size limit of 300 nucleotides. However, large ribosomal RNA was a minor portion of the contaminants, which could be reduced in the future by application of ribosomal depletion. Since the largest portion of the sequenced msRNAs belong to 5.8S RNA (44% for HEK293 and 35% for B16-F10, Figure 12A), the reduction of ribosomal RNA could also benefit the generation of sequencing reads for other msRNAs as reported for mRNA-Seqs (Culviner et al. 2020). Additionally, lowering the stringency on read length for genomic coverage below 40 nucleotides certainly increase

detection of msRNAs substantially, as a fraction of reads already contained less than 40 nucleotides post the adapter-trimming procedure. Naturally, adaptor-ligation procedures also introduce a bias for the msRNA-Seq in terms of absolute quantification, due to incomplete removal of 5'- or 3'-end modification. Hence, the sequencing reads generated with this protocol are still biased due to natural capability of some transcripts to be more efficiently ligated to adaptors (e.g. snoRNAs). Nonetheless, the efficiency of treating extracted RNAs with RppH has been confirmed in another publication, which reported increased sequencing reads for U2 snRNA with a trimethylguanosine at the 5'-end and 5S rRNA that contains a triphosphate by 400-fold and 100-fold, respectively (Mori and Ichiyanagi 2021). Furthermore, recent developments in small RNA-Sequencing suggest more efficient adaptor-ligation methods utilizing random sequence adaptors, 3'-adaptor ligation and circularization as well as ligation-free adaptor amplification with polyadenylation and template switching (Dard-Dascot et al. 2018). However, while these methods improve sequencing rates for small RNAs like miRNAs, each of these methods contain still their own bias. Thus, sequencing results could be improved for msRNAs with unmodified 5'-ends by altering adaptor-ligation protocols, but it remains challenging for other msRNA species with 5'-modifications and secondary structure formation at the 5'-end. Nonetheless, the presented msRNA-Seq pipeline is suitable for relative msRNA quantification and identified several msRNA species including 5.8S RNA, snoRNAs and snRNAs (Figure 12A). Mature tRNAs exist in the form of aminoacyl-tRNAs and are thereby only sequenced in the intermediary state before coupled to an amino acid. This results in an only small fraction of mature tRNAs detected in the msRNA-Seq. However, in future analyses the yield could be improved by tRNA de-acylation enhancing adaptor-ligation and subsequent sequencing. The importance for tRNAs and tRNA derived fragments (tRFs) in human cancer has recently gained scientific interest (Yu et al. 2020; Goodarzi et al. 2016). Differential expression analyses of tRNAs and tRFs could therefore provide a useful tool, but remain challenging due to base modifications and secondary structures inherent to tRNAs, further complicating cDNA synthesis (Torres et al. 2019). In addition to providing a protocol for quantitative msRNA-Sequencing, the established pipeline also identified previously unannotated msRNA transcripts (HEK293: 16% and B16-F10 22%), which have been analyzed under the premise of RNA-polymerase III based transcription. Though, the majority could be linked to reported tRNA sequences, three novel msRNAs – POLAR, PAMIR and COIR could be identified (Figure 13). To evaluate, if these transcripts could in principle encode functional msRNAs or represent transcriptional background, subcellular distribution and BDP1-dependent regulation were characterized (Figure 14). Furthermore, the msRNA POLAR revealed predominant localization into paraspeckles with co-localization of PSP-proteins NONO, PSF (SFPQ) and PSP1, which also associate with *in vitro* transcribed POLAR RNA in addition to RBM14 and the La-protein (Figure 16 and 17). The association of paraspeckle

proteins with the Neat1 long-non coding RNA leads to sequestration of proteins and RNAs with trapping of SFPQ in paraspeckles resulting in transcriptional regulation (Hirose et al. 2014; Hirose et al. 2023). In view of the association of POLAR with PSP-proteins it could be speculated that the RNA acts in a similar manner to keep RNA-binding proteins in paraspeckles in addition to sequestration mediated by Neat. The functional adaptation of medium-sized non-coding RNAs as scaffolds has been shown for other POLIII transcripts, like 7SL RNA recruiting proteins forming the signal recognition particle (Akopian et al. 2013). Additionally, the 7SK RNA actively recruits MePCE, LARP7 and HEXIM1 to stall translation elongation by P-ETFb through inhibition of the phosphorylation by CDK9 (Nguyen et al. 2001; Yang et al. 2001; Täuber et al. 2019). Hence, the POLAR RNA might also act as a nuclear scaffolding RNA in paraspeckles. However, the physiological importance of paraspeckles remains yet to be elucidated and thus also additional studies are required to shed light on the function of the newly identified msRNA POLAR. In summary, the here presented sequencing protocol enhances detection of msRNAs in a quantitative manner and allows identification of unknown non-coding RNAs. The removal of 5'-end modifications and triphosphates by RppH treatment enables sequencing of previously unobtainable RNA species and in combination with the refined computational pipeline for multi-mapped reads provides a platform to develop sequencing protocols for distinct RNA species. The continuous development of adaptor-ligation methods and the increasing availability of full length (long read) RNA-Sequencing devices will certainly increase the detection of previously unidentified non-coding RNAs and enable more precise quantitative analyses combined with the msRNA-Sequencing protocol.

4.2 The intriguing relation of 4.5S_H RNA and SRSF1

The identification of unknown ncRNAs is an essential step to unravel their potential function and how they contribute to regulating cellular pathways. However, thorough investigation is required to characterize already known transcripts and to identify a potential function of the RNA. The 4.5S_H RNA has been detected in the msRNA-Seq from B16-F10 cells and was the subject of study in this thesis hereinafter. Though this RNA has been initially characterized, the functional role of this RNA remained largely unknown. The 4.5S_H RNA has been described as a nuclear POLIII transcript that displays a narrow expression pattern in specific rodent lineages, which could also be confirmed in this thesis (Harada and Kato 1980; Gogolevskaya et al. 2005). The transcription by RNA polymerase III was based on the existence of A- and B-boxes in the 4.5S_H gene and could be linked to the association of the RNA polymerase III components Gtf3c1 and Rpc155 with 4.5S_H gene loci detected in the msRNA seq analysis. The 4.5S_H RNA is mainly localized in the nucleus and had been shown to interact with Nucleolin and the La protein among other unidentified proteins detected in an electro mobility shift assay (Hirose and Harada 2008). Furthermore, immunoprecipitation

experiments in this study found 4.5S_H RNA to be enriched in the nucleolus and nucleoplasm fraction. However, another publication indicated a co-localization with the RS-domain containing protein SRSF1 inside nuclear speckles (Ishida et al. 2015) which could also be observed in this thesis with SC-35 as speckle marker in B16-F10 cells. The organization of nuclear speckles involves RBPs and RNAs related to transcription and splicing and also harbors pre-mRNAs in close vicinity to the speckle 'core' (Galganski et al. 2017; Fei et al. 2017; Spector and Lamond 2011). Accordingly, the LC-MS/MS analyses of proteins enriched with 4.5S_H RNA revealed a significant fraction of mRNA processing and splicing associated factors including several members of the SR-protein family. These proteins are important factors for constitutive and alternative splicing (Lin and Fu 2007) and have been reported to recruit the U1 snRNP complex to poorly conserved splice sites acting as exonic splicing enhancers (Zahler and Roth 1995; Valcárcel and Green 1996). For SRSF1, CLIP-Seq analyses in HEK293 revealed that the majority of transcripts recognized by SRSF1 are protein-coding genes with preferential binding in exons and in close proximity to exon-exon or exon-intron junctions (Sanford et al. 2009; Änkö 2014). Among the non-coding RNA genes, SRSF1 recognizes snoRNAs and microRNA precursors as well as Malat1 and Xist long-non coding RNAs. In addition to the exonic enhancer sites, these transcripts share a purine-rich sequence motif that has been identified as the preferred SRSF1 binding site (Sanford et al. 2009; Cook et al. 2011; Liu et al. 1998). Conveniently, the 4.5S_H RNA features an extended sequence region containing GAGG repeats that potentially mediate SRSF1 association, as deletion of this region decreased binding of SRSF1 and other SR-proteins specifically. Furthermore, analysis of a CLIP-Seq in the mouse thymus revealed SRSF1 binding peaks in the genomic loci where 4.5S_H RNA expression could be detected in the msRNA-Seq. However, the neighboring loci displays 4.5S_I RNA expression, but SRSF1 binding could not be detected despite the existence of purine-rich regions in the RNA. Notably, the RNA folding prediction of the 4.5S_H RNA implies the putative SRSF1 binding site in the single stranded loop region, which might enable binding.

The domain structure of SR-proteins is composed of one or two RRM domains and an extended RS-tandem repeat sequence of varying length (Blanco and Bernabéu 2012). Post-translational phosphorylation of the RS-domain by SRPK1 is required for active transport of SRSF1 into the nucleus by transportin-SR and subcellular localization inside nuclear speckles in cooperation with CLK1 (Aubol et al. 2013). The importance of the arginines and serines for correct localization behavior has been investigated in mutation studies, revealing replacement of RS with RE and RD as well as reducing the RS-domain to 10 residues still leads to localization in nuclear speckles. Additionally, these SRSF1 constructs were able to complement HeLa S100 splicing extracts in *in vitro* splicing assays indicating sufficient functionality (Cazalla et al. 2002). On the other hand, KS, RT, GS and RG mutants disperse

the SRSF1 in the cyto- and nucleoplasm and impair SRSF1 mediated splicing function *in vitro* (Cazalla et al. 2002; Cáceres and Krainer 1993). Hence, the RS-domain contributes to SRSF1 mediated cellular functions, however the RS-domain appears to be dispensable for RNA binding as deletion of the RS-domain did not affect association with 4.5S_H RNA and did not alter binding specificity. Intriguingly, the loss of the RS-domain disturbs association of other RS-domain containing proteins apparently mediating protein-protein interaction as it has been suggested in earlier studies (Wu and Maniatis 1993). In detail, SRSF1 mediates U1 and U2 snRNP association to 5'- and 3'-splice site with the RS-domain enhancing recruitment of U1-70K and U2AF proteins (Cho et al. 2011; Shen et al. 2004; Shen and Green 2006). However, here depletion of the RS-domain of SRSF1 also reduces interaction with other SRSF-family members. This might be related to the observation that nuclear speckles display an enrichment in RS-domain containing proteins and recent investigations have revealed the capability of RS-domains to form condensates *in vitro* (Greig et al. 2020). Hence, RS-domain mediated protein-protein interactions could drive liquid-liquid phase separation *in cellulo*. While there is a strong indication that the SRSF1 protein is directly associating with the 4.5S_H RNA, it cannot be excluded that a certain amount is also associated via protein-protein interaction, as it could be shown for the other SRSF-family members upon SRSF1 depletion. Further *in vitro* reconstitution experiments involving the 4.5S_H RNA and purified SR-proteins could resolve the RNP complex composition. The association of SR-proteins with the RNA could be tested in pulldown setups for all family members to identify direct and indirect binders.

4.3 Cooperation for nuclear speckle integrity - 4.5S_H and Malat1

Besides the interaction with proteins described in this thesis, the 4.5S_H RNA also was reported to form RNA-RNA complexes based on its high homology to the SINE B1 retrotransposon element (Ishida et al. 2015; Labuda and Zietkiewicz 1994). The RNA duplex is supposedly formed with transcripts containing an antisense sequence of the SINE B1 element, which hybridizes with the 4.5S_H RNA as shown in luciferase reporter assays. Moreover, this complex formation was suggested to lead to nuclear retention of RNAs with asB1 after 4.5S_H RNA knockdown in Neuro2A cells. These findings were tested in this thesis with the Malat1 transcript, which also contains an asB1 element potentially able to hybridize with 4.5S_H RNA. Although an enrichment of the Malat1 RNA could be confirmed, the association might not be exclusively related to an RNA-RNA interaction. The Malat1 RNA has been reported to be an important factor in splicing speckles and is recognized by the SRSF1 protein (Hutchinson et al. 2007; Tripathi et al. 2010; Gutschner et al. 2013). The experimental conditions also allowed RNA-protein interactions and the observed effect could be a result of SRSF1 binding to the 4.5S_H RNA and recruiting other SR-proteins via the RS-domain which in turn bind Malat1 molecules. However, the association of the 4.5S_H RNA with Malat1 might also

imply a functional dependence in mouse. In fact, the mouse Malat1 and human transcript display differences in transcript length (Hutchinson et al. 2007) and relevance for cellular function in the respective organism (Tripathi et al. 2010; Nakagawa et al. 2012). This discrepancy could be explained by differences in the Malat1 transcript between these organisms or more intriguingly the existence of an additional non-coding RNA with similar functions to Malat1 in terms of subcellular localization and the property to associate with SR-proteins. In regard to potential differences in the Malat1 RNA, the 5'-end extension of the human transcript likely originates from an alternative transcription start site, as the remaining transcript displays a remarkable homology and conservation when compared to mouse Malat1 (Hutchinson et al. 2007). Additionally, SRSF1 binding and localization to nuclear speckles in human cells is mainly facilitated in these conserved regions with only 1 additional minor binding site to be present in the human transcript (Fish et al. 2019; Tripathi et al. 2010). Hence, the capability of SRSF1 binding to the Malat1 RNA is presumably not affected based on the sequence variation in human and mouse, but might be explained by differences in transcriptional output in these organisms. The MALAT1 transcript expression at steady state levels is considered in the same range as highly transcribed housekeeping genes in various human cell lines (Djebali et al. 2012). This is due to positive regulation of MALAT1 through the antisense transcript TALAM1 (Zong et al. 2016) and the 3'-end processing of the transcript resulting in the short transcript known as mascRNA and formation of a triple loop structure further stabilizing the RNA in human and mouse (Wilusz et al. 2012; Brown et al. 2014). Nonetheless, Malat1 RNA expression varied greatly in analyzed mouse (B16-F10, P19 and NSC34) and human cancer cell lines (CAL-62, U2-OS and MV-3) when amplified in the conserved RNA region. Arguably, the expression of the Malat1 RNA has been shown to differ across diverse tissues in human and mouse (Hutchinson et al. 2007), implying a bias on the chosen tissue origin of the cell lines (B16-F10 melanoma/skin, P19 embryonic germ cell tumor, NSC34 fused neuroblastoma/motor neurons, CAL-62 thyroid anaplastic carcinoma, U2-OS osteosarcoma, MV-3 melanoma/lymph node). However, none of the mouse cell lines display Malat1 RNA abundances as seen in all of the human cell lines indicating that human Malat1 RNA overall is expressed at higher levels. While this observation has to be re-evaluated with additional cell lines, it is tempting to assume mouse Malat1 levels to be lower expressed when compared to the human transcript, hence developing a non-coding RNA with similar features to compensate for this. Regardless of this, Malat1 RNA was shown to be relevant in human HeLa cells for proper localization of SRSF1, U2AF65, SF3a60 and B¹U2snRNP as knockdown of Malat1 disperses these proteins into the nucleoplasm (Tripathi et al. 2010). In mice however, depletion of Malat1 by knockout did not alter the distribution of SRSF1, SRSF2 and U6 snRNA outside nuclear speckles (Nakagawa et al. 2012). These observations could imply a different mode of action for Malat1 RNA in these two organisms or hint towards functional redundancy

of the Malat1 RNA in mouse cells. The depletion of 4.5S_H RNA in the B16-F10 cells led to upregulation of Malat1, while reduction of Malat1 levels resulted in elevated expression of 4.5S_H suggesting a potential compensatory effect and thus supports the hypothesis of functional redundancy. The importance of the 4.5S_H RNA could also be shown in the condensation behavior of SC-35, which displays smaller nuclear speckles and less fluorescent intensity indicating a role of the 4.5S_H RNA in maintaining speckle integrity as it was shown for the Malat1 RNA in HeLa cells (Tripathi et al. 2010). Incidentally, the SC-35 antibody frequently utilized for nuclear speckle staining does not specifically target the SRSF2 protein, but recognizes the RS-domain containing protein Srrm2 (Ilik et al. 2020). This protein has been suggested to be an essential protein for splicing speckle formation and integrity in cooperation with SON (Ilik et al. 2020; Fei et al. 2017). Notably, the MS/MS analysis also revealed enrichment of this speckle 'core' protein with the 4.5S_H RNA further implying that the RNA may be actively involved in maintaining splicing speckle integrity in mouse, in line with the observation of splicing speckle size and local fluorescence intensity of the Srrm2 protein upon 4.5S_H RNA depletion. The phosphorylated RS-domain represents a mixed charged domain (MCD) that recently has been characterized to provide nuclear speckle residency and enable condensate formation (Greig et al. 2020). Given that the 4.5S_H RNA recruits SR-proteins by association of SRSF1 and protein-protein interaction mediated by the RS-domain, it could positively regulate liquid-liquid phase separation of RS-domain containing proteins. Tethering experiments fixating LacI-SRSF1 to a labeled LacI locus revealed recruitment of several RS-domain containing proteins and splicing factors as well as Malat1 RNA leading to *de novo* splicing speckle assembly mediated by SRSF1 (Tripathi et al. 2012). The same study suggested Malat1 RNA to promote splicing speckle assembly tethering MS2-Malat1 to a LacI locus followed by co-localization with splicing speckle core protein SC-35 (Srrm2). Underlining this hypothesis, the reduction of the Malat1 RNA in human U2-OS cells led to significantly reduced speckle size and fluorescence intensity, confirming Malat1 to be required for correct localization of splicing factors in human cells as suggested previously (Tripathi et al. 2010; Fei et al. 2017). However, depletion in the mouse cells did not interfere with SC-35 localization and in contrast splicing speckles appear more condensed, which in turn could be related to the upregulation of 4.5S_H RNA in these cells. Moreover, abolishing this upregulation in Malat1 knockdown cells by co-transfection of 4.5S_H RNA LNAs reveals an even more drastic dispersion of SC-35 emphasizing the importance of the 4.5S_H RNA for murine splicing speckle integrity. However, it has to be noted that the 4.5S_H knockdown efficiency was typically around 50-60%. Thereby it cannot be ruled out that some cellular effects (e.g. accumulation of pre-mRNAs upon 4.5S_H knockdown) could not be assessed entirely.

Besides the assembly and maintenance of splicing speckles, Malat1 RNA also has been implicated to be relevant for SRSF1 mediated splicing. It has been shown that Malat1

binds the chromatin of active gene sites (West et al. 2014), is involved in recruitment of SRSF1 to transcribed gene loci and enabling SRSF1 splicing regulation (Bernard et al. 2010; Engreitz et al. 2014). For mouse Malat1, only one study investigated changes in pre-mRNA splicing by exon-inclusion rates in mouse liver and brain reporting only minor changes when Malat1 was knocked out (Zhang et al. 2012). In contrast to this, ASO mediated knockdown of Malat1 revealed increased amounts of intron-retention in the analyzed transcripts (Eif5a, Cfl1, Rplp2, Eef2 and Ppia) indicating the RNA to be relevant for splicing regulation also in mouse cells. In line with this observation, the reduction of 4.5S_H levels did not affect intron retention to the same extent. On one side, this might suggest that 4.5S_H RNA is not required for recruitment of SRSF1 to active splice sites and solely functions as an RNA involved in speckle integrity as shown in this thesis. On the other side, the depletion of 4.5S_H RNA increased Malat1 levels which could potentially rescue some of the splicing defects, as increased intron retention is still detected for Eif5a, Rplp2, Eef2 and Ppia transcripts. Nonetheless, the role of splicing speckles remains ambiguous with the most recent model suggesting the splicing speckle to be a gene expression hub, which promote transcription and co-transcriptional processes such as splicing in the 'shell' in close proximity to active genes enabling rapid recycling of factors stored in the nuclear speckle (reviewed in (Hirose et al. 2023)). This hypothesis also encompasses the nuclear speckle as a storage organelle that appear larger and rounder upon transcriptional inhibition actively buffering concentrations of factors involved in transcription and mRNA processing from the nucleoplasm. Thus, preventing RNA processing in an uncontrolled fashion throughout the nucleoplasm, which might explain why dissolving the splicing speckles as seen in 4.5S_H and Malat1 double knockdown experiments display intron retention to a lesser degree when compared to Malat1 knockdown only. Additionally, the splicing speckles appearing rounder and more condensed in the B16-F10 cells with decreased Malat1 levels might be a result of transcriptional inhibition as Malat1 RNA has been implied to be relevant for gene transcription (West et al. 2014). On the contrary, this could not be seen in the human U2-OS cells upon Malat1 depletion, further indicating that the 4.5S_H RNA might be responsible for the nuclear speckle appearance and a relevant factor for splicing speckle integrity in mouse.

In summary, the 4.5S_H RNA is bound by RS-domain containing protein SRSF1 and is associated with other SR-proteins and splicing factors localized in nuclear speckles. The depletion of the RNA results in dispersion of nuclear speckle marker protein SC-35 indicating an impact on splicing speckle integrity. This observation could be based on the recruitment of RS-domain containing proteins to the 4.5S_H RNA effectively enhancing condensate formation as it has been shown for Malat1 RNA and SRSF1 protein in human (Tripathi et al. 2012). Additionally, the depletion of the 4.5S_H RNA resulted in an accumulation of transcripts with intron-retention indicating a potential regulatory effect on constitutive splicing. However, if this effect is related to active recruitment of the SRSF1 protein to splice sites by the 4.5S_H RNA as

it has been described for Malat1 (Bernard et al. 2010; Engreitz et al. 2014) remains to be elucidated.

4.4 Alternative focal adhesions and Rbfox1 – factors for multinucleation

The process of alternative splicing enables higher organisms to create a diversity of transcript and protein isoforms contributing to complex cellular functions in a tissue specific manner and to modify molecular pathways. The development of high-throughput sequencing and the increasing availability of bioinformatic tools allows analyses of alternatively spliced isoforms in the context of cardiac development and disease (Hasimbegovic et al. 2021). During the development, the heart undergoes physiological and morphological changes that require narrowly regulated isoform expression directed by a network of splicing factors. The role of alternative splicing is essential to heart muscle development and deregulation of this process was shown to be involved in cardiac pathologies impairing heart function like myocardial infarction (Kalsotra et al. 2010; Giudice et al. 2014; Wang et al. 2016). In recent studies, genome-wide analyses of gene expression and alternatively spliced isoform usage have identified several mis-spliced transcripts due to the downregulation of certain splicing factors in heart pathologies (Park et al. 2011). The cardiomyocyte-specific knockout of splicing factors SRSF1 and SRSF2 resulted in the development of dilated cardiomyopathy (DCM) after 6 and 5 weeks, respectively (Xu et al. 2005; Ding et al. 2004). RNA-sequencing analysis of human hearts with DCM revealed a substantial amount of differentially spliced genes which were enriched for genes implicated to function in the contractile machinery as GO-term analysis displayed classifications of “actin filament organization”, “Z disc” and “I band” (Heinig et al. 2017). In pathological heart hypertrophy, alternative splicing patterns resemble gene expression observed in fetal hearts (Ames et al. 2013). The same observation could be made in Rbfox1 depleted mice during cardiac hypertrophy implicating Rbfox1 to be relevant for cardiac development and to regulate important alternative splicing processes for maintaining proper heart function (Gao et al. 2016; Frese et al. 2015).

In line with previous findings, Rbfox1 showed a significant downregulation in heart failure (MI) and cardiac hypertrophy models (EGFR-KO, TAC). Further investigations for the potential target spectrum of Rbfox1 led to the discovery of genes to be potentially regulated by Rbfox1 based on the Rbfox1 binding motif in intronic regions. Interestingly, clustering of potential Rbfox1 target genes revealed enrichment for genes encoding focal adhesion proteins. The focal adhesion complexes have been shown to be crucial structures required for myoblast cell fusion and mediating force propagation in cardiac muscle cells (Samarel 2014; McCain et al. 2012). Among the *in silico* predicted genes to be regulated by Rbfox1, the focal adhesion genes Paxillin, Vinculin and Actinin1 and their alternatively spliced isoforms have been described to be important for muscle cell function (Foley and Young 2013; Kanoldt et al.

2020). The expression of Rbfox1 protein is connected to the differentiation status of the muscle cell with increasing levels during the differentiation process from myoblast to myotubes in skeletal and cardiac muscle cells (Pedrotti et al. 2015). In line with the observation that Rbfox1 depletion results in a fetal-like isoform expression (Gao et al. 2016), Rbfox1 binding sites are enriched in introns of genes regarded as developmentally regulated and alternatively spliced in the heart muscle (Kalsotra et al. 2008). During the differentiation process of the H9C2 cells, an upregulation of Rbfox1 could be confirmed alongside increased expression of Metavinculin (mVcl) and an uncharacterized Paxillin isoform (ePxn).

The Metavinculin isoform has been identified and characterized previous to this work. The transcript contains an additional exon 19 contributing 68 amino acids to the C-terminal tail domain of the vinculin isoform (Rangarajan et al. 2010). Additionally, the expression of mVcl could only be detected in cardiac, skeletal and smooth muscle cells in contrast to the ubiquitous expression of the Vcl isoform (Feramisco et al. 1982; Belkin et al. 1988; Kim et al. 2016). The increasing expression of mVcl during the differentiation into cardiac and muscle cells while absent in undifferentiated cells emphasizes these findings. Overexpression of Rbfox1 led to a significant increase of the mVcl isoform indicating a potential regulation of this transcript variant by Rbfox1. This observation could be confirmed in the minigene analysis with point mutations in all putative Rbfox1 binding sites to result in a decrease in exon 19 inclusion into the final transcript. It has been previously reported that introns adjacent to Rbfox1 regulated exons often contain several Rbfox1 binding sites, which can be detected in both, mVcl and ePxn (Conboy 2017). Notably, Rbfox1 binding sites downstream of the alternative exons have been implicated to promote exon inclusion, while upstream recognition sites tend to lead to exon skipping (Sun et al. 2012; Lovci et al. 2013). However, the molecular mechanisms for position-dependent exon splicing are not well understood. The mutation of the downstream putative Rbfox1 site in the mVcl minigenes resulted in the exclusion of exon 19 in the final transcript in line with published analyses. On the contrary, point mutations in the upstream putative Rbfox1 motifs also decreased exon 19 inclusion to a lower extent. One study demonstrated Rbfox2 binding to the consensus motif positively regulated 5'-ss selection by recruitment of U1 snRNP to weak splice sites in minigene analyses with a downstream Rbfox binding site in close proximity (Huang et al. 2012). Given the positions of the putative Rbfox1 binding sites in the upstream introns to be in close proximity to potential 5'- and 3'- splice sites, one could speculate to have a similar mechanism enhancing exon inclusion even with upstream Rbfox1 binding sites. Thus, the reduction of the exon inclusion in the ePxn minigenes BS2 mut and BS3 mut could also be explained by the close proximity of the Rbfox1 binding motifs to the 5'ss of exon 7, actively recruiting the splicing machinery. On the other side, it has been previously suggested that additional binding sites in close vicinity enhance Rbfox1 mediated splicing (Das et al. 2007). Consistent with this report, the mutation of either of the two binding sites resulted

in a reduction of exon inclusion in the same range (~0.18 exon inclusion compared to the wild type minigene) indicating that they mutually regulate Rbfox1 mediated alternative splicing of the exon. Due to the nature of the mutation only introducing a nucleotide switch from GCATG to GACTG it is unlikely to affect other potential splicing proteins, however secondary effects independent from Rbfox1 mediated regulation cannot be excluded.

The ePxn isoform was identified and characterized in this thesis to be regulated by the Rbfox1 protein. Despite the increased size with the exon inclusion mediated by Rbfox1, the ePxn variant also localizes to focal adhesions (data not shown). In regard to the alternatively spliced exons translating to a highly disordered amino acid sequence this could imply potential regulatory functions in focal adhesions. Recently, intrinsically disordered proteins became the subject of study in a variety of biological fields as they lack defined tertiary structures enabling the engagement into a multitude of biological functions (Uversky et al. 2000; Oldfield and Dunker 2014). The N-terminal region of Paxillin also contains large and flexible intrinsically disordered regions (IDRs) spacing the functional motifs that have been shown to interact with the Focal Adhesion Kinase (FAK), Vinculin and the Src proteins (Brown et al. 2005; Hu et al. 2014). These flexible IDRs often feature phosphorylation sites that can modulate the interaction network of proteins observable in focal adhesions. This was shown for tissue-specific alternative splicing of FAK introducing additional phosphorylation sites recognized by the Src family kinases and activating signaling pathways (Ingley 2008). In focal adhesions, Paxillin represents an important adaptor protein that recruits additional proteins to these sites forming the focal adhesion complex (Brown and Turner 2004). According to the predictions made for the alternatively spliced exons the ePxn isoform contains several SH3 domain binding motifs that could provide amino acids for post-translational modifications thus modulating focal adhesion kinetics or assembly. Moreover, altered alternative splicing of proteins in focal adhesion and the actin cytoskeleton due to Rbfox1 mediated regulation might contribute to the differences in morphology observed in the H9C2 cells upon Rbfox1 overexpression. The Metavinculin isoform was reported to modulate binding and flexibility of actin filaments ultimately altering filament organization in presence of vinculin (Tolbert et al. 2013; Oztug Durer et al. 2015). Thus, the rearrangement of the cytoskeleton could be mediated by the alternative isoforms mVcl and ePxn due to altered engagement of binding partners present in the focal adhesion sites. However, the putative target genes identified in this thesis also include other focal adhesion proteins with a potential impact on focal adhesion dynamics. Hence, Rbfox1 regulated alternative splicing events could also occur in these transcripts and contribute to the morphological changes observed in these cells.

The increasing Rbfox1 expression during differentiation in addition to the reports of fetal-like expression patterns in Rbfox1 depleted mouse hearts (Gao et al. 2016) emphasizes the role of Rbfox1 for cardiac development. In this regard, Rbfox1 depletion prior to the

initiating the differentiation leads to a significant decrease in multinucleated cells indicating an alteration of the differentiation process. Furthermore, expression analyses of differentiation markers revealed a minor but significant reduction in myogenic transcription factors upon Rbfox1 knockdown. Thus, the initial differentiation process remains unaffected by Rbfox1 depletion but is perturbed during the process of multinucleation at later stages. This comprises alternative splicing changes due to Rbfox1 levels in focal adhesion genes like mVcl and ePxn which might not only alter focal adhesion dynamics but are also required for the differentiation process. These findings are supported by similar observations upon inhibition of FAK in another study leading to disturbed multinucleation behavior but no aberrant effect on the myogenic terminal differentiation program (Quach et al. 2009). In summary, the results presented in this thesis suggest tissue-specific splicing factor Rbfox1 is potentially able to promote fusion-competent myoblasts leading to a higher degree of multinucleation. Additionally, morphological changes upon Rbfox1 modulation indicate a contribution of focal adhesion transcripts that have been shown to be alternatively spliced by Rbfox1 (mVcl and ePxn).

5. Summary

RNA-sequencing provides a proficient method to investigate the transcriptome and enables a broad spectrum of analyzes depending on library preparation, sequencing method and subsequent processing of the reads using bioinformatic tools. The first part of the thesis aimed to improve the library generation of medium-sized ncRNAs for high-throughput RNA-Sequencing and to adapt the bioinformatic pipeline for the analyses of obtained reads. The enzymatic cleavage of 5'-modifications and triphosphates enhanced adapter ligation and consequently sequencing of the non-coding RNA species. The method presented in this study successfully identified msRNAs from several ncRNA families including 5.8S RNA, snRNAs, snoRNAs as well as premature tRNAs. The transcripts generated in this msRNA-Seq were classified as msRNA genes emphasizing the specificity of the method. Additionally, identified sequences could be referred to several genomic loci in both, human and mouse, by inclusion of multi-mapped reads highlighting the analytical capabilities of this method. Furthermore, lowering the threshold of 50 reads with minimal 40 nucleotide length for genomic mapping and implementing rRNA depletion may have further enhanced msRNA detection with this protocol. Nonetheless, three previously unknown msRNA genes (POLAR, PAMIR, COIR) could be identified and were found to be transcribed by RNA-polymerase III. The initial characterization of the POLAR RNA revealed an association with paraspeckle proteins and subcellular localization in these MLOs, indicating a potential role in paraspeckle function. Thus, this method provides a tool for the identification of novel medium-sized non-coding RNAs in addition to the quantitative analyses of these RNAs. Furthermore, the msRNA dataset generated from the mouse B16-F10 also provided expression data for the second part of the thesis - the characterization of the non-coding RNA 4.5S_H. The narrow expression profile in the myodonta clade and the subcellular localization inside nuclear speckles could be confirmed according to previously published data. In line with the co-localization of the 4.5S_H RNA and splicing factor SC-35, a global analysis of associated proteins with the RNA in a LC-MS/MS analysis revealed enrichment for proteins involved in splicing and mRNA processing via the spliceosome. Furthermore, the RNA displayed strong interactions with proteins of the SRSF-family and other proteins containing phosphorylated disordered RS-domains as it could be shown by pulldown analyses. Based on a secondary structure prediction the RNA is folded into a stem-loop structure that contains purine-rich sequence elements in the extended loop region that are required for the association of these proteins as the deletion resulted in diminished detection of bound RS-proteins. Moreover, the RS-domain appears to mediate protein-protein interaction as the overexpression of SRSF1 without an RS-domain failed to recruit the other SR-proteins to the RNA but still associated with endogenous SRSF1 in pulldown experiments. Further analyses indicated the SRSF1 protein to be the main

associated protein among the SRSF-family members in line with the mass spectrometry analysis. Given these results, the 4.5S_H RNA displays similar properties to the Malat1 RNA that has been closely associated with SRSF1 function in human cells. Intriguingly, depletion of the 4.5S_H RNA increased expression of the Malat1 RNA and vice versa, indicating a potential redundancy of these two RNAs in mouse. Moreover, the depletion of the 4.5S_H RNA displayed impaired splicing speckle appearance based on fluorescence imaging of splicing speckle marker SC-35, a quality that previously has been attributed to the Malat1 RNA in human cells and could also be observed in this thesis. However, the Malat1 RNA was observed to be less relevant for the splicing speckle integrity in the mouse cells in line with previously published data. In contrast, knockdown of Malat1 resulted in the accumulation of pre-mRNA indicative of a disturbed overall splicing process, which occurred to a lesser extent with 4.5S_H RNA depletion. Thus, these results indicate a redundancy of the 4.5S_H RNA for SRSF1 binding and maintaining splicing speckle integrity, but with limited capability of modulating SRSF1 mediated splicing as observed with Malat1 RNA. Consequently, the existence of the 4.5S_H RNA has gone unnoticed in previous mouse Malat1 models and potentially masked the effects of the Malat1 modulation. Notably, future investigations in this direction should at least consider the partial redundancy of these RNAs to gain new insights into splicing speckle morphology and splicing regulation.

The Rbfox1 protein has previously been investigated as a major splicing regulator in heart and neuronal tissue with a substantial contribution in promoting alternatively spliced isoforms in a tissue-specific manner. Here, two alternatively spliced isoforms of the focal adhesion proteins Paxillin and Vinculin have been identified to be regulated by the Rbfox1 protein. In detail, Rbfox1 promotes expression of the previously published Metavinculin isoform and an uncharacterized Pxn isoform that has been termed ePxn (for extended Paxillin). The relevance of the Rbfox1 binding sites has been investigated in minigene reporters containing flanking exons and alternatively spliced exon 19 and exons 7 to 10 for Metavinculin and ePxn, respectively. Co-expression of Rbfox1 increased exon inclusion for both isoforms (Metavinculin and ePxn) generated from the minigenes. However, the putative Rbfox1 binding sites each contribute to overall exon inclusion, as mutation of the individual GCATG motifs resulted in decreased expression of the alternatively spliced isoforms for both transcripts. Hence, the tissue-specific isoforms of Vinculin and Paxillin are regulated in an Rbfox1-dependent manner. Furthermore, the Rbfox1 protein might regulate additional focal adhesion isoforms thereby modulating focal adhesion complexes and the actin cytoskeleton as Rbfox1 overexpression resulted in altered morphological appearances of these cells. Additionally, Rbfox1 depletion interfered with the cardiac differentiation process of H9C2 cells, which lead to observable morphological changes and reduced multinucleation in later stages of the differentiation. Consequently, Rbfox1 mediated splicing changes of focal adhesion proteins

including Metavinculin and ePxn could not only affect focal adhesion complex composition and dynamics, but might also be crucial for the differentiation process.

6. References

- Abmayr, Susan M.; Balagopalan, Lakshmi; Galletta, Brian J.; Hong, Sue-Jean (2003): Cell and molecular biology of myoblast fusion. In: *International review of cytology* 225, S. 33–89. DOI: 10.1016/s0074-7696(05)25002-7.
- Akopian, David; Shen, Kuang; Zhang, Xin; Shan, Shu-ou (2013): Signal recognition particle: an essential protein-targeting machine. In: *Annual review of biochemistry* 82, S. 693–721. DOI: 10.1146/annurev-biochem-072711-164732.
- Ames, E. G.; Lawson, M. J.; Mackey, A. J.; Holmes, J. W. (2013): Sequencing of mRNA identifies re-expression of fetal splice variants in cardiac hypertrophy. In: *Journal of molecular and cellular cardiology* 62, S. 99–107. DOI: 10.1016/j.yjmcc.2013.05.004.
- Anczuków, Olga; Rosenberg, Avi Z.; Akerman, Martin; Das, Shipra; Zhan, Lixing; Karni, Rotem et al. (2012): The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. In: *Nature structural & molecular biology* 19 (2), S. 220–228. DOI: 10.1038/nsmb.2207.
- Änkö, Minna-Liisa (2014): Regulation of gene expression programmes by serine-arginine rich splicing factors. In: *Seminars in cell & developmental biology* 32, S. 11–21. DOI: 10.1016/j.semcdb.2014.03.011.
- Arun, Gayatri; Diermeier, Sarah; Akerman, Martin; Chang, Kung-Chi; Wilkinson, J. Erby; Hearn, Stephen et al. (2016): Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. In: *Genes & development* 30 (1), S. 34–51. DOI: 10.1101/gad.270959.115.
- Aubol, Brandon E.; Adams, Joseph A. (2011): Applying the brakes to multisite SR protein phosphorylation: substrate-induced effects on the splicing kinase SRPK1. In: *Biochemistry* 50 (32), S. 6888–6900. DOI: 10.1021/bi2007993.
- Aubol, Brandon E.; Plocinik, Ryan M.; Hagopian, Jonathan C.; Ma, Chen-Ting; McGlone, Maria L.; Bandyopadhyay, Reeti et al. (2013): Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. In: *Journal of molecular biology* 425 (16), S. 2894–2909. DOI: 10.1016/j.jmb.2013.05.013.
- Babski, Julia; Maier, Lisa-Katharina; Heyer, Ruth; Jaschinski, Katharina; Prasse, Daniela; Jäger, Dominik et al. (2014): Small regulatory RNAs in Archaea. In: *RNA biology* 11 (5), S. 484–493. DOI: 10.4161/rna.28452.
- Bacrot, Séverine; Doyard, Mathilde; Huber, Céline; Alibeu, Olivier; Feldhahn, Niklas; Lehalle, Daphné et al. (2015): Mutations in SNRPB, encoding components of the core splicing

machinery, cause cerebro-costo-mandibular syndrome. In: *Human mutation* 36 (2), S. 187–190. DOI: 10.1002/humu.22729.

Banani, Salman F.; Lee, Hyun O.; Hyman, Anthony A.; Rosen, Michael K. (2017): Biomolecular condensates: organizers of cellular biochemistry. In: *Nature reviews. Molecular cell biology* 18 (5), S. 285–298. DOI: 10.1038/nrm.2017.7.

Baralle, Francisco E.; Giudice, Jimena (2017): Alternative splicing as a regulator of development and tissue identity. In: *Nature reviews. Molecular cell biology* 18 (7), S. 437–451. DOI: 10.1038/nrm.2017.27.

Basu, Swaraj; Müller, Ferenc; Sanges, Remo (2013): Examples of sequence conservation analyses capture a subset of mouse long non-coding RNAs sharing homology with fish conserved genomic elements. In: *BMC bioinformatics* 14 Suppl 7 (Suppl 7), S14. DOI: 10.1186/1471-2105-14-S7-S14.

Belkin, A. M.; Ornatsky, O. I.; Kabakov, A. E.; Glukhova, M. A.; Koteliansky, V. E. (1988): Diversity of vinculin/meta-vinculin in human tissues and cultivated cells. Expression of muscle specific variants of vinculin in human aorta smooth muscle cells. In: *The Journal of biological chemistry* 263 (14), S. 6631–6635.

Ben-Hur, Vered; Denichenko, Polina; Siegfried, Zahava; Maimon, Avi; Krainer, Adrian; Davidson, Ben; Karni, Rotem (2013): S6K1 alternative splicing modulates its oncogenic activity and regulates mTORC1. In: *Cell reports* 3 (1), S. 103–115. DOI: 10.1016/j.celrep.2012.11.020.

Berg, Matthew D.; Brandl, Christopher J. (2021): Transfer RNAs: diversity in form and function. In: *RNA biology* 18 (3), S. 316–339. DOI: 10.1080/15476286.2020.1809197.

Bernard, Delphine; Prasanth, Kannanganattu V.; Tripathi, Vidisha; Colasse, Sabrina; Nakamura, Tetsuya; Xuan, Zhenyu et al. (2010): A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. In: *The EMBO journal* 29 (18), S. 3082–3093. DOI: 10.1038/emboj.2010.199.

Bertram, Karl; Agafonov, Dmitry E.; Dybkov, Olexandr; Haselbach, David; Leelaram, Majety N.; Will, Cindy L. et al. (2017): Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. In: *Cell* 170 (4), 701-713.e11. DOI: 10.1016/j.cell.2017.07.011.

Binas, Stephanie; Knyrim, Maria; Hupfeld, Julia; Kloeckner, Udo; Rabe, Sindy; Mildenerberger, Sigrid et al. (2020): miR-221 and -222 target CACNA1C and KCNJ5 leading to altered cardiac ion channel expression and current density. In: *Cellular and molecular life sciences : CMLS* 77 (5), S. 903–918. DOI: 10.1007/s00018-019-03217-y.

- Blanco, Francisco Javier; Bernabéu, Carmelo (2012): The Splicing Factor SRSF1 as a Marker for Endothelial Senescence. In: *Frontiers in physiology* 3, S. 54. DOI: 10.3389/fphys.2012.00054.
- Borodulina, Olga R.; Kramerov, Dmitri A. (2008): Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. In: *RNA (New York, N.Y.)* 14 (9), S. 1865–1873. DOI: 10.1261/rna.1006608.
- Boutz, Paul L.; Stoilov, Peter; Li, Qin; Lin, Chia-Ho; Chawla, Geetanjali; Ostrow, Kristin et al. (2007): A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. In: *Genes & development* 21 (13), S. 1636–1652. DOI: 10.1101/gad.1558107.
- Brown, Jessica A.; Bulkley, David; Wang, Jimin; Valenstein, Max L.; Yario, Therese A.; Steitz, Thomas A.; Steitz, Joan A. (2014): Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. In: *Nature structural & molecular biology* 21 (7), S. 633–640. DOI: 10.1038/nsmb.2844.
- Brown, Michael C.; Cary, Leslie A.; Jamieson, Jennifer S.; Cooper, Jonathan A.; Turner, Christopher E. (2005): Src and FAK kinases cooperate to phosphorylate Paxillin kinase linker, stimulate its focal adhesion localization, and regulate cell spreading and protrusiveness. In: *Molecular biology of the cell* 16 (9), S. 4316–4328. DOI: 10.1091/mbc.e05-02-0131.
- Brown, Michael C.; Turner, Christopher E. (2004): Paxillin: adapting to change. In: *Physiological reviews* 84 (4), S. 1315–1339. DOI: 10.1152/physrev.00002.2004.
- Burke, James M.; Kincaid, Rodney P.; Nottingham, Ryan M.; Lambowitz, Alan M.; Sullivan, Christopher S. (2016): DUSP11 activity on triphosphorylated transcripts promotes Argonaute association with noncanonical viral microRNAs and regulates steady-state levels of cellular noncoding RNAs. In: *Genes & development* 30 (18), S. 2076–2092. DOI: 10.1101/gad.282616.116.
- Cáceres, J. F.; Krainer, A. R. (1993): Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. In: *The EMBO journal* 12 (12), S. 4715–4726. DOI: 10.1002/j.1460-2075.1993.tb06160.x.
- Cáceres, J. F.; Misteli, T.; Sreaton, G. R.; Spector, D. L.; Krainer, A. R. (1997): Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. In: *The Journal of cell biology* 138 (2), S. 225–238. DOI: 10.1083/jcb.138.2.225.

- Cáceres, J. F.; Stamm, S.; Helfman, D. M.; Krainer, A. R. (1994): Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. In: *Science (New York, N.Y.)* 265 (5179), S. 1706–1709. DOI: 10.1126/science.8085156.
- Calvo Sánchez, Jaime; Köhn, Marcel (2021): Small but Mighty-The Emerging Role of snoRNAs in Hematological Malignancies. In: *Non-coding RNA* 7 (4). DOI: 10.3390/ncrna7040068.
- Cao, W.; Jamison, S. F.; Garcia-Blanco, M. A. (1997): Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. In: *RNA (New York, N.Y.)* 3 (12), S. 1456–1467.
- Cazalla, Demian; Zhu, Jun; Manche, Lisa; Huber, Elisabeth; Krainer, Adrian R.; Cáceres, Javier F. (2002): Nuclear export and retention signals in the RS domain of SR proteins. In: *Molecular and cellular biology* 22 (19), S. 6871–6882. DOI: 10.1128/MCB.22.19.6871-6882.2002.
- Chan, Patricia P.; Lowe, Todd M. (2016): GtRNADB 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. In: *Nucleic acids research* 44 (D1), D184-9. DOI: 10.1093/nar/gkv1309.
- Cho, Suhyung; Hoang, Amy; Sinha, Rahul; Zhong, Xiang-Yang; Fu, Xiang-Dong; Krainer, Adrian R.; Ghosh, Gourisankar (2011): Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. In: *Proceedings of the National Academy of Sciences of the United States of America* 108 (20), S. 8233–8238. DOI: 10.1073/pnas.1017700108.
- Christov, Christo P.; Gardiner, Timothy J.; Szüts, Dávid; Krude, Torsten (2006): Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. In: *Molecular and cellular biology* 26 (18), S. 6993–7004. DOI: 10.1128/MCB.01060-06.
- Cléry, Antoine; Sinha, Rahul; Anczuków, Olga; Corrionero, Anna; Moursy, Ahmed; Daubner, Gerrit M. et al. (2013): Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (30), E2802-11. DOI: 10.1073/pnas.1303445110.
- Conboy, John G. (2017): Developmental regulation of RNA processing by Rbfox proteins. In: *Wiley interdisciplinary reviews. RNA* 8 (2). DOI: 10.1002/wrna.1398.
- Conesa, Ana; Madrigal, Pedro; Tarazona, Sonia; Gomez-Cabrero, David; Cervera, Alejandra; McPherson, Andrew et al. (2016): A survey of best practices for RNA-seq data analysis. In: *Genome biology* 17, S. 13. DOI: 10.1186/s13059-016-0881-8.

- Cook, Kate B.; Kazan, Hilal; Zuberi, Khalid; Morris, Quaid; Hughes, Timothy R. (2011): RBPDB: a database of RNA-binding specificities. In: *Nucleic acids research* 39 (Database issue), D301-8. DOI: 10.1093/nar/gkq1069.
- Cuellar-Partida, Gabriel; Buske, Fabian A.; McLeay, Robert C.; Whittington, Tom; Noble, William Stafford; Bailey, Timothy L. (2012): Epigenetic priors for identifying active transcription factor binding sites. In: *Bioinformatics (Oxford, England)* 28 (1), S. 56–62. DOI: 10.1093/bioinformatics/btr614.
- Culviner, Peter H.; Guegler, Chantal K.; Laub, Michael T. (2020): A Simple, Cost-Effective, and Robust Method for rRNA Depletion in RNA-Sequencing Studies. In: *mBio* 11 (2). DOI: 10.1128/mbio.00010-20.
- Damianov, Andrey; Black, Douglas L. (2010): Autoregulation of Fox protein expression to produce dominant negative splicing factors. In: *RNA (New York, N.Y.)* 16 (2), S. 405–416. DOI: 10.1261/rna.1838210.
- Dard-Dascot, Cloelia; Naquin, Delphine; d'Aubenton-Carafa, Yves; Alix, Karine; Thermes, Claude; van Dijk, Erwin (2018): Systematic comparison of small RNA library preparation protocols for next-generation sequencing. In: *BMC genomics* 19 (1), S. 118. DOI: 10.1186/s12864-018-4491-6.
- Darty, Kévin; Denise, Alain; Ponty, Yann (2009): VARNA: Interactive drawing and editing of the RNA secondary structure. In: *Bioinformatics (Oxford, England)* 25 (15), S. 1974–1975. DOI: 10.1093/bioinformatics/btp250.
- Das, Debopriya; Clark, Tyson A.; Schweitzer, Anthony; Yamamoto, Miki; Marr, Henry; Arribere, Josh et al. (2007): A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing. In: *Nucleic acids research* 35 (14), S. 4845–4857. DOI: 10.1093/nar/gkm485.
- Das, Shipra; Krainer, Adrian R. (2014): Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. In: *Molecular cancer research : MCR* 12 (9), S. 1195–1204. DOI: 10.1158/1541-7786.MCR-14-0131.
- Diederichs, Sven (2014): The four dimensions of noncoding RNA conservation. In: *Trends in genetics : TIG* 30 (4), S. 121–123. DOI: 10.1016/j.tig.2014.01.004.
- Ding, Jian-Hua; Xu, Xiangdong; Yang, Dongmei; Chu, Pao-Hsien; Dalton, Nancy D.; Ye, Zhen et al. (2004): Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. In: *The EMBO journal* 23 (4), S. 885–896. DOI: 10.1038/sj.emboj.7600054.

- Djebali, Sarah; Davis, Carrie A.; Merkel, Angelika; Dobin, Alex; Lassmann, Timo; Mortazavi, Ali et al. (2012): Landscape of transcription in human cells. In: *Nature* 489 (7414), S. 101–108. DOI: 10.1038/nature11233.
- Driedonks, Tom A. P.; Nolte-'t Hoen, Esther N. M. (2018): Circulating Y-RNAs in Extracellular Vesicles and Ribonucleoprotein Complexes; Implications for the Immune System. In: *Frontiers in immunology* 9, S. 3164. DOI: 10.3389/fimmu.2018.03164.
- DuBridge, R. B.; Tang, P.; Hsia, H. C.; Leong, P. M.; Miller, J. H.; Calos, M. P. (1987): Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. In: *Molecular and cellular biology* 7 (1), S. 379–387. DOI: 10.1128/mcb.7.1.379-387.1987.
- Engreitz, Jesse M.; Sirokman, Klara; McDonel, Patrick; Shishkin, Alexander A.; Surka, Christine; Russell, Pamela et al. (2014): RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. In: *Cell* 159 (1), S. 188–199. DOI: 10.1016/j.cell.2014.08.018.
- Expert-Bezançon, Alain; Sureau, Alain; Durosay, Patrice; Salesse, Roland; Groeneveld, Herman; Lecaer, Jean Pierre; Marie, Joëlle (2004): hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. In: *The Journal of biological chemistry* 279 (37), S. 38249–38259. DOI: 10.1074/jbc.M405377200.
- Fei, Jingyi; Jadaliha, Mahdieh; Harmon, Tyler S.; Li, Isaac T. S.; Hua, Boyang; Hao, Qinyu et al. (2017): Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. In: *Journal of cell science* 130 (24), S. 4180–4192. DOI: 10.1242/jcs.206854.
- Feramisco, J. R.; Smart, J. E.; Burridge, K.; Helfman, D. M.; Thomas, G. P. (1982): Co-existence of vinculin and a vinculin-like protein of higher molecular weight in smooth muscle. In: *The Journal of biological chemistry* 257 (18), S. 11024–11031.
- Fidler, I. J. (1975): Biological behavior of malignant melanoma cells correlated to their survival in vivo. In: *Cancer research* 35 (1), S. 218–224.
- Fish, Lisa; Navickas, Albertas; Culbertson, Bruce; Xu, Yichen; Nguyen, Hoang C. B.; Zhang, Steven et al. (2019): Nuclear TARBP2 Drives Oncogenic Dysregulation of RNA Splicing and Decay. In: *Molecular cell* 75 (5), 967-981.e9. DOI: 10.1016/j.molcel.2019.06.001.
- Foley, Kate S.; Young, Paul W. (2013): An analysis of splicing, actin-binding properties, heterodimerization and molecular interactions of the non-muscle α -actinins. In: *The Biochemical journal* 452 (3), S. 477–488. DOI: 10.1042/BJ20121824.

- Fox, Archa H.; Lam, Yun Wah; Leung, Anthony K. L.; Lyon, Carol E.; Andersen, Jens; Mann, Matthias; Lamond, Angus I. (2002): Paraspeckles: a novel nuclear domain. In: *Current biology* : CB 12 (1), S. 13–25. DOI: 10.1016/s0960-9822(01)00632-7.
- Frankish, Adam; Diekhans, Mark; Ferreira, Anne-Maud; Johnson, Rory; Jungreis, Irwin; Loveland, Jane et al. (2019): GENCODE reference annotation for the human and mouse genomes. In: *Nucleic acids research* 47 (D1), D766-D773. DOI: 10.1093/nar/gky955.
- Frese, Karen S.; Meder, Benjamin; Keller, Andreas; Just, Steffen; Haas, Jan; Vogel, Britta et al. (2015): RNA splicing regulated by RBFOX1 is essential for cardiac function in zebrafish. In: *Journal of cell science* 128 (16), S. 3030–3040. DOI: 10.1242/jcs.166850.
- Frith, Martin C.; Pheasant, Michael; Mattick, John S. (2005): The amazing complexity of the human transcriptome. In: *European journal of human genetics* : EJHG 13 (8), S. 894–897. DOI: 10.1038/sj.ejhg.5201459.
- Fu, Xiang-Dong (2014): Non-coding RNA: a new frontier in regulatory biology. In: *National science review* 1 (2), S. 190–204. DOI: 10.1093/nsr/nwu008.
- Galganski, Lukasz; Urbanek, Martyna O.; Krzyzosiak, Wlodzimierz J. (2017): Nuclear speckles: molecular organization, biological function and role in disease. In: *Nucleic acids research* 45 (18), S. 10350–10368. DOI: 10.1093/nar/gkx759.
- Gao, Chen; Ren, Shuxun; Lee, Jae-Hyung; Qiu, Jinsong; Chapski, Douglas J.; Rau, Christoph D. et al. (2016): RBFOX1-mediated RNA splicing regulates cardiac hypertrophy and heart failure. In: *The Journal of clinical investigation* 126 (1), S. 195–206. DOI: 10.1172/JCI84015.
- Ge, H.; Manley, J. L. (1990): A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. In: *Cell* 62 (1), S. 25–34. DOI: 10.1016/0092-8674(90)90236-8.
- Gehring, Niels H.; Wahle, Elmar; Fischer, Utz (2017): Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. In: *Trends in biochemical sciences* 42 (5), S. 369–382. DOI: 10.1016/j.tibs.2017.02.004.
- Gerstberger, Stefanie; Meyer, Cindy; Benjamin-Hong, Sigi; Rodriguez, Joe; Briskin, Daniel; Bognanni, Claudia et al. (2017): The Conserved RNA Exonuclease Rexo5 Is Required for 3' End Maturation of 28S rRNA, 5S rRNA, and snoRNAs. In: *Cell reports* 21 (3), S. 758–772. DOI: 10.1016/j.celrep.2017.09.067.
- Giudice, Jimena; Xia, Zheng; Wang, Eric T.; Scavuzzo, Marissa A.; Ward, Amanda J.; Kalsotra, Auinash et al. (2014): Alternative splicing regulates vesicular trafficking genes in

- cardiomyocytes during postnatal heart development. In: *Nature communications* 5, S. 3603. DOI: 10.1038/ncomms4603.
- Gogolevskaya, Irina K.; Koval, Anastasia P.; Kramerov, Dmitri A. (2005): Evolutionary history of 4.5SH RNA. In: *Molecular biology and evolution* 22 (7), S. 1546–1554. DOI: 10.1093/molbev/msi140.
- Goodarzi, Hani; Nguyen, Hoang C. B.; Zhang, Steven; Dill, Brian D.; Molina, Henrik; Tavazoie, Sohail F. (2016): Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. In: *Cell* 165 (6), S. 1416–1427. DOI: 10.1016/j.cell.2016.05.046.
- Gozani, O.; Feld, R.; Reed, R. (1996): Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. In: *Genes & development* 10 (2), S. 233–243. DOI: 10.1101/gad.10.2.233.
- Gray, T. A.; Smithwick, M. J.; Schaldach, M. A.; Martone, D. L.; Graves, J. A.; McCarrey, J. R.; Nicholls, R. D. (1999): Concerted regulation and molecular evolution of the duplicated SNRPB'/B and SNRPN loci. In: *Nucleic acids research* 27 (23), S. 4577–4584. DOI: 10.1093/nar/27.23.4577.
- Greig, Jamie A.; Nguyen, Tu Anh; Lee, Michelle; Holehouse, Alex S.; Posey, Ammon E.; Pappu, Rohit V.; Jedd, Gregory (2020): Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. In: *Molecular cell* 77 (6), 1237-1250.e4. DOI: 10.1016/j.molcel.2020.01.025.
- Grosshans, Helge; Filipowicz, Witold (2008): Molecular biology: the expanding world of small RNAs. In: *Nature* 451 (7177), S. 414–416. DOI: 10.1038/451414a.
- Gunawan, Felix; Gentile, Alessandra; Fukuda, Ryuichi; Tsedeke, Ayele Taddese; Jiménez-Amilburu, Vanesa; Ramadass, Radhan et al. (2019): Focal adhesions are essential to drive zebrafish heart valve morphogenesis. In: *The Journal of cell biology* 218 (3), S. 1039–1054. DOI: 10.1083/jcb.201807175.
- Gutschner, Tony; Hämmerle, Monika; Diederichs, Sven (2013): MALAT1 -- a paradigm for long noncoding RNA function in cancer. In: *Journal of molecular medicine (Berlin, Germany)* 91 (7), S. 791–801. DOI: 10.1007/s00109-013-1028-y.
- Hafner, Markus; Landgraf, Pablo; Ludwig, Janos; Rice, Amanda; Ojo, Tolulope; Lin, Carolina et al. (2008): Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. In: *Methods (San Diego, Calif.)* 44 (1), S. 3–12. DOI: 10.1016/j.ymeth.2007.09.009.

- Hamilton, T. C.; Young, R. C.; McKoy, W. M.; Grotzinger, K. R.; Green, J. A.; Chu, E. W. et al. (1983): Characterization of a human ovarian carcinoma cell line (NIH:OVCAR-3) with androgen and estrogen receptors. In: *Cancer research* 43 (11), S. 5379–5389.
- Harada, F.; Kato, N. (1980): Nucleotide sequences of 4.5S RNAs associated with poly(A)-containing RNAs of mouse and hamster cells. In: *Nucleic acids research* 8 (6), S. 1273–1285. DOI: 10.1093/nar/8.6.1273.
- Harada, F.; Kato, N.; Hoshino, H. (1979): Series of 4.5S RNAs associated with poly(A)-containing RNAs of rodent cells. In: *Nucleic acids research* 7 (4), S. 909–917. DOI: 10.1093/nar/7.4.909.
- Hasimbegovic, Ena; Schweiger, Victor; Kastner, Nina; Spannbaauer, Andreas; Traxler, Denise; Lukovic, Dominika et al. (2021): Alternative Splicing in Cardiovascular Disease-A Survey of Recent Findings. In: *Genes* 12 (9). DOI: 10.3390/genes12091457.
- Hastings, M. L.; Krainer, A. R. (2001): Pre-mRNA splicing in the new millennium. In: *Current opinion in cell biology* 13 (3), S. 302–309. DOI: 10.1016/s0955-0674(00)00212-x.
- Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K. et al. (2001): Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. In: *DNA research : an international journal for rapid publication of reports on genes and genomes* 8 (1), S. 11–22. DOI: 10.1093/dnares/8.1.11.
- Haynes, Chad; Iakoucheva, Lilia M. (2006): Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. In: *Nucleic acids research* 34 (1), S. 305–312. DOI: 10.1093/nar/gkj424.
- Heinig, Matthias; Adriaens, Michiel E.; Schafer, Sebastian; van Deutekom, Hanneke W. M.; Lodder, Elisabeth M.; Ware, James S. et al. (2017): Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. In: *Genome biology* 18 (1), S. 170. DOI: 10.1186/s13059-017-1286-z.
- Hirose, Tetsuro; Ninomiya, Kensuke; Nakagawa, Shinichi; Yamazaki, Tomohiro (2023): A guide to membraneless organelles and their various roles in gene regulation. In: *Nature reviews. Molecular cell biology* 24 (4), S. 288–304. DOI: 10.1038/s41580-022-00558-8.
- Hirose, Tetsuro; Virnicchi, Giorgio; Tanigawa, Akie; Naganuma, Takao; Li, Ruohan; Kimura, Hiroshi et al. (2014): NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. In: *Molecular biology of the cell* 25 (1), S. 169–183. DOI: 10.1091/mbc.E13-09-0558.

- Hirose, Yutaka; Harada, Fumio (2008): Mouse nucleolin binds to 4.5S RNAh, a small noncoding RNA. In: *Biochemical and biophysical research communications* 365 (1), S. 62–68. DOI: 10.1016/j.bbrc.2007.10.117.
- Hirth, Sofia; Bühler, Anja; Bührdel, John B.; Rudeck, Steven; Dahme, Tillman; Rottbauer, Wolfgang; Just, Steffen (2016): Paxillin and Focal Adhesion Kinase (FAK) Regulate Cardiac Contractility in the Zebrafish Heart. In: *PloS one* 11 (3), e0150323. DOI: 10.1371/journal.pone.0150323.
- Hu, T.; Miller, C. M.; Ridder, G. M.; Aardema, M. J. (1999): Characterization of p53 in Chinese hamster cell lines CHO-K1, CHO-WBL, and CHL: implications for genotoxicity testing. In: *Mutation research* 426 (1), S. 51–62. DOI: 10.1016/s0027-5107(99)00077-9.
- Hu, Ying-Li; Lu, Shaoying; Szeto, Kai W.; Sun, Jie; Wang, Yingxiao; Lasheras, Juan C.; Chien, Shu (2014): FAK and Paxillin dynamics at focal adhesions in the protrusions of migrating cells. In: *Scientific reports* 4, S. 6024. DOI: 10.1038/srep06024.
- Hu, Zhi-Yan; Wang, Xiao-Yan; Guo, Wen-Bin; Xie, Lin-Ying; Huang, Yu-Qi; Liu, Yan-Ping et al. (2016): Long non-coding RNA MALAT1 increases AKAP-9 expression by promoting SRPK1-catalyzed SRSF1 phosphorylation in colorectal cancer cells. In: *Oncotarget* 7 (10), S. 11733–11743. DOI: 10.18632/oncotarget.7367.
- Huang, Da Wei; Sherman, Brad T.; Lempicki, Richard A. (2009): Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. In: *Nature protocols* 4 (1), S. 44–57. DOI: 10.1038/nprot.2008.211.
- Huang, Shu-Ching; Ou, Alexander C.; Park, Jennie; Yu, Faye; Yu, Brian; Lee, Angela et al. (2012): RBFOX2 promotes protein 4.1R exon 16 selection via U1 snRNP recruitment. In: *Molecular and cellular biology* 32 (2), S. 513–526. DOI: 10.1128/MCB.06423-11.
- Hutchinson, John N.; Ensminger, Alexander W.; Clemson, Christine M.; Lynch, Christopher R.; Lawrence, Jeanne B.; Chess, Andrew (2007): A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. In: *BMC genomics* 8, S. 39. DOI: 10.1186/1471-2164-8-39.
- Ilik, İbrahim Avşar; Malszycki, Michal; Lübke, Anna Katharina; Schade, Claudia; Meierhofer, David; Aktaş, Tuğçe (2020): SON and SRRM2 are essential for nuclear speckle formation. In: *eLife* 9. DOI: 10.7554/eLife.60579.
- Ingle, Evan (2008): Src family kinases: regulation of their activities, levels and identification of new pathways. In: *Biochimica et biophysica acta* 1784 (1), S. 56–65. DOI: 10.1016/j.bbapap.2007.08.012.

- Ishida, Kentaro; Miyauchi, Kenjyo; Kimura, Yuko; Mito, Mari; Okada, Shunpei; Suzuki, Tsutomu; Nakagawa, Shinichi (2015): Regulation of gene expression via retrotransposon insertions and the noncoding RNA 4.5S RNAH. In: *Genes to cells : devoted to molecular & cellular mechanisms* 20 (11), S. 887–901. DOI: 10.1111/gtc.12280.
- Jelinek, W.; Leinwand, L. (1978): Low molecular weight RNAs hydrogen-bonded to nuclear and cytoplasmic poly(A)-terminated RNA from cultured Chinese hamster ovary cells. In: *Cell* 15 (1), S. 205–214. DOI: 10.1016/0092-8674(78)90095-8.
- Ji, Ping; Diederichs, Sven; Wang, Wenbing; Böing, Sebastian; Metzger, Ralf; Schneider, Paul M. et al. (2003): MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. In: *Oncogene* 22 (39), S. 8031–8041. DOI: 10.1038/sj.onc.1206928.
- Jin, Yui; Suzuki, Hitoshi; Maegawa, Shingo; Endo, Hitoshi; Sugano, Sumio; Hashimoto, Katsuyuki et al. (2003): A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. In: *The EMBO journal* 22 (4), S. 905–912. DOI: 10.1093/emboj/cdg089.
- Johnsson, Per; Lipovich, Leonard; Grandér, Dan; Morris, Kevin V. (2014): Evolutionary conservation of long non-coding RNAs; sequence, structure, function. In: *Biochimica et biophysica acta* 1840 (3), S. 1063–1071. DOI: 10.1016/j.bbagen.2013.10.035.
- Jühling, Frank; Mörl, Mario; Hartmann, Roland K.; Sprinzl, Mathias; Stadler, Peter F.; Pütz, Joern (2009): tRNADB 2009: compilation of tRNA sequences and tRNA genes. In: *Nucleic acids research* 37 (Database issue), D159-62. DOI: 10.1093/nar/gkn772.
- Jumper, John; Evans, Richard; Pritzel, Alexander; Green, Tim; Figurnov, Michael; Ronneberger, Olaf et al. (2021): Highly accurate protein structure prediction with AlphaFold. In: *Nature* 596 (7873), S. 583–589. DOI: 10.1038/s41586-021-03819-2.
- Kalsotra, Auinash; Wang, Kun; Li, Pei-Feng; Cooper, Thomas A. (2010): MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. In: *Genes & development* 24 (7), S. 653–658. DOI: 10.1101/gad.1894310.
- Kalsotra, Auinash; Xiao, Xinshu; Ward, Amanda J.; Castle, John C.; Johnson, Jason M.; Burge, Christopher B.; Cooper, Thomas A. (2008): A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. In: *Proceedings of the National Academy of Sciences of the United States of America* 105 (51), S. 20333–20338. DOI: 10.1073/pnas.0809045105.
- Kanoldt, Verena; Kluger, Carleen; Barz, Christiane; Schweizer, Anna-Lena; Ramanujam, Deepak; Windgasse, Lukas et al. (2020): Metavinculin modulates force transduction in cell

- adhesion sites. In: *Nature communications* 11 (1), S. 6403. DOI: 10.1038/s41467-020-20125-z.
- Karni, Rotem; Stanchina, Elisa de; Lowe, Scott W.; Sinha, Rahul; Mu, David; Krainer, Adrian R. (2007): The gene encoding the splicing factor SF2/ASF is a proto-oncogene. In: *Nature structural & molecular biology* 14 (3), S. 185–193. DOI: 10.1038/nsmb1209.
- Keshava Prasad, T. S.; Goel, Renu; Kandasamy, Kumaran; Keerthikumar, Shivakumar; Kumar, Sameer; Mathivanan, Suresh et al. (2009): Human Protein Reference Database--2009 update. In: *Nucleic acids research* 37 (Database issue), D767-72. DOI: 10.1093/nar/gkn892.
- Kiledjian, M.; Dreyfuss, G. (1992): Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. In: *The EMBO journal* 11 (7), S. 2655–2664. DOI: 10.1002/j.1460-2075.1992.tb05331.x.
- Kim, Laura Y.; Thompson, Peter M.; Lee, Hyunna T.; Pershad, Mihir; Campbell, Sharon L.; Alushin, Gregory M. (2016): The Structural Basis of Actin Organization by Vinculin and Metavinculin. In: *Journal of molecular biology* 428 (1), S. 10–25. DOI: 10.1016/j.jmb.2015.09.031.
- Kimes, B. W.; Brandt, B. L. (1976): Properties of a clonal muscle cell line from rat heart. In: *Experimental cell research* 98 (2), S. 367–381. DOI: 10.1016/0014-4827(76)90447-x.
- Köhn, Marcel; Ihling, Christian; Sinz, Andrea; Krohn, Knut; Hüttelmaier, Stefan (2015): The Y3** ncRNA promotes the 3' end processing of histone mRNAs. In: *Genes & development* 29 (19), S. 1998–2003. DOI: 10.1101/gad.266486.115.
- Köhn, Marcel; Pazaitis, Nikolaos; Hüttelmaier, Stefan (2013): Why YRNAs? About Versatile RNAs and Their Functions. In: *Biomolecules* 3 (1), S. 143–156. DOI: 10.3390/biom3010143.
- Kopp, Florian; Mendell, Joshua T. (2018): Functional Classification and Experimental Dissection of Long Noncoding RNAs. In: *Cell* 172 (3), S. 393–407. DOI: 10.1016/j.cell.2018.01.011.
- Krainer, A. R.; Conway, G. C.; Kozak, D. (1990a): Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. In: *Genes & development* 4 (7), S. 1158–1171. DOI: 10.1101/gad.4.7.1158.
- Krainer, A. R.; Conway, G. C.; Kozak, D. (1990b): The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. In: *Cell* 62 (1), S. 35–42. DOI: 10.1016/0092-8674(90)90237-9.
- Kramerov, Dimitri A.; Vassetzky, Nikita S. (2005): Short retroposons in eukaryotic genomes. In: *International review of cytology* 247, S. 165–221. DOI: 10.1016/S0074-7696(05)47004-7.

- Kuehner, Jason N.; Pearson, Erika L.; Moore, Claire (2011): Unravelling the means to an end: RNA polymerase II transcription termination. In: *Nature reviews. Molecular cell biology* 12 (5), S. 283–294. DOI: 10.1038/nrm3098.
- Kuroyanagi, Hidehito (2009): Fox-1 family of RNA-binding proteins. In: *Cellular and molecular life sciences : CMLS* 66 (24), S. 3895–3907. DOI: 10.1007/s00018-009-0120-5.
- Labuda, D.; Zietkiewicz, E. (1994): Evolution of secondary structure in the family of 7SL-like RNAs. In: *Journal of molecular evolution* 39 (5), S. 506–518. DOI: 10.1007/BF00173420.
- Lai, M. C.; Lin, R. I.; Huang, S. Y.; Tsai, C. W.; Tarn, W. Y. (2000): A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. In: *The Journal of biological chemistry* 275 (11), S. 7950–7957. DOI: 10.1074/jbc.275.11.7950.
- Lambert, Nicole; Robertson, Alex; Jangi, Mohini; McGeary, Sean; Sharp, Phillip A.; Burge, Christopher B. (2014): RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. In: *Molecular cell* 54 (5), S. 887–900. DOI: 10.1016/j.molcel.2014.04.016.
- Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J. et al. (2001): Initial sequencing and analysis of the human genome. In: *Nature* 409 (6822), S. 860–921. DOI: 10.1038/35057062.
- Landim-Vieira, Maicon; Schipper, Joslyn M.; Pinto, J. Renato; Chase, P. Bryant (2020): Cardiomyocyte nuclearity and ploidy: when is double trouble? In: *Journal of muscle research and cell motility* 41 (4), S. 329–340. DOI: 10.1007/s10974-019-09545-7.
- Leinwand, L. A.; Wydro, R. M.; Nadal-Ginard, B. (1982): Small RNA molecules related to the Alu family of repetitive DNA sequences. In: *Molecular and cellular biology* 2 (11), S. 1320–1330. DOI: 10.1128/mcb.2.11.1320-1330.1982.
- Li, Bingzong; Chen, Ping; Qu, Jing; Shi, Lei; Zhuang, Wenyue; Fu, Jinxiang et al. (2014): Activation of LTBP3 gene by a long noncoding RNA (lncRNA) MALAT1 transcript in mesenchymal stem cells from multiple myeloma. In: *The Journal of biological chemistry* 289 (42), S. 29365–29375. DOI: 10.1074/jbc.M114.572693.
- Lin, Shengrong; Fu, Xiang-Dong (2007): SR proteins and related factors in alternative splicing. In: *Advances in experimental medicine and biology* 623, S. 107–122. DOI: 10.1007/978-0-387-77374-2_7.
- Liu, H. X.; Zhang, M.; Krainer, A. R. (1998): Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. In: *Genes & development* 12 (13), S. 1998–2012. DOI: 10.1101/gad.12.13.1998.

- Livak, K. J.; Schmittgen, T. D. (2001): Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. In: *Methods (San Diego, Calif.)* 25 (4), S. 402–408. DOI: 10.1006/meth.2001.1262.
- Lovci, Michael T.; Ghanem, Dana; Marr, Henry; Arnold, Justin; Gee, Sherry; Parra, Marilyn et al. (2013): Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. In: *Nature structural & molecular biology* 20 (12), S. 1434–1442. DOI: 10.1038/nsmb.2699.
- Lunde, Bradley M.; Moore, Claire; Varani, Gabriele (2007): RNA-binding proteins: modular design for efficient function. In: *Nature reviews. Molecular cell biology* 8 (6), S. 479–490. DOI: 10.1038/nrm2178.
- Maeder, Corina; Kutach, Alan K.; Guthrie, Christine (2009): ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. In: *Nature structural & molecular biology* 16 (1), S. 42–48. DOI: 10.1038/nsmb.1535.
- Malakar, Pushkar; Shilo, Asaf; Mogilevsky, Adi; Stein, Ilan; Pikarsky, Eli; Nevo, Yuval et al. (2017): Long Noncoding RNA MALAT1 Promotes Hepatocellular Carcinoma Development by SRSF1 Upregulation and mTOR Activation. In: *Cancer research* 77 (5), S. 1155–1167. DOI: 10.1158/0008-5472.CAN-16-1508.
- Manley, James L.; Krainer, Adrian R. (2010): A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). In: *Genes & development* 24 (11), S. 1073–1074. DOI: 10.1101/gad.1934910.
- Mao, Yuntao S.; Zhang, Bin; Spector, David L. (2011): Biogenesis and function of nuclear bodies. In: *Trends in genetics : TIG* 27 (8), S. 295–306. DOI: 10.1016/j.tig.2011.05.006.
- Maraia, Richard J.; Mattijssen, Sandy; Cruz-Gallardo, Isabel; Conte, Maria R. (2017): The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. In: *Wiley interdisciplinary reviews. RNA* 8 (6). DOI: 10.1002/wrna.1430.
- Martin, Arnold; Schneider, Susanne; Schwer, Beate (2002): Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. In: *The Journal of biological chemistry* 277 (20), S. 17743–17750. DOI: 10.1074/jbc.M200762200.
- Mattick, John S.; Makunin, Igor V. (2006): Non-coding RNA. In: *Human molecular genetics* 15 Spec No 1, R17-29. DOI: 10.1093/hmg/ddl046.
- McCain, Megan L.; Lee, Hyungsuk; Aratyn-Schaus, Yvonne; Kléber, André G.; Parker, Kevin Kit (2012): Cooperative coupling of cell-matrix and cell-cell adhesions in cardiac muscle. In: *Proceedings of the National Academy of Sciences of the United States of America* 109 (25), S. 9881–9886. DOI: 10.1073/pnas.1203007109.

- McGinn, Jon; Czech, Benjamin (2014): Small RNA library construction for high-throughput sequencing. In: *Methods in molecular biology (Clifton, N.J.)* 1093, S. 195–208. DOI: 10.1007/978-1-62703-694-8_16.
- Ménard, C.; Pupier, S.; Mornet, D.; Kitzmann, M.; Nargeot, J.; Lory, P. (1999): Modulation of L-type calcium channel expression during retinoic acid-induced differentiation of H9C2 cardiac cells. In: *The Journal of biological chemistry* 274 (41), S. 29063–29070. DOI: 10.1074/jbc.274.41.29063.
- Merendino, L.; Guth, S.; Bilbao, D.; Martínez, C.; Valcárcel, J. (1999): Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. In: *Nature* 402 (6763), S. 838–841. DOI: 10.1038/45602.
- Miyagawa, Ryu; Tano, Keiko; Mizuno, Rie; Nakamura, Yo; Ijiri, Kenichi; Rakwal, Randeep et al. (2012): Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. In: *RNA (New York, N.Y.)* 18 (4), S. 738–751. DOI: 10.1261/rna.028639.111.
- Montemayor, Eric J.; Didychuk, Allison L.; Yake, Allyson D.; Sidhu, Gurnimrat K.; Brow, David A.; Butcher, Samuel E. (2018): Architecture of the U6 snRNP reveals specific recognition of 3'-end processed U6 snRNA. In: *Nature communications* 9 (1), S. 1749. DOI: 10.1038/s41467-018-04145-4.
- Mori, Yoshinobu; Ichiyanagi, Kenji (2021): miRNA-seq for Expression Analysis of SINE RNAs and Other Medium-Length Non-Coding RNAs. In: *Mobile DNA* 12 (1), S. 15. DOI: 10.1186/s13100-021-00245-z.
- Nakagawa, Shinichi; Ip, Joanna Y.; Shioi, Go; Tripathi, Vidisha; Zong, Xinying; Hirose, Tetsuro; Prasanth, Kannanganattu V. (2012): Malat1 is not an essential component of nuclear speckles in mice. In: *RNA (New York, N.Y.)* 18 (8), S. 1487–1499. DOI: 10.1261/rna.033217.112.
- Nakahata, Shingo; Kawamoto, Sachiyo (2005): Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. In: *Nucleic acids research* 33 (7), S. 2078–2089. DOI: 10.1093/nar/gki338.
- Neerathilingam, Muniasamy; Bairy, Sneha G.; Mysore, Sumukh (2016): Deciphering Mode of Action of Functionally Important Regions in the Intrinsically Disordered Paxillin (Residues 1-313) Using Its Interaction with FAT (Focal Adhesion Targeting Domain of Focal Adhesion Kinase). In: *PLoS one* 11 (2), e0150153. DOI: 10.1371/journal.pone.0150153.

- Nguyen, V. T.; Kiss, T.; Michels, A. A.; Bensaude, O. (2001): 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. In: *Nature* 414 (6861), S. 322–325. DOI: 10.1038/35104581.
- Nomura, Yuichiro; Roston, Daniel; Montemayor, Eric J.; Cui, Qiang; Butcher, Samuel E. (2018): Structural and mechanistic basis for preferential deadenylation of U6 snRNA by Ubs1. In: *Nucleic acids research* 46 (21), S. 11488–11501. DOI: 10.1093/nar/gky812.
- Ogle, J. M.; Brodersen, D. E.; Clemons, W. M.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. (2001): Recognition of cognate transfer RNA by the 30S ribosomal subunit. In: *Science (New York, N.Y.)* 292 (5518), S. 897–902. DOI: 10.1126/science.1060612.
- Oldfield, Christopher J.; Dunker, A. Keith (2014): Intrinsically disordered proteins and intrinsically disordered protein regions. In: *Annual review of biochemistry* 83, S. 553–584. DOI: 10.1146/annurev-biochem-072711-164947.
- Oztug Durer, Zeynep A.; McGillivray, Rebecca M.; Kang, Hyeran; Elam, W. Austin; Vizcarra, Christina L.; Hanein, Dorit et al. (2015): Metavinculin Tunes the Flexibility and the Architecture of Vinculin-Induced Bundles of Actin Filaments. In: *Journal of molecular biology* 427 (17), S. 2782–2798. DOI: 10.1016/j.jmb.2015.07.005.
- Pandit, Shatakshi; Zhou, Yu; Shiue, Lily; Coutinho-Mansfield, Gabriela; Li, Hairi; Qiu, Jinsong et al. (2013): Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. In: *Molecular cell* 50 (2), S. 223–235. DOI: 10.1016/j.molcel.2013.03.001.
- Park, Ji Yeon; Li, Wencheng; Zheng, Dinghai; Zhai, Peiyong; Zhao, Yun; Matsuda, Takahisa et al. (2011): Comparative analysis of mRNA isoform expression in cardiac hypertrophy and development reveals multiple post-transcriptional regulatory modules. In: *PloS one* 6 (7), e22391. DOI: 10.1371/journal.pone.0022391.
- Park, Jong-Lyul; Lee, Yeon-Su; Kunkeaw, Nawapol; Kim, Seon-Young; Kim, In-Hoo; Lee, Yong Sun (2017): Epigenetic regulation of noncoding RNA transcription by mammalian RNA polymerase III. In: *Epigenomics* 9 (2), S. 171–187. DOI: 10.2217/epi-2016-0108.
- Paz, Sean; Ritchie, Anastasia; Mauer, Christopher; Caputi, Massimo (2021): The RNA binding protein SRSF1 is a master switch of gene expression and regulation in the immune system. In: *Cytokine & growth factor reviews* 57, S. 19–26. DOI: 10.1016/j.cytogfr.2020.10.008.
- Pedrotti, Simona; Giudice, Jimena; Dagnino-Acosta, Adan; Knoblauch, Mark; Singh, Ravi K.; Hanna, Amy et al. (2015): The RNA-binding protein Rbfox1 regulates splicing required for skeletal muscle structure and function. In: *Human molecular genetics* 24 (8), S. 2360–2374. DOI: 10.1093/hmg/ddv003.

- Podnar, Jessica; Deiderick, Heather; Huerta, Gabriella; Hunicke-Smith, Scott (2014): Next-Generation Sequencing RNA-Seq Library Construction. In: *Current protocols in molecular biology* 106, 4.21.1-4.21.19. DOI: 10.1002/0471142727.mb0421s106.
- Quach, Navaline L.; Biressi, Stefano; Reichardt, Louis F.; Keller, Charles; Rando, Thomas A. (2009): Focal adhesion kinase signaling regulates the expression of caveolin 3 and beta1 integrin, genes essential for normal myoblast fusion. In: *Molecular biology of the cell* 20 (14), S. 3422–3435. DOI: 10.1091/mbc.e09-02-0175.
- Rangarajan, Erumbi S.; Lee, Jun Hyuck; Yogesha, S. D.; Izard, Tina (2010): A helix replacement mechanism directs metavinculin functions. In: *PloS one* 5 (5), e10679. DOI: 10.1371/journal.pone.0010679.
- Rieder, Dietmar; Ploner, Christian; Krogsdam, Anne M.; Stocker, Gernot; Fischer, Maria; Scheideler, Marcel et al. (2014): Co-expressed genes prepositioned in spatial neighborhoods stochastically associate with SC35 speckles and RNA polymerase II factories. In: *Cellular and molecular life sciences : CMLS* 71 (9), S. 1741–1759. DOI: 10.1007/s00018-013-1465-3.
- Riggs, Claire L.; Kedersha, Nancy; Ivanov, Pavel; Anderson, Paul (2020): Mammalian stress granules and P bodies at a glance. In: *Journal of cell science* 133 (16). DOI: 10.1242/jcs.242487.
- Rochlin, Kate; Yu, Shannon; Roy, Sudipto; Baylies, Mary K. (2010): Myoblast fusion: when it takes more to make one. In: *Developmental biology* 341 (1), S. 66–83. DOI: 10.1016/j.ydbio.2009.10.024.
- Roscigno, R. F.; Garcia-Blanco, M. A. (1995): SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. In: *RNA (New York, N. Y.)* 1 (7), S. 692–706.
- Sakshi, Shukla; Jayasuriya, Ravichandran; Ganesan, Kumar; Xu, Baojun; Ramkumar, Kunka Mohanram (2021): Role of circRNA-miRNA-mRNA interaction network in diabetes and its associated complications. In: *Molecular therapy. Nucleic acids* 26, S. 1291–1302. DOI: 10.1016/j.omtn.2021.11.007.
- Samarel, Allen M. (2014): Focal adhesion signaling in heart failure. In: *Pflugers Archiv : European journal of physiology* 466 (6), S. 1101–1111. DOI: 10.1007/s00424-014-1456-8.
- Sanford, Jeremy R.; Wang, Xin; Mort, Matthew; Vanduyne, Natalia; Cooper, David N.; Mooney, Sean D. et al. (2009): Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. In: *Genome research* 19 (3), S. 381–394. DOI: 10.1101/gr.082503.108.
- Scheckel, Claudia; Drapeau, Elodie; Frias, Maria A.; Park, Christopher Y.; Fak, John; Zucker-Scharff, Ilana et al. (2016): Regulatory consequences of neuronal ELAV-like protein

binding to coding and non-coding RNAs in human brain. In: *eLife* 5. DOI: 10.7554/eLife.10421.

Schoeniger, L. O.; Jelinek, W. R. (1986): 4.5S RNA is encoded by hundreds of tandemly linked genes, has a short half-life, and is hydrogen bonded in vivo to poly(A)-terminated RNAs in the cytoplasm of cultured mouse cells. In: *Molecular and cellular biology* 6 (5), S. 1508–1519. DOI: 10.1128/mcb.6.5.1508-1519.1986.

Schramm, Laura; Hernandez, Nouria (2002): Recruitment of RNA polymerase III to its target promoters. In: *Genes & development* 16 (20), S. 2593–2620. DOI: 10.1101/gad.1018902.

Schreier, Barbara; Rabe, Sindy; Schneider, Bettina; Bretschneider, Maria; Rupp, Sebastian; Ruhs, Stefanie et al. (2013): Loss of epidermal growth factor receptor in vascular smooth muscle cells and cardiomyocytes causes arterial hypotension and cardiac hypertrophy. In: *Hypertension (Dallas, Tex. : 1979)* 61 (2), S. 333–340. DOI: 10.1161/HYPERTENSIONAHA.112.196543.

Scotti, Marina M.; Swanson, Maurice S. (2016): RNA mis-splicing in disease. In: *Nature reviews. Genetics* 17 (1), S. 19–32. DOI: 10.1038/nrg.2015.3.

Selmer, Maria; Dunham, Christine M.; Murphy, Frank V.; Weixlbaumer, Albert; Petry, Sabine; Kelley, Ann C. et al. (2006): Structure of the 70S ribosome complexed with mRNA and tRNA. In: *Science (New York, N.Y.)* 313 (5795), S. 1935–1942. DOI: 10.1126/science.1131127.

Séraphin, B. (1995): Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. In: *The EMBO journal* 14 (9), S. 2089–2098. DOI: 10.1002/j.1460-2075.1995.tb07200.x.

Shen, Haihong; Green, Michael R. (2006): RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. In: *Genes & development* 20 (13), S. 1755–1765. DOI: 10.1101/gad.1422106.

Shen, Haihong; Kan, Julie L. C.; Green, Michael R. (2004): Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. In: *Molecular cell* 13 (3), S. 367–376. DOI: 10.1016/s1097-2765(04)00025-5.

Sheth, Nihar; Roca, Xavier; Hastings, Michelle L.; Roeder, Ted; Krainer, Adrian R.; Sachidanandam, Ravi (2006): Comprehensive splice-site analysis using comparative genomics. In: *Nucleic acids research* 34 (14), S. 3955–3967. DOI: 10.1093/nar/gkl556.

Shimba, S.; Reddy, R. (1994): Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for gamma-monomethyl-capped small RNAs. In: *The Journal of biological chemistry* 269 (17), S. 12419–12423.

- Shin, Yongdae; Brangwynne, Clifford P. (2017): Liquid phase condensation in cell physiology and disease. In: *Science (New York, N.Y.)* 357 (6357). DOI: 10.1126/science.aaf4382.
- Shopland, Lindsay S.; Johnson, Carol V.; Byron, Meg; McNeil, John; Lawrence, Jeanne B. (2003): Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. In: *The Journal of cell biology* 162 (6), S. 981–990. DOI: 10.1083/jcb.200303131.
- Siliciano, J. D.; Craig, S. W. (1982): Meta-vinculin--a vinculin-related protein with solubility properties of a membrane protein. In: *Nature* 300 (5892), S. 533–535. DOI: 10.1038/300533a0.
- Simons, F. H.; Rutjes, S. A.; van Venrooij, W. J.; Pruijn, G. J. (1996): The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA. In: *RNA (New York, N.Y.)* 2 (3), S. 264–273.
- Singh, Ravinder; Valcárcel, Juan (2005): Building specificity with nonspecific RNA-binding proteins. In: *Nature structural & molecular biology* 12 (8), S. 645–653. DOI: 10.1038/nsmb961.
- Sinha, K. M.; Gu, J.; Chen, Y.; Reddy, R. (1998): Adenylation of small RNAs in human cells. Development of a cell-free system for accurate adenylation on the 3'-end of human signal recognition particle RNA. In: *The Journal of biological chemistry* 273 (12), S. 6853–6859. DOI: 10.1074/jbc.273.12.6853.
- Spector, David L. (2006): SnapShot: Cellular bodies. In: *Cell* 127 (5), S. 1071. DOI: 10.1016/j.cell.2006.11.026.
- Spector, David L.; Lamond, Angus I. (2011): Nuclear speckles. In: *Cold Spring Harbor perspectives in biology* 3 (2). DOI: 10.1101/cshperspect.a000646.
- Staknis, D.; Reed, R. (1994): SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. In: *Molecular and cellular biology* 14 (11), S. 7670–7682. DOI: 10.1128/mcb.14.11.7670-7682.1994.
- Stark, Rory; Grzelak, Marta; Hadfield, James (2019): RNA-Sequencing: the teenage years. In: *Nature reviews. Genetics* 20 (11), S. 631–656. DOI: 10.1038/s41576-019-0150-2.
- Statello, Luisa; Guo, Chun-Jie; Chen, Ling-Ling; Huarte, Maite (2021): Gene regulation by long non-coding RNAs and its biological functions. In: *Nature reviews. Molecular cell biology* 22 (2), S. 96–118. DOI: 10.1038/s41580-020-00315-9.
- Steitz, Thomas A. (2008): A structural understanding of the dynamic ribosome machine. In: *Nature reviews. Molecular cell biology* 9 (3), S. 242–253. DOI: 10.1038/nrm2352.

- Storz, Gisela; Vogel, Jörg; Wassarman, Karen M. (2011): Regulation by small RNAs in bacteria: expanding frontiers. In: *Molecular cell* 43 (6), S. 880–891. DOI: 10.1016/j.molcel.2011.08.022.
- Sun, Feng-Jie; Caetano-Anollés, Gustavo (2009): The evolutionary history of the structure of 5S ribosomal RNA. In: *Journal of molecular evolution* 69 (5), S. 430–443. DOI: 10.1007/s00239-009-9264-z.
- Sun, Shuying; Zhang, Zuo; Fregoso, Oliver; Krainer, Adrian R. (2012): Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. In: *RNA (New York, N.Y.)* 18 (2), S. 274–283. DOI: 10.1261/rna.030486.111.
- Tatosyan, K. A.; Koval, A. P.; Gogolevskaya, I. K.; Kramerov, D. A. (2017): 4.5SI and 4.5SH RNAs: Expression in various rodent organs and abundance and distribution in the cell. In: *Molekuliarnaia biologii* 51 (1), S. 142–149. DOI: 10.7868/S0026898417010177.
- Täuber, Hendrik; Hüttelmaier, Stefan; Köhn, Marcel (2019): POLIII-derived non-coding RNAs acting as scaffolds and decoys. In: *Journal of molecular cell biology* 11 (10), S. 880–885. DOI: 10.1093/jmcb/mjz049.
- Tolbert, Caitlin E.; BurrIDGE, Keith; Campbell, Sharon L. (2013): Vinculin regulation of F-actin bundle formation: what does it mean for the cell? In: *Cell adhesion & migration* 7 (2), S. 219–225. DOI: 10.4161/cam.23184.
- Torres, Adrian Gabriel; Reina, Oscar; Stephan-Otto Attolini, Camille; Ribas de Pouplana, Lluís (2019): Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. In: *Proceedings of the National Academy of Sciences of the United States of America* 116 (17), S. 8451–8456. DOI: 10.1073/pnas.1821120116.
- Trinchieri, G.; Aden, D. P.; Knowles, B. B. (1976): Cell-mediated cytotoxicity to SV40-specific tumour-associated antigens. In: *Nature* 261 (5558), S. 312–314. DOI: 10.1038/261312a0.
- Tripathi, Vidisha; Ellis, Jonathan D.; Shen, Zhen; Song, David Y.; Pan, Qun; Watt, Andrew T. et al. (2010): The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. In: *Molecular cell* 39 (6), S. 925–938. DOI: 10.1016/j.molcel.2010.08.011.
- Tripathi, Vidisha; Song, David Y.; Zong, Xinying; Shevtsov, Sergey P.; Hearn, Stephen; Fu, Xiang-Dong et al. (2012): SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. In: *Molecular biology of the cell* 23 (18), S. 3694–3706. DOI: 10.1091/mbc.E12-03-0206.
- Trippe, Ralf; Guschina, Elena; Hossbach, Markus; Urlaub, Henning; Lührmann, Reinhard; Benecke, Bernd-Joachim (2006): Identification, cloning, and functional analysis of the human

- U6 snRNA-specific terminal uridylyl transferase. In: *RNA (New York, N.Y.)* 12 (8), S. 1494–1504. DOI: 10.1261/rna.87706.
- Turowski, Tomasz W.; Tollervey, David (2016): Transcription by RNA polymerase III: insights into mechanism and regulation. In: *Biochemical Society transactions* 44 (5), S. 1367–1375. DOI: 10.1042/BST20160062.
- Turunen, Janne J.; Niemelä, Elina H.; Verma, Bhupendra; Frilander, Mikko J. (2013): The significant other: splicing by the minor spliceosome. In: *Wiley interdisciplinary reviews. RNA* 4 (1), S. 61–76. DOI: 10.1002/wrna.1141.
- Twyffels, Laure; Gueydan, Cyril; Kruys, Véronique (2011): Shuttling SR proteins: more than splicing factors. In: *The FEBS journal* 278 (18), S. 3246–3255. DOI: 10.1111/j.1742-4658.2011.08274.x.
- Uversky, V. N.; Gillespie, J. R.; Fink, A. L. (2000): Why are "natively unfolded" proteins unstructured under physiologic conditions? In: *Proteins* 41 (3), S. 415–427. DOI: 10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7.
- Valcárcel, J.; Gaur, R. K.; Singh, R.; Green, M. R. (1996): Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA corrected. In: *Science (New York, N.Y.)* 273 (5282), S. 1706–1709. DOI: 10.1126/science.273.5282.1706.
- Valcárcel, J.; Green, M. R. (1996): The SR protein family: pleiotropic functions in pre-mRNA splicing. In: *Trends in biochemical sciences* 21 (8), S. 296–301.
- Vassetzky, Nikita S.; Ten, Oleg A.; Kramerov, Dmitri A. (2003): B1 and related SINEs in mammalian genomes. In: *Gene* 319, S. 149–160. DOI: 10.1016/s0378-1119(03)00805-9.
- Visa, N.; Puvion-Dutilleul, F.; Bachellerie, J. P.; Puvion, E. (1993): Intranuclear distribution of U1 and U2 snRNAs visualized by high resolution in situ hybridization: revelation of a novel compartment containing U1 but not U2 snRNA in HeLa cells. In: *European journal of cell biology* 60 (2), S. 308–321.
- Wahl, Markus C.; Will, Cindy L.; Lührmann, Reinhard (2009): The spliceosome: design principles of a dynamic RNP machine. In: *Cell* 136 (4), S. 701–718. DOI: 10.1016/j.cell.2009.02.009.
- Wan, Ruixue; Yan, Chuangye; Bai, Rui; Wang, Lin; Huang, Min; Wong, Catherine C. L.; Shi, Yigong (2016): The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. In: *Science (New York, N.Y.)* 351 (6272), S. 466–475. DOI: 10.1126/science.aad6466.

- Wang, Eric T.; Sandberg, Rickard; Luo, Shujun; Khrebtkova, Irina; Zhang, Lu; Mayr, Christine et al. (2008): Alternative isoform regulation in human tissue transcriptomes. In: *Nature* 456 (7221), S. 470–476. DOI: 10.1038/nature07509.
- Wang, He; Chen, Yanmei; Li, Xinzhong; Chen, Guojun; Zhong, Lintao; Chen, Gangbing et al. (2016): Genome-wide analysis of alternative splicing during human heart development. In: *Scientific reports* 6, S. 35520. DOI: 10.1038/srep35520.
- Wang, Ke; Wang, Lantian; Wang, Jianshu; Chen, Suli; Shi, Min; Cheng, Hong (2018): Intronless mRNAs transit through nuclear speckles to gain export competence. In: *The Journal of cell biology* 217 (11), S. 3912–3929. DOI: 10.1083/jcb.201801184.
- Wang, Mingcong; Herrmann, Christina J.; Simonovic, Milan; Szklarczyk, Damian; Mering, Christian von (2015): Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. In: *Proteomics* 15 (18), S. 3163–3168. DOI: 10.1002/pmic.201400441.
- Wang, Xin; Juan, Liran; Lv, Junjie; Wang, Kejun; Sanford, Jeremy R.; Liu, Yunlong (2011): Predicting sequence and structural specificities of RNA binding regions recognized by splicing factor SRSF1. In: *BMC genomics* 12 Suppl 5 (Suppl 5), S8. DOI: 10.1186/1471-2164-12-S5-S8.
- West, Jason A.; Davis, Christopher P.; Sunwoo, Hongjae; Simon, Matthew D.; Sadreyev, Ruslan I.; Wang, Peggy I. et al. (2014): The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. In: *Molecular cell* 55 (5), S. 791–802. DOI: 10.1016/j.molcel.2014.07.012.
- West, Jason A.; Mito, Mari; Kurosaka, Satoshi; Takumi, Toru; Tanegashima, Chiharu; Chujo, Takeshi et al. (2016): Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. In: *The Journal of cell biology* 214 (7), S. 817–830. DOI: 10.1083/jcb.201601071.
- Weyn-Vanhentenryck, Sebastien M.; Feng, Huijuan; Ustianenko, Dmytro; Duffié, Rachel; Yan, Qinghong; Jacko, Martin et al. (2018): Precise temporal regulation of alternative splicing during neural development. In: *Nature communications* 9 (1), S. 2189. DOI: 10.1038/s41467-018-04559-0.
- Wickramasinghe, Vihandha O.; González-Porta, Mar; Perera, David; Bartolozzi, Arthur R.; Sibley, Christopher R.; Hallegger, Martina et al. (2015): Regulation of constitutive and alternative mRNA splicing across the human transcriptome by PRPF8 is determined by 5' splice site strength. In: *Genome biology* 16 (1), S. 201. DOI: 10.1186/s13059-015-0749-3.

- Wilkinson, Max E.; Charenton, Clément; Nagai, Kiyoshi (2020): RNA Splicing by the Spliceosome. In: *Annual review of biochemistry* 89, S. 359–388. DOI: 10.1146/annurev-biochem-091719-064225.
- Wilusz, Jeremy E.; JnBaptiste, Courtney K.; Lu, Laura Y.; Kuhn, Claus-D; Joshua-Tor, Leemor; Sharp, Phillip A. (2012): A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. In: *Genes & development* 26 (21), S. 2392–2407. DOI: 10.1101/gad.204438.112.
- Wohlgemuth, Ingo; Beringer, Malte; Rodnina, Marina V. (2006): Rapid peptide bond formation on isolated 50S ribosomal subunits. In: *EMBO reports* 7 (7), S. 699–703. DOI: 10.1038/sj.embor.7400732.
- Wolin, Sandra L.; Cedervall, Tommy (2002): The La protein. In: *Annual review of biochemistry* 71, S. 375–403. DOI: 10.1146/annurev.biochem.71.090501.150003.
- Wu, Huang; Yang, Li; Chen, Ling-Ling (2017): The Diversity of Long Noncoding RNAs and Their Generation. In: *Trends in genetics : TIG* 33 (8), S. 540–552. DOI: 10.1016/j.tig.2017.05.004.
- Wu, J. Y.; Maniatis, T. (1993): Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. In: *Cell* 75 (6), S. 1061–1070. DOI: 10.1016/0092-8674(93)90316-i.
- Xiang, Shengqi; Gapsys, Vytautas; Kim, Hai-Young; Bessonov, Sergey; Hsiao, He-Hsuan; Möhlmann, Sina et al. (2013): Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. In: *Structure (London, England : 1993)* 21 (12), S. 2162–2174. DOI: 10.1016/j.str.2013.09.014.
- Xu, Xiangdong; Yang, Dongmei; Ding, Jian-Hua; Wang, Wang; Chu, Pao-Hsien; Dalton, Nancy D. et al. (2005): ASF/SF2-regulated CaMKII δ alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. In: *Cell* 120 (1), S. 59–72. DOI: 10.1016/j.cell.2004.11.036.
- Yaffe, D.; Saxel, O. (1977): Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. In: *Nature* 270 (5639), S. 725–727. DOI: 10.1038/270725a0.
- Yamazaki, Tomohiro; Nakagawa, Shinichi; Hirose, Tetsuro (2019): Architectural RNAs for Membraneless Nuclear Body Formation. In: *Cold Spring Harbor symposia on quantitative biology* 84, S. 227–237. DOI: 10.1101/sqb.2019.84.039404.
- Yamazaki, Tomohiro; Souquere, Sylvie; Chujo, Takeshi; Kobelke, Simon; Chong, Yee Seng; Fox, Archa H. et al. (2018): Functional Domains of NEAT1 Architectural lncRNA Induce

- Paraspeckle Assembly through Phase Separation. In: *Molecular cell* 70 (6), 1038-1053.e7. DOI: 10.1016/j.molcel.2018.05.019.
- Yang, Cheng; Wu, Po-Yen; Tong, Li; Phan, John H.; Wang, May D. (2015): The impact of RNA-seq aligners on gene expression estimation. In: *ACM-BCB ... : the ... ACM Conference on Bioinformatics, Computational Biology and Biomedicine. ACM Conference on Bioinformatics, Computational Biology and Biomedicine 2015*, S. 462–471. DOI: 10.1145/2808719.2808767.
- Yang, Liuqing; Lin, Chunru; Liu, Wen; Zhang, Jie; Ohgi, Kenneth A.; Grinstein, Jonathan D. et al. (2011): ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. In: *Cell* 147 (4), S. 773–788. DOI: 10.1016/j.cell.2011.08.054.
- Yang, Yuan; Eichhorn, Catherine D.; Wang, Yaqiang; Cascio, Duilio; Feigon, Juli (2019): Structural basis of 7SK RNA 5'- γ -phosphate methylation and retention by MePCE. In: *Nature chemical biology* 15 (2), S. 132–140. DOI: 10.1038/s41589-018-0188-z.
- Yang, Z.; Zhu, Q.; Luo, K.; Zhou, Q. (2001): The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. In: *Nature* 414 (6861), S. 317–322. DOI: 10.1038/35104575.
- Yu, Mengqian; Lu, Bingjian; Zhang, Jisong; Ding, Jinwang; Liu, Pengyuan; Lu, Yan (2020): tRNA-derived RNA fragments in cancer: current status and future perspectives. In: *Journal of hematology & oncology* 13 (1), S. 121. DOI: 10.1186/s13045-020-00955-6.
- Zahler, A. M.; Roth, M. B. (1995): Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. In: *Proceedings of the National Academy of Sciences of the United States of America* 92 (7), S. 2642–2646. DOI: 10.1073/pnas.92.7.2642.
- Zhang, Bin; Arun, Gayatri; Mao, Yuntao S.; Lazar, Zsolt; Hung, Gene; Bhattacharjee, Gourab et al. (2012): The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. In: *Cell reports* 2 (1), S. 111–123. DOI: 10.1016/j.celrep.2012.06.003.
- Zhang, Chaolin; Frias, Maria A.; Mele, Aldo; Ruggiu, Matteo; Eom, Taesun; Marney, Christina B. et al. (2010): Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. In: *Science (New York, N.Y.)* 329 (5990), S. 439–443. DOI: 10.1126/science.1191150.
- Zhang, Chaolin; Zhang, Zuo; Castle, John; Sun, Shuying; Johnson, Jason; Krainer, Adrian R.; Zhang, Michael Q. (2008): Defining the regulatory network of the tissue-specific splicing

factors Fox-1 and Fox-2. In: *Genes & development* 22 (18), S. 2550–2563. DOI: 10.1101/gad.1703108.

Zhou, Ling; Xu, De-Yu; Sha, Wen-Gang; Shen, Lei; Lu, Guo-Yuan (2018): Long non-coding RNA MALAT1 interacts with transcription factor Foxo1 to regulate SIRT1 transcription in high glucose-induced HK-2 cells injury. In: *Biochemical and biophysical research communications* 503 (2), S. 849–855. DOI: 10.1016/j.bbrc.2018.06.086.

Zhou, Sihang; van Bortle, Kevin (2023): The Pol III transcriptome: Basic features, recurrent patterns, and emerging roles in cancer. In: *Wiley interdisciplinary reviews. RNA*, e1782. DOI: 10.1002/wrna.1782.

Zong, Xinying; Nakagawa, Shinichi; Freier, Susan M.; Fei, Jingyi; Ha, Taekjip; Prasanth, Supriya G.; Prasanth, Kannanganattu V. (2016): Natural antisense RNA promotes 3' end processing and maturation of MALAT1 lncRNA. In: *Nucleic acids research* 44 (6), S. 2898–2908. DOI: 10.1093/nar/gkw047.

Zorn, Peter; Misiak, Danny; Gekle, Michael; Köhn, Marcel (2021): Identification and initial characterization of POLIII-driven transcripts by msRNA-sequencing. In: *RNA biology* 18 (11), S. 1807–1817. DOI: 10.1080/15476286.2020.1871216.

7. Appendix

7.1 List of publications

Zorn P, Misiak D, Gekle M, Köhn M. Identification and initial characterization of POLIII-driven transcripts by msRNA-sequencing. *RNA Biol.* 2021 Nov;18(11):1807-1817. doi: 10.1080/15476286.2020.1871216. Epub 2021 Jan 18. PMID: 33404286; PMCID: PMC8583065.

Piersimoni L, Abd El Malek M, Bhatia T, Bender J, Brankatschk C, Calvo Sánchez J, Dayhoff GW, Di Ianni A, Figueroa Parra JO, Garcia-Martinez D, Hesselbarth J, Köppen J, Lauth LM, Lippik L, Machner L, Sachan S, Schmidt L, Selle R, Skalidis I, Sorokin O, Ubbiali D, Voigt B, Wedler A, Wei AAJ, **Zorn P**, Dunker AK, Köhn M, Sinz A, Uversky VN. Lighting up Nobel Prize-winning studies with protein intrinsic disorder. *Cell Mol Life Sci.* 2022 Jul 26;79(8):449. doi: 10.1007/s00018-022-04468-y. PMID: 35882686.

Zorn P, Calvo Sánchez J, Alakhras T, Schreier B, Gekle M, Köhn M. Rbfox1 controls splicing decisions of focal adhesion genes to modulate cardiac muscle cell fate. (2023)

This manuscript has been submitted to Journal of Molecular Cell Biology (JMCB) on the 5th of April 2023. The revised manuscript was re-submitted on the 28th of September 2023 awaiting final decision.

7.2 List of abbreviations

5S rRNA - 5S Ribosomal RNA

7SK RNA - 7SK Small Nuclear RNA

7SL RNA - 7SL Ribonucleic Acid

A-Box - Activation Box

AG - Adenosine-Guanosine

B1-SINE - Short Interspersed Nuclear Elements of B1 family

B-Box - Binding Box

BC200 RNA - Brain Cytoplasmic 200 RNA

BDP1 - B double prime 1

BPS - Branch Point Sequence

BRF1 - B-related factor 1

BRF2 - B-related factor 2

cAMP - Cyclic Adenosine Monophosphate

ChIP-seq - Chromatin Immunoprecipitation Sequencing

circRNAs - Circular RNAs

COIR - C1QTNF6 interspersed RNA

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

DNA - Deoxyribonucleic Acid

DUSP11 - Dual-specificity phosphatase 11

ER - Endoplasmic Reticulum

ESE - Exonic Splicing Enhancer

ESS - Exonic Splicing Silencer

FISH - Fluorescence In Situ Hybridization

GO - Gene Ontology

hY1 - Human Y1 RNA

hY3 - Human Y3 RNA

hY4 - Human Y4 RNA

hY5 - Human Y5 RNA

ICG - Inter-Chromatin Granule

IP - Immunoprecipitation
ISE - Intronic Splicing Enhancer
ISS - Intronic Splicing Silencer
KH - K-homology
LNAs - Locked Nucleic Acids
lncRNA - Long Non-Coding RNA
MALAT1 - Metastasis-Associated Lung Adenocarcinoma Transcript 1
MEF2 - Myocyte Enhancer Factor 2
miRNAs - MicroRNAs
MLO - Membraneless Organelle
mRNA - Messenger RNA
msRNA - Medium-sized RNA
Neat1 - Nuclear Paraspeckle Assembly Transcript 1
NONO - Non-POU Domain Containing Octamer Binding
nts - nucleotides
PAMIR - PAMR1 interspersed RNA
POLAR - POLR3E antisense RNA
POLR3E - RNA Polymerase III Subunit E
PPT - Polypyrimidine Tract
PSPC1 - Paraspeckle Component 1
Rbfox1 - RNA-Binding Fox-1 Homolog 1
Rbfox2 - RNA-Binding Fox-1 Homolog 2
Rbfox3 - RNA-Binding Fox-1 Homolog 3
RBM14 - RNA Binding Motif Protein 14
RBP - RNA-Binding Protein
RNP - Ribonucleoprotein
RRMs - RNA Recognition Motifs
RS - Arginine-Serine (in RS-domain)
SBP - Streptavidin binding peptide
SFPQ - Splicing Factor, Glutamine-Rich
SINE - Short Interspersed Nuclear Element

SmB - Small Nuclear Ribonucleoprotein Polypeptides B

SRPK1 - SR-Protein Kinase 1

SR-proteins - Serine/Arginine-Rich Proteins

SRSF1 - Serine/Arginine-Rich Splicing Factor 1

SRSF3 - Serine/Arginine-Rich Splicing Factor 3

SRSF5 - Serine/Arginine-Rich Splicing Factor 5

SS - Splice Site

STAT - Signal Transducer and Activator of Transcription

TFIIIA - Transcription Factor IIIA

TFIIIB - Transcription Factor IIIB

TFIIIC - Transcription Factor IIIC

TFs - Transcription Factors

TREX - Transcription-Export Complex

TSS - Transcription start site

U1 - U1 Small Nuclear Ribonucleoprotein

U2 - U2 Small Nuclear Ribonucleoprotein

U3 - U3 Small Nuclear Ribonucleoprotein

U4 - U4 Small Nuclear Ribonucleoprotein

U5 - U5 Small Nuclear Ribonucleoprotein

U6 - U6 Small Nuclear Ribonucleoprotein

UTR - Untranslated Region

XIST - X-Inactive Specific Transcript

Y RNAs - Y Ribonucleic Acids

7.3 List of Figures

Figure 1. Schematic illustration of RNA Polymerase III transcription with promoter types I, II and III.	3
Figure 2. Membraneless organelles (MLOs) formed in eukaryotic cells.	7
Figure 3. Nuclear speckle composition according to the ,core‘ and ,shell‘ hypothesis.	9
Figure 4 Paraspeckle composition according to the ,core‘ and ,shell‘ hypothesis.	11
Figure 5. The domain organization of SRSF1.	14
Figure 6. SRSF1 and hnRNPA/B mediated recruitment of U1 and U2 snRNP.	15
Figure 7. The RS-domain of SRSF1 directs U2AF35 and U1-70k.	16
Figure 8. Conserved sequence elements required for splicing recognition in metazoans.	18
Figure 9. Cross-intron assembly and disassembly of the major spliceosome.	19
Figure 10. Regulation of alternative splicing by <i>trans</i> -acting proteins.	20
Figure 11. The msRNA-Sequencing pipeline.	45
Figure 12. Classification of the coverage regions from the msRNA-Sequencing.	46
Figure 13. Identification of msRNAs transcribed by RNA-Polymerase III	47
Figure 14. Initial characterization of the ncRNAs identified by the msRNA-Sequencing.	48
Figure 15. Initial characterization of the POLAR msRNA.	49
Figure 16. Analysis of cellular localization of POLAR in OVCAR-3 cells.	50
Figure 17. Western blot analysis of an RNA pulldown using OVCAR-3 cell lysate with biotinylated <i>in vitro</i> transcribed POLAR RNA.	51
Figure 18. Genomic view of one msRNA-region of the mouse genome (chr6:128,759,326-128,764,958).	52
Figure 19. The 4.5S _H RNA sequence and predicted RNA fold.	53
Figure 20. The 4.5S _H RNA expression and subcellular distribution.	54
Figure 21. Fluorescence <i>in situ</i> hybridization (FISH) of 4.5S _H RNA combined with immunostaining.	55
Figure 22. Proteins identified by pulldown and LC MS/MS analyses to be associated with 4.5S _H RNA.	56
Figure 23. Functional clustering and association with spliceosomal proteins.	56
Figure 25. SRSF1 pulldown for associated RS-domain containing proteins and RNAs.	59
Figure 26. SRSF1 recruits SRSF-family members to the 4.5S _H RNA.	60
Figure 27. Quantitative real-time PCR of RNAs associated with 4.5S _H and 4.5S _I RNA.	61
Figure 28. The Malat1 RNA in human and mouse.	62
Figure 29. LNA mediated knockdown of 4.5S _H and Malat1 in B16-F10.	63
Figure 30. Splicing speckle integrity in human U2-OS cells upon ncRNA modulation.	64
Figure 31. Splicing speckle integrity in mouse B16-F10 cells upon ncRNA modulation.	65
Figure 32. Analysis of pre-mRNA splicing upon modulation of 4.5S _H and Malat1.	66
Figure 33. RNA Expression of Rbfox1 in RNA-Seq datasets. Expression is analysed	68
Figure 34. Computational analysis of putative Rbfox1 target genes.	69
Figure 35. Rbfox1 controls alternative splicing of focal adhesion genes Vinculin and Paxillin.	70
Figure 36. Vcl Minigene constructs in pEGFP-C1 vectors containing Vcl exon 18-20.	71
Figure 37. qRT-PCR analyses determining alternatively spliced transcripts from the mVcl minigene constructs.	72
Figure 38. The ePxn isoform provides a large and highly disordered amino acid extension.	73
Figure 39. Pxn Minigene constructs in pEGFP-C1 vectors containing Pxn exon 6 to 11.	74

7.4 List of Tables

Table 1. Cultivated cell lines.....	24
Table 2. Oligonucleotides for siRNA-mediated inhibition of gene expression	25
Table 3. Modified oligonucleotides for LNA mediated inhibition of gene expression	25
Table 4. Commercial plasmids	25
Table 5. Plasmids used for eukaryotic expression	26
Table 6. Oligonucleotides used for insert generation and cloning	27
Table 7. Oligonucleotides for amplification in RT-qPCR analyses	27
Table 8. Oligonucleotides for in vitro transcription	29
Table 9. Fluorescently-labeled anti-sense probes for Northern blot analysis	29
Table 10. primary antibodies for Western blot detection and immunofluorescence	29
Table 11. Fluorescently-conjugated secondary antibodies	30
Table 12. Commercial Kits and Assays.....	30
Table 13. Chemical composition of buffers	31
Table 14. Manufacturer’s list of used devices	32

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich meine Dissertationsschrift selbstständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen habe ich als solche kenntlich gemacht.

Mit der vorliegenden Arbeit bewerbe ich mich erstmals um die Erlangung des Doktorgrades.

Halle (Saale), 09. Oktober 2023

Peter Zorn

Danksagung

- Aus datenschutzrechtlichen Gründen entfernt -

- Aus datenschutzrechtlichen Gründen entfernt -

Curriculum Vitae

Persönliche Angaben

Name **Peter Zorn**
geboren in Halle / Saale

Berufspraxis

- 11.2016 - 04.2023 **Wissenschaftlicher Mitarbeiter**
Projekte: Forschung zur Promotion (siehe unten),
dazu Grundlagenforschung für die Humanmedizin zu einem
Protein und seinem Einfluss die Funktionsweise des Herzens,
Praktikumsbetreuer Masterstudenten Biochemie,
Zahnmediziner
Medizinische Fakultät, Institut für Molekulare Medizin
Martin-Luther-Universität Halle-Wittenberg
mehrfach befristet
- 08.2014 - 06.2015 **Wissenschaftliche Hilfskraft**
Forschungsassistenz bei Versuchen mit Mäusen: Präparation,
Genotypisierung, Isolation RNA und Proteine
Medizinische Fakultät, Institut für Molekulare Medizin
Martin-Luther-Universität Halle-Wittenberg
begleitende Studienfinanzierung
- 11.2012 - 07.2014 **Kurierfahrer**
Auslieferung von Medikamenten
Apotheke im Neustadt Centrum, Halle / Saale
begleitende Studienfinanzierung
- 09.2009 - 05.2010 **Medizinische Hilfskraft**
Logistik Operationsbesteck

Zentrale Sterilisations- und Versorgungsabteilung

Uniklinikum Halle / Saale

Zivildienst

Bildungsweg

11.2016 -

Promotionsstudium

Molekulare Zellbiologie; Grundlagenforschung, RNA Biologie

Promotionsarbeit: From sequencing to function: A triad of RNA-guided processes in the nucleus - medium-sized ncRNAs, nuclear speckles and Rbfox1-mediated splicing

Medizinische Fakultät, Institut für Molekulare Medizin

Martin-Luther-Universität Halle-Wittenberg

10.2013 - 10.2016

Masterstudium

Biochemie

Masterarbeit: Physiologische und molekulare Phänotypen in IGF2BP2 überexprimierenden Mäusen

Martin-Luther-Universität Halle-Wittenberg

Abschluss: Master of Science

10.2010 - 09.2013

Bachelorstudium

Biochemie

Bachelorarbeit: Einfluss des murinen Cytomegalovirus Glycoproteins ,gp40' auf zellbiologische und immunologische Funktionen in murinen embryonalen Fibroblasten

Martin-Luther-Universität Halle-Wittenberg

Abschluss: Bachelor of Science

09.1997 - 08.2009

Schulbildung

bis 08.2003 Grund- und Mittelschule

Georg-Cantor-Gymnasium Halle / Saale

Abschluss: Abitur

Weiterbildung

08.2020	GMP basic course
04.2018	GMP Statisikkurs
03.2015	FELASA-B Zertifizierung für Kleinterversuche Martin-Luther-Universität Halle-Wittenberg

IT (Auswahl), Fremdsprachen, Interessen, Mobilität

MS Office, OpenOffice
OS bis Windows 11
Adobe Photoshop
IG Viewer, SnapGene, PyMOL, BioRender, Image Studio,
Primer-BLAST
Englisch C1
Ausgleichsport (Klettern, Radfahren, Volleyball u.a.), Familie
Führerschein Klasse B, Fahrpraxis