

Charles University in Prague

Faculty of Science

Department of Genetics and Microbiology

PhD Study Program: Molecular and Cellular Biology, Genetics and Virology



Quality control in snRNP biogenesis

PhD Thesis

Adriana Roithová

2018

Supervisor: doc. Mgr. David Staněk, PhD
Department of RNA Biology, Institute of Molecular Genetics AS CR, v. v. i.

DISCLAIMER:

I hereby declare that this thesis is my own original work and has not been submitted before to obtain another academic degree. I have acknowledged all sources used and have cited these in the reference section.

Prague, 27.08.2018

Signature:

Adriana Roithová

ACKNOWLEDGEMENT

I would like to express my thanks to my supervisor, David Staněk, for his ideas, advices and big support during my studies. I would like to thank all my colleagues from Laboratory of RNA Biology, IMG ASCR, for creating an amazing funny and inspiring working atmosphere. Thanks to my family for their generous support and encouragement. Special thanks go to my husband Matyáš and my children Sebastian and Filip for keeping me sane and for simply everything, as always.

Contents

Abbreviations	5
Abstract (English)	9
Abstrakt v češtině	10
Introduction.....	12
Aims	14
1. Literary review	15
1.1.Spliceosome assembly and splicing	15
1.2.snRNPs	20
1.3.snRNP biogenesis.....	23
1.4.SMN complex	29
1.5.Nuclear phase and Cajal bodies.....	34
1.6.Quality control of snRNP assembly.....	36
1.7.Degradation pathways in the cytoplasm	37
1.8.Quality control of snRNP assembly in the nucleus	42
2. Materials and methods	43
3. Results	57
1.9.3.1. The Sm-core mediates the retention of partially-assembled spliceosomal snRNPs in Cajal bodies until their full maturation	59
3.2 Identifying role of Gemin3 in Sm ring assembly.....	83
3.3 The recognition of the immature particles and their targeting into the P bodies	92
4. Discussion	98

5. Summary.....	104
6. References	106

Abbreviations

3'ss	3'splice site
5'ss	5'splice site
aa	Amino acid
Ago2	Argonaute 2 protein
CB	Cajal body
CBC	Cap binding complex
CLIP	cross-linking immunoprecipitation
Crm1	chromosome region maintenance 1
CTD	Carboxy terminal domain
Dcp1	Decapping enzyme 1
DDX6	DEAD-box helicase 6
DIS3L2	DIS3 mitotic control homolog-like2
DMEM	Dulbecco's modified Eagle medium
DSE	Distal sequence element
Egr2	early growth response protein 2
FISH	fluorescence <i>in situ</i> hybridization
FOXL2	forkhead transcription factor
Gem	Gemini of Cajal bodies
GFP	Green fluorescent protein
GR repeats	Gly/Arg rich repeats

hnRNP	heterogeneous nuclear RNP
HSP90	Heat shock protein 90
IBB domain	Importin beta binding domain
IP	immunoprecipitation
IRES	internal ribosome entry site
KD	Knock down
LSm	Like-Sm proteins
NC	Negative control
NLS	Nuclear localization signal
NSS	Near Sm site stem
nt	nucleotide
P body	Processing body
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer saline
PCR	Polymerase chain reaction
Phax	phosphorylated adaptor complex for RNA export
pICln	chloride ion current inducer protein
PRMT5	protein arginine methylase 5
Prp22	pre-mRNA processing protein 22
Prp5	pre-mRNA processing protein 5
PSE	Proximal sequence element

RNA	ribonucleic acid
RT-PCR	reverse transcription followed by polymerase chain reaction
SDS	sodium dodecylsulphate
SL	Stem loop
SMA	Spinal muscular atrophy
SMN	Survival of motor neurons
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particle
SPN1	Snurportin 1
SR proteins	Ser/Arg rich proteins
SRP RNA	Signal recognition particle RNA
TGS1	trimethylguanosine synthase 1
TMG	trimethylguanosine cap
Tris	2-amino-2-hydroxymethylpropan-1,3-diol
TUT4	Terminal uridyly transferase 4
U1-70K	U1 small nuclear ribonucleoprotein 70 kDa
U1A	U1 small nuclear ribonucleoprotein A
U2AF	U2 auxiliary factor
WB	western blotting
wt	Wild type
Xrn1	Exoribonuclease 1

YFP

Yellow fluorescent protein

Abstract (English)

snRNPs are key components of the spliceosome. During their life, they are found in the cytoplasm and also in the nucleus, where they carry out their function. There are five major snRNPs named according to the RNA they contain: U1, U2, U4, U5, and U6. Each snRNP consists of RNA, a ring of seven Sm or LSM proteins, and additional proteins specific for each snRNP. Their biogenesis starts in the nucleus, where they are transcribed. Then they are transported into the cytoplasm. During their cytoplasmic phase, the SMN complex forms the Sm ring around the specific sequence on snRNA, and the cap is trimethylated. These two modifications are the signals for reimport of snRNA into the nucleus, where they accumulate in the nuclear structures called Cajal bodies (CBs), where the final maturation steps occur.

There are several quality control points during snRNP biogenesis that ensure that only fully assembled particles reach the spliceosome. The first checkpoint is in the nucleus immediately after the transcription, when the export complex is formed. The second checkpoint is in the cytoplasm and proofreads Sm ring assembly. If the Sm ring formation fails, the defective snRNPs are degraded in the cytoplasm by Xrn1 exonuclease. However, it is still unclear how the cell distinguishes between normal and defective snRNAs. The last checkpoint occurs in CBs. However, signals that target and retain snRNPs in CBs have yet to be described.

In this work, I analyzed the main role of the Sm ring in the quality control of snRNA in the nucleus and the cytoplasm.

First, we identified Sm protein motifs important for targeting of snRNPs into CBs and proposed a model, where Sm proteins play an important role in quality control in CBs.

Second, we explored a role of the component of the SMN complex, Gemin3, in the Sm ring assembly. My data suggest that Gemin3 is involved in unwinding of the secondary structure of snRNA prior to Sm ring formation.

Finally, we investigated the defective snRNAs which failed to acquire the Sm ring in the cytoplasm. We found that immature snRNAs are localized in P bodies and identified a new role for the LSM1 protein in snRNA degradation.

Abstrakt v češtině

snRNP patří k nejdůležitějším částem sestřihového komplexu. Jejich životní cyklus se odehrává v cytoplasmě, kde probíhají první fáze jejich biogeneze, a také v jádře, kde plní svoji hlavní funkci. Všechny snRNP jsou složeny z krátké nekódující RNA, z Sm či LSm proteinů tvořící 7-členný kruh a z proteinů specifických pro každý snRNP. Jejich životní cyklus začíná v jádře, kde jsou transkribovány RNA polymerázou II nebo III. Poté jsou transportovány do cytoplasmy. Během své cytoplasmatické fáze se formuje Sm kruh kolem specifické sekvence na RNA pomocí SMN komplexu a následně se trimetyluje čepička na 5'konci snRNA. Tyto 2 úpravy jsou signálem, že je snRNP připravena na transport do jádra, kde je hromaděna v jaderných strukturách nazývajících se Cajalova tělíska. V Cajalových tělískách probíhá finální část jejich zrání.

Průběh snRNP biogeneze je průběžně kontrolován. První kontrola probíhá v jádře ihned po jejich transkripci a následuje vytvoření exportního komplexu. Druhý kontrolní bod je v cytoplasmě a zahrnuje tvorbu Sm kruhu. Víme, že Sm kruh je tvořen SMN komplexem ale detailní mechanismus je stále neznámý. Pokud snRNA neprojde těmito kontrolními body, tak je v cytoplasmě degradována. Avšak, jak buňka rozlišuje mezi normálními a defektními snRNA se stále neví.

Třetí a poslední kontrolní bod se nachází v Cajalových tělískách. Signál, který navádí snRNP do těchto struktur je stále nepopsaný.

V mé práci jsme se soustředili na roli Sm kruhu v kontrole kvality během životního cyklu snRNP v jádře i v cytoplasmě.

Nejprve jsme našli důležitý motiv v Sm proteinech, který je zodpovědný za lokalizaci snRNP do Cajalových tělísek a zároveň jsme navrhli model, kde Sm kruh hraje důležitou roli při kontrole kvality zrání snRNP v Cajalových tělískách.

V druhé části jsme zkoumali roli proteinu Gemin3, který je součástí SMN komplexu zodpovědného za skládání Sm kruhu v cytoplasmě. Naše výsledky ukazují, že Gemin3 je RNA helikáza, jejíž funkcí v životním cyklu snRNP je rozplétat sekundární strukturu snRNA pro lepší přístup Sm vazebného místa pro tvorbu Sm kruhu.

Nakonec jsme se soustředili na osud nezralých snRNP, které nebyly schopny získat Sm kruh v cytoplasmě. Ukázaly jsme, že nezralé částice jsou hromaděny v cytoplasmatických P tělíkách. Také jsme ukázaly novou roli LSm1 proteinu v degradaci špatných snRNA.

Introduction

In all cells, the guideline for life is encoded in DNA, where the information is divided into discrete functional units called genes. The main cellular components – proteins, encoded in protein-coding genes, are converted into a protein sequence by transcription and translation. In eukaryotes, these two processes are spatially separated by the nuclear membrane. In the first step, the DNA is transcribed into the pre-mRNA. Before the pre-mRNA is translated, it has to be modified in the nucleus. These modifications are adding a monomethyl cap at the 5' end, removing of non-coding sequences (introns) and joining of coding sequences (exons), polyadenylation at 3' end and sometimes base editing. The removing of introns from pre-mRNA is called splicing. This well-controlled process is catalyzed by the spliceosome, which is a large complex composed from more than 150 proteins but the essential core is formed from 5 small nuclear ribonucleoprotein particles (U1, U2, U4, U5 and U6 snRNPs). These particles are composed of short non-coding, non-polyadenylated RNA, Sm or LSm proteins and proteins, which are specific for each snRNP.

My work is focused mainly on the biogenesis of U1, U2, U4 and U5 snRNPs. They are transcribed in the nucleus by RNA polymerase II and transported into the cytoplasm where Sm proteins are loaded on RNA and form an Sm ring around the specific sequence. After this step, the monomethyl cap is trimethylated, and snRNPs are prepared for transport back to the nucleus, where snRNPs first accumulate in the nuclear structures called Cajal bodies (CBs). However, the precise mechanism of how is the maturation of snRNPs controlled and how are the snRNPs targeted into the CBs is still unknown. My thesis is composed of three closely connected projects, where I tried to improve our understanding of snRNP biogenesis and how this process is controlled.

In the first chapter, we used biochemical methods, such as co-immunoprecipitation, *in vitro* transcription of mutated snRNAs coupled with microinjection into human cultured cells and using advanced fluorescent microscopy

approaches. We established Sm ring as a novel Cajal body targeting signal which is responsible for navigation of snRNPs into Cajal bodies.

In second chapter, we microinjected and expressed mutated forms of snRNA in the cells depleted of Gemin3. Our data suggest a role of a putative helicase Gemin3 in the unwinding of snRNA secondary structure prior assembly of the Sm ring.

In the last chapter, we analyzed the fate of snRNAs without Sm ring. We depleted proteins important for Sm ring formation by the siRNA mediated RNA interference and analysed snRNA localization by fluorescence *in situ* hybridization and microinjection of *in vitro* transcribed RNAs. We aimed to address, how are immature snRNAs recognized and degraded in the cytoplasm. We found out that immature snRNAs accumulate in the cytoplasmic structures called P bodies and LSM1-7 ring play a role in the transport of immature snRNAs into these structures.

Aims

The primary goal of my project is to gain a deeper insight into the biogenesis of snRNPs, which are critical components of the spliceosomal complex. I focus on three specific aims:

- To decipher localization signal, which navigates the snRNP into nuclear structures called Cajal body, the place of final snRNP maturation
- To describe the role of the SMN complex component, Gemin3, in the Sm ring formation in the cytoplasm)
- To identify factors responsible for the recognition and degradation of incomplete snRNPs in the cytoplasm.

1. Literary review

Most of the eukaryotic genes are transcribed as pre-mRNA containing noncoding sequences (introns), which have to be removed in a process called pre-mRNA splicing. The splicing is catalyzed by the spliceosome. This large complex is composed of more than 150 proteins, but the catalytical core is formed of three snRNAs-U2, U5 and U6.

1.1. Spliceosome assembly and splicing

The spliceosome complex is formed from the snRNPs in a stepwise manner directly on pre-mRNA. Spliceosomal snRNA function is driven by base pairing with short conserved motifs located at the junctions between the expressed exon sequences and the intervening introns of target pre-mRNAs (Matera and Wang, 2014b). The boundary of the intron is defined by 5' splice site (5'ss) at the beginning and 3' splice site (3'ss) is located on the end of the intron. 15-50 nucleotides upstream of 3'ss is present the branch point adenosine (Fig. 1)

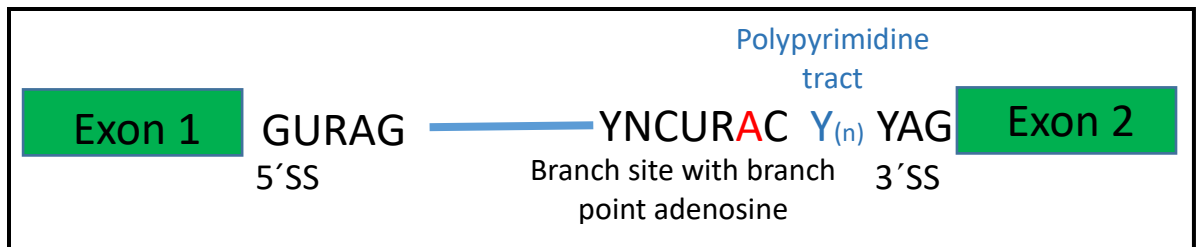


Figure 1: Conserved sequences defining the introns in metazoan pre-mRNA. Y stands for pyrimidine, R for purine. Adapted from (Will and Luhrmann, 2011).

During the first step, U1 snRNP interacts with the 5'ss resulting in the formation of the early complex (E-complex, Fig. 2) (Mount et al. 1983). This interaction is promoted by RNA polymerase CTD domain which interacts directly with U1 snRNP, but the functional role of this interaction is still under the debate (Gornemann et al., 2011; Morris and Greenleaf, 2000). The 3'ss is recognized by the U2 snRNP and associated factors SF1 and

U2AF. U2 snRNP recognizes the sequence near to the branch point site and interacts with U1 snRNP resulting in the formation of pre-spliceosome A-complex (Valcarcel et al., 1996) (Fig. 2). This process is catalyzed by DExD/H helicases (Gozani et al., 1996). A further step is binding of pre-assembled U4/U6·U5 tri-snRNP to form the B-complex (Bindereif and Green, 1987). In the next steps, the compositional and conformational rearrangements catalyzed by several helicases such as Brr2 and Prp28 lead to release of U1 and U4 snRNPs from the spliceosome and catalytically active B complex (B*complex) is established. The activation of the B complex leads to the formation of U2/U6 snRNA structure that catalyzes the splicing reaction (Sun and Manley, 1995). After the first catalytic step, the complex C is generated (Bessonov et al., 2008; Konarska et al., 2006). This complex contains the free exon 1 and the intron-exon 2 lariat intermediate (Fig. 2). It undergoes further rearrangements essential to carry out the second catalytic step of splicing. It results in a post-spliceosomal complex that contains the lariat intron and spliced exons. Finally, the DExD/H helicase Prp22 catalyzes the release of spliced product from the spliceosome and U2, U5, and U6 snRNPs are recycled for additional rounds of splicing (Ilagan et al., 2013; Schwer and Gross, 1998).

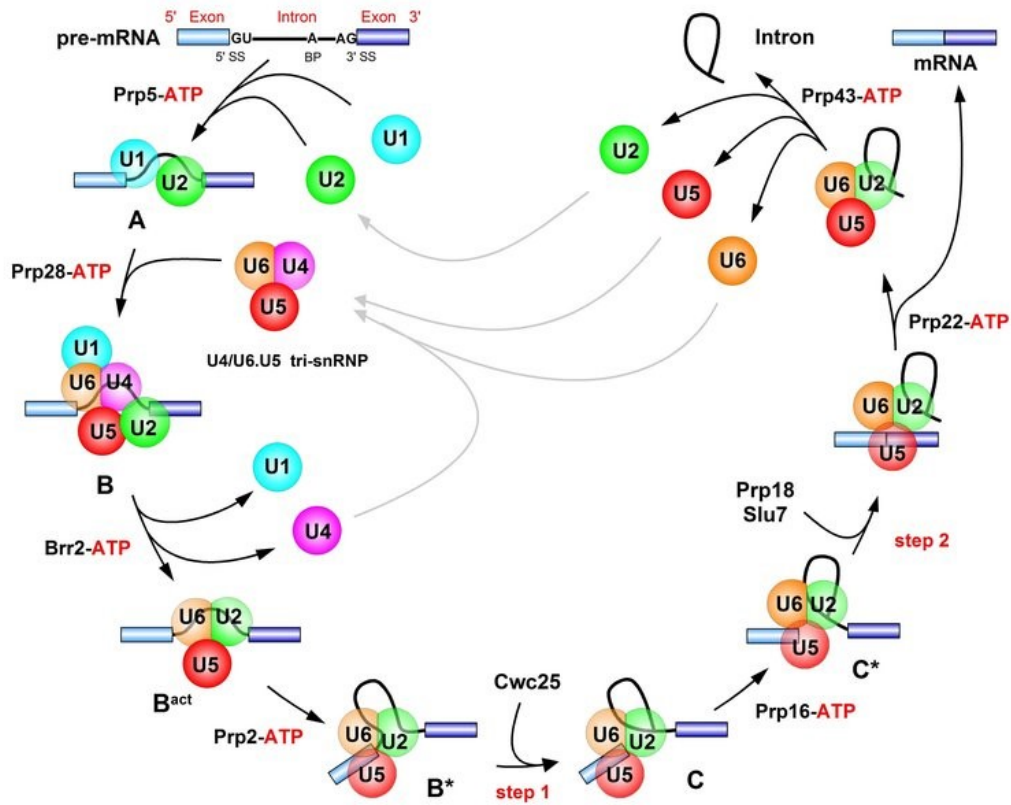


Figure 2: Pre-mRNA splicing. Canonical cross-intron assembly and disassembly pathway. Exon and intron sequences are indicated by boxes and lines. Adapted from (Will and Luhrmann, 2011)

1.1.1. Exon and intron definition

A typical transcript in animal cells has very long introns and short exons. The opposite situation is in lower eukaryotes where genes primarily consist of large exon interrupted by small introns (De Conti et al., 2013). When the introns are long, and exons are short, the splicing machinery is first formed across an exon and only later converted into the cross-intron complex. This hypothesis is called exon definition model (Fig. 3) (Robberson et al., 1990). It was previously shown that splice sites alone are insufficient to define exon/intron borders and there are additional *cis*-acting elements within the pre-mRNA which modulate splicing (Reed and Maniatis, 1986). These RNA elements (exonic/intronic splicing enhancers or silencers) are bound by regulatory proteins, which recruit or repulse other parts of splicing machinery. The enhancers are bound by members of Ser/Arg-rich protein family (SR

proteins) (Zhong et al., 2009). SR proteins can interact with each other, with U1-70K (specific protein of U1 snRNP) and with U2AF35 splicing factor through their RS domain. These findings led to the proposal that SR proteins facilitate splicing by forming the interaction across the exons and introns (Blencowe, 2000). The negative regulation of exon recognition is carried out by hnRNP protein family, which bind the exon or intron splicing silencers (ESS or ISS). hnRNPs are less defined than SR proteins and also the mechanism of their function has not been fully uncovered. It is thought that they inhibit splicing through the steric obstruction or by promoting the formation of inhibiting RNA secondary structures (Busch and Hertel, 2012; De Conti et al., 2013; Spellman and Smith, 2006; Tange et al., 2001).

In lower eukaryotes where genome architecture is characterized by short introns and large exons the intron definition model, where splice sites are paired across introns rather than exons, is predominant (Fig.3)(Berget, 1995). This model was recently described in *S. pombe*, where SR-related protein Rsd1 was identified as a bridging factor between U1 and U2 snRNPs. It interacts with U1A (specific protein of U1 snRNP) and at the same time through the Prp5 ATPase with U2 snRNP core protein SF3b.

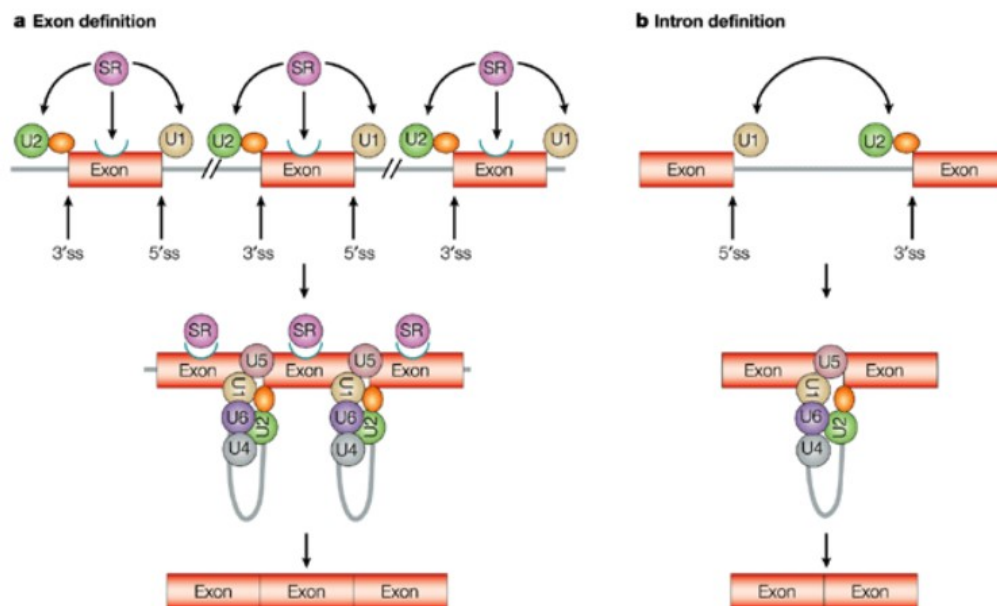


Figure 3: a, **Exon definition:** SR proteins (purple) bind to exonic splicing enhancers (ESE; blue), recruiting U1 to the downstream 5' splice site and the splicing factor U2AF (orange) to the upstream polypyrimidine tract and 3' splice site. Therefore, when the SR proteins bind the ESE, they promote the formation of a 'cross-exon' recognition complex by placing the basal splicing machinery in the splice sites that flanked the same exon. b, **Intron definition:** the binding of U1 to the upstream 5' splice site (ss) and U2AF and U2 to the downstream polypyrimidine tract and branch site, respectively, of the same intron. Therefore, intron definition selects pairs of splice sites located on both ends of the same intron, and SR proteins can also mediate this process. Adapted from (Ast, 2004)..

1.1.2. Splicing regulation and alternative splicing

Alternative splicing plays an indispensable role in the expression of various forms of proteins that determine the identity of the cell and whole tissues. Moreover, nearly 95% of mammalian genes undergo alternative splicing to produce multiple protein isoforms with different functions (Matera and Wang, 2014a; Pan et al., 2008). The spliceosome is responsible for both constitutive and alternative splicing, and regulation of its assembly is a critical control point in these processes.

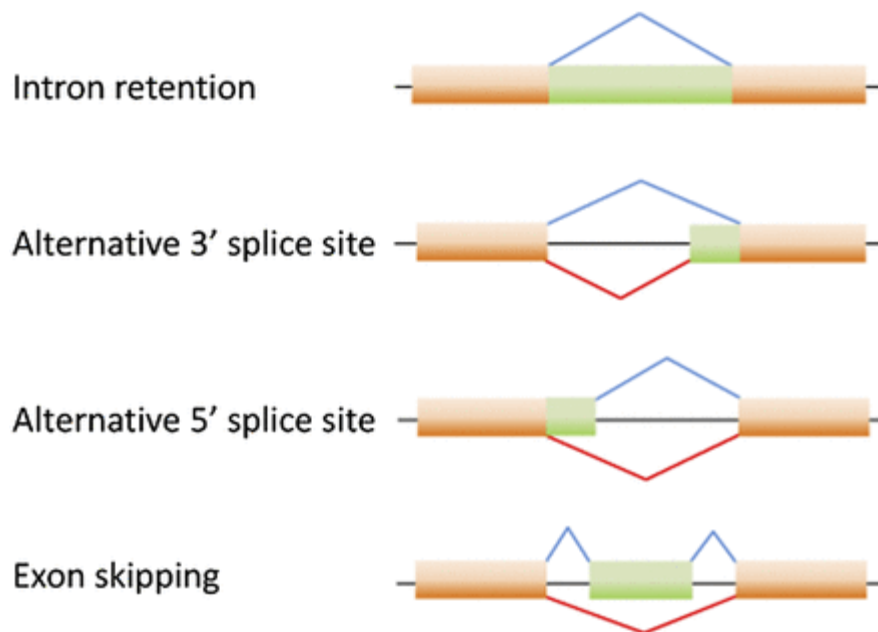


Figure 4: A schematic representation of alternative splicing: different types of alternative splicing: exon inclusion or skipping, alternative splice-site selection, mutually exclusive exons, and intron retention. For an individual pre-mRNA, different alternative exons often show different types of alternative-splicing patterns. Adapted from (Li et al., 2013)

Alternative exon shares the same sequence features as a constitutive exon, but are “weaker”, which means that they have a lowered affinity to the spliceosome resulting in reduced recognition. The sequences surrounding weak sites diverge subtly from strong sites, yet weak sites are frequently conserved between mouse and human (Sugnet et al., 2004). As mentioned before, SR proteins and hnRNPs play a role in exon recognition. As a general rule, SR-proteins promote exon inclusion, and hnRNPs antagonize exon inclusion, and the interplay between them defines the final decision about alternative exon inclusion or exclusion. However, there are many exceptions to this rule. For example the GTPase Rac1, where a critical exon is regulated by the competition of two SR-proteins, SRp20 and ASF/SF2 (Goncalves et al., 2009). Genome-wide analyses of splicing factor interactions combined with RNA expression analyses detecting the alternatively spliced versions enabled the formulation of the putative splicing code, which can predict whether the exon is constitutive or alternative (reviewed in Kornblihtt et al., 2013).

However, these analyses are still not able to predict the alternative splicing outcomes with the absolute accuracy, because in this process play roles other regulations mainly the coupling of splicing with transcription, the influence of histone marks or nucleosome position.

snRNPs

Five major snRNPs U1, U2, U4, U5, and U6 and four minor snRNPs U11, U12, U4atac and U6atac are present in the eukaryotic cell. They are composed of short uridine-rich noncoding non-polyadenylated RNA, seven Sm or LSm proteins and proteins, which are specific for each snRNP. Based on their sequence and protein composition we can distinguish two types of snRNAs. The first type is the Sm-type U1, U2, U4, U5, U11, U12, U4atac. In this case, the Sm ring is formed in the cytoplasm around a specific sequence called Sm binding site. The second case is LSm-type of snRNAs U6 and U6atac. They contain LSm 2-8 ring formed in the nucleus. My projects are focused only on the major

Sm-type of snRNAs U1, U2, U4 and U5 and I will not discuss the LSm-type snRNAs or minor snRNAs in my thesis.

1.1.3. Composition of snRNPs

1.1.3.1. Sm proteins

In the eukaryotic cell, there are 7 Sm proteins named SmB, SmD1, SmD2, SmD3, SmE, SmF and SmG (Fig.5). They are evolutionary conserved. Furthermore, biochemical characterization of the protein composition of snRNPs in various species revealed their presence in yeast, plants, fruit fly, mouse, and human. The Sm proteins belong to the large family of Sm and LSm proteins, defined by the presence of two conserved motifs – Sm1 and Sm2. These homologous regions are 32 and 14 amino acids long, and they are important for Sm protein-protein interactions (Hermann et al., 1995). Sm proteins are assembled in a step-wise manner into a ring-shaped structure whose positively charged interior is thought to interact with the Sm site (PuAU₄₋₆GPu) of the U snRNPs (reviewed in Urlaub et al. 2001). The main function of Sm proteins is stabilization and protection of snRNA against nucleases.

The three Sm proteins SmB/B', SmD1 and SmD3, contain unique Glycine-Arginine rich domain (GR repeats) in their C-terminal parts and Arginines in GR repeats are symmetrically dimethylated by the PRMT5 complex. These repeats fulfill the special functions. At first, they are essential for interaction with SMN complex, which binds them through its Tudor domain. Other studies have shown that positively charged C-termini of SmB, D1, and D3 can serve as a nuclear localization signal and can also bind the pre-mRNA within the 5' splice site region and help to stabilize spliceosomal complex (Bordonné, 2000; Zhang et al., 2001).

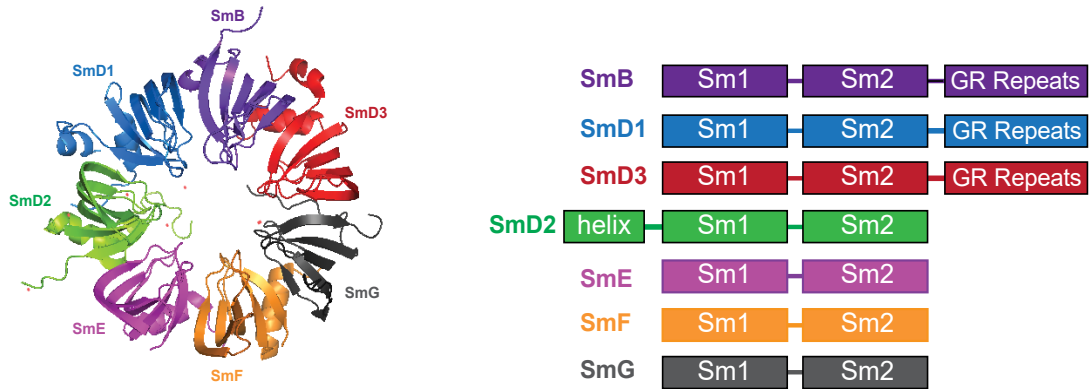


Figure 5: Structure and protein composition of Sm ring. Adapted from PDB with the letter code 4pjo.

1.1.3.2. snRNP specific proteins

Besides of Sm proteins, snRNPs contain many other proteins which are specific for each of them (Fig. 6). It is not precisely clear, where and in what order all proteins are added to the core snRNP and which proteins are involved in this process. Previous studies have shown evidence that the HSP90/R2TP complex participates in the U5 and U4 snRNP assembly (Bizarro et al., 2015; Malinova et al., 2017).

Some of the snRNP specific proteins are RNA helicases such as, Brr2, and Prp8. These three proteins are components of U4/U6•U5 tri-snRNP and play a crucial role in activation of the spliceosome and formation of the active site (reviewed in Will and Luhrmann, 2011). Another specific proteins such as SF3A1 or SF3B1 are essential in early steps of spliceosome assembly and help in splice site recognition (Shao et al., 2012). In the splicing of pre-mRNA are involved more proteins than I specified here, but in my thesis I will not focus on the snRNA specific proteins.

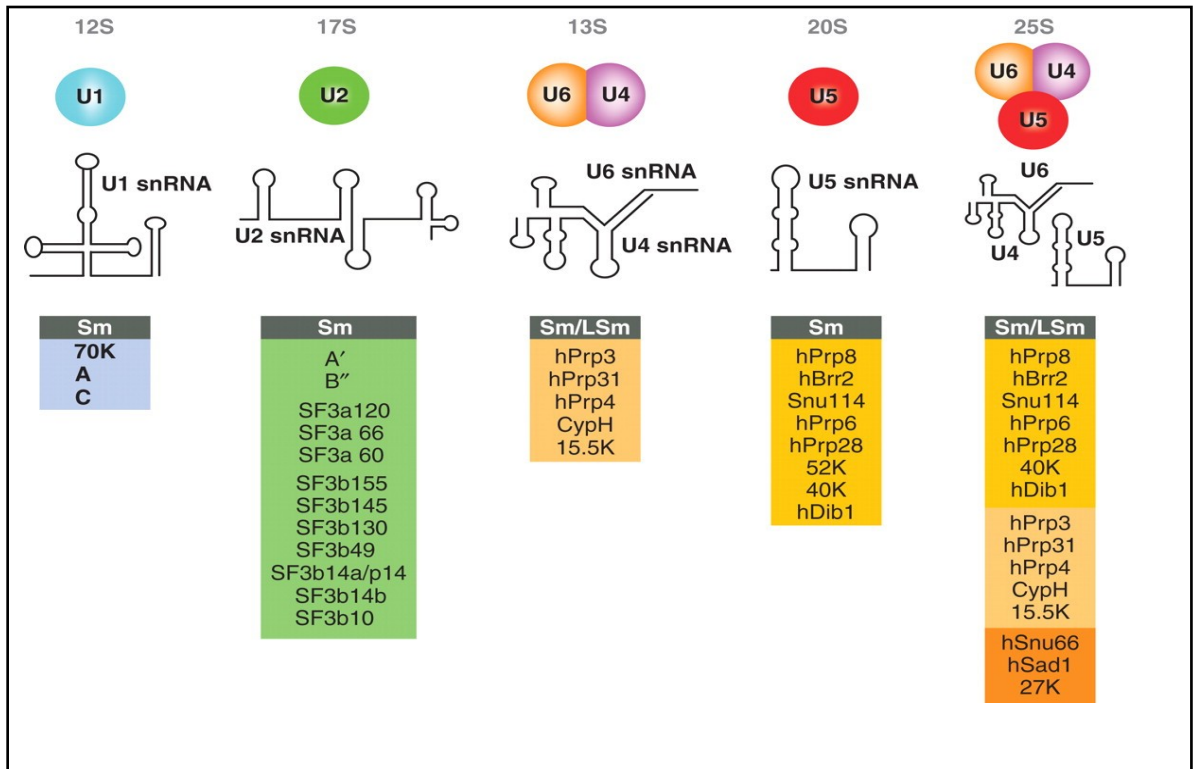


Figure 6: Protein composition and snRNA structures of the major spliceosomal snRNPs. Sm proteins and LSm proteins are indicated by „Sm“ or „LSm“ at the top of the boxes showing the proteins associated with each snRNP. Adapted from (Will and Luhrmann, 2011).

snRNP biogenesis

snRNPs are assembled in a step-wise manner (Fig.7). The biogenesis takes place in the nucleus and in the cytoplasm. This distribution into two separate compartments can be the part of the quality control mechanism, to prevent the incomplete snRNPs meeting the final substrate. The disturbances in snRNPs maturation can lead to diseases such as Spinal Muscular Atrophy (SMA), Retinitis Pigmentosa, chronic lymphocytic leukemia or myelodysplasia (Eggert et al., 2006; Krausova and Stanek, 2018; Matera, 1999; Pellizzoni et al., 1999).

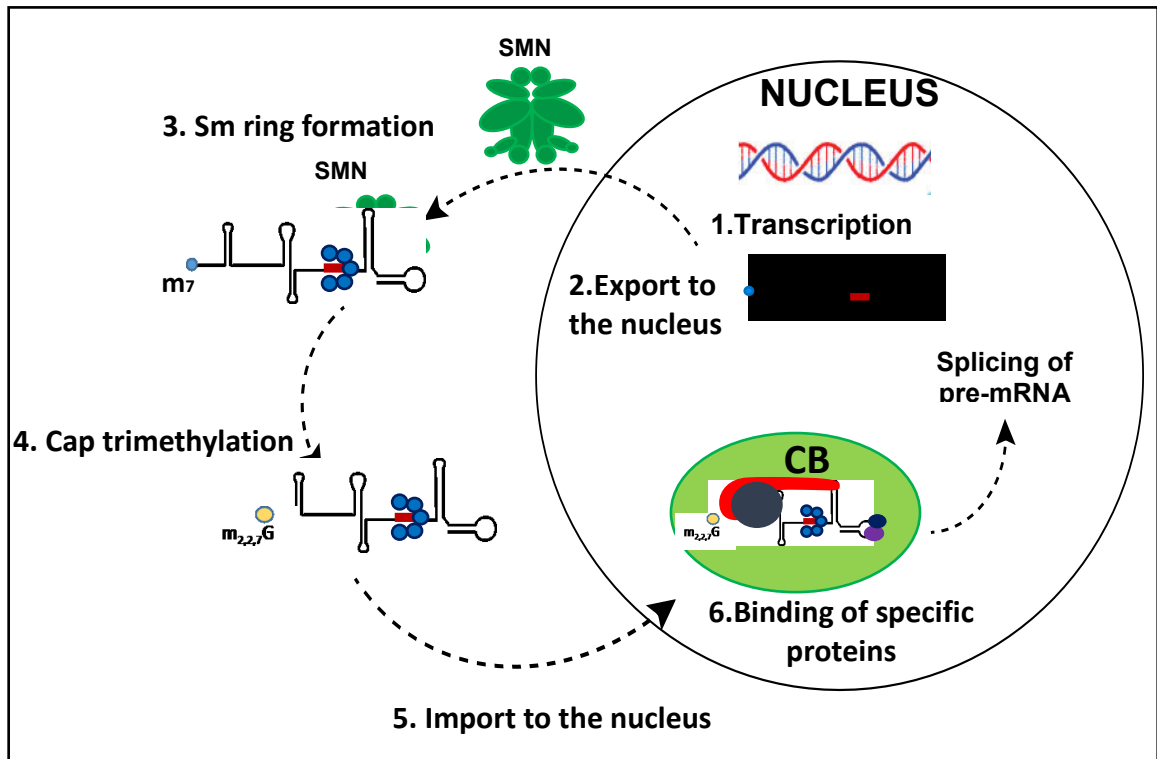


Figure 7: snRNP biogenesis. 1. Transcription of pre-snRNA, 2. Formation of export complex and transport into the cytoplasm, 3. Loading of SMN complex on pre-snRNA and formation of Sm ring around Sm site, 4. Cap trimethylation by TGS1, 5. Transport snRNP core into the nucleus, 6. Final maturation of snRNPs in the Cajal bodies and releasing fully matured particle into the nucleoplasm where becomes a part of the spliceosome.

1.1.4. Transcription of Sm-class snRNAs

The biogenesis of snRNPs starts in the nucleus, where they are transcribed by RNA polymerase II. Transcription starts with the promoter, which contains the proximal and distal regulatory element (PSE and DSE) and enhancer sequences (Fig. 8). The transcription is activated by DSE, which contains various binding sites for multiple transcription factors. One of them is octamer sequence ATGCAAAT, which recruits transcription activator Oct-1 (Ford et al., 1998). This factor recruits another initiation factors such general transcription factors (TFIIA, TFIIB, TFIIE, and TFIIF) and the pentameric factor called snRNA-activating protein complex (SNAPc) to form a stable transcription initiation complex. This complex recognizes the proximal regulatory element (PSE), which is located upstream of position -40 and is essential and sufficient

to direct basal levels of transcription (Sadowski et al., 1993; Sadowski et al., 1996). RNA polymerase II transcribes snRNA gene beyond 3' box located 9-19 nucleotides downstream of the 3' end of the mature snRNA (Cuello et al., 1999; Yuo et al., 1985). This element is required for transcription termination of snRNA genes. 3' box is recognized by a multi-subunit factor called the Integrator complex, which endonucleolytically cleaves nascent transcript (Baillat et al., 2005). Its subunits IntS9 and IntS11 have sequence similarity to cleavage factors CPSF73 and CPSF100 involved in the 3' end processing of mRNA (reviewed in Matera and Wang 2014).

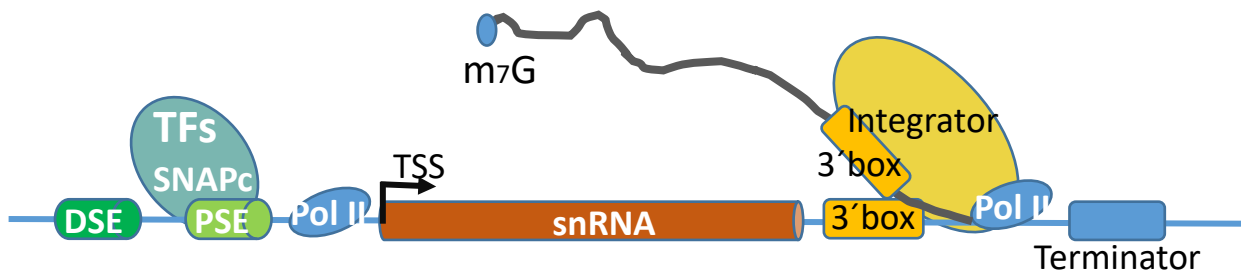


Figure 8: Transcription and processing of Sm-class snRNA genes. Adapted from (Matera and Wang, 2014a)

1.1.5. Export to the cytoplasm

After the transcription, the Sm-class snRNAs have to be transported into the cytoplasm. Because they are transcribed by RNA polymerase II, the 5' monomethyl cap structure is added on the 5' end, and bound by the Cap binding complex (CBC) (Izaurralde et al., 1995). Then the phosphorylated adaptor complex for RNA export (PHAX) binds CBC and recruits the Crm1-RanGTP complex through its leucine-rich nuclear export signal (Ohno et al., 2000a). Crm1 (exportin1) is a protein which interacts with nuclear pore and with the cooperation of RanGTP promotes the export of snRNAs into the cytoplasm (Fornerod et al., 1997).

The mRNAs have the similar export pathway as a snRNAs. Both contain 5' monomethyl cap on 5' end which is bound by CBC. It gives rise the question of how the cell recognizes

between these two types of RNA (snRNA and mRNA). It was shown that the distinguish is based on their length and their association with heterogeneous nuclear RNP (hnRNP) C1-C2 proteins, which recognizes transcripts longer than 200-400 nucleotides and funnels them into mRNA export pathway inhibiting PHAX binding (Masuyama et al., 2004; Ohno, 2012).

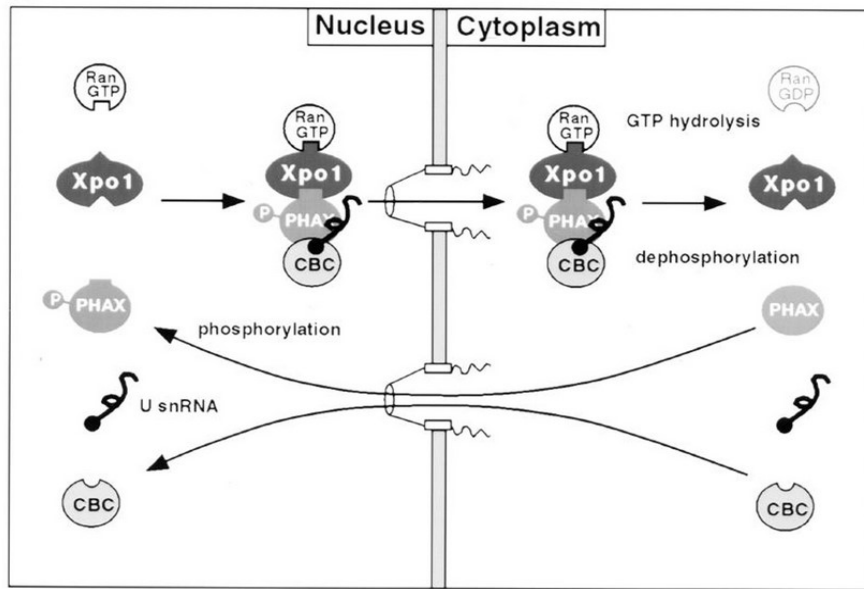


Figure 9: Phosphorylation/Dephosphorylation cycle of PHAX protein. Adapted from (Ohno et al., 2000a).

1.1.6. Sm ring formation by SMN complex

The cytoplasmic phase of the biogenesis comprises Sm ring formation, 3' end trimming, and cap trimethylation. When the pre-snRNA is translocated into the cytoplasm, the export complex is dissociated by dephosphorylation of PHAX protein (Kitao et al., 2008) and snRNA is ready for acquiring the Sm ring. After translation, Sm proteins are immediately bound by pICln protein to increase their specificity for snRNAs (Prusty et al., 2017). Together with pICln form 2 types of complexes 6S complex (D1, D2, E, F, G-pICln) and heterotrimer (B/B', D3-pICln)(Chari et al., 2008) (Fig.10). pICln protein also associates with PRMT5 N-methyltransferase and together with the protein WD40

forms PRMT5 complex, which is responsible for creating symmetrical dimethylated Arginines (sDMA) on C-terminus of SmB/B', SmD1 and SmD3 (Friesen et al., 2001a; Meister et al., 2001). After this modification, the 6S complex and heterotrimer are released from the PRMT5 complex, and Sm complexes are delivered to the SMN complex. Gemin2, the component of SMN complex, binds 6S particle forming 8S assembly intermediate. This binding triggers arrangements leading to stabilization of opened ring (Grimm et al., 2013). The pre-snRNA is recognized by Gemin5, another component of SMN complex. Gemin5 recognizes the Sm site in the snRNA and delivers SMN complex with pre-assembled Sm proteins on snRNA (Battle et al., 2006a; Tang et al., 2016; Xu et al., 2016; Yong et al., 2002a). In the end, the SmB/B' and SmD3 are incorporated into the Sm ring and closing it.

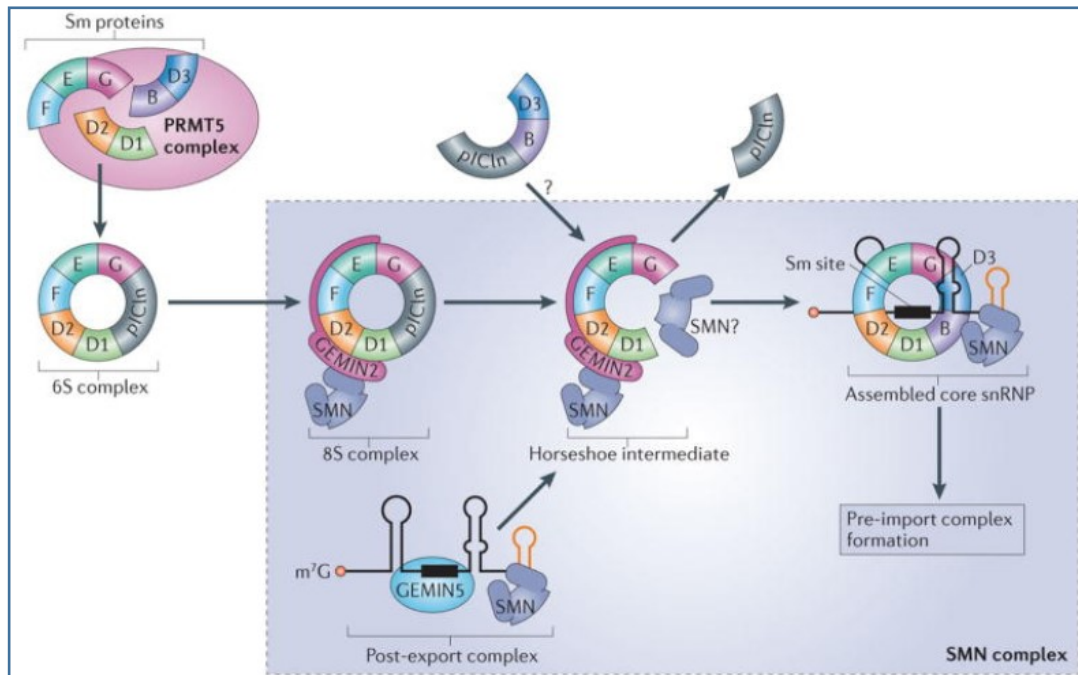


Figure 10: Assisted Sm ring assembly. Sm proteins are sequestered and symmetrically dimethylated by the PRMT5 complex. Once formed, the 6S complex of the Sm (D1-D2-F-E-G) pentamer and pICln is thought to be released from PRMT5 complex as a separate particle. This 6S complex is delivered to the SMN complex, which provides the overall platform for subsequent assembly steps. Gemin2, the heterodimeric binding partner of SMN, binds to the 6S complex, forming an early 8S assembly intermediate. In parallel, the SMN complex, including Gemin5, recognizes specific sequence elements (the Sm-site and the adjacent 3' stem-loop) within the post-export snRNA. Adapted from (Matera and Wang, 2014a)

1.1.7. Cap trimethylation by TGS1 and nuclear transport

Formation of Sm ring protects and stabilizes snRNA and initiates further steps of biogenesis. In second cytoplasmic phase, the SMN complex binds and recruits RNA methyltransferase TGS1 that modifies 5' end of snRNA to a 2,2,7-trimethylguanosine cap structure (Matera and Wang, 2014a; Mouaikel et al., 2003). It was shown that TGS1 directly binds SMN protein, but it can also bind the C-terminus of SmB, which also interacts with SMN protein (Fig. 11) (Mouaikel et al., 2003; Plessel et al., 1994). The interaction of TGS1 with the C-terminus of SmB can trigger conformational changes and SMN can be released from SmB.

After this cap modification, the 3' end of snRNA is trimmed. Previous studies have shown the presence of snRNA precursors in the cytoplasm, which is a few nucleotides longer than snRNAs (Madore et al., 1984). It is still unclear when exactly the 3' end of snRNA is processed and the enzyme which is responsible for shortened 3' end is not still known, but the biggest candidate is the exonuclease Dis3L2, which was shown to interact with snRNAs and degrade the extended RNAs (Huang et al., 1997; Labno et al., 2016; Ustianenko et al., 2016).

The cap hypermethylation triggers the snRNP import to the nucleus. Snurportin 1 (SPN1) binds the m₃G cap and through its importin β binding domain (IBB domain) binds importin β . It was shown, that SMN also directly binds the importin β and can be the second NLS receptor (Narayanan et al., 2002). Besides, the Importin 7 was shown to be necessary for snRNP nuclear transport in *Drosophila melanogaster* (Natalizio and Matera, 2013). When all modifications of snRNA (Sm ring formation, cap trimethylation, 3' end trimming) are finished, the snRNA is transported into the nucleus.

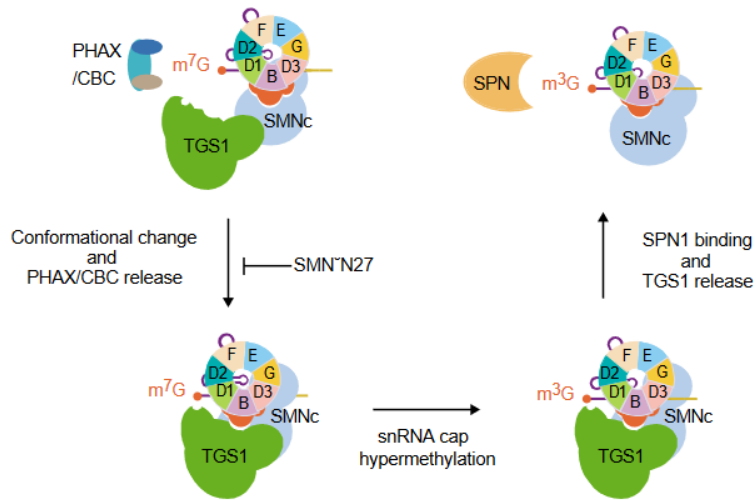


Figure 11: Model of TGS1 and SMN interactions in cytoplasmic snRNP biogenesis. Filled red circles represent C-terminal extensions of SmB/B', SmD1 and SmD3 proteins. m⁷G – monomethylguanosin cap, m³G- m_{2,2,7}G trimethylated cap, SPN1- Snurportin 1, TGS1- methyltransferase. Adapted from (Mouaikel et al., 2003)

SMN complex

The SMN complex is composed from the SMN protein, from Gemin proteins 2-8 and unrip protein (Baccon et al., 2002; Carissimi et al., 2005; Carissimi et al., 2006; Charroux et al., 1999; Charroux et al., 2000; Gubitz et al., 2002; Liu and Dreyfuss, 1996a; Pellizzoni et al., 2002). The SMN complex is dispersed in the cytoplasm and also in the nucleus, where it SMN complex accumulates in the nuclear foci called Gemini of coiled bodies – Gems (Liu and Dreyfuss, 1996b). SMN complex function in the cytoplasm is mainly in formation of the Sm ring and it might also participates in reimport of core snRNPs back into the nucleus (Massenet et al., 2002). In the cytoplasm binds preferentially SmB/B'-D3 dimers and drives them to the 6S particle (Sm D1, D2, E, F, G, pICln) holding together by Gemin2 and helps to close the Sm ring around the Sm site of snRNA (Pellizzoni et al., 1999). It was shown that the nuclear SMN complex associates with the Sm proteins, early spliceosome and U1 snRNP components but the exact role is still unknown (Meister et al., 2000; Stejskalova and Stanek, 2014).

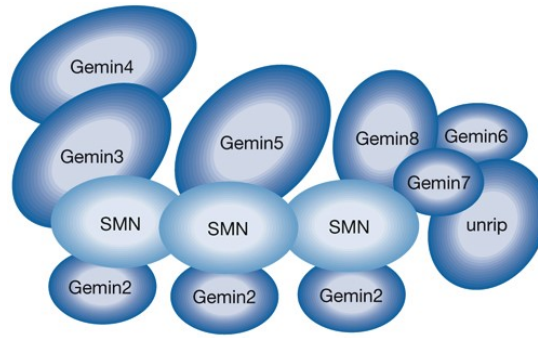


Figure 12: Schematic illustration of the survival motor neuron (SMN) complex. The survival motor neuron (SMN) protein binds Gemin2, Gemin3, Gemin5, Gemin7 and Gemin8, whereas Gemin4 and Gemin6 associate with SMN through interactions with Gemin3 and Gemin7, respectively. Gemin8 also binds the Gemin6–Gemin7 heterodimer, and mediates the association of Gemin6, Gemin7 and unrip with SMN (Pellizzoni, 2007).

1.1.8. SMN protein

Human SMN protein is 294 amino acids long polypeptide, which harbors multiple domains, including N-terminal nucleic-acid binding domain, a central Tudor domain and C-terminal proline-rich and YG domains. Mutations in all domains have been linked to Spinal Muscular Atrophy (SMA). SMA is neuromuscular disorder caused by a decreased level of the SMN protein (reviewed in Lanfranco, Vassallo, and Cauchi 2017).

As a substrate for binding of SMN serves RG rich domain where the Arginines are symmetrically dimethylated (Friesen et al., 2001b). These GR repeats are found in the C-termini of SmB/B', SmD1, SmD3, but also in other proteins such as a LSm4, fibrillarin or hnRNP and are able to interact with SMN protein (BRAHMS et al., 2001; Paushkin et al., 2002). Deletion of GR repeats completely abolishes the binding of SMN complex (Fig. 13) (Friesen and Dreyfuss, 2000).

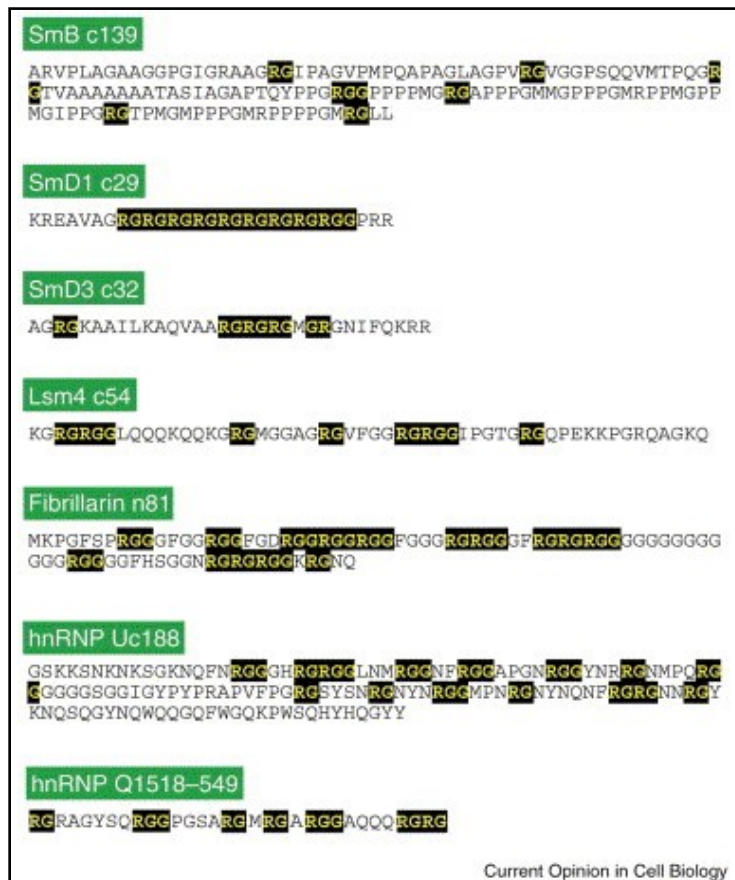


Figure 13: Schematic representation of arginine- and glycine-rich domains of the SMN complex substrates. RG-rich domains are typed in yellow with black background (Paushkin et al., 2002).

For the interaction with GR repeats is necessary the C-terminus of SMN protein, spanning exon 6 and 7, and its deletion completely disrupts binding of SMN to all of its known substrates and also its ability to oligomerize (Pellizzoni et al., 1999). It also contains QNQKE motif serving as a nuclear localization signal (Carrel et al., 2006). The C-terminus of SMN is highly conserved and over 96% of SMA patients show homozygous absence of either *SMN1* in exons 7 and 8 or exon 7 alone, showing that this domain is important for physiological function of the SMN protein (Brunhilde, 2000).

The N-terminus of SMN appears to be specific to mammals; it harbors binding sites for several critical interacting partners. SMN mutant lacking 27 N-terminal amino acids, displays a dominant negative effect on various SMN functions, including splicing, snRNP

reorganization, telomerase activity and hyper methylation by TGS1 activity (Mouaikel et al., 2003; Pellizzoni et al., 1998).

As I mentioned before, the main function of SMN complex is in snRNP assembly, but it undertakes more functions in the cell. It plays role in telomerase and snoRNA biogenesis (Bachand et al., 2002), in 3' end processing of histone mRNAs (Pillai et al., 2003), in translation (Sanchez et al., 2016) and also in DNA damage and repair (Takizawa et al., 2010). Here, I just mention only some important Gemin (Gemin2, Gemin3 and Gemin5), which have a significant role in the snRNP core assembly.

1.1.9. Gemin 2/SIP1

The Gemin2 is 32kDa protein and together with the SMN protein, is the most conserved component of the SMN complex. It forms Gemin2-SMN heterodimer. In the past, attention has been paid primarily to the SMN protein, which was thought to be responsible for forming the Sm ring around the snRNA. It was big surprise when several studies showed that SMN protein is not primary architect of the snRNP core assembly. Two crystallographic studies demonstrated that Gemin2 binds directly five Sm proteins (D1, D2, E, F and G) and holds them in proper orientation for subsequent snRNA binding and closing the Sm ring (Fig. 6) (Zhang et al., 2011). Consistent with this, the ubiquitously expressed Gemin2 is essential for viability of all eukaryotic organisms. Gemin2 gene deletion in the mouse causes embryonic lethality, at even earlier stage than in case of SMN gene deletion (Jablonka et al., 2002; Paushkin et al., 2000).

1.1.10. Gemin 3/Ddx20/DP103

Gemin3 is 99 kDa protein and contain DEAD box, which is found in a family of RNA helicases and it is therefore assumed that Gemin 3 is RNA helicase. It was shown that Gemin3 interacts directly with SMN, Gemin2, Gemin4 and with several Sm proteins (B/B', D2 and D3) *in vivo* and *in vitro* (Charroux et al., 1999; Charroux et al., 2000). The

interaction with the SMN protein is mediated by the C-terminus of Gemin3 (Fig. 14). The Gemin3 was originally isolated as a cellular factor that associates with Epstein-Barr virus nuclear proteins EBNA2 and EBNA3C, which play a role in the transcriptional regulation of latent viral and cellular genes (Grundhoff et al., 1999). Because of its role in the transcription the later studies showed its ability to interact with and modulate the activity of various transcription factors including steroidogenic factor 1 (SF-1), early growth response protein 2 (Egr2) or forkhead transcription factor (FOXO2)(Gilliam and Svaren, 2004; Lee et al., 2005; Ou et al., 2001; Yan et al., 2003). In addition to interactions with transcription factors, Gemin3 also forms SMN-independent complex with Argonaute 2 protein (Ago2) and numerous microRNAs (miRNAs)(Mourelatos et al., 2002). The role of Gemin3 in the snRNP assembly is not clear.

The disruption of Gemin3 gene leads to lethality in both vertebrates and invertebrates. The *gemin3*-null mice die at embryonic stage, but heterozygous mice are viable and fertile. It suggests that small amount of Gemin3 is sufficient for its function (Mouillet et al., 2008). It was also analyzed in *Drosophila melanogaster* and *Caenorhabditis elegans*, where the loss of Gemin3 results in defects in several aspects of development (Cauchi et al., 2008; Minasaki et al., 2009).

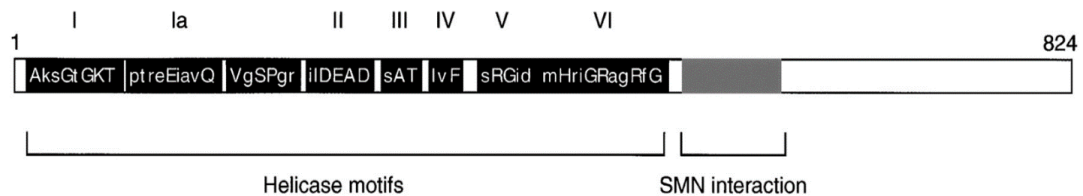


Figure 14: Gemin3 encodes a DEAD box containing RNA helicase. A, Schematic representation of the modular structure of Gemin3. The seven helicase motifs (I, Ia, II, III, IV, V, and VI) are represented by boxes with conserved amino acids in white. Upper cases are for the highly conserved residues, lower cases for the less conserved ones. The helicase motifs are boxed in black. The SMN interacting domain (amino acids 456–547) is boxed in gray. Adapted from (Charroux et al., 2000).

1.1.11. Gemin5

The Gemin5 is 170 kDa tryptophan-aspartic acid (WD) repeat protein that binds to SMN *in vivo* and *in vitro*. Like the other components of the SMN complex, Gemin5 is found in the cytoplasm and in gems (Gubitz et al., 2002). Gemin5 functions as the snRNA

binding protein of the SMN complex, which allows to distinguish snRNAs from other cellular RNAs (Battle et al., 2006a). Recent crystallographic studies showed that Gemin5 interacts directly with m⁷GpppG cap and the Sm site through its N-terminal WD motifs (Tang et al., 2016; Wahl and Fischer, 2016). In addition to the fact that Gemin5 plays one of the key role in the snRNP biogenesis, it has been shown that Gemin5 can bind viral internal ribosome entry site (IRES) elements. This interaction is mediated via the stem loop placed at the 3' end. It results in downregulating of translation (Fernandez-Chamorro et al., 2014; Pacheco et al., 2009). Further possible role of Gemin5 is in quality control of snRNP assembly in the cytoplasm. Last studies showed that Gemin5 is able to bind m⁷GpppG but not to trimethylated cap. It could indicate a new role of Gemin5 in the transport of incomplete snRNP to the cytoplasmic structures called P bodies, where the defective snRNPs are sequestered (Ishikawa et al., 2014; Xu et al., 2016).

Nuclear phase and Cajal bodies

Upon nuclear reentry, snRNPs appears in the nuclear structures called Cajal bodies (CBs). CBs are non-membrane nuclear structures discovered in 1903 by Ramón y Cajal in neurons (Fig. 15). In 1990 was discovered its scaffold protein called coilin (Raska et al., 1990). Coilin is 80kDa protein, which consists from N-terminal self-interacting domain, the central part containing NLS signals and C-terminal part containing Tudor like structure (Fig. 16). It was shown that Tudor domains in other proteins bind methylated amino acids such as a methyl-lysine or methyl-arginine, which we can find in the Sm proteins. The experiments with coilin's Tudor domain did not show any binding properties to these amino acids (Shanbhag et al., 2010). However, the C-terminal part was shown to interact with Sm fold of Sm proteins, but the molecular mechanism of coilin-Sm proteins binding is still unclear (Xu et al., 2005).

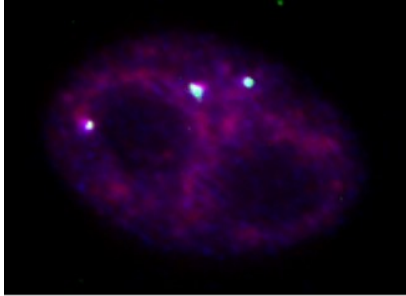


Figure 15: Cajal bodies in HeLa cell: coilin, marker of CBs labeled by Alexa-488 (green), U2 snRNA (red), DNA stained by DAPI (blue)

Coilin has been analyzed in *Mus musculus*, *Danio reiro*, *Arabidopsis thaliana* and *Drosophila melanogaster*. In all cases, coilin depletion causes CB disappearance but in plants and flies coilin knockout does not affect viability and fertility (Collier et al., 2006; Liu et al., 2009). By contrast, coilin depletion is lethal within 24 hours of development in zebrafish embryo and coilin^{-/-} mice are significantly less fertile (Minasaki et al., 2009; Strzelecka et al., 2010; Tucker et al., 2001; Walker et al., 2009).

CBs are place of several snRNP specific proteins binding with the snRNPs (Nesic et al., 2004). Moreover, CBs play also role in *de novo* formation of U4/U6 di-snRNP and U4/U6·U5 tri-snRNP as well as their post-splicing recycling (Novotny et al., 2011; Stanek et al., 2003). The mathematical modeling and measurement of snRNPs kinetic in CBs revealed that the higher concentration of snRNPs in CBs increases the snRNPs assembly rate by factor of 10 (Klingauf et al., 2006; Novotny et al., 2011).

In the CBs we can also find scaRNAs which are responsible for pseudouridylation and 2'-O-methylation of snRNPs. This modifications are confined to snRNA regions involved in formation of RNA-RNA or RNA-protein interactions that are crucial for spliceosome function, suggesting that they have beneficial effects on the efficacy or fidelity of pre-mRNA splicing (Jady et al., 2003).

Despite the important role of CBs in final maturation of snRNPs, it is still unknown, how are the snRNPs navigated to the CBs.

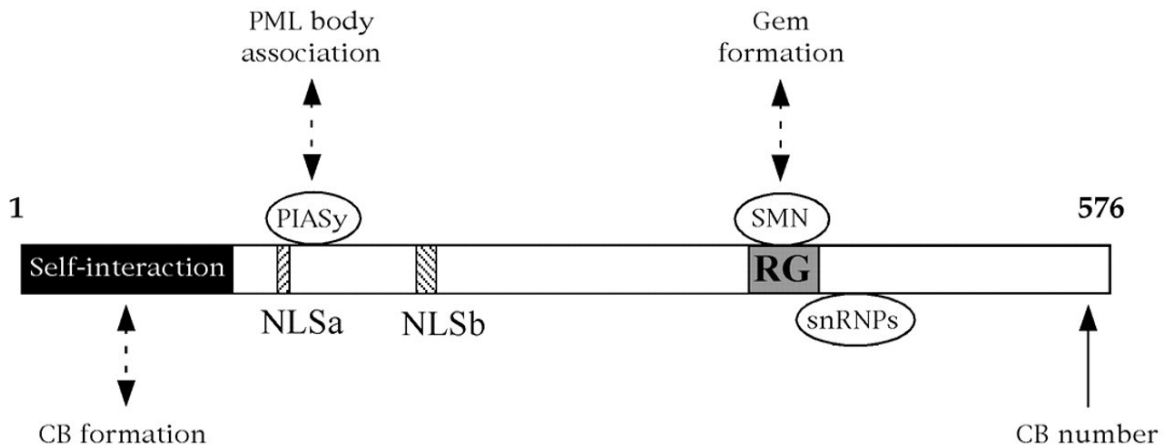


Figure 16: Schematic summary of coilin and its role in nuclear organization. The N-terminus contain Self-interaction domain (black box), C-terminus contain Glycine-Arginine rich domain (RG- grey box), NLS- nuclear localization signal. Adapted from (Sun et al., 2005).

Quality control of snRNP assembly

The snRNPs play an essential role in the splicing of pre-mRNA. For the cells and whole organisms is necessary to have the spliceosome correctly assembled. Otherwise, the cells have problems with protein composition and their functions which can lead to the cell death. Thus snRNP assembly has to be under precise control. The first check point is in the nucleus immediately after pre-snRNA transcription. This checkpoint controls assembly of the export complex comprising of CBC, PHAX, CRM1 and Ran-GTP on the 5' end of pre-snRNA (Fornerod et al., 1997; Izaurralde et al., 1995; Ohno et al., 2000b). If the export complex assembly fails defective snRNAs are retained in CBs (Suzuki et al., 2010). After this first control the pre-snRNA is exported into the cytoplasm where is second check point involving the Sm ring assembly. snRNAs that do not acquire the Sm ring fail to reimport back to the nucleus (Fischer et al., 1993). The previous studies have shown that immature snRNPs lacking Sm ring are targeted into the P bodies and degraded (Ishikawa et al., 2014; Ishikawa et al., 2018; Shukla and Parker, 2014).

However, the mechanism how the cells recognize the immature snRNPs is still unknown. Third quality control checkpoint, which occurs in the CBs, has been identified in our laboratory. snRNPs that do not form mature functional particles are sequestered in the CBs. However, the mechanism, which holds the immature snRNPs in the CBs has to be solved.

Degradation pathways in the cytoplasm

The RNA degradation is an important control point in the regulation of the gene expression. Generally, RNA degradation is initiated by deprotection of the transcript ends. In the case of mRNA, the ends are protected by 5' cap and 3' polyadenylated end. snRNAs contain the 5' cap but 3' end protection has not been identified. However, the Sm ring is a strong snRNA stabilizer and protects snRNAs against exo- and endonucleases (Fig. 17). When the transcripts lose the protection, the degradation proceeds via three different ways described below.

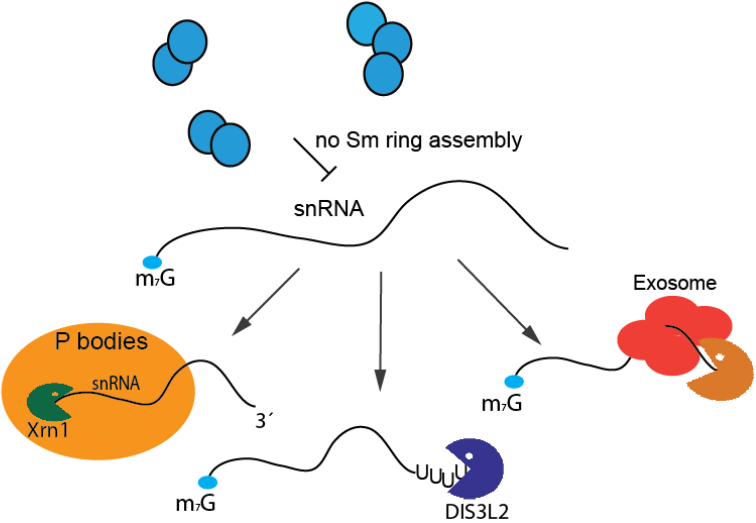


Figure 17: Three pathways of mRNA degradation in the cytoplasm. mRNAs can be degraded in the 5'-3' direction by Xrn1 exonuclease, or in the 3'-5' direction by either the exosome complex or Dis3L2 exonuclease, which acts independently from the exosome.

1.1.12. 5'-3' degradation pathway and P bodies

The two main 5'-3' exoribonucleases are Xrn1 and Xrn2. They differ in localization. Xrn2 is mainly in the nucleus and is involved in ribosomal RNA maturation, in transcription termination and telomere maintenance (El Hage et al., 2008; Luke et al., 2008; Wang and Pestov, 2011).

In this section, I will focus only on Xrn1 exoribonuclease, which is mainly cytoplasmic and play a role in degradation of 5' monophosphorylated RNA, such as a decapped or cleaved RNAs. This degradation pathway was best described in case of mRNA. The lifetime of mRNA is mainly determined by the length of poly(A) tail. The shortening of this tail is crucial for activation of both 5'-3' and 3'-5' degradation pathways. Deadenylation mRNAs are further processed by the decapping enzymes. The best characterized decapping enzyme is Dcp2, which collaborates together with decapping activator called Dcp1 to remove 7-methylguanosin cap from 5' end by hydrolysis (STEIGER et al., 2003). The interaction between mRNA and Dcp2 is enhanced by LSm1-7 ring formed around the shortened poly(A) tail (Nissan et al., 2010). After decapping the mRNA is prepared for degradation of Xrn1. Xrn1 can also degrades the snRNAs, which failed to acquire functional Sm ring (Shukla and Parker, 2014).

LSm1-7 ring

The LSm1-7 ring is composed from seven Sm-like proteins. All Sm proteins contain the characteristic Sm-fold and form the doughnut shaped structure. The LSm1-7 proteins are highly enriched in the cytoplasmic foci called P bodies (Ingelfinger et al., 2002). The main function of this LSm ring is in the stabilization of decapping enzymes (Dcp1 and Dcp2) on mRNA targeted for degradation (Nissan et al., 2010). The crystal structure revealed the highest binding preferences to the octa-U oligonucleotide and with lower affinity to the octa-A oligonucleotide at the 3' end of RNA (Zhou et al., 2014b). The mutational analysis also showed the direct interaction between the

extended C-terminus of LSm1 and RNA (Chowdhury et al., 2016). The LSm1-7 ring acts together with the Pat1 protein, which is the main activator of the decapping enzyme and is essential for the normal rates of mRNA decapping *in vivo* (Totaro et al., 2011). It also significantly enhances the RNA binding activity of LSm1-7 ring (Chowdhury et al., 2014).

P bodies

These cytoplasmic bodies were discovered more than ten years ago and over the past few years, the numbers of proteins detected in P bodies has increased exponentially. In addition to proteins involved in RNA degradation, P body components include proteins with roles in RNA surveillance, RNA interference and translation repression (reviewed in Eulalio et al., 2007).

Because the P bodies contain many factors important for mRNA decay, it arised the question, whether they serve as a place of RNA degradation or as a storage of these enzymes. Several lines of evidence indicate that RNA is degraded in these bodies. First, the inhibition of transcription with actinomycin D, or the exposure of cells to ribonuclease A, leads to P-body loss, which indicates that P-body assembly is dependent on RNA (Cougot et al., 2004; Teixeira et al., 2005). On the other hand, blocking mRNA decay in later stage results in an increase of P body numbers and size (Sheth and Parker, 2003). The P bodies are highly dynamic structures and their number and size change in different physiological conditions and also during the cell cycle (Yang et al., 2004).

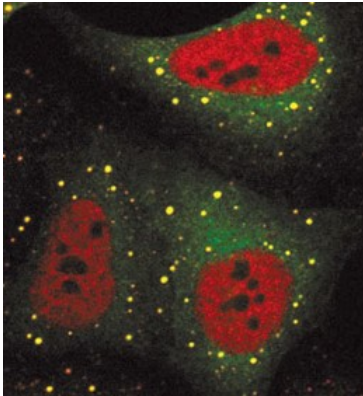


Figure 19: The image shows P-bodies (yellow) in HeLa cells expressing green fluorescent protein (GFP) fused to DCP1. Cells were stained with antibodies cross-reacting with EDC4 and a human nuclear antigen (red). Adapted from (Tritschler et al., 2010)

1.1.13. 3'-5' degradation pathways – exosome and Dis3L2

In the cytoplasm, we can find two independent 3'-5' degradation pathways. One is the cytoplasmic exosome complex which degrades mRNA with shortened poly(A) tail and second is Dis3L2 which prefers oligouridylated RNAs.

Exosome

The eukaryotic exosome is composed from eleven subunits. Two subunits (Rrp6 and Dis3) are nucleases and the rest of the proteins helping to navigate the substrate to the active sites (Vincent and Deutscher, 2009). It appears that the exosome rarely engages its substrates directly and that its activity *in vivo* is suppressed by depriving it of eligible substrates. Different classes of exosome substrates are recognized by specialized adapters and brought to the exosome. The exosome itself is rarely involved in the decision process that commits a molecule for degradation and in most cases it can be viewed as a common effector of the many pathways of RNA processing/degradation/surveillance (reviewed in Chlebowski et al., 2013)).

Dis3L2

The second poly(U) tail 3'-5' exonuclease in the cytoplasm is DIS3L2, which acts independently on the exosome. Mutations in the Dis3L2 gene were found in patients

with Perlman's syndrome. These mutations were suggested to inhibit the exonucleolytic activity of DIS3L2 and lead to deregulation of cell-cycle genes resulting in faster cell growth (Astuti et al., 2012). The discovery of the DIS3L2 exonuclease also showed that uridylation of 3' end can lead to RNA degradation. In mammalian cells two roles of the U-tailing were described. First is the biogenesis of pre-let7 miRNA (Heo et al., 2008; Heo et al., 2009; Heo et al., 2012) and second is in the degradation of replication-dependent histone mRNAs in the late S-phase of the cell cycle (Mullen and Marzluff, 2008; Schmidt et al., 2011). However, the uridylation need not always lead to the RNA degradation. The monouridylation of pre-miRNAs facilitates the DICER processing (Heo et al., 2012) and oligouridylation of U6 snRNA is a key part of its biogenesis (Lund and Dahlberg, 1992; Tazi et al., 1993).

In the mammalian cells at least four uridyltransferases (TUTs)- TUT1/4/7 and GLD-2 are known, which can add the non-templated UMPs to the 3' termini (Scott and Norbury, 2013). Recent studies suggested a broader spectrum of DIS3L2 targets. Upon viral infection, TUTases and DIS3L2 are involved in template-dependent miRNA degradation (TDMD) (Haas et al., 2016) and in the decay of improperly processed ncRNAs (Eckwahl et al., 2015). Using the catalytic non-active mutant of DIS3L2 D391N the *in vivo* targets of DIS3L2 were identified by crosslinking and immunoprecipitation followed by sequencing (CLIP-seq). The targets including RNAs include snRNAs, rRNAs, tRNAs and long non-coding RNAs (lncRNAs) (Ustianenko et al., 2016). Another study also showed snRNAs as a DIS3L2 targets suggesting that DIS3L2 can be the exonuclease responsible for 3' end trimming during snRNA biogenesis (Labno et al., 2016).

Quality control of snRNP assembly in the nucleus

After the Sm ring is formed, the mono-methyl cap is trimethylated and transported into the nucleus. It is directly navigated to the Cajal bodies, where the final snRNP maturation occur. Interestingly, inhibition of proteins involved in the U4/U6·U5 tri-snRNP formation or snRNP specific proteins lead to di-snRNP or incomplete snRNP accumulation (Novotny et al., 2015; Schaffert et al., 2004; Tanackovic and Kramer, 2005). Thus, it seems that role of CBs is not only to increase the efficiency of snRNP assembly, but also in precise control of final steps of snRNP maturation. However, the molecular mechanism of this final quality control are missing.

2. Materials and methods

2.1. Cell culture

HeLa cells were provided by Karla Neugebauer (Yale University, USA). HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 4.5 g glucose/L (Sigma) supplemented with 10% fetal bovine serum (FBS), 1% penicillin and streptomycin (Gibco). HeLa cells were incubated in 37 °C and with 5% CO₂.

2.2. Antibodies

For indirect immunostaining we used following antibodies:

- anti-coilin (5P10) antibody, kindly provided by M. Carmo-Fonseca (Institute of Molecular Medicine, Lisboa) (Almeida et al., 1998),
- anti-DDX6 antibody (Promega)

Secondary anti-mouse antibodies conjugated with Alexa-647 (Invitrogen) were used.

For immunoprecipitation we used following antibodies:

- goat anti-GFP antibodies obtained from David Drechsel (MPI-CBG Dresden, Germany)

For Western blotting we used the following antibodies:

- mouse anti-GFP (Santa-Cruz),
- mouse anti-U2B'' (Progen),
- rabbit anti-SmD1 (Abcam),
- rabbit anti-SmG (Abcam),
- rabbit anti-SF3a60/SF3A3 (Abcam),
- mouse anti-SF3b49 (Abcam),
- rabbit anti-PRPF31 (Abcam),
- mouse anti-GAPDH (Abcam),
- rabbit anti-β actin (Abcam),

- mouse anti-SMN (Sigma),
- rabbit anti-Flag (Sigma),
- mouse anti-tubulin, kindly provided by P. Draber (Institute of Molecular Genetics, CAS)
- anti-Sm antibody (Y12) was produced from a hybridoma cell line (a gift from Karla Neugebauer, Yale University, New Haven, USA) at the Antibody facility (Institute of Molecular Genetics, CAS).

Secondary goat anti-mouse or anti-rabbit antibodies conjugated with horseradish peroxidase (Jackson ImmunoResearch Laboratories) were used.

2.3. Plasmids

The U4-MS2 RNA construct, where the U4 snRNA is under the control of endogenous U4 promoter elements and a single MS2 loop is inserted into U4 stem loop II, was obtained from Edouard Bertrand (IGMM, CNRS, Montpellier (Bizarro et al., 2015)). The U2 snRNA full length, which includes the promoter sequence (bp 563 nt upstream of the U2 transcription start site), was amplified from HeLa genomic DNA using specific primers:

F: 5'AGTCGGATCCGGCAGAGGAACTCCAGCCCCT3',

R: 5'- ATAGGAATTCCAAGCCCGCCCCGCAGGTGCTACC- 3' and cloned into the pcDNA3 vector without CMV promoter using EcoRI/BamHI restriction sites. The U2 snRNA sequence is identical to the transcript ID ENST00000616345. The MS2 loop (5'-TAACATGAGGATCACCCATGTTTT- 3') was inserted into the stem loop IIb by site-directed mutagenesis using the primers

F: 5'-AGGAGAACAAATCCGAGGACAATATATTAAT -3' and

R: 5'- TTATAGACTATGCAGGAGATAACAAGGGTAA -3'.

The deletion construct lacking stem loops I and IIa,b (U2 Δ SLI+IIa,b-MS2) and lacking Sm site (U2 Δ Sm) were created by site-directed mutagenesis using the following primers:
U2 Δ SLI+IIa,b-MS2:

(F:5'-TAACATGAGGATCACCCATGT-3', R:5'- GCGCTCGCCTTCGCGCCCGCCGTCA- 3') and U2ΔSm (F: 5'-CCATTTAATATATTGTCCTCGG 3', R: 5'GAGCAGGGAGATGGAATAG -3').

The U4 Δ1-64-MS2 deletion construct (done by Klára Klimešová) was created from the U4-MS2 wt plasmid by site-directed mutagenesis using primers F: 5'-AAAACTTTTCCCAATACCCCGC-3', R: 5'-GGAAAGGCTTTATTCGCGCC-3'. All plasmids were co-transfected with A1-MS2-YFP (Brody and Shav-Tal, 2011), with MS2 coat protein containing the NES/NLS shuttling signal, using Lipofectamine3000 (Invitrogen) according to the manufacturer's protocol. Transfected HeLa cells were cultured for 24h at 37°C. SmB/B'-YFP and SmD1-GFP plasmids (Sleeman and Lamond, 1999) were provided by A. Lamond (University of Dundee, United Kingdom). SmD3-GFP plasmid was prepared from the total RNA of HeLa cells by RT followed by PCR using specific primers (Table 1) and cloned into the GFP-N1 vector (Clontech) using EcoRI/BamHI restriction sites. Deletion constructs (SmBΔCtail, SmD3ΔCtail, SmD1Δ1/4GR, SmD1Δ1/2GR and SmD1ΔGR) were created by PCR using primers listed in Table S1. D3Ala-GFP, BAla-GFP and D1Ala-GFP constructs were created by site-directed mutagenesis using specific primers (Table 1) and verified by DNA sequencing.

Lsm1-GFP construct was prepared from the total RNA of HeLa cells by RT followed by PCR using specific primers (Table 1) and cloned into the GFP-N2 vector (Clontech) using EcoRI/BamHI restriction sites. SMN-YFP plasmid was provided by M. Dundr (Rosalind Franklin University, USA).

U2snRNA more stable, U2 snRNA U bulge and U2 snRNA no hairpin constructs were prepared from U2 snRNA wt by site directed mutagenesis PCR using specific primers (Table 1) and verified by DNA sequencing.

Primers:

List of primers:	
T7 promoter	5'TAATACGACTCACTATAGGG 3'
U2wt	F: 5'TAATACGACTCACTATAGGG/ATCGCTTCTCGGCCTTTTGG3' R: 5'TGGTGCACCGTTCCTGGAGGT3'
U2ΔSLI	F: 5'TAATACGACTCACTATAGGG/TGTAGTATCTGTTCATCAG3' R: 5'TGGTGCACCGTTCCTGGAGGT3'
U2ΔSLIV	F: 5'TAATACGACTCACTATAGGG/ATCGCTTCTCGGCCTTTTGG3' R: 5'GGAGTGGACGGAGCAAGCTC3'
U2ΔSm	F: 5'GAGCAGGGAGATGGAATAG3' R: 5'CCATTTAATATATTGTCCTCGG3'
U2altSLIII	F: 5'TAATACGACTCACTATAGGG/ATCGCTTCTCGGCCTTTTGG3' R: 5'CGATTGCGTGGAGTATCTCCCTGCTCCAAAAATCCATTTAAT3'
U2ΔSLI,SLIIa,b	F: 5'TAATACGACTCACTATAGGG/ATATTAATGGATTTTTGGAACAG3' R: 5'TGGTGCACCGTTCCTGGAGGT3'
U2U1Sm	F: 5'GAGCAGGGAGATGGAATAG3' R:5' CACAAATTCATTTAATATATTGTCCT3'
U1wt	F: 5'TAATACGACTCACTATAGGG/ATACTTACTGGCAGGGGAG 3' R: 5'CAGGGGAAAGCGCGAACGCAG3'
U1ΔSm	F: 5'TAGTGGGGGACTGCGTTCGCG3' R: 5'ATGCAGTCGAGTTTCCACAT3'
U4wt	F: 5'TAATACGACTCACTATAGGG/AGCTTTGCGCAGTGGCAGTAT3' R: 5'CAGTCTCCGTAGAGACTGTCA3'
U4ΔSm	F: 5'TAATACGACTCACTATAGGG/AGCTTTGCGCAGTGGCAGTAT3' R: 5'-CAGTCTCCGTAGAGACTGTGGCCGGCCCAATGCCGAC-3'
U5wt	F: 5'TAATACGACTCACTATAGGG/ATACTCTGGCTTCTCTTCAGAT3' R: 5'AGTGCTGGATTAGCCTTGCCAA3'
U5ΔSm	F: 5'CACAAACGTGCCTTGCCTTGG3' R: 5'GGGTTAAGACTCAGAGTTGTTCT3'
7SK wt + T7	F: 5'TAATACGACTCACTATAGGG/GGATGTGAGGGCGATCTGGCTG3' R: 5'AGAAAGGCAGACTGCCACATGC3'
7SKSm	R: AGAAAGGCAGACTGCCACATGCAGCGCCTCATTGGATGTGCAAAAATCT3' 5'
7SKSMN	R: 5'TGGTACCGGTCATCATATTTACACCCAGTACCTAC3'
7SKSm+SMN	R: 5'TGGTACCGGTCATCATATTTACACCCAGTACCTACAAAAATTGGT3'
Alu wt + T7	F:5'TAATACGACTCACTATAGGG/CTCCCGCAACGCTACTCTCGT3' R: 5'AGTAGAGACGGGGTTTCACCATGTT3'
Alu + Sm	R: 5'TACCTACAAAAATTGGTCAGCATGGGGGCCCTGCCAGCTACAT 3'
Alu + Sm + SMN	R: 5'TGGTACCGGTCATCATATTTACACCCAG TACCTACAAAAATTGGTCAGCA3'

SRP wt + T7	F: 5' TAATACGACTCACTATAGGG/CTCCCGCAACGCTACTCTCGT3' R: 5' TGGGGGCCCTGCCAGCTACAT 3'
SRP+Sm	R: 5' TACCTACAAAAATTGGTCAGCATGGGGGCCCTGCCAGCTACAT 3'
SRP+Sm+SMN	R: 5' TGGTACCGGTCATCATATTTACACCCAG TACCTACAAAAATTGGTCAGCA 3'
RT-qPCR primers	
U1 wt	F: 5' AACTTACCTGGCAGGGGAG 3' R: 5' CAGGGGAAAGCGCGAACGCA 3'
U2 wt	F: 5' CTCGGCCTTTTGGCTAAGAT 3' R: 5' CGTTCCTGGAGGTACTGCAA 3'
U4 wt	F: 5' TGGCAGTATCGTAGCCAATG 3' R: 5' CTGTCAAAAATTGCCAGTGC 3'
U5 wt	F: 5' ACTCTGGCTTCTCTTCAGATCA 3' R: 5' GCCATTCTAACTGGCATGAG 3'
Sm proteins	
SmD1	Δ GR F: 5' AGCGAATTCTGATGACCCTGAAGAACAGAGAACCT3' R: 5' GCGGGATCCTTCCTGCAACAGCTTCCCTTTTCTTA3'
	Δ 1/4GR F: 5' AGCGAATTCTGATGACCCTGAAGAACAGAGAACCT3' R: 5' ATAGGATCC T TCTTCCTCTGCCACGGCCACG3'
	Δ 1/2GR F: 5' AGCGAATTCTGATGACCCTGAAGAACAGAGAACCT3' R: 5' ATAGGATCC T GTCCTCTTCCTCTTCCTCTTCCTC3'
SmD2	wt F: 5' AGCGAATTCTGATGAGCCTCCTCAACAAGCCCA3' R: 5' GCGGGATCCTCTTGCCGGCGATGAGCGGGTT3'
	Δ helix F: 5' AGCGAATTCTGATGCAATACCCAAGTGCTCATCAA3' R: 5' GCGGGATCCTCTTGCCGGCGATGAGCGGGTT3'
	Δ 1-24 F: 5' AGCGAATTCTGATGAACACCGGTCCACTCTCTGTGC3' R: 5' GCGGGATCCTCTTGCCGGCGATGAGCGGGTT3'
	Δ 111-118 F: 5' AGCGAATTCTGATGAGCCTCCTCAACAAGCCCA3' R: 5' GCGGGATCCTCCGCAGGACCACGATGACTG3'
SmD3	wt F: 5' AGCGAATTCATGTCTATTGGTGTGCCGATT3' R: 5' GCGGGATCCGTTCTTCGCTTTTGAAAGATG3' Δ Ctail F: 5' AGCGAATTCATGTCTATTGGTGTGCCGATT3' R: 5' GCGGGATCCGTTCAAGGCCCAAGTGGCCGCA3'
	Ala F: 5' GCAATGGCAGCGGCAAACATGTTTCAAACCGAAGA3' R: 5' TGCTGCTGCTGCTGCTGCGGCCACTTGGGCCTTGAGAATA3'
	Δ Ctail F: 5' AGCGAATTCTGATGACGGTGGGCAAGAGCAGCA3' R: 5' GCGGGATCCTCAGGTGGGTACTGGGTTGGAG3'
SmB/B'	Δ Ctail F: 5' AGCGAATTCTGATGACGGTGGGCAAGAGCAGCA3' R: 5' GCGGGATCCTCAGGTGGGTACTGGGTTGGAG3'
MS2 constructs	
MS2 loop	5' TAACATGAGGATCACCCATGTTTT 3'
U2 wt FL	F: 5' AGTCGGATCCGGCAGAGGAACTCCAGCCCCT3' R: 5' ATAGGAATCCAAGCCGCCCGCAGGTGCTACC3'

U2 wt FL MS2	F: 5'AGGAGAACAAATCCGAGGACAATATATTAAT 3' R: 5'TTATAGACTATGCAGGAGATACAAGGGTAA3'
U2ΔSLI+IIa,b-MS2	F: 5TAACATGAGGATCACCCATGT3' R: 5'GCGCTCGCCTTCGCGCCCGCCGTCA3'
U4Δ1-64-MS2	F: 5'AAAACTTTTCCCAATACCCCGC3' R: 5'GGAAAGGCTTTATTCGCGCC3'

NSS mutants:

U2 stableNSS	F: 5'CCTCTATCCGAGGACAATATATTAATGGAT 3' R: 5'ACGTATCAGATATTAATTTTTAGGAACAGATACT 3'
U2 noNSS	F: 5'GGAAAAGGGAGATGGAATAGGAGCTTGCTCCGT 3' R: 5'AAAATTCATTTAATATATTGTCCTCGGATA 3'
U2stableNSS-MS2	F: 5'CCTCTATCCTAACATGAGGATCACCCATGT 3' R: 5'ACGTATCAGATATTAATTTTTAGGAACAGATACT 3'
U2 no NSS-MS2	F: 5'GGAAAAGGGAGATGGAATAGGAGCTTGCTCCGT 3' R: 5'AAAATTCATTTAATATATTGTCCTCAAACATGGG 3'
LSm1	
LSm1	F: 5'GCGGAATTCTGATGAACTATATGCCTGGCAC 3' R: 5'GCGGGATCCTGTACTCATCAAGAGTATCTGCT 3'

Table 1: List of primers

2.4. Heterokaryon preparation

Hela cells were cultured on 3 cm Petri dish to 90% confluency, washed with 1x PBS and treated with 50% Polyethylen Glycol 8000 (Sigma Aldrich) for 5 min at room temperature. Then cells were washed 3 times with 1x PBS, cell culture medium was added and cells were incubated at 37°C for 2h before injection.

2.5. RNAi

For RNA interference experiments, 30-50% confluent cells were transfected using these siRNAs (life technologies, SMN- Sigma Aldrich)

siRNA	Sequence	Catalogue number	Working conc.	Incub. time
SmB/B'	5'-UCUACUGUCAUUGAGACCAga-3'	Silencer select s13219	20nM	48h

SmD1	5'-UUAGGUUCAACAUCCACAAgt-3'	Silencer select s13230	20nM	72h
SmG	5'UACUAAUUUCCUCGUAUUACca3'	Silencer select s13243	20nM	48h
Xrn1	5'-GAGAGUAUAUUGACUAUGAtt-3'	Silencer select s29015	20nM	72h
LSm1	5'GAAGGACACUUAUAGGCUUtt 3'	Silencer select s26063	20nM	48h
Gemin3	5'GCAUACAUAUGGUUAUAGCAtt 3'	Silencer select s22143	20nM	72h
SMN	5'CCAGAGCGATGATTCTGACATTTGGGATG 3'	EHU 148811	35 nM	72h
Gemin5	5'GAAAUACGGCAACACGAAAtt 3'	Silencer select s24773	20 nM	48 h
Gemin2	5'AUAUCUGAGUAAUUGGUUUt3'	Silecer select s16119	20nM	48h
NC5		Silencer select am4642	20nM	

Table 2: List of siRNAs

siRNAs were transfected with Oligofectamine (Invitrogen) according to the manufacturer's protocol.

2.6. RT-qPCR analysis

RNA was isolated from immunoprecipitation by phenol/chloroform extraction and was analyzed by RT-qPCR. Reverse transcription was performed using the SuperScriptIII (Invitrogen) and random hexamers (Eastport). The synthesis was performed as follows: 5 min at 65 °C, 5 min at 25 °C and 1h at 50 °C. The inactivation was 15 min at 75 °C. The

cDna was analyzed by qPCR using a Roche Light Cycler 480 standard protocol (45 cycles, 60 °C annealing). Complete list of RT-qPCR primers is available in Table 1.

2.7. *In vitro* transcription

All DNA templates for *in vitro* transcription were prepared by PCR using Phusion polymerase (Biolab). The primers are listed in Table 1. All forward primers contained the T7 promoter sequence (5'TAATACGACTCACTATAGGG 3'). ΔSm site and U2withU1Sm mutants were created by site-directed mutagenesis using ΔSm and U2withU1Sm primers (Table 1). Plasmids containing full-length U1, U2, U4 and U5 snRNAs (a gift from Karla Neugebauer, Yale U., New Haven, USA) were used as templates. 7SK RNA and Alu cytoplasmic RNA were cloned from total HeLa RNA isolated by TRIZOL reagent (Invitrogen) according to the manufacturer's protocol and cDNA was synthesized using specific primer:

7SK -R: 5' - AGAAAGGCAGACTGCCACATGC -3'

Alu - R: 5' - AGTAGAGACGGGGTTTCACCATGTT - 3'

by SuperScriptIII (Invitrogen). SRP RNA was isolated from total *Escherichia coli* as previously described (Hnilicova et al., 2014) and cDNA was synthesized using specific primer R: 5' - TGGGGGCCCTGCCAGCTACAT - 3' by SuperScriptIII (Invitrogen). WT and Sm+SMN mutants (7SK, Alu and SRP) were prepared by PCR using Phusion polymerase (Biolab). Primers are listed in Table 1.

Fluorescently labelled RNAs were prepared by *in vitro* transcription using MegashortscriptIII kit (Invitrogen) using the manufacture protocol. The mixture contained additional UTP-Alexa 488 (Invitrogen) in ration 1:4 (UTP-Alexa488 : UTP) and trimethylated cap analog ($m_3^{2,2,7}G(5')ppp(5')G$ (Jena Bioscience) in ratio 1:4 (cap analog : GTP). I incubated the mixture in 37 °C over night. After synthesis, RNA was isolated by phenol/chloroform extraction, precipitated and dissolved in nuclease-free water. RNA was diluted in solution containing dextran-TRITC 70-kDa (Sigma-Aldrich) to final concentration 200ng/μl.

2.8. Prediction of snRNA secondary structure

Secondary structure of all U2 snRNA mutants was analyzed by mathematical modeling). Structure analysis was carried out using the Vienna RNA package (Lorenz et al., 2011). Minimum free energy RNA secondary structures were used, generated by both constrained and unconstrained prediction. For the first, RNAfold was used. The latter was accomplished using constrained RNAfold (RNAfold -C). Structures were plotted using RNAplot (done by Josef Pánek, Microbiology Institute ASCR, Prague)

2.9. Microinjection

HeLa cells were grown on glass coverslips for 24h and RNA was microinjected using InjectMan coupled with FemtoJet (Eppendorf). The injection pressure (p_i) was 150 hPa and compensation pressure (p_c) was 50 hPa. Cells were then rinsed twice in PBS and and processed for indirect immunofluorescence.

2.10. Indirect immunofluorescence and image acquisition

HeLa cells grown on coverslips were fixed in 4% paraformaldehyde in piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES) for 15 minutes. Cells were permeabilized with 0.2% Triton X-100 (Sigma Aldrich) for 5 minutes and blocked with 5% normal goat serum in PBS (Jackson ImmunoResearch) for 10 minutes. Then cells were incubated with primary and secondary antibodies each for 1 hour. After washing steps the coverslips were mounted to microscope slides using Fluoromount G containing 4,6-diamidino-2-phenylindole (DAPI) (Southern Biotech) for DNA staining. Images were acquired using the DeltaVision microscopic system (Applied Precision) coupled to an Olympus IX70 microscope equipped with an oil immersion 63x objective/1.42 NA, Photometrics CoolSNAP HQ2 camera (Princeton Instruments) and acquisition software SoftWorx (Applied Precision). Stacks of 20 z-sections with 200 nm z steps were collected per sample and subjected to mathematical deconvolution using SoftWorx software.

Maximal projections of deconvolved pictures were generated by SoftWorx and are presented. For high-content microscopy, samples were scanned using automated acquisition driven by the Acquisition Scan^R program using Scan^R system (Olympus) equipped with an oil-immersion objective (60x/1.35 NA). A total of 225 images were taken per sample. Several hundreds of cells were collected per sample. Each image was reconstructed from stacks of ten optical sections with 300 nm z step and automatically restored using a measured point spread function implemented in the Analysis Scan^R software (Olympus). Cellular compartments were automatically identified based on fluorescence intensity combined with compartment edge detection. Cell nuclei were visualized using 40 ,6-diamidino-2-phenylindole (DAPI) staining, and anti-coilin antibody was used to visualize CBs. Total intensities, areas, and counts for each cellular object were obtained using the Analysis Scan^R software. The ratio of fluorescence in CBs versus the nucleoplasm was calculated according to:

$$R = \frac{\frac{\sum total IF_{CB \text{ per nucleus}}}{\sum Area_{CB \text{ per nucleus}}}{total IF_{nucleus} - \sum total IF_{CB \text{ per nucleus}}}{Area_{nucleus} - \sum Area_{CB \text{ per nucleus}}}$$

where the mean of CB intensities per nucleus was calculated and then divided by the mean fluorescence of the rest of the nucleus. The mean and SEM of three biological experiments were calculated and plotted. For high-content microscopy, samples were scanned as described previously (Novotný et al. 2015). Mean and SEM of three biological experiments were calculated and plotted. Statistical significance was analyzed by the Student T-test.

4% paraformaldehyde (PFA) in PIPES
4% (w/v) PFA (Sigma Aldrich)
0.1M PIPES pH 6.9 (Sigma Aldrich)
2 mM MgCl ₂ (Sigma Aldrich)
1mM EGTA pH 8 (Sigma Aldrich)

2.11. Fluorescence *in situ* hybridization – FISH

FITC labeled DNA probes (Table S2) were used against U1, U2 and U4 snRNAs. Forty-eight hours after siRNA transfection cells were fixed in 4% paraformaldehyde/PIPES (Sigma) for 15 min, permeabilized with 0.5% Triton X-100 for 5 min, and incubated with anti-DDX6 antibodies as a marker of P bodies followed by incubation with secondary antibody conjugated with Alexa-647 (Life Technologies). Cells were again fixed in 4% paraformaldehyde/PIPES for 5 min, quenched for 5 min in 0.1 M glycine/0.2 M Tris, pH 7.4, and incubated with FITC-labeled U1, U2 or U4 probes in 2× SSC/50% formamide/10% dextran sulfate/1% BSA for 60 min at 37°C. After washing in 2× SSC/50% formamide, 2× SSC and 1× SSC. Coverslips were mounted to microscope slides using Fluoromount G containing DAPI (Southern Biotech) for DNA staining. Images were collected using a DeltaVision microscope system as described above.

Tris buffer:	20xSSC (saline sodium citrate) buffer
Trizma base (Sigma Aldrich)	3M NaCl (Sigma Aldrich)
pH 7.4 adjusted with HCl (Penta)	300mM sodium citrate (Sigma Aldrich)
	pH 7 adjusted with HCl (Penta)

DNA probes	sequences
U1	5' Cy3-CCTTCGTGATCATGGTATCTCCCCTGCCAGGTAAGTAT 3'
U2	5' Cy3-GAACAGATACTACACTTGATCTTAGCCAAAAGGCCGAGAAGC3'
U4	5' Cy3-TCACGGCGGGGTATTGGGAAAAGTTTTCAATTAGCAATAATCGCGCCT 3'

Table 3: DNA probes sequences for FISH

2.12. snRNP precipitation

HeLa cells were grown on 10 cm Petri dish were placed on ice, washed three times with ice cold Mg-PBS and harvested into NET-2 buffer supplemented with a complete

mix of protease inhibitors (Roche) and with the RNAsin (Promega). Then cells were pulse-sonicated for 50 s on ice. Cell extracts were centrifuged at 20 000g and the supernatant incubated with Protein-G Agarose beads (Santa Cruz) coated with goat anti-GFP antibody for 4 h at 4°C. Captured complexes were extracted by bead incubation in 2x protein sample buffer for 5 min at 95°C and the precipitated proteins were detected by Western blotting. RNA was isolated by phenol/chloroform extraction and precipitated for 1h at -20°C with 100% ethanol and analyzed by RT-qPCR or resolved on 7M urea denaturing polyacrylamide gel and silver stained (all used chemicals from Sigma Aldrich).

NET-2 buffer
50mM Tris-HCL pH 7.5
150 mM NaCl (Sigma Aldrich)
0.05% Nonidet P40 (Sigma Aldrich)
dH ₂ O

2x Sample buffer
4% SDS (sodium dodecyl sulphate, Sigma Aldrich)
10% 2-mercaptoethanol (Sigma Aldrich)
20% glycerol (Lechner)
0.004% bromphenol blue (Sigma Aldrich)
125 mM Tris-HCL pH 6.8
dH ₂ O

2.13. SDS-PAGE and Western blot (WB)

Cells were harvested from 12-well plates into 100 µl 2x Sample buffer and denaturated at 95 °C for 5 min. Samples from IP were prepared as previously described. Protein samples in the form of either cell lysates or samples from IP were separated by polyacrylamide gel electrophoresis in SDS running buffer (SDS-PAGE). For testing siRNA efficiency were prepared cell lysates. The Western blot was done as is described previously (Huranova et al., 2009).

The membrane stain by Poncaeu S solution (0.1% w/v in 5% acetic acid, Sigma Aldrich) for 10 min. After staining the membrane was cut into smaller pieces and blocked with 10% (w/v) low-fat milk in PBST for 30 min. After blocking the membrane was incubated with primary antibody diluted in 1% (w/v) low-fat milk in PBST

supplemented with 0.1% sodium azide (Sigma Aldrich) for 60 min. Then the membrane was washed 3x 10min in PBST and incubated for 60 min with anti-mouse or anti-rabbit secondary antibody conjugated with horseradish peroxidase (HRP) diluted 1:10000 in 1% (w/v) low-fat milk in PBST. After washing (3x10min in PBST), the SuperSignal™West Femto or Pico Maximum Sensitivity Substrate (Life technologies) was poured over the membrane and chemiluminescence was detected by LAS-3000 Imager (Fujifilm).

2.14. Core-snRNP *in vitro* reconstitution

(done by Cyrille Girard, Max Planck Institute for Biophysical Chemistry, Germany)

snRNP reconstitution were carried out as described in (Malatesta et al., 1999; Segault et al., 1995; Sumpster et al., 1992). Typically, 15 pmol of *in vitro* transcribed snRNA were assembled with 20 µg of native snRNP proteins (TP's) in 30µl of reconstitution buffer 20mM HEPES-KOH pH 7.9, 50mM NaCl, 5mM MgCl₂. Reconstituted snRNPs were microinjected without further purification.

2.15. Preparation of U2 snRNPs, SF3a and SF3b

(done by Cyrille Girard, Max Planck Institute for Biophysical Chemistry, Germany)

Human SF3a and SF3b complexes were affinity-purified from HeLa nuclear extract (Dignam et al., 1983) in G buffer (20 mM Hepes, pH 7.9, 1.5 mM MgCl₂, 10% (w/v) glycerol, 0.5 mM DTE, 0.5 mM PMSF) containing 250 mM NaCl, that was first passed over an anti-m3G immunoaffinity column. The NaCl concentration was increased to 600 mM to ensure the complete dissociation of SF3a and SF3b from U2 snRNPs, and the extract was applied to affinity columns with covalently bound anti-peptide antibodies against human SF3B1 (amino acids 99-113) or SF3A2 (amino acids 444-458). Bound complexes were eluted with an excess of the cognate peptide in G buffer containing 600 mM NaCl, further purified by gel filtration on a Superose 6 column (SF3b) or on a Superdex 200 column (SF3a), and then concentrated by ultracentrifugation. To isolate

12S U2 snRNPs, a mixture of anti-m3G affinity purified spliceosomal snRNPs were first separated by 10-30% (v/v) glycerol gradient centrifugation in G buffer containing 150 mM NaCl. The 12S peak (containing both 12S U1 and U2 snRNPs) was subsequently applied to a HiTrap Heparin HP column (GE Healthcare) and the bound snRNPs were eluted with a linear salt gradient (50 to 1000 mM NaCl in buffer containing 20 mM Tris, pH 7.9, 1.5 mM MgCl₂, 1 mM DTT). The eluted 12S U2 snRNPs were subsequently concentrated by ultracentrifugation. 15S U2 snRNPs were generated by combining equal molar amounts of purified 12S U2 and SF3b, whereas 17S U2 snRNPs were generated by combining equal molar amounts of purified 12S U2, SF3b and SF3a, and then incubating for 1h on ice. Gradient centrifugation confirmed that the vast majority of the 12S U2 snRNPs were converted to 15S or 17S complexes, under these conditions (data not shown).

3. Results

This section is composed of our published and unpublished data divided into three projects. If not stated otherwise in the Materials and methods section, I performed the described experiments myself.

- **1. Project: Molecular mechanism of the quality control of snRNPs biogenesis in Cajal bodies**

We focus on the targeting of snRNP into the Cajal bodies and their role in the quality control of snRNP biogenesis.

Figures done by me: 21, 22, 23, 24, 26 (A-RNA gel, IP1), 27, 28, 29, 30 (A, B, D), 31

Figures done by Klára Klimešová: 26 A (IP2), B, 31C

Figures done by Josef Pánek (IM, ASCR, Prague): 20

Figures done by Cyrille Girard (MPIBPC, Germany): 30, 32

Figure done by Ivan Novotný (IMG, Prague): 33

Two articles were published:

“SART3-dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal bodies”

Novotný I., Malinová A., Stejskalová E., Matějů D., Klimešová K., Roithová A., Švéda M., Knejzlík Z., Staněk D. (2015) Cell Reports, 10(3), 429-440

This first article was published in the Cell reports journal (IF 2016/2017: 8.282) in 2015. I was a co-author and I contributed to the experiments showing accumulation of incomplete snRNP in the Cajal bodies, namely the microinjection of WT and mutated U4 snRNA (Fig. 33).

“The Sm-core mediates the retention of partially-assembled spliceosomal snRNPs in Cajal bodies until their full maturation”

Roithová A., Klimešová K., Pánek J., Will CL., Luhrmann R., Staněk D., Girard C. (2018) *Nucleic Acid Res.*, 46(7), 3774-3790

The second article published in *The Nucleic Acid Research* (IF 2016/2017: 10.162) in 2018. I was the first author of this article and I performed all the microinjection experiments and prepared all *in vitro* transcribed constructs for microinjection and all U2-MS2, SmB/B', SmD3 and SmD1 constructs for Western blot. I did WB with SmB/B', SmD3 constructs and with U2-MS2 constructs. I also prepared the figures for this article and I participated in the manuscript writing.

- **2. Project: “Identifying role of Gemin3 in Sm ring assembly”**

We focus on the role of Gemin3 protein (a component of SMN complex) in the Sm ring formation in the cytoplasm. This project has not been published yet. I performed all the experiments except the prediction of the secondary structures of snRNAs (done by Josef Pánek, Microbiology Institute ASCR, Prague).

- **3. Project: “The recognition of defective snRNAs and their targeting into the P bodies”**

This project has not been published yet.

I performed all experiments and figures presented here.

3.1. The Sm-core mediates the retention of partially-assembled spliceosomal snRNPs in Cajal bodies until their full maturation

The Cajal bodies are a place the final snRNP maturation (Nesic et al., 2004; Stanek et al., 2003). However, the signal, which navigates the snRNP into the Cajal bodies has remained unknown. In my thesis, I characterized the Sm ring as a Cajal body targeting signal and proposed its role in the quality control.

Sm and SMN binding sites are necessary for targeting of microinjected U2 snRNA to Cajal bodies

To determine snRNA sequences that are necessary for the targeting of snRNAs to Cajal bodies, we utilized the U2 snRNA as a model RNA molecule. We prepared several deletion mutants. The effects of mutations on snRNA structure were estimated by mathematical modeling (Fig. 20). I transcribed mutated snRNAs *in vitro* in the presence of UTP-Alexa488 and the trimethyl cap. *In vitro* transcribed snRNAs were microinjected into HeLa cells (into the nucleus or the cytoplasm) together with TRITC-labeled dextran-70kDa that does not cross the nuclear membrane and serves as a marker of nuclear or cytoplasmic injection (Fig. 21). Cells were incubated for 60 min following injection, fixed and, a marker of CBs, coilin was detected by indirect immunofluorescence. WT U2 snRNA accumulated in CBs after both cytoplasmic and nuclear injection (Fig. 21A). Nuclear localization of cytoplasmatically injected RNAs showed that WT U2 snRNAs acquired the Sm-ring and were imported into the nucleus.

Then we mapped the importance of different snRNA domains for CB targeting. We first deleted stem-loop I (U2 Δ SLI), which is important for binding of the SF3A3 (SF3a60) protein from the SF3a complex (Brosi et al., 1993). We observed that the U2 snRNA

without the stem-loop I was highly accumulated in CBs (Fig. 21B). Deletion of the stem-loop IV (U2 Δ SLIV), which contains a part of the U2 SMN binding motif and interacts with the U2-specific dimer, SNRNPA1 (U2A)/ SNRPB2 (U2B'')(Boelens et al., 1991; Pellizzoni et al., 2002), decreased CB localization in comparison to WT (Fig. 21A). Next, we removed both 3' end stem loops III and IV (U2 Δ SLIII+IV), which together form the binding platform for the SMN complex (Yong et al., 2002b). The deletion of both stem loops completely inhibited CB targeting (Fig. 21D). The U2 Δ SLIII+IV RNA remained in the cytoplasm after cytoplasmic injection, indicating that the Sm ring was not formed and that snRNA without the Sm ring was not able to cross the nuclear membrane. In native snRNAs, the Sm site is found between the two stem loops, which is a spatial organization missing in the U2 Δ SLIII+IV RNA. Therefore, we replaced stem loops III and IV with a shortened stem-loop III (U2altSLIII) that lacked 18 central nucleotides, 120-137, which were previously shown to bind the SMN complex (Battle et al., 2006b). After cytoplasmic injection, the U2altSLIII RNA was partially retained in the cytoplasm while CB accumulation was significantly reduced. The U2altSLIII RNA injected into the nucleus failed to accumulate in CBs (Fig. 21E).

These data suggest that the SMN complex is, directly or indirectly, via Sm-ring assembly required for CB localization of the U2 snRNA. To further analyze CB targeting, we prepared U2 snRNA lacking the Sm site (U2 Δ Sm site) (Fig. 21F). Consistent with the lack of the Sm ring, the U2 Δ Sm site RNA injected into the cytoplasm was not imported into the nucleus. The U2 Δ Sm site RNA that was injected into the nucleus mimicked the behavior of RNAs that lacked the SMN binding site (U2 Δ SLIII+IV); the deletion of the Sm site completely abolished CB localization. These data collectively demonstrate that Sm and SMN binding sites are together necessary for U2 snRNA targeting to CBs.

To test whether Sm and SMN binding sites are also sufficient in the targeting of the U2 snRNA to CBs, we deleted the first 94 nt of U2 snRNA containing stem loops I and IIa,b and leaving only the Sm and SMN sites (U2 Δ SLI+IIa,b). This RNA was localized to CBs similarly to WT U2 snRNA (Fig. 21G). The stem-loop IV binds U2B'' and U2A, which might still target U2 snRNA to CBs. Therefore, we further deleted stem-loop IV

(U2 Δ SLI+IIa,b+IV). This minimal RNA of only 55 nucleotides, was localized to the CB, but CB accumulation was lower than in case of WT or U2 Δ SLI+IIa,b RNAs. The reduced CB accumulation was more pronounced in the case of nuclear microinjection (Fig.21H). These results together suggest that the minimal sequence which targets U2 snRNAs to CBs comprised of Sm and SMN binding sites.

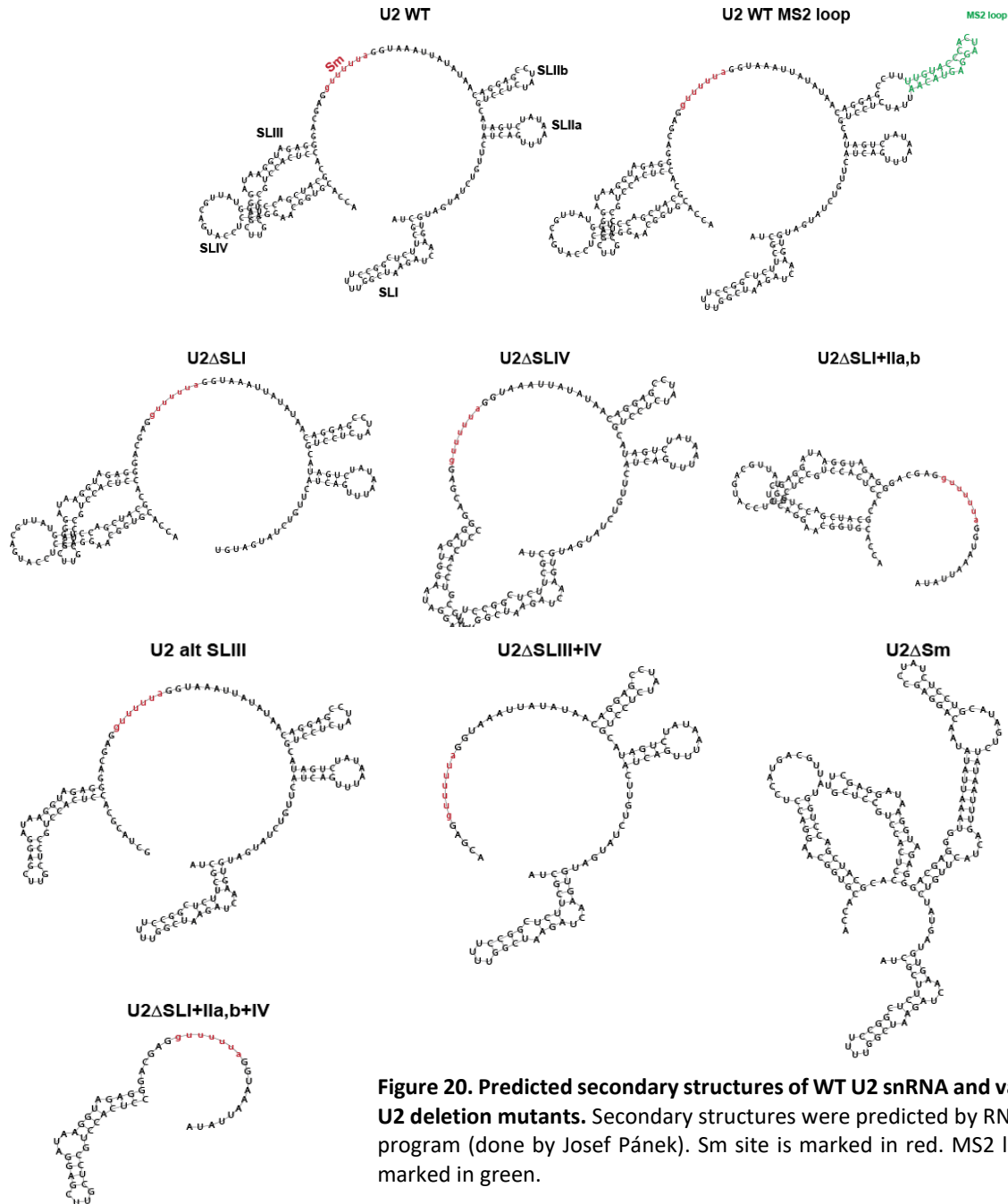


Figure 20. Predicted secondary structures of WT U2 snRNA and various U2 deletion mutants. Secondary structures were predicted by RNA fold program (done by Josef Pánek). Sm site is marked in red. MS2 loop is marked in green.

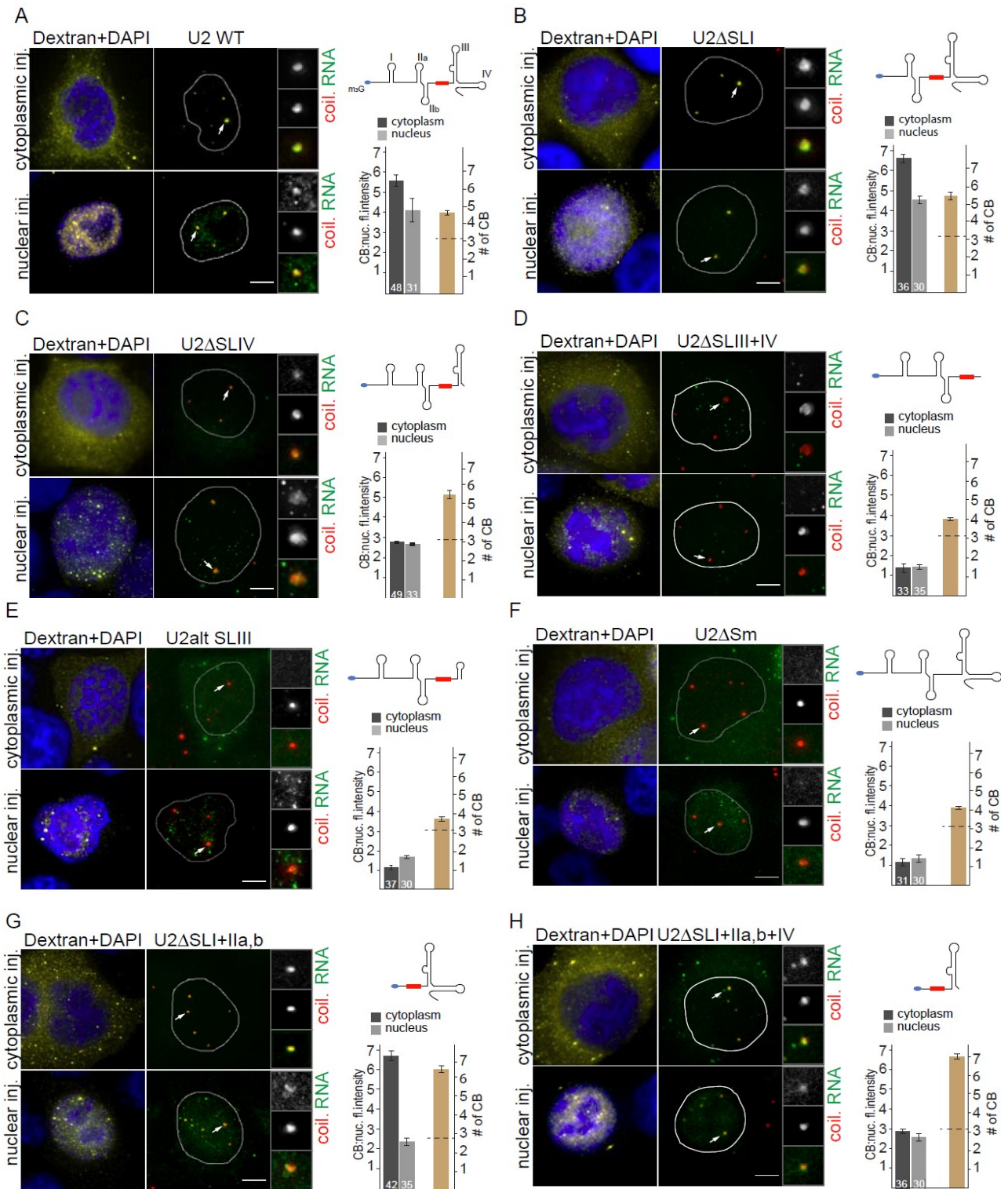


Figure 21. Sm and SMN binding sites are necessary to target the microinjected U2 snRNA into Cajal bodies

(A-H) *In vitro* transcribed WT U2 snRNA or deletion mutants thereof were microinjected into the cytoplasm or the nucleus of HeLa cells. U2 snRNA was labeled with UTP-Alexa-488 (green), coilin, a marker of CBs, was immunolabeled by Alexa-647 (red). Dextran-TRITC 70kDa (yellow) was used to monitor nuclear or

cytoplasmic injection, DNA was stained by DAPI (blue). Small red box in U2 snRNA scheme represents the Sm site. The intensity of the RNA signal in CBs vs. the nucleoplasm was determined for each CB, and the average and SEM are shown in graphs next to the micrographs (number of microinjected cells is indicated in the bar). Number of CBs in microinjected cells was counted and plotted. A dotted line shows a number of CBs in control non-injected cells. The scale bar represents 10 μm .

The Sm site is a general snRNA CB targeting sequence

Next, we tested whether the Sm site is required for CB targeting of other snRNAs (U1, U4, and U5). We microinjected WT snRNAs and snRNAs without the Sm site and found that WT snRNAs were targeted to CBs (Fig. 22). In all cases, snRNAs microinjected into the nucleus exhibited weaker CB localization than snRNAs microinjected into the cytoplasm. This effect was most pronounced in the case of U5 snRNA. Surprisingly, microinjected U1 snRNA also accumulated in CBs while it has been shown that endogenous U1 snRNA does not accumulate in this nuclear compartment but rather localized to gems (Stejskalova and Stanek, 2014). None of the snRNAs which lacked the Sm site were targeted to CBs. In the case of cytoplasmic microinjection, snRNAs ΔSm were retained in the cytoplasm, which is consistent with the inhibition of Sm ring formation and subsequent nuclear import (Fischer et al., 1993) (Fig. 22). Following nuclear injection, snRNAs lacking Sm sites did not accumulate in CBs but remained in the nucleus, which indicates that snRNAs injected into the nucleus are not exported to the cytoplasm. To confirm that RNAs injected into the nucleus do not cycle through the cytoplasm, we injected WT U2 snRNA into the nucleus of a heterokaryon cell (Fig. 23C). After a 1h incubation period, we did not observe any RNA signal in the other nuclei within the heterokaryon cell (Fig. 23C).

Previous work has shown that inhibition of the final U2 snRNP maturation by the knockdown of SF3A3 resulted in the accumulation of U2 snRNA in CBs (Tanackovic and Kramer, 2005). To test whether CB localization of the U2 snRNA lacking the SF3A3 binding site (U2 ΔSLI) also depends on the Sm site, we prepared RNA that lacked both the SF3A3 and Sm binding sites (U2 $\Delta\text{SLI}+\text{Sm}$) and injected this RNA into the nucleus as well as the cytoplasm (Fig. 23A). We did not observe any CB accumulation, which

demonstrates that the Sm site is indispensable for CB targeting of mutated snRNAs that are unable to form the mature particle.

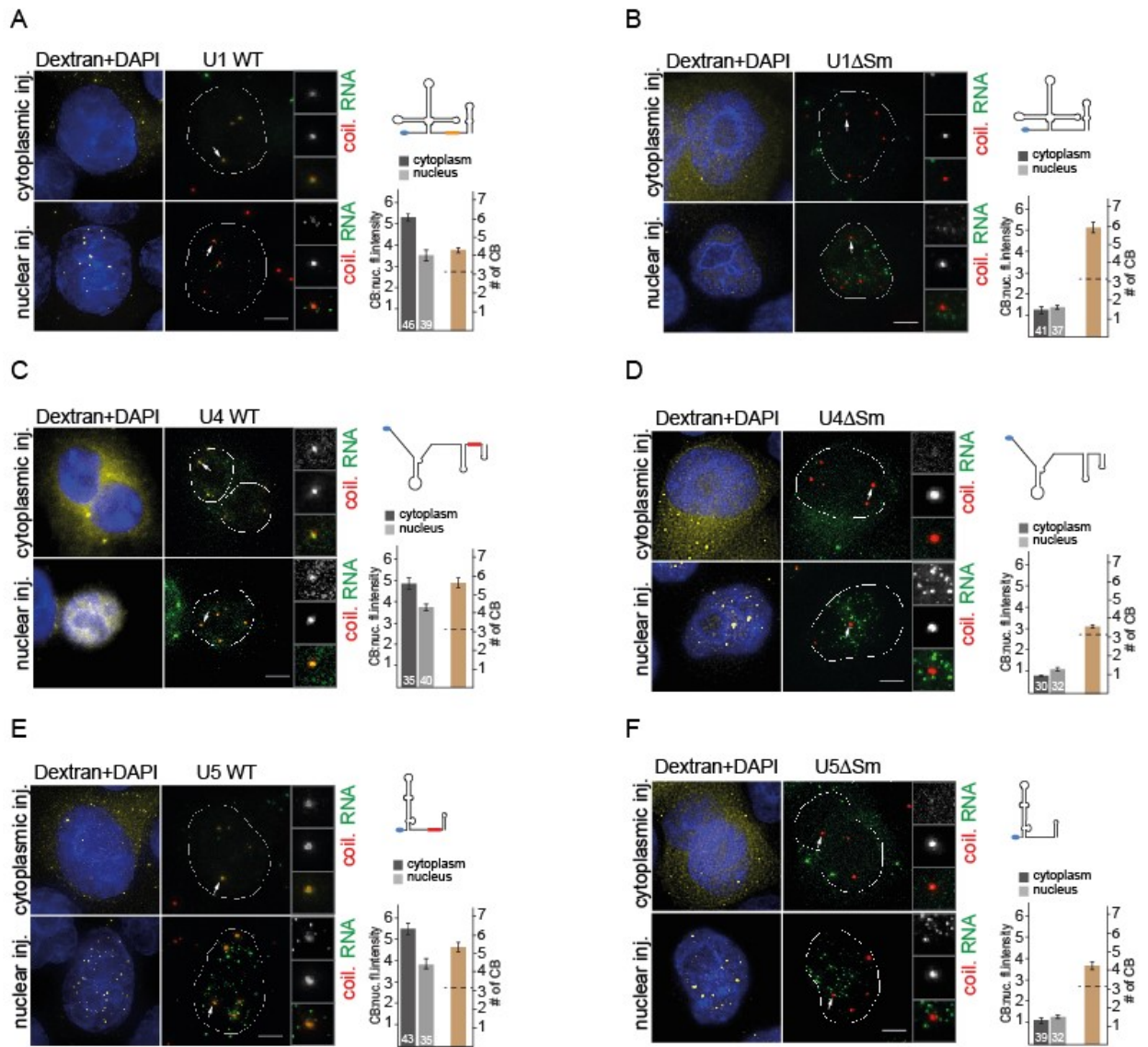


Figure 22. The Sm site is necessary for Cajal body targeting of U1, U4 and U5 snRNAs

(A-D) WT or Δ Sm U1, U4 and U5 snRNAs were transcribed *in vitro* and microinjected into the cytoplasm or the nucleus of HeLa cells (green). CBs are marked by coilin immunolabeling (red). Small red boxes in snRNA schemes represent the canonical Sm site. Dextran-TRITC 70kDa (yellow) was used to monitor nuclear or cytoplasmic injection, DNA was stained by DAPI (blue). The intensity of the RNA signal in CBs vs. the nucleoplasm was determined for each CB, and the average and SEM are shown in graphs next to the micrographs (number of microinjected cells is indicated in the bar). Number of CBs in microinjected cells was counted and plotted. A dotted line shows a number of CBs in control non-injected cells. The scale bar represents 10 μ m.

The U1 snRNA Sm site differs from the canonical Sm site found in U2, U4, and U5 snRNAs. To test whether the non-canonical U1 Sm site can act as a CB targeting signal in the context of other snRNAs we replaced the U2 Sm site with the U1 Sm site (U2withU1Sm). This chimeric snRNA localized to CBs only after cytoplasmic microinjection. In the case of nuclear microinjection, the U2withU1Sm RNA did not localize to CBs but was diffusely distributed in the nucleoplasm (Fig. 23B). This finding suggests that the U1 Sm site cannot entirely replace the canonical Sm site, at least when the U1 Sm site has been inserted into U2 snRNA.

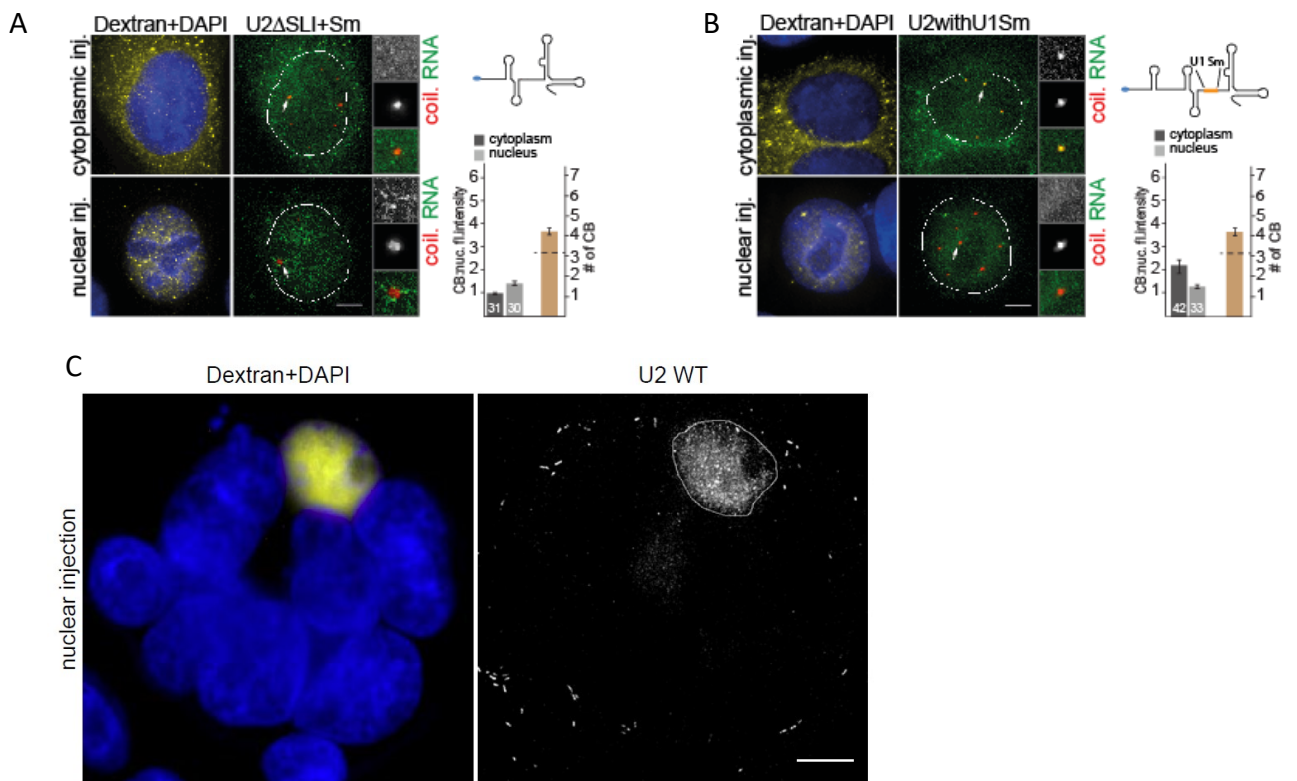


Figure 23: Microinjection of U2 snRNA WT and mutants :

(A) U2 Δ SII+Sm and (B) U2 snRNA with U1-like Sm site were injected into the nucleus and the cytoplasm of HeLa cells. The intensity of the RNA signal in CBs vs. the nucleoplasm was determined for each CB, and the average and SEM are shown in graphs next to the micrographs (number of microinjected cells is indicated in the bar). Number of CBs in microinjected cells was counted and plotted. A dotted line shows a number of CBs in control non-injected cells. (C) Microinjection of U2 snRNA WT into one nucleus of a heterokaryon cell. U2 snRNA was labeled with UTP-Alexa-488 (right panel). Dextran-TRITC 70 kDa is a marker of injection (yellow). DNA was stained with DAPI (blue). The scale bar represents 10 μ m.

Sm and SMN sites are sufficient to target non-coding RNAs into Cajal bodies

To test whether Sm and SMN sites are sufficient to target also the non-CB RNAs to the CBs, we added the Sm and SMN sites to several non-coding RNAs, including human 7SK and Alu RNAs, and *E. coli* SRP RNA. First, we utilized the 7SK RNA, which is a nuclear non-coding RNA. It plays a role in the regulation of RNA polymerase II and under physiological conditions does not accumulate in CBs (Matera and Ward, 1993; Peterlin et al., 2012). We prepared three different mutants of 7SK RNAs: (1) 7SK+Sm site RNA containing the consensus Sm site (AUUUUUG) inserted between two stem-loops at the 3' end; (2) 7SK+SMNsite RNA where the 3' end stem-loop was replaced with a stem-loop from Herpes saimiri virus (HSUR1) RNA, which binds the SMN complex (Golembe et al., 2005); and (3) 7SK+Sm+SMNsites RNA containing both Sm and SMN sites. All 7SK RNAs were microinjected into the nucleus or the cytoplasm of HeLa cells (Fig. 24A-D). We found that neither WT nor chimeric RNAs that contained either the Sm site or the SMN site accumulated in CBs (Fig. 24A-C). However, the 7SK+Sm+SMNsites RNA which possessed both Sm and SMN binding sequences, efficiently localized in CBs after both nuclear or cytoplasmic injections (Fig. 24D). To further confirm this finding, we attached Sm + SMN sequences to Alu RNA, a 120nt RNA component of the cytoplasmic Alu RNP, which is involved in translation regulation (Maraia et al., 1993). While WT Alu RNA is not present normally in CBs (Fig. 24E), mutant Alu+Sm+SMN RNA was accumulated in CBs after nuclear and cytoplasmic injections (Fig. 24F). Finally, we used 114nt non-coding SRP RNA from bacteria *E.coli*. Microinjected WT SRP RNA was not accumulated in CBs (Fig. 24G), but the insertion of Sm+SMN sites targeted this RNA to CBs (Fig. 24H). This experiment shows that SMN and Sm sites are together sufficient to target various non-coding RNAs to CBs.

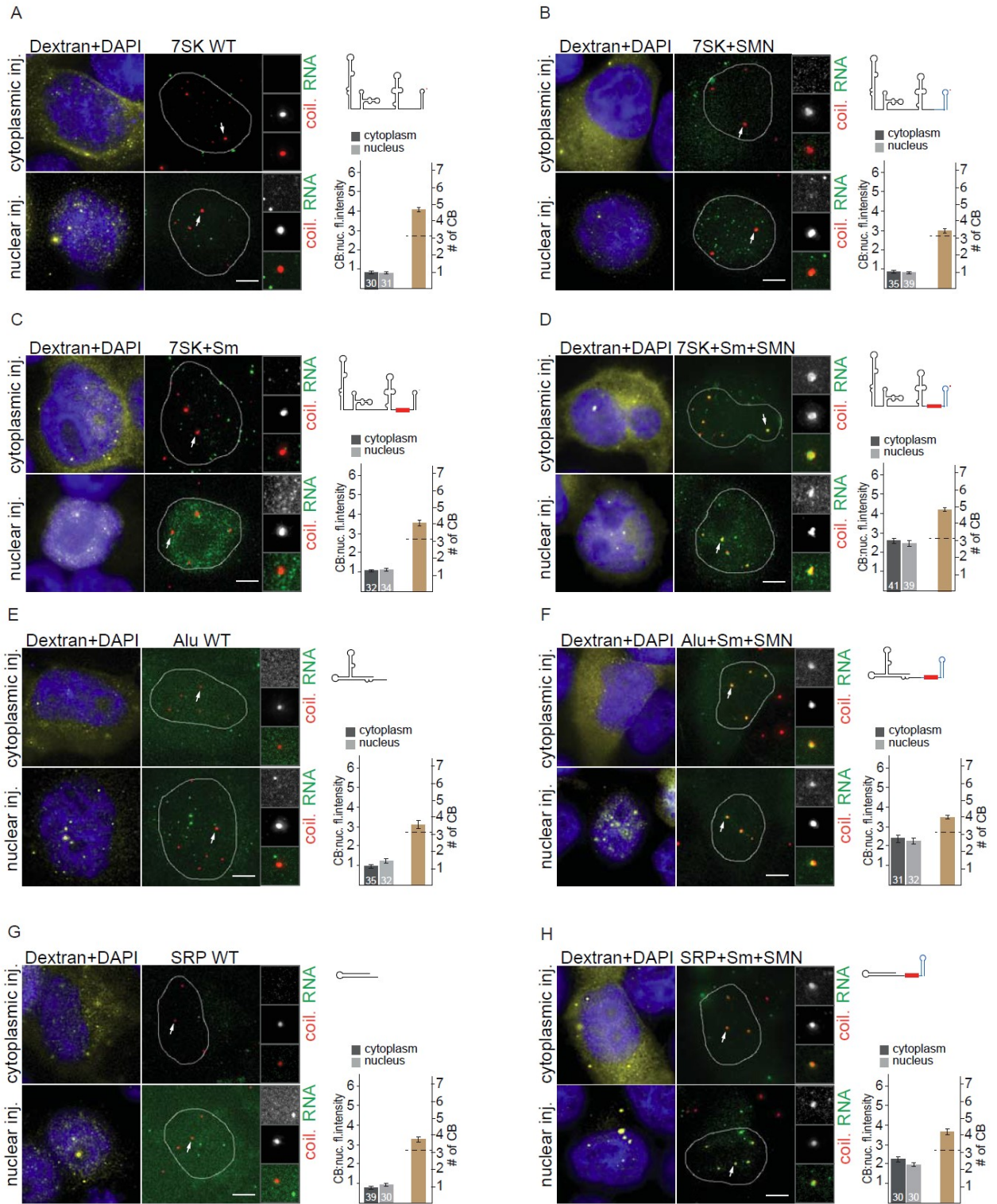


Figure 24. Sm and SMN sites are sufficient to target non-coding RNAs into Cajal bodies

(A-D) *In vitro* transcribed 7SK RNAs (WT or chimeras containing Sm, SMN or both sites) were microinjected into the nucleus or cytoplasm of HeLa cells. WT Alu or Alu+Sm+SMN sites RNAs were microinjected into

the nucleus or cytoplasm of HeLa cells. WT E. coli SRP RNA or SRP+Sm+SMN sites RNA were microinjected into the nucleus or cytoplasm of HeLa cells. Dextran-TRITC 70kDa (yellow) was used to monitor nuclear or cytoplasmic injection, DNA was stained by DAPI (blue). The intensity of the RNA signal in CBs vs. the nucleoplasm was determined for each CB, and the average and SEM are shown in graphs next to the micrographs (number of microinjected cells is indicated in the bar). Number of CBs in microinjected cells was counted and plotted. A dotted line shows a number of CBs in control non-injected cells. The scale bar represents 10 μ m.

Minimal ectopically expressed snRNAs containing Sm and SMN sites accumulate in Cajal bodies

Our experiments showed the importance of Sm and SMN sites for CB accumulation of microinjected RNAs. To test whether these sequences are important also for CB targeting of ectopically expressed snRNAs we used a system devised by Edouard Bertrand and colleagues in which the MS2 binding site was inserted to detect ectopically expressed U4 snRNAs (Hallais et al., 2013b). The U4-MS2 RNA under the control of endogenous U4 promoter elements was obtained from Edouard Bertrand (IGMM, CNRS, Montpellier). To detect the ectopically expressed U2 snRNA we cloned the MS2 site into the stem loop IIb of U2 snRNA driven by the U2 promoter (U2-MS2). To prepare minimized U2 snRNA, we deleted the first 94 nucleotides from the U2-MS2 RNA leaving only the MS2 loop followed by the Sm site and the SMN site containing stem loops III and IV (U2 Δ SLI-IIa,b-MS2). Similarly, the first 64 nucleotides were deleted from U4-MS2 creating a minimized U4 RNA (U4 Δ 1-64-MS2)(done by Klára Klimešová) that contains the Sm site between stem loops III and IV, which together serve as the SMN binding site (Yong et al., 2004).

U2-MS2 and U4 -MS2 constructs were co-transfected with plasmid containing MS2-YFP protein and after 24 h cells were subjected to immunoprecipitation using anti-GFP antibodies. Immunoprecipitation followed by RNA detection showed that the deletion mutants were expressed at levels similar to WT snRNAs (Fig. 25A and 26). Western blot analysis of proteins co-precipitating with MS2 snRNAs revealed the association of U2-MS2 with Sm proteins, SNRNP2 (U2B^{''}), SF3A3 and SF3B4 (SF3b49) (Fig. 25A) and the U4-MS2 RNA with Sm proteins and Prpf31 (Fig. 25B). Although, only a small subset of snRNP-

specific proteins has been tested, these results suggest that the insertion of MS2 insertion does not block the association of snRNAs with snRNP-specific proteins. This data also confirms that U2 snRNA lacking stem loops I and II does not interact with neither SF3a nor SF3b complexes. Next, MS2-tagged snRNAs were expressed together with MS2-YFP, fixed after 24h and coilin was detected by indirect immunofluorescence (Fig. 26). Both minimized U2 and U4 snRNAs localized to CBs in a similar manner as WT snRNAs, which was consistently observed in our microinjection experiments, hence showing that Sm + SMN sites are sufficient for targeting of snRNAs into CBs. It should be noted that we cannot test the role of Sm and SMN sites in CB targeting directly by their depletion because ectopically expressed snRNA without these sequences would be retained in the cytoplasm and would not reach the nucleus.

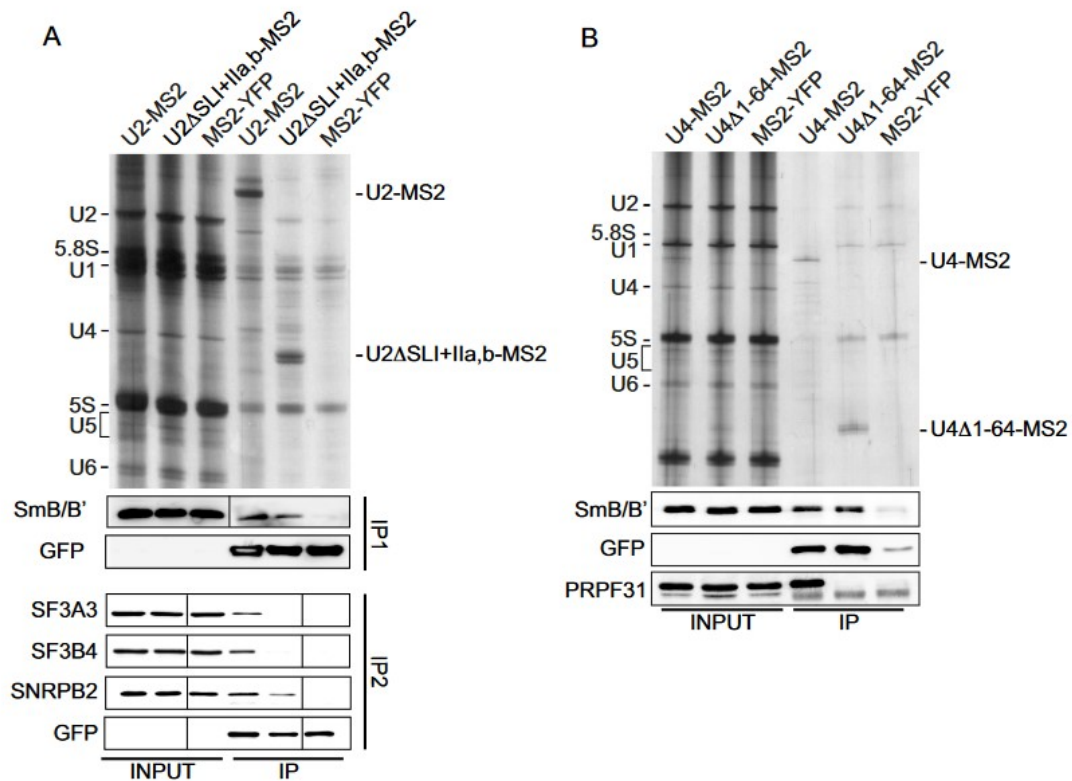


Figure 25. Transiently-expressed, truncated snRNAs containing Sm and SMN binding are able to bind Sm and specific proteins

(A) Immunoprecipitation of U2-MS2 and the deletion mutant U2ΔSLI+IIa,b-MS2. RNAs were immunoprecipitated via MS2-YFP by anti-GFP antibodies and detected by silver staining. Co-precipitated proteins were analyzed by Western blotting (done by Klára Klimešová). With the full-length U2-MS2 RNA we detected all tested proteins (SmB/B', SF3A3, SNRPB2 and SF3B4), while only SmB/B' and SNRPB2 co-precipitated with the U2ΔSLI+IIa,b-MS2 RNA.

(B) Immunoprecipitation of U4-MS2 and the deletion mutant U4 Δ 1-64-MS2. Legend as in (A). With the full-length U4-MS2 RNA we detected both tested proteins (SmB/B' and PRPF31), while only SmB/B' co-precipitated with U4 Δ 1-64-MS2 RNA (done by Klára Klimešová)

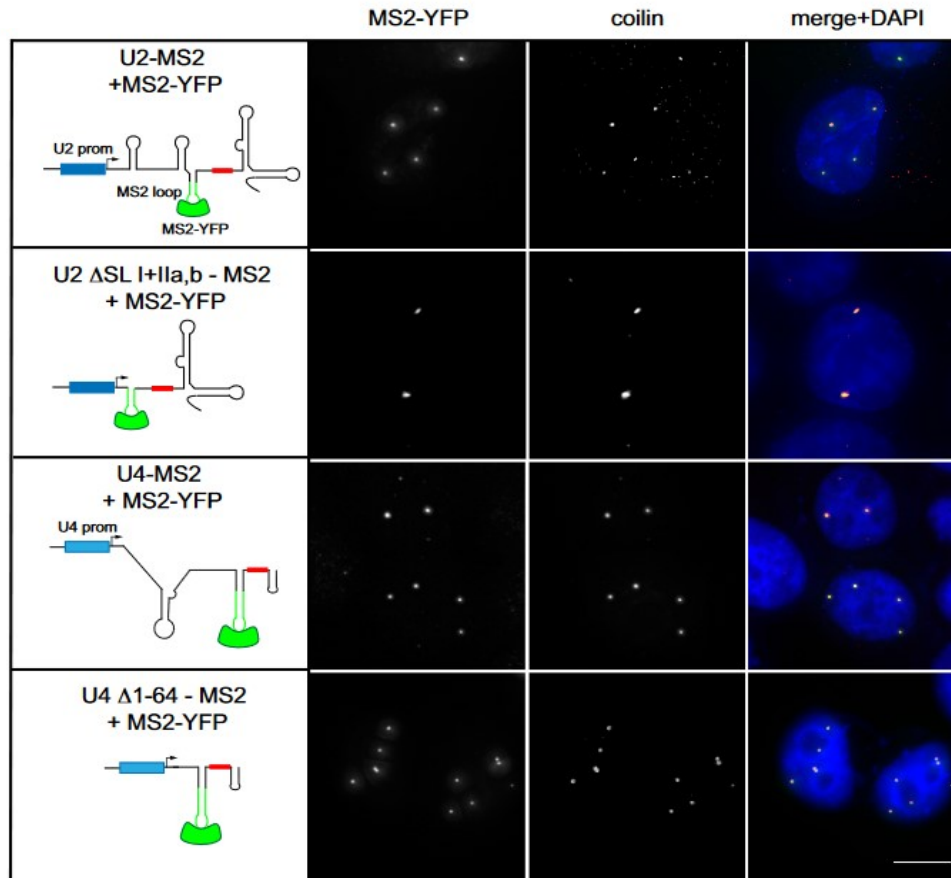


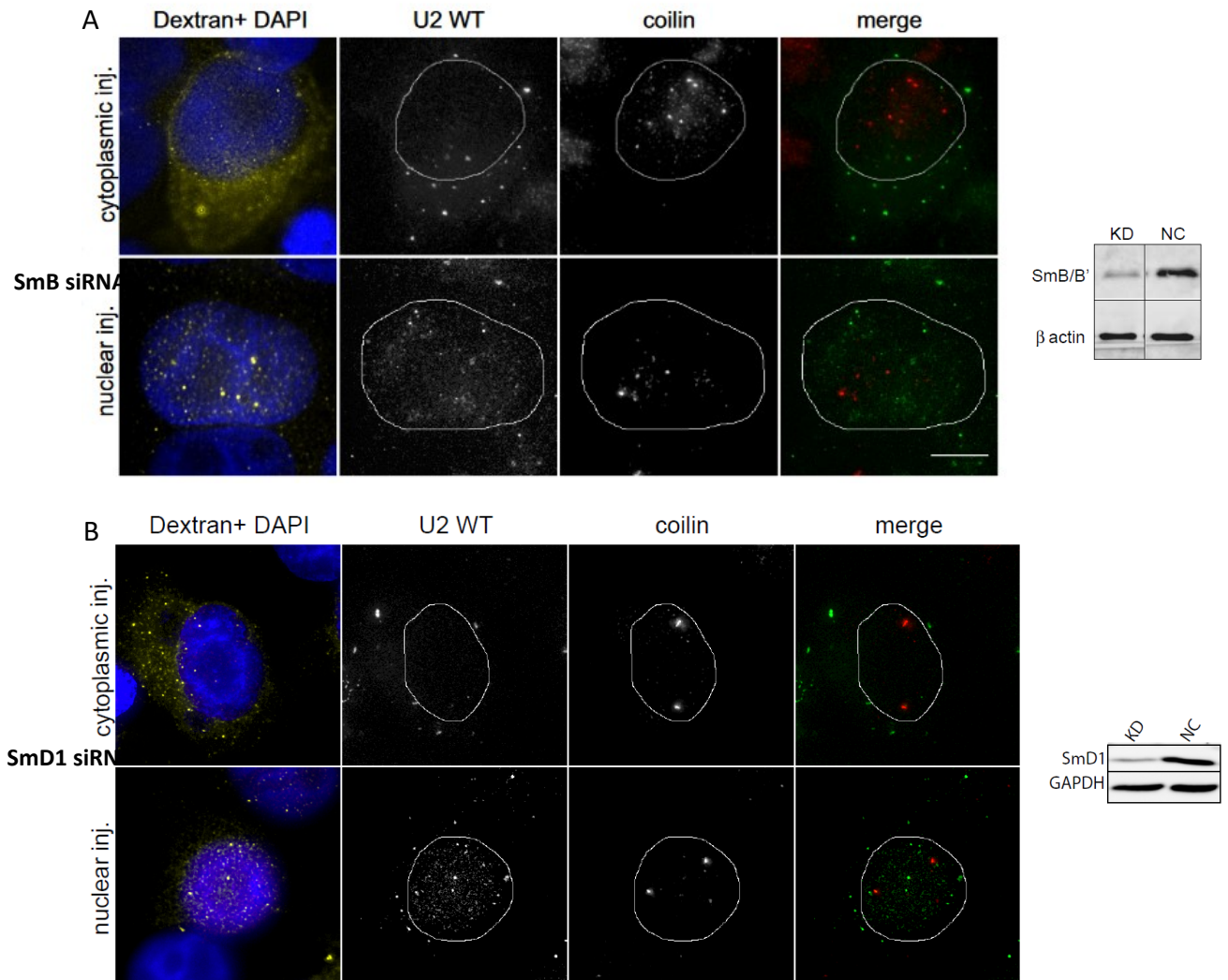
Figure 26. Transiently-expressed, truncated snRNAs containing Sm and SMN binding sites accumulate in Cajal bodies

HeLa cells were co-transfected with U2 or U4 snRNAs containing the MS2 loop (green stem loop) and MS2-YFP (green). Coilin was used as a marker of CBs (red). Small red box in snRNA schemes represents the canonical Sm site and blue boxes represent endogenous U2 or U4 promoters. DNA was stained by DAPI. The scale bar represents 10 μ m.

Sm proteins are essential for snRNA Cajal body targeting

Our experiments provide evidence that SMN and Sm sites are essential for snRNA targeting into CBs. These minimal sequences were previously shown to be sufficient for SMN complex binding and Sm ring assembly (Golembe et al., 2005; Yong et al., 2004). To test whether the snRNA sequence *per se* or the Sm ring is essential for targeting

snRNAs to CBs, we depleted several of the Sm proteins, SmB/B', SmD1 or SmG by RNA interference (Fig. 27A-C). When WT U2 snRNA was microinjected into the cytoplasm, SmB/B', SmD1 and SmG knockdown completely inhibited its CB localization and the U2 snRNA was retained in the cytoplasm, confirming the efficiency of the knockdown and the inhibition of Sm ring assembly (Fig. 27). After nuclear microinjection, WT U2 snRNA remained in the nucleus but did not accumulate in CBs. We did not observe this phenotype in the cells treated with a negative control siRNA (Fig. 27D). It should be noted that depletion of Sm proteins had a negative effect on CB integrity and in a fraction of cells, coilin partially accumulated in nucleoli. However, CBs were still present in a significant number of cells treated for 48h with the anti-SmB/B' siRNA. These experiments suggest that Sm proteins assembled around the Sm site are recognized as a CB targeting signal.



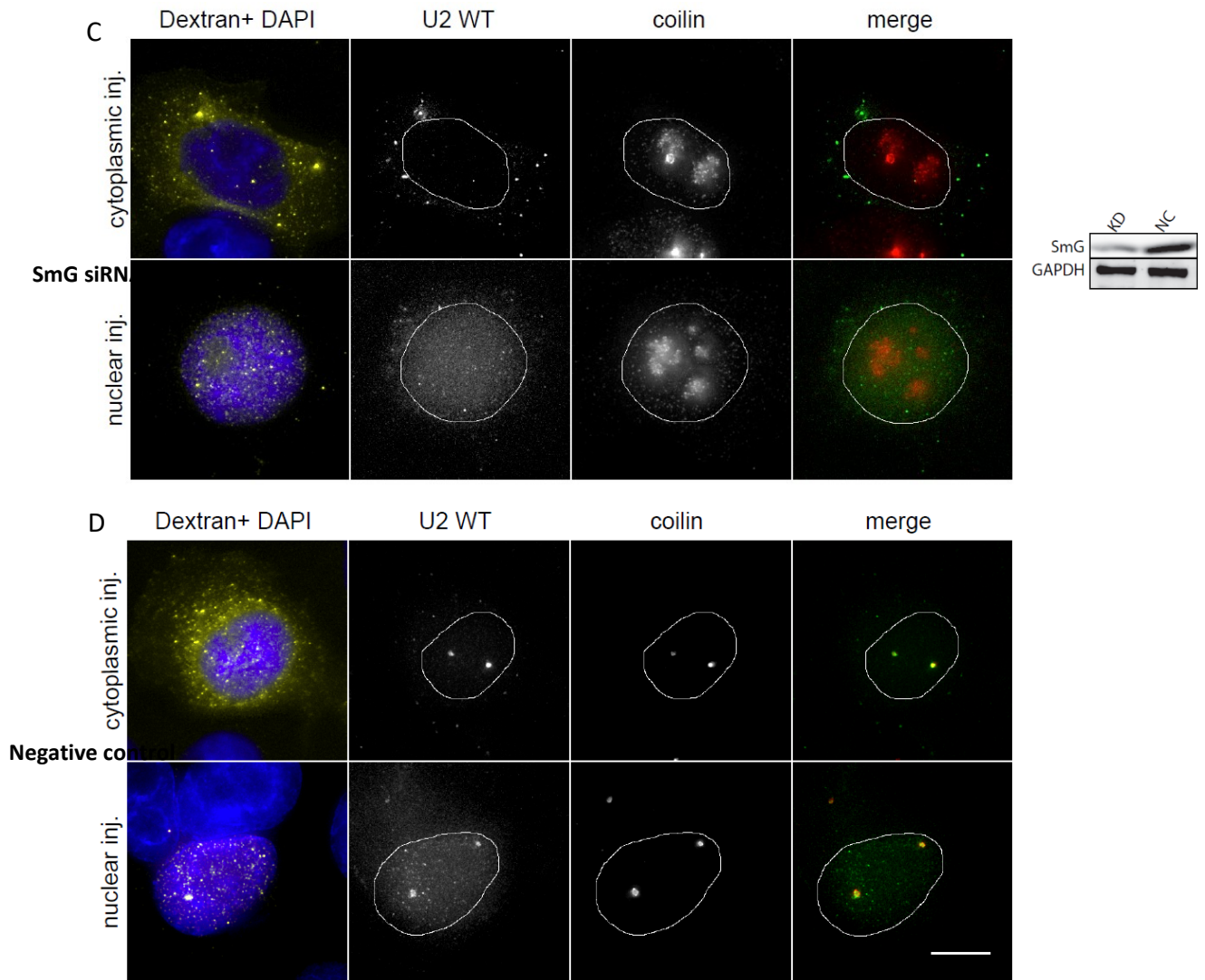


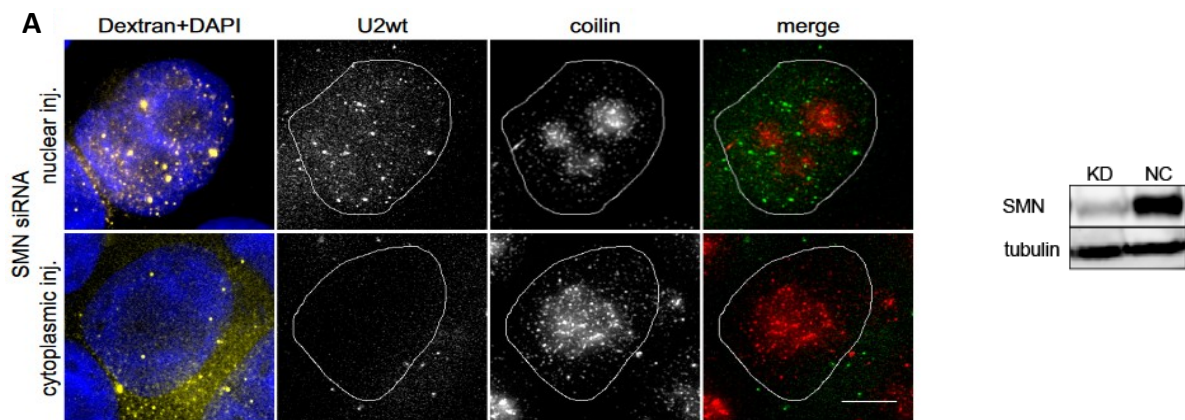
Figure 27: Sm proteins are essential for snRNP targeting into Cajal bodies

(A, B, C) Depletion of the SmB/B', SmD1 and SmG proteins disrupts targeting of microinjected snRNAs into the Cajal body. Sm proteins were depleted by RNAi and Alexa-488 labeled WT U2 snRNA (green) subsequently microinjected into HeLa cells. Dextran-TRITC 70 kDa was used as a marker of microinjection (yellow). (D) Microinjection of U2 WT into HeLa cell after depletion of Negative control. DNA was stained by DAPI (blue). Test of siRNA efficiency. GAPDH was used as a loading control. SmD1 protein was stained by anti-SmD1 antibody and SmG protein was stained by anti-SmG antibody. DNA was stained with DAPI (blue). The scale bar represents 10 μ m.

Sm proteins but not the SMN protein is essential for Cajal body targeting of snRNA

Both SMN and Sm binding sites are recognized by the SMN complex (Chari et al., 2008; Xu et al., 2016). Therefore, we decided to test whether the SMN complex is also directly required for snRNP targeting to CBs. We knocked down the SMN protein and two components of SMN complex – Gemin2 and Gemin5. The depletion of SMN protein

resulted in CB disappearance, but depletion of Gemin2 or Gemin5 did not result in CB dissolving (Fig. 28). We microinjected *in vitro* synthesized U2 snRNA WT into the nucleus and to the cytoplasm and measured the accumulation of snRNA in the CBs. The depletion of Gemin5 had almost no effect on U2 snRNA targeting into CBs (Fig. 28C). In contrast, depletion of Gemin2 significantly reduced U2 snRNA accumulation in CBs after nuclear and cytoplasmic microinjection (Fig. 28B). From these results, we suggest that SMN complex play an important role in snRNP targeting into CBs. However, the SMN complex is involved in Sm ring assembly, which we showed above as a CB targeting signal (Fig. 21E, D). To test whether the SMN complex is directly involved in snRNP transport into the CB, we established a collaboration with Cyril Girard from Max Planck Institute for Biophysical Chemistry. He depleted SMN protein by RNAi and microinjected pre-assembled core U2 snRNPs. Microinjected U2 core into SMN depleted cells restored CBs and injected snRNPs accumulated in CBs (Fig. 29) (done by Girard Cyril). These data show the essential role of Sm proteins in targeting U2 snRNA to CBs, and suggest that the SMN complex is not directly involved in snRNP CB localization.



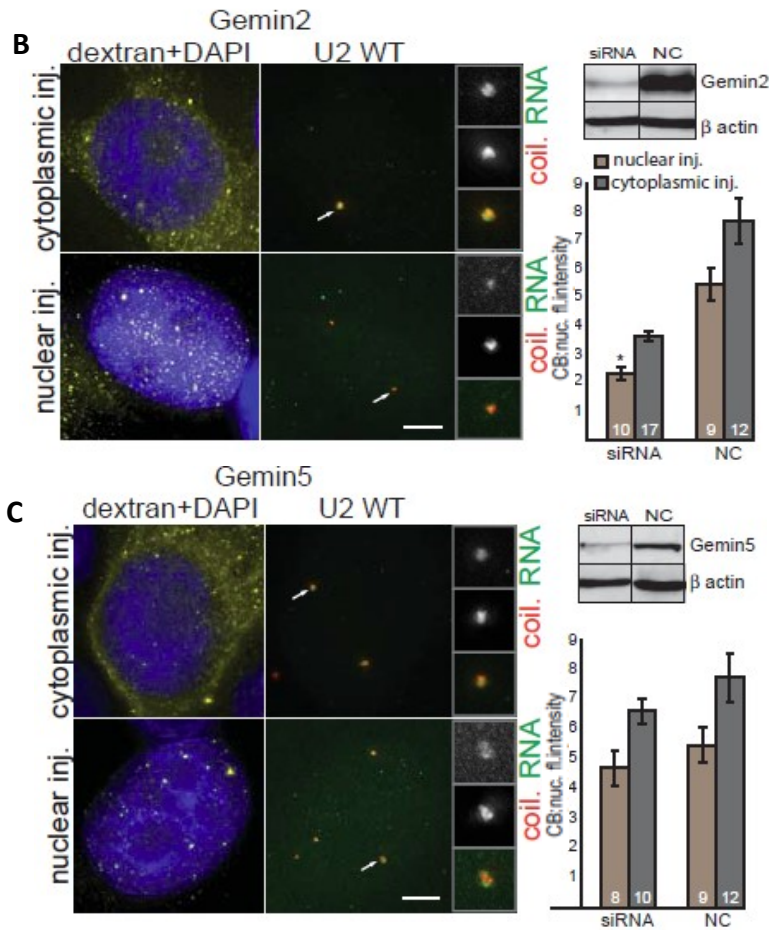


Figure 28: Sm proteins but not the SMN protein is essential for Cajal body targeting of snRNA
 (A, B, C) SMN protein, Gemin2, and Gemin5 were depleted by siRNA and either Alexa-488 labeled WT U2 snRNA. (B) core U2 snRNP was microinjected into HeLa cells. The intensity of the RNA signal in CBs vs. the nucleoplasm was determined for each CB, and the average and SEM are shown in graphs next to the micrographs (number of microinjected cells is indicated in the bar). Number of CBs in microinjected cells was counted and plotted. Depletion of the SMN protein disrupts CBs, which can be rescued by cytoplasmic microinjection of core U2 snRNP (red) (done by Cyril Girard). Coilin was used as a marker of CB (green). The scale bar represents 10 μ m.

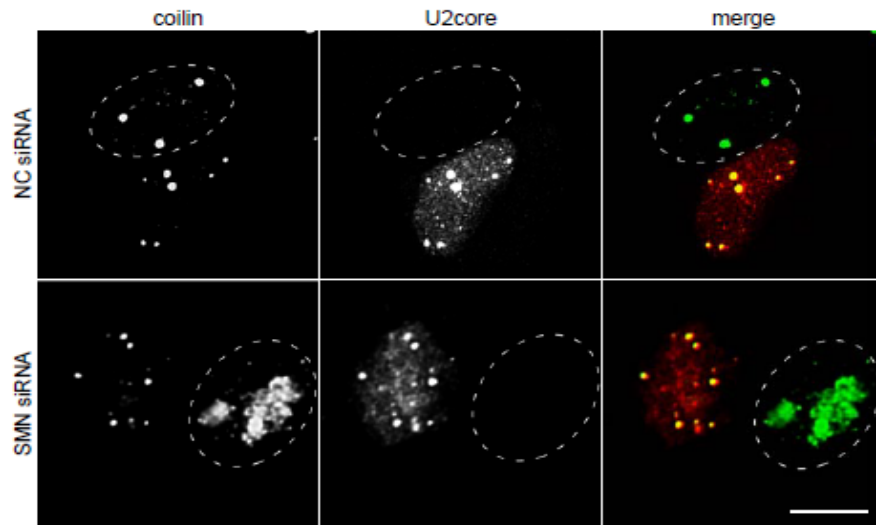


Figure 29: Microinjection of U2 core particle

SMN was depleted by siRNA. Depletion of the SMN protein disrupts CBs, which can be rescued by cytoplasmic microinjection of core U2 snRNP (red) (done by Cyril Girard). Coilin was used as a marker of CB (green). The scale bar represents 10 μ m.

C-terminal GR-rich tails of SmB, Smd1 and Smd3 are important for Cajal body localization

To specify Sm protein motifs, which are important for CB targeting, we deleted several non-Sm fold domains, namely the GR domain of Smd1 and C-terminal tails of Smd3 and SmB, which contain several GR dipeptides (Figs. 30, 31). We tagged the proteins with GFP, expressed them transiently in HeLa cells and assayed their interaction with snRNAs by immunoprecipitation and CB localization by fluorescence microscopy. Cells were transfected with the same amount of DNA, but some constructs exhibited better expression than others likely due to better protein and/or mRNA stability. Deletion of the C-terminal tails from Smd3 (aa 110-126) and SmB (aa 170-231) did not prevent Smd3 and SmB interactions with snRNAs (Fig. 30A and B) but significantly reduced CB localization (Fig. 31). To test whether GR repeats are important for CB targeting, we substituted Smd3 amino acids 110-120 containing several GR repeats (GRGRGMGRGN) with a corresponding stretch of alanine residues (D3Ala) and two GR

repeats (107-108 and 111-112) predicted to be methylated by PRMT5 (UniProt) in SmB (BAla). The GR substitution exhibited the same phenotype as the C-tail deletion. The D3Ala mutant immunoprecipitated snRNA, albeit to a lesser extent than WT SmD3, while its CB localization was strongly reduced (Figs. 30A and 31). Similarly, the BAla mutant precipitated the same amount of snRNAs as wild-type SmB (Fig. 31B) but accumulated less efficiently in CBs (Fig. 31). Finally, we deleted the GR domain from SmD1 (aa 97-119), which reduced interaction with snRNAs and CB localization (Figs. 30 and 31). Therefore, we removed a half (aa 107-119) (D1 Δ 1/2GR) or a quarter (aa 113 - 119)(D1 Δ 1/4GR) of the GR domain. Partial deletion of GR repeats did not block association with snRNAs (Fig. 30C) and only partial reduction in CB localization was observed in case of SmD1 Δ 1/2GR (Fig. 31). Finally, we mutated last C-terminal 23 amino acids (aa 97-119) of SmD3, which mostly consists of GR repeats, into alanines (D3Ala). This mutant associated with snRNAs to a similar extent as SmD3 wt (Fig. 31A), but its CB targeting was significantly reduced (Figs. 31). These data demonstrate that GR repeats in C-terminal domains of SmB, D1 and SmD3 are important for CB targeting of Sm proteins and likely the whole snRNP.

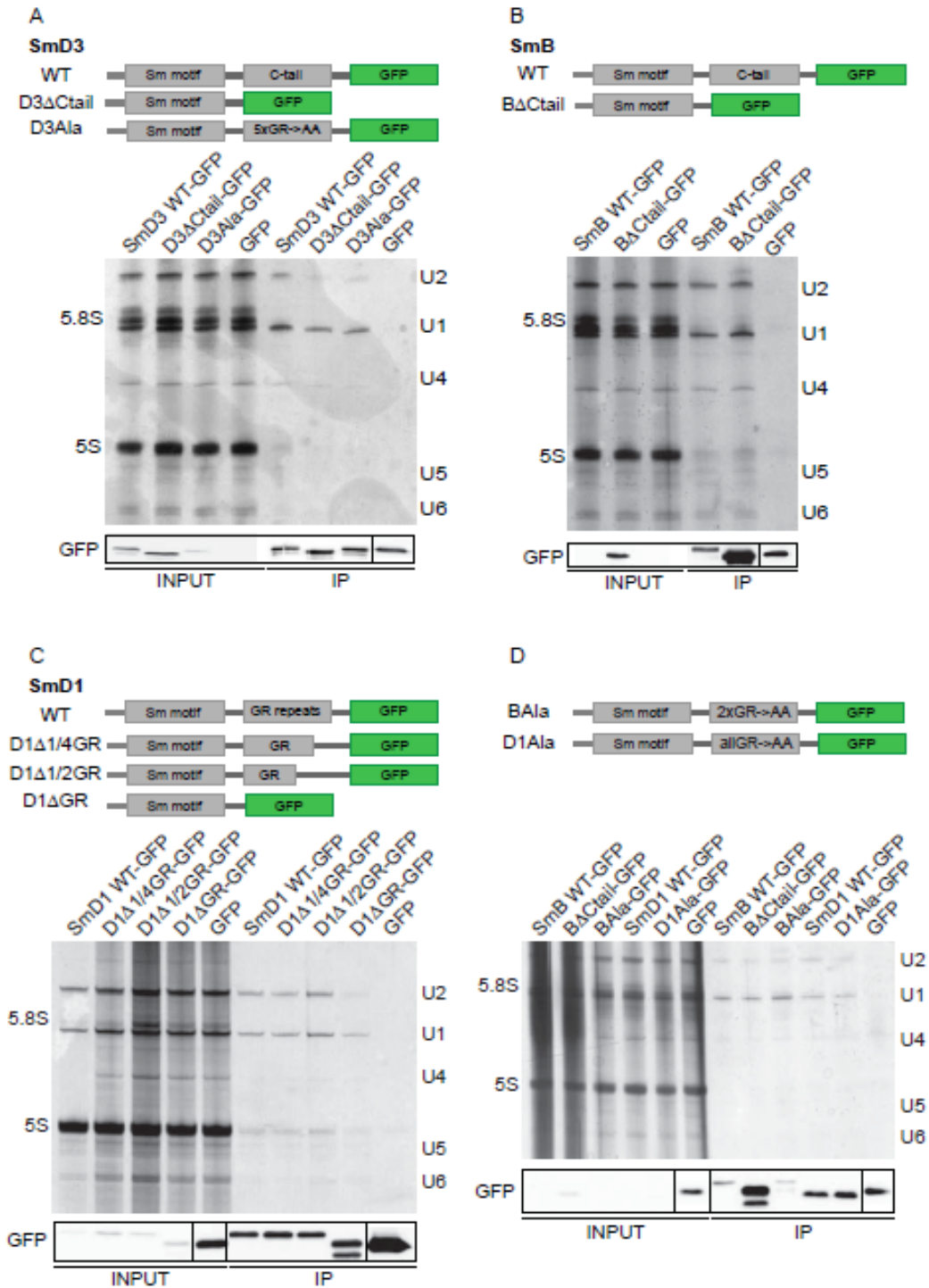
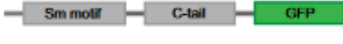
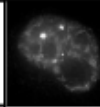
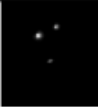
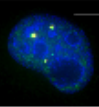

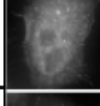
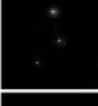
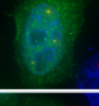
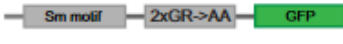
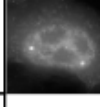

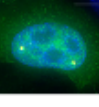

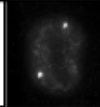

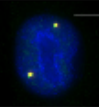

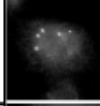

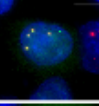

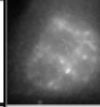

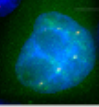

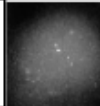

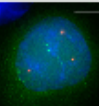
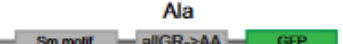
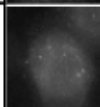
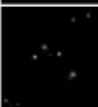
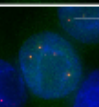
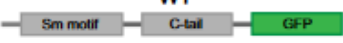
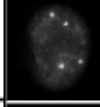
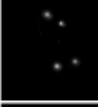
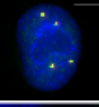

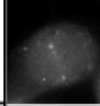

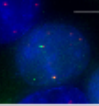
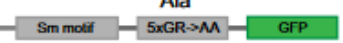
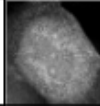

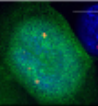


Figure 30. Deletion mutants of SmB, SmD1 and SmD3 are able to bind snRNA. Immunoprecipitation of (A) WT SmD3-GFP and deletion mutants thereof, (B) SmBGFP and the deletion mutant SmBΔCtail-GFP, (C) SmB and SmD1 GR substitution mutants and (D) SmD1-GFP and GR deletion mutants, was performed using anti-GFP antibodies. Precipitated proteins were detected by Western blotting using anti-GFP antibodies (bottom) and co-precipitated RNAs were resolved on a polyacrylamide gel and visualized by silver staining.

SmB protein	Sm protein	coilin	merge +DAPI	rel.CB localization
 WT				1
 ΔCtail				0.61 ± 0.03
 Ala				0.63 ± 0.07
SmD1 protein				
 WT				1
 Δ1/4GR				1.01 ± 0.03
 Δ1/2GR				0.84 ± 0.11
 ΔGR				0.52 ± 0.02
 Ala				0.53 ± 0.16
SmD3 protein				
 WT				1
 ΔCtail				0.57 ± 0.04
 Ala				0.53 ± 0.03

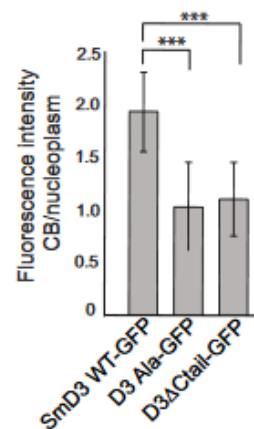
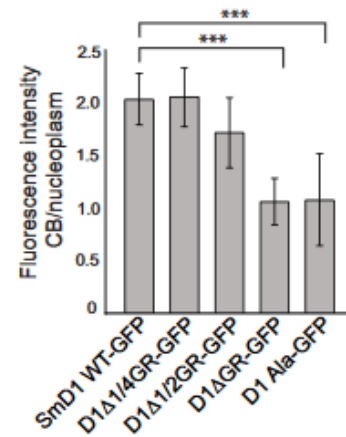
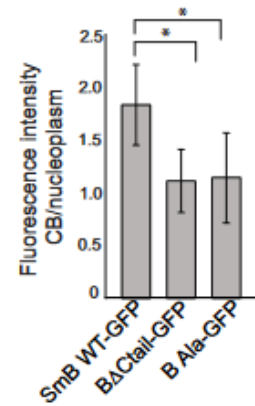


Figure 31. C-terminal tails of SmB, D1 and D3 are important for Cajal body localization

Cells were transfected with plasmids expressing SmB, D1 or D3 protein variants tagged with GFP. Coilin was used as a marker of CBs (red). DNA was stained with DAPI. The scale bar represents 10 μ m. The intensity of GFP signal in CBs vs. the nucleoplasm was determined by high-content microscopy. Values normalized to the WT proteins are shown in the table next to the micrographs, and non-normalized values are shown in graphs. The average of three experiments and SEM are shown. Statistical significance was assayed by the two-tailed t-test and data with a p-value < 0.1 are marked by * and p < 0.001 by ***.

Partially-assembled but not mature U1 and U2 snRNP are targeted to CBs (experiments done by Cyril Girard)

Our previous experiments show that the Sm ring is essential for targeting snRNPs to CBs and that the snRNA containing only the Sm ring (core snRNP) can form new CBs when the snRNP biogenesis was inhibited. To test whether also the mature snRNPs can restore CB formation, we induced CB dissociation by siRNA-mediated knockdown of the TGS1 protein. The TGS1 is the methyltransferase responsible for hypermethylation of the 5' cap of snRNAs, and its depletion leads to the disintegration of canonical CBs and redistribution of coilin into multiple small foci scattered throughout the nucleoplasm and also coilin accumulation in nucleoli (Mouaikel et al., 2003) (Fig.32). To obtain fully- and partially-assembled U2 snRNPs, we purified endogenous 12S U2 snRNP and reconstituted *in vitro* partially-assembled 15S and mature 17S U2 snRNP particles with either the SF3b complex (15S) or SF3a and SF3b complexes (17S). Analysis of their protein composition by SDS PAGE after gradient centrifugation confirmed that the reconstituted particles corresponded to 15S and 17S U2 snRNPs, respectively (Fig. 32A). Next, the 12S, 15S and 17S U2 snRNPs were microinjected into TGS1-depleted cells along with FITC-labelled Dextran. We observed the newly assembled coilin positive foci in case of 12S and 15S particles (Fig. 32B). On the other hand, we did not observe any coilin positive foci upon 17S U2 snRNP microinjection, and its distribution showed the same phenotype like in the non-microinjected cells, coilin remained dispersed in the nucleoplasm and enriched in the nucleolus. Similar results were obtained with the U1 snRNPs; microinjection of an *in vitro*-reconstituted core U1 snRNP (U1 snRNA+Sm proteins) resulted in the formation of CBs in TGS1-depleted cells (Fig. 32C), while microinjection of an endogenous purified mature U1 snRNP did not (Fig. 32D). Taken together these data show that only partially-assembled snRNP can induce CB formation and localize to CBs.

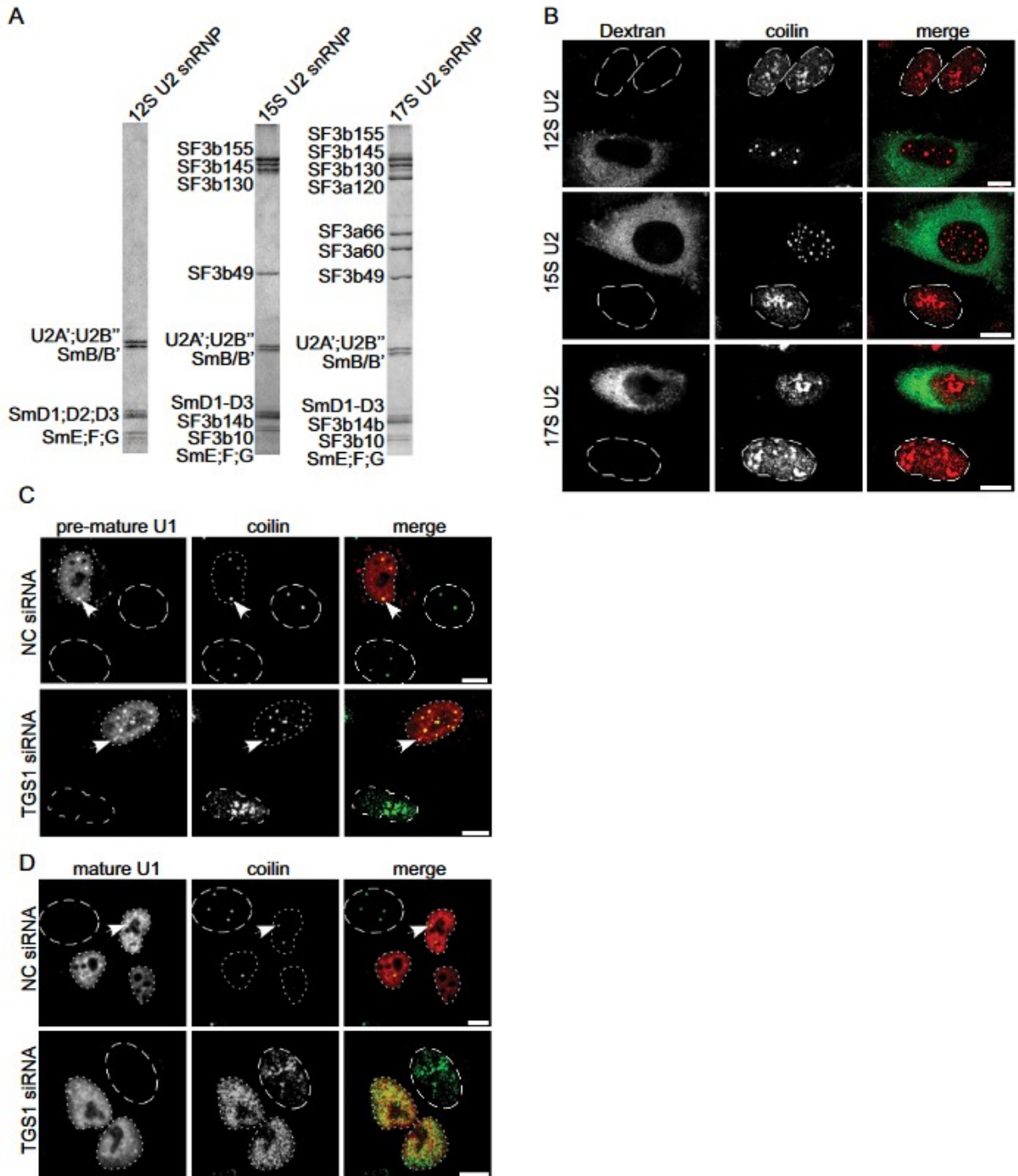


Figure 32. Partially-assembled snRNP particles induce formation of CBs.
(done by Cyril Girard)

(A) Purified 12S U2 snRNP and *in vitro* reconstituted 15S, and 17S U2 snRNPs were analyzed by SDS-PAGE and proteins visualized by Coomassie staining.

(B) TGS1 was knocked down by siRNA and cells were microinjected into the cytoplasm with native 12S U2 snRNP (top panel), *in vitro* reconstituted 15S U2 snRNP (middle panel) or mature 17S U2 snRNP (bottom

panel). FITC-Dextran served as microinjection marker (green), and coilin was visualized by immunostaining (red). The dashed line indicates non-microinjected cell nucleus. (C,D) TGS1 was knocked down by siRNA and cells were microinjected into the cytoplasm with digoxigenin-labeled *in vitro*-reconstituted U1 snRNP (C) or a native Cyan3-labelled mature U1 snRNP (D) and examined by immunofluorescence 2h post microinjection using the anti-coilin antibody (C, D) and anti-digoxigenin antibodies (C). Scale bars: 10 μ m.

Incomplete snRNPs are sequestered in Cajal bodies (published in (Novotny et al., 2015))

Injection of various U2 particles indicates that cells discriminate between complete and incomplete snRNPs and that only immature particles are sequestered in CBs. To test this hypothesis, we mutated U4 snRNA to prevent its base-pairing with U6 snRNA and thus block the formation of the U4/U6 di-snRNP.

We prepared U4 snRNA WT and U4 snRNA lacking the U6 binding site (mut U4) by *in vitro* transcription. Both RNAs were microinjected into the cytoplasm of HeLa cells, and their accumulation was analyzed in CBs over time. We observed that after 30 min the snRNAs (WT and mutant) were localized in the CBs to the same extent. However, after longer incubation time the accumulation of mutU4 was significantly higher than U4 WT (Fig. 33). These data are consistent with the model that CB sequester preferentially in incomplete or defective snRNP particles.

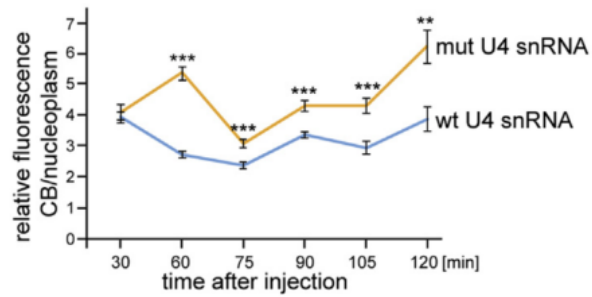
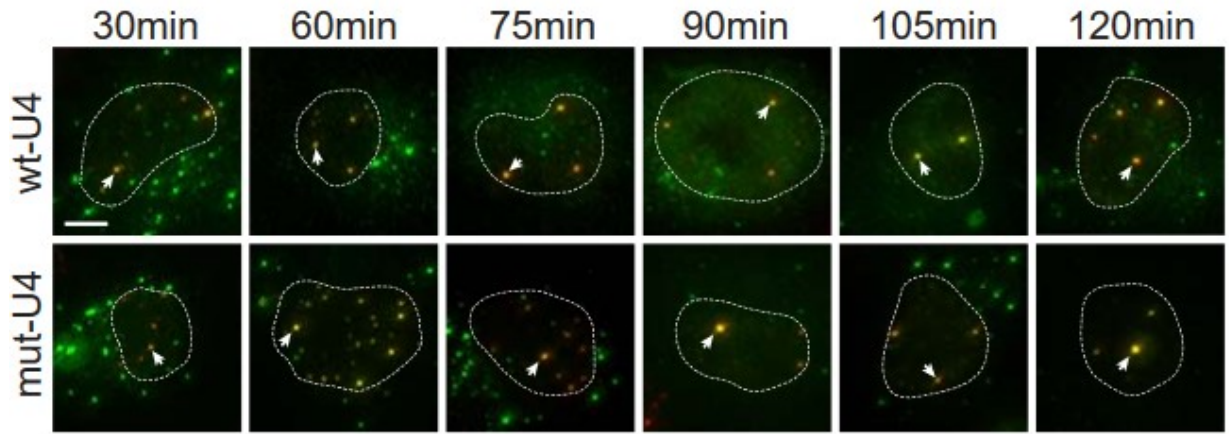


Figure 33: (A) Accumulation of microinjected U4 snRNA in CBs. Fluorescently labeled U4snRNA (wild-type– wt or mutant lacking the U6-base-pairing domain – mut; green) was microinjected into the cytoplasm, cells incubated for a given time period, fixed and the CB marker coilin visualized by immunodetection (red). Nuclear contours are

marker by a dotted line. Representative CBs are marker by arrows. Scale bar represents 5 μ m (B) The graphical depiction of the microinjection experiments. The average of 40–80 CBs (15–25 cells) is shown together with the SEM. The significance was assayed by t- test against wild-type (WT) U4 snRNA; **p%0.01 and ***p%0.001.

3.2 Identifying role of Gemin3 in Sm ring assembly

During the mathematical modeling of U2 snRNA structures, we uncovered the alternative snRNA structure, where the Sm site is surrounded by an extensive stem, which we named Near Sm site Stem (NSS) (Fig.34). This structure is not observed in snRNA containing Sm ring, which can prevent the formation of the NSS. However, the NSS might form on newly transcribed snRNAs before the Sm ring is assembled. The SMN complex, which is responsible for Sm ring formation, contains a putative DEAD-box RNA helicase Gemin3. Its role in Sm ring assembly is still unknown.

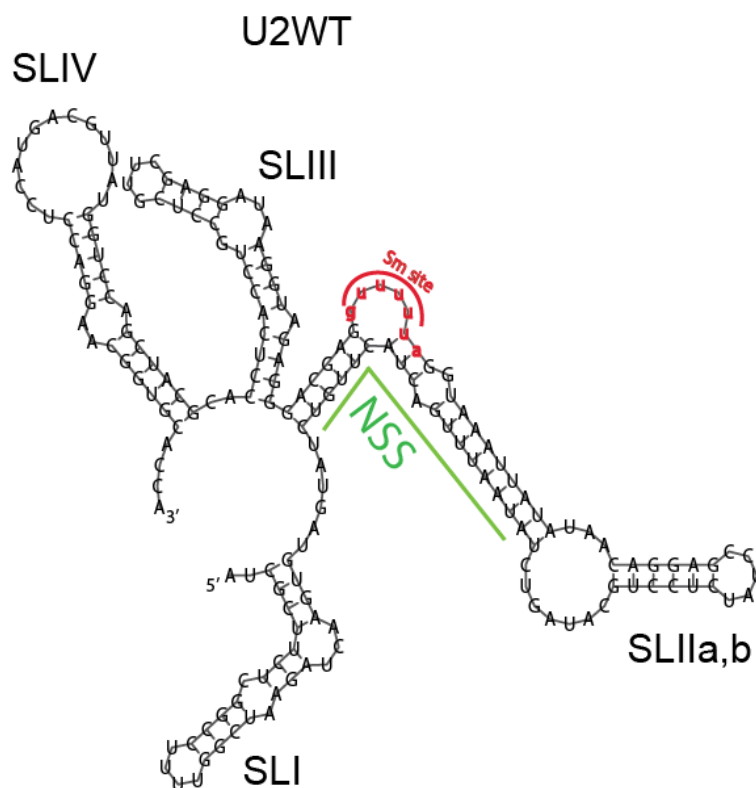


Figure 34: U2 WT secondary structure

The secondary structure was predicted with RNA fold (done by Josef Pánek). Sm site is marked in red. Near Sm site Stem (NSS) is marked with green lines. The stem loops of U2 snRNA are marked by SLI, SLIIa,b, SLIII and SLIV.

Depletion of Gemin 3 leads to decrease snRNA accumulation in CBs

To test whether the Gemin3 plays an important role in snRNP assembly we depleted this protein by RNA interference. Then we microinjected *in vitro* transcribed U1, U4 or U2 WT snRNAs labeled by Alexa 488 into the cytoplasm of Gemin3 depleted cells and observed lower accumulation of snRNAs in the Cajal bodies. WT U1 and U2 snRNAs were mainly localized in the cytoplasm, which suggests that Sm ring was not assembled resulting in stalling of snRNAs in the cytoplasm (Fig. 35 A, B). WT U4 snRNA was localized in the cytoplasm and the nucleus, but it was not targeted into the CBs (Fig. 35 C). We did not observe this phenotype in cells treated with negative control siRNA (Fig. 35 A, B and C). It should be noted that depletion of Gemin3 did not have substantial effect on the Cajal body integrity, e.g. as a depletion of Sm proteins (Fig. 27), which indicates that the biogenesis of snRNPs and their targeting into the CBs was not completely abolished (e.g. as in the case of TGS1 or SMN knockdowns Figs. 28, 29 and 32).

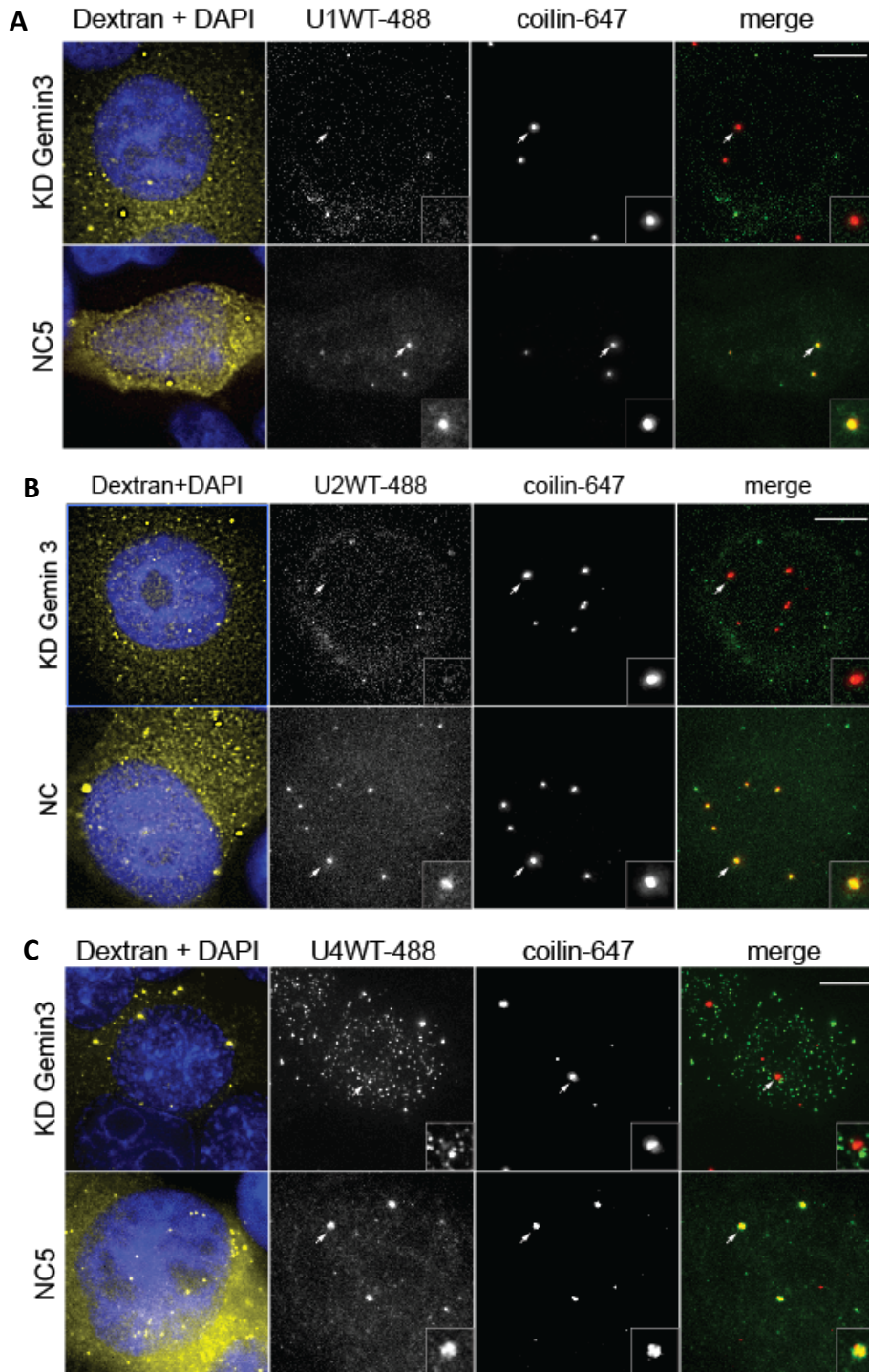


Figure 35: Depletion of Gemin 3 leads to decrease snRNA accumulation in CB

(A,B,C) *In vitro* transcribed U1 WT (A), U2 WT (B) or U4 WT (C) were microinjected into the cytoplasm of Gemin3 depleted HeLa cells or HeLa cells treated by Negative control (NC) siRNA. U1, U2, and U4 snRNAs were labeled with UTP-Alexa-488 (green), coilin, a marker of CBs, was immunolabeled by Alexa-647 (red). Dextran-TRITC 70kDa (yellow) was used to monitor cytoplasmic microinjection, DNA was stained by DAPI (blue). The arrows indicate CBs that are enlarged in insets. The scale bar represents 10 μ m.

The changes in the secondary structure lead to the snRNA targeting into CBs in the Gemin3 depletion cells

Next, we decided to test, whether the strengths of the putative NSS has an affects on snRNP biogenesis. We designed mutations in the U2 snRNA WT sequence, which increased or decreased the stability of the NSS structure. Mutations U2 stableNSS increase the stability of NSS extension of the stem through the Sm site. In contrast, mutations in U2-noNSS destabilize the NSS resulting in open structure (Fig. 36).

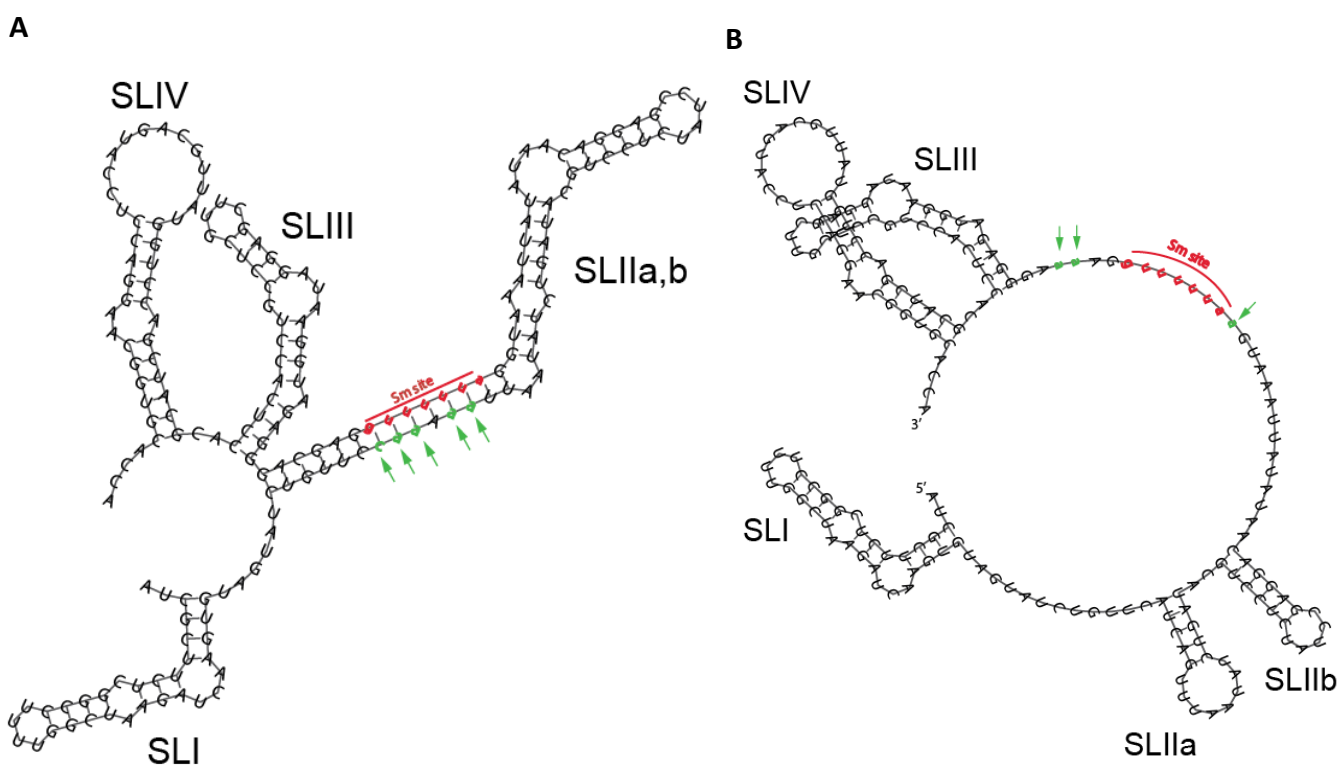


Figure 36: U2 stableNSS and U2 noNSS secondary structures

The secondary structures were predicted with the program RNA fold (done by Josef Pánek). (A) U2 stableNSS mutant with more stable NSS. (B) U2 noNSS mutant with less stable NSS. Sm site is marked in red. The green arrows mark the point mutations. The stem loops of U2 snRNA are marked by SLI, SLIIa,b, SLIII and SLIV.

We microinjected *in vitro* synthesized U2 stableNSS RNA labeled by Alexa-488 into the cytoplasm of the HeLa cells (Fig. 37A). We observed the same phenotype as in Gemin3

depleted cells. U2 stableNSS snRNA mostly stayed in the cytoplasm and its localization into CBs was reduced, which indicates the problem with Sm ring assembly. In contrast, microinjected U2 noNSS snRNA into the cytoplasm reached the nucleus and CBs in control cells as well as in cells depleted of Gemin3 (Fig. 37B). These findings indicate that the structure around the Sm site is important for snRNP biogenesis and specifically for the Sm ring assembly.

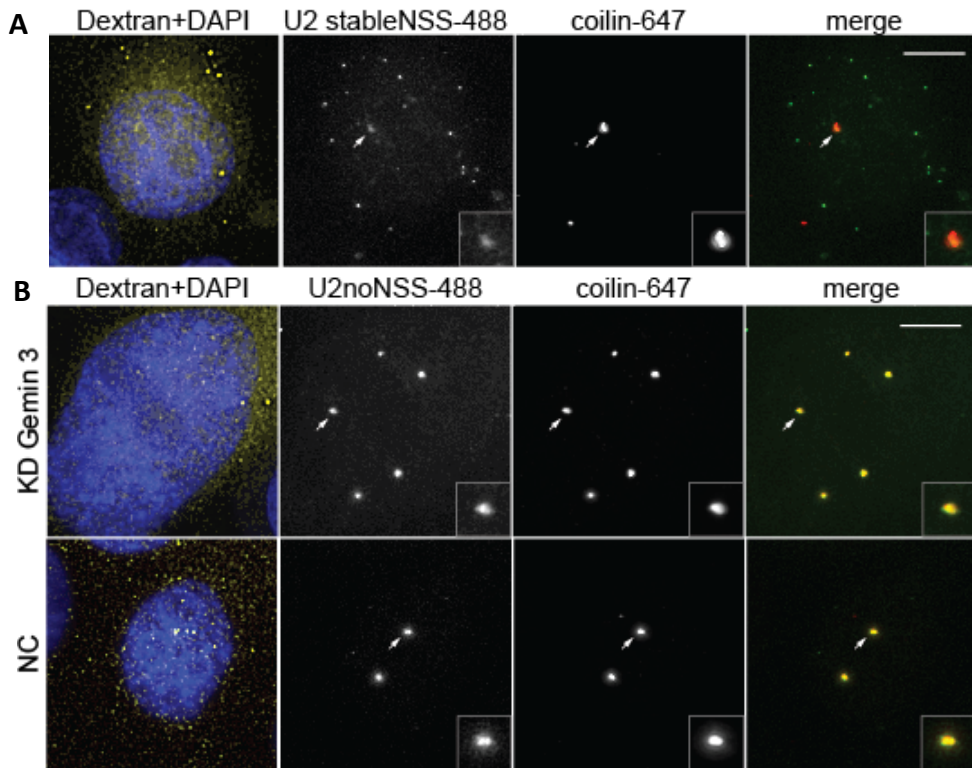


Figure 37: Microinjection of U2 U bulge and U2 no hairpin mutants

A) *In vitro* transcribed U2 U bulge snRNA was microinjected into the cytoplasm of HeLa cells. (B) Gemin3 was depleted by siRNA and *in vitro* transcribed U2 no hairpin snRNA was microinjected into the cytoplasm of Gemin3 depleted HeLa cells or treated with Negative control (NC). snRNAs were labeled by Alexa-488 (green) and coilin, a marker of Cajal body, was immunostained by Alexa-647 (red). Dextran-Tritc (70kDa) was used as a marker for the localization of microinjection (yellow). The arrows indicate the position of insets. The scale bar represents 10 μ m.

To test that the structure of snRNA is important for the snRNP maturation we prepared WT U2 snRNA by *in vitro* transcription and denatured it in 80°C for 90 s before microinjection into the cytoplasm of Gemin3 depleted cells (Fig. 38A). In contrast to non-denatured snRNA (Fig. 38B), denatured the U2 snRNA was targeted into the CBs even if the Gemin3 was depleted. These experiments confirm that releasing the secondary structure (either by mutations or heat denaturation) rescue the phenotype observed after Gemin3 downregulation and suggest that Gemin3 is important to open the snRNA secondary structure around the Sm site and allow Sm ring assembly.

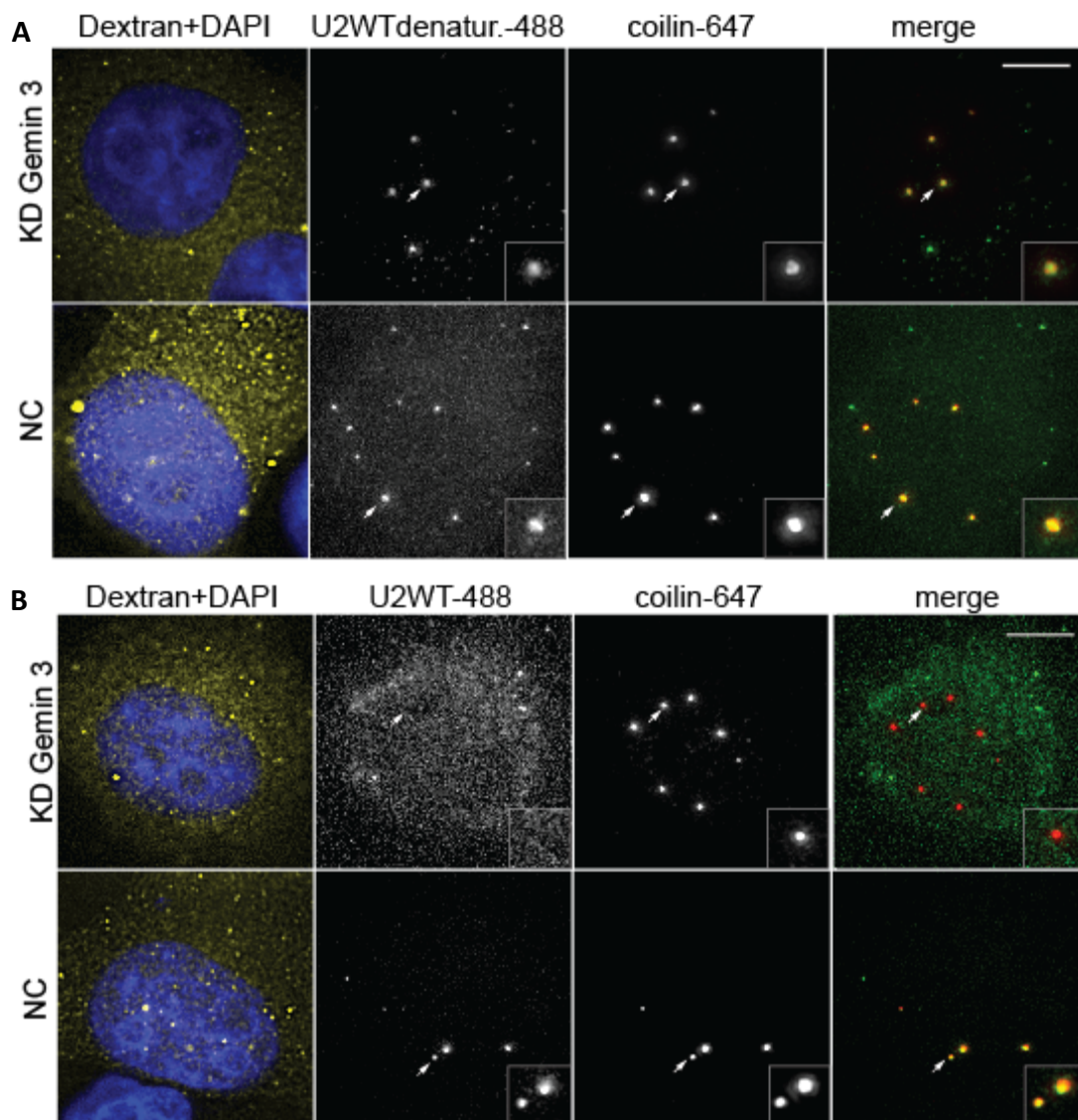


Figure 38: Microinjection of denatured WT U2 snRNA

(A) *In vitro* transcribed WT U2 snRNA was denatured and microinjected into the cytoplasm of HeLa cells. (B) Non-denatured *in vitro* transcribed WT U2snRNA was microinjected into the cytoplasm. WT U2 snRNA was labeled by Alexa-488. The coilin, marker of CBs, was immunostained by Alexa-647 (red). Dextran-Tritic (70kDa) was used as a marker of the localization of microinjection (yellow). The arrows indicate position of the insets. The scale bar represents the 10 μ m.

The ectopic expression of U2 stableNSS or U2 noNSS mutants

Our experiments showed the importance of U2 snRNA secondary structure in the Sm ring assembly and sequential targeting into the nucleus and to CBs. To test whether this effect is specific for microinjected snRNAs we analyzed ectopically expressed U2 snRNAs containing the MS2-binding site (see above Fig. 26). In the cells co-transfected with U2WT MS2 construct and MS2-YFP, the targeting into the CBs was less efficient in the Gemin3 depleted cell, but not completely abolished (Fig. 39A). We did not observe these phenotype in the cells treated with negative control (NC) (Fig. 39A), where the U2WT MS2 was targeted into the CBs without any accumulation in the cytoplasm. We prepared the mutants (U2 stableNSS-MS2 and U2 noNSS-MS2) by site directed mutagenesis and co-transfected U2 noNSS-MS2 or U2WT-MS2) with plasmid containing MS2-YFP protein into control or the Gemin3 depleted cells. We co-transfected mutant U2 stableNSS-MS2 with MS2-YFP into non-treated cells. 24h after transfection cells were fixed and coilin was detected by immunofluorescence (Fig. 39 A, B, C). U2 NoNSS-MS2 construct was nicely targeted into the CBs both Gemin3 depleted cells and cells treated with negative control (NC) (Fig. 39C). We did not observe any cytoplasmic accumulation. In contrast, strengthening the NSS in U2 stableNSS construct strongly reduced nuclear and CB accumulation in non-treated cells suggesting that strong base-pairing around the Sm site inhibits Sm ring formation. This construct localized mainly in the cytoplasm (Fig. 39B).

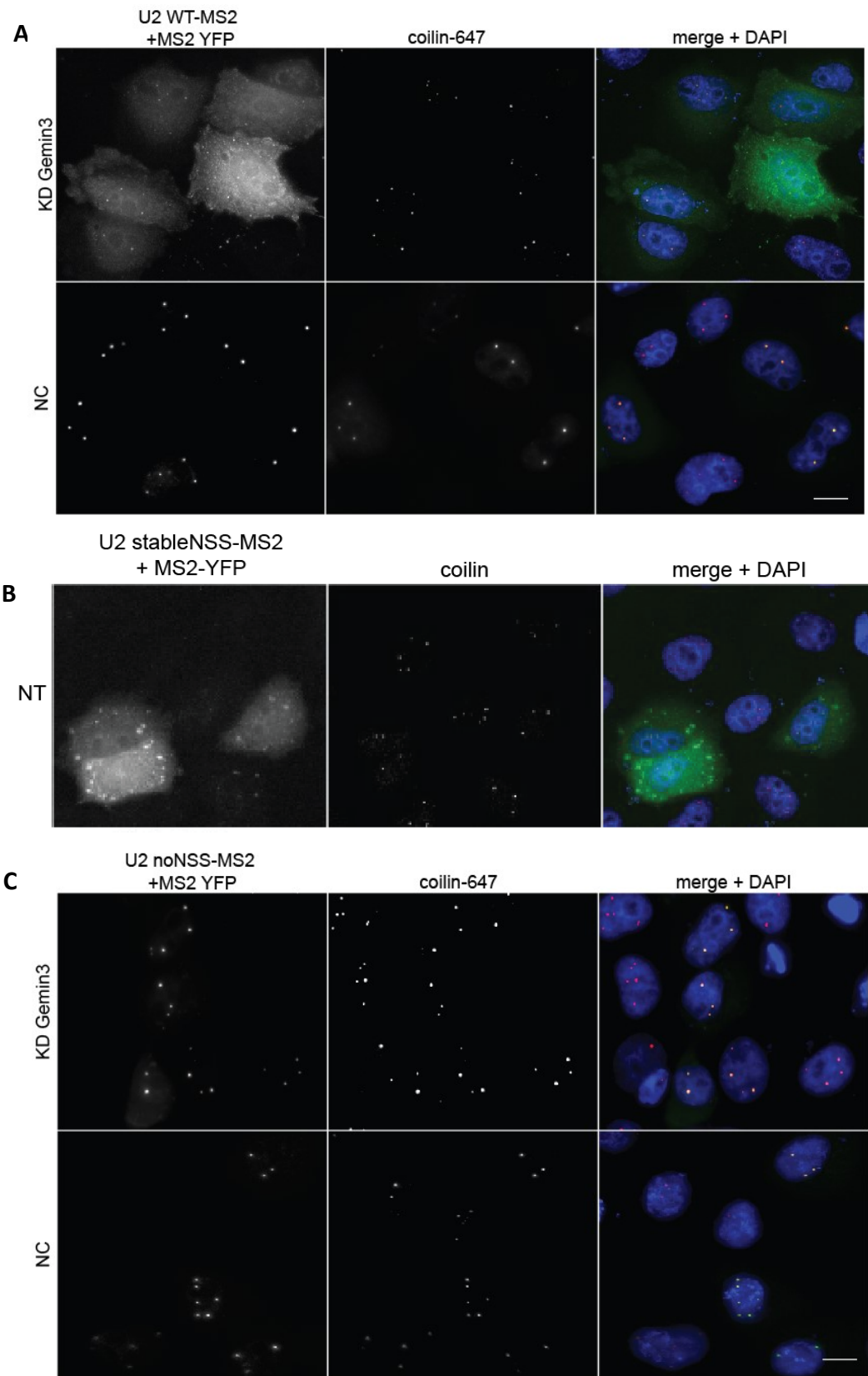


Figure 39: The ectopic expression of U2 U bulge MS2 or U2 no hairpin MS2 mutants and U2 WT MS2

(A,B,C) Gemin3 was depleted by RNAi, and subsequently, the HeLa cells were co-transfected with U2 constructs (U2 WT MS2 (A), U2 stableNSS-MS2 (B) and U2 noNSS-MS2 (C) containing the MS2 loop and MS2-YFP (green). Coilin was used as a marker of CBs (red). NT means non-treated cells. DNA was stained by DAPI. The scale bar represents 10 μ m.

3.3 The recognition of the immature particles and their targeting into the P bodies

Disruption of Sm ring assembly leads to localization of snRNAs into P bodies

While analyzing snRNA targeting to the nucleus and CBs, we noticed that treatments preventing Sm ring assembly result in accumulation of snRNA in bright cytoplasmic foci (Figs.21, 22, and 27). Last years it was shown that truncated form of U1 snRNA lacking the Sm site accumulates in the cytoplasmic structures called P bodies (Ishikawa et al., 2014). Therefore, we wanted to test, whether these cytoplasmic foci correspond to P bodies. We prepared U2 snRNA lacking the Sm site (U2 Δ Sm) and microinjected *in vitro* synthesized RNA into the cytoplasm of HeLa cells. We detected DDX6, marker of P bodies, with immunofluorescence and checked the colocalization with U2 Δ Sm. The mutant was not targeted into the nucleus but accumulated in several cytoplasmic foci that co-localized with DDX6 (Fig. 40).

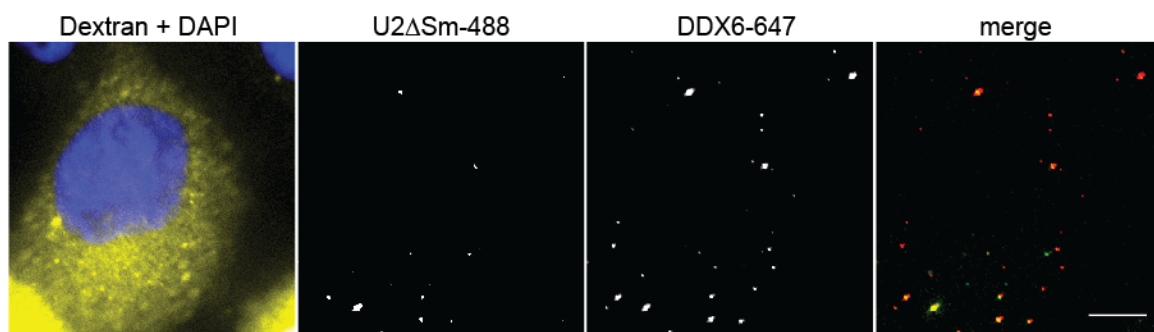
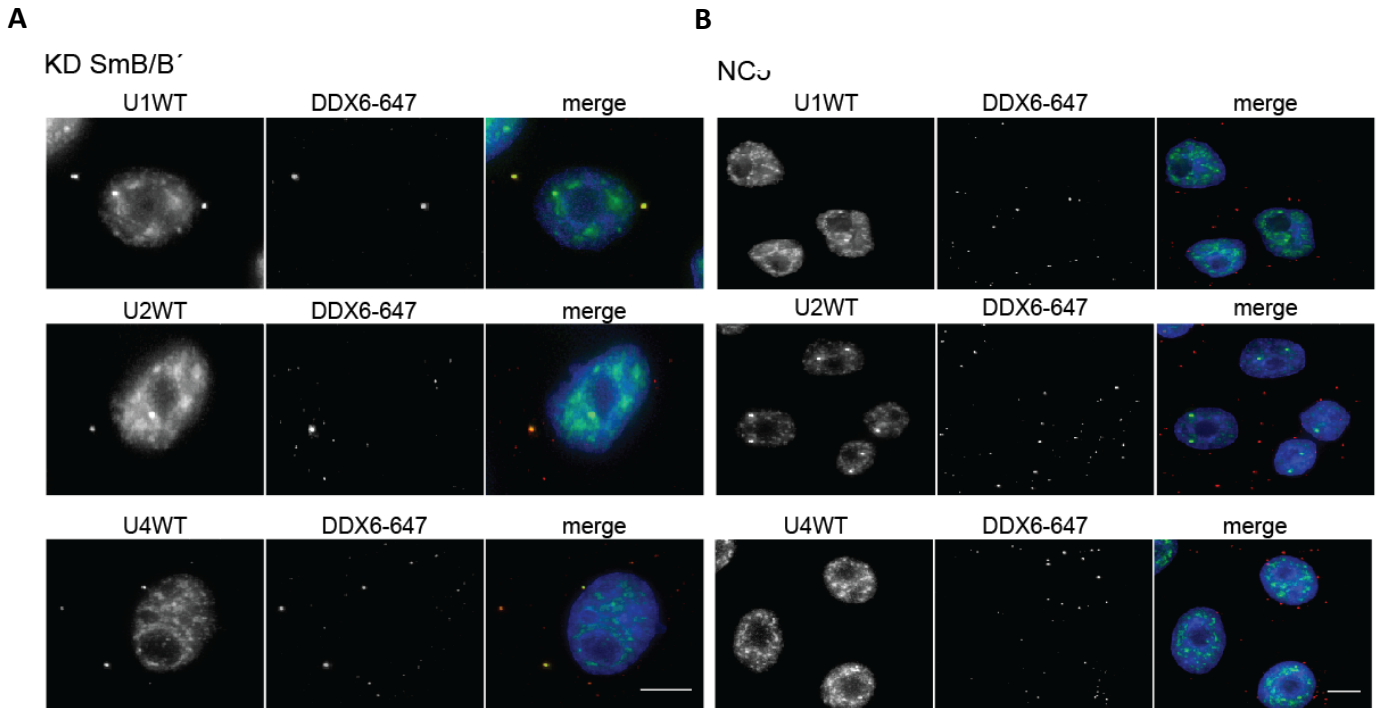


Figure 40: U2 Δ Sm snRNAs is accumulated in the P bodies

In vitro transcribed U2 Δ Sm snRNA was microinjected into the cytoplasm of HeLa cell. snRNA was labeled by Alexa-488 (green). DDX6, a marker of P bodies, was immunostained by Alexa-647 (red). Dextran-Tritc (70kDa) was used as an indicator of localization of microinjection (yellow). DNA was stained by DAPI (blue). The scale bar represents 10 μ m.

To further test, whether the disruption of the Sm ring induces P body accumulation of snRNAs, we depleted the SmB/B' protein which inhibits the formation of Sm ring in cells. The U1, U2, and U4 snRNAs were detected by fluorescence *in situ* hybridization and DDX6, by immunostaining. After the depletion of the SmB/B' protein the U1, U2, and U4 snRNAs

were accumulated in P bodies and a nuclear signal of snRNAs was decreased in comparison with the cells treated with negative control (Fig. 41 A, B).



C

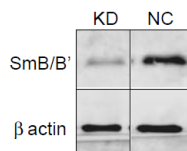


Figure 41: Endogenous U1, U2 and U4 snRNAs are accumulated in P bodies after Sm ring disruption

Cells were treated either with anti-SmB/B' siRNA (A) or negative control (B) and U1, U2 and U4 snRNAs were detected by fluorescence *in situ* hybridization using DNA probes labeled by Cy3 (green). DDX6, marker of P bodies, was immunostained by Alexa-647 (red). DNA was stained by DAPI (blue). The scale bars represent the 10 μ m. (C) Test of SmB/B' siRNA efficiency. β actin was used as a loading control.

The SMN protein colocalize with immature snRNAs in P bodies

It has been shown that Sm ring is formed by the SMN complex (Paushkin et al., 2002). Therefore, we decided to test, whether the SMN complex is bound on the immature snRNAs or whether the inhibition of Sm ring formation also abolishes the SMN binding of snRNA. We used the same U2-MS2 constructs as in previous chapters. We prepared the U2 Δ Sm MS2 construct, which lacks the Sm site and transfected it to cells together with the MS2-YFP. U2 Δ Sm MS2 construct was accumulated in the P bodies (Fig. 42A). Immunostaining of the SMN protein showed that U2 Δ Sm MS2 positive P bodies also

contain the SMN protein (Fig. 42B) while in control cells the SMN protein does not accumulate in P bodies (Fig. 42C).

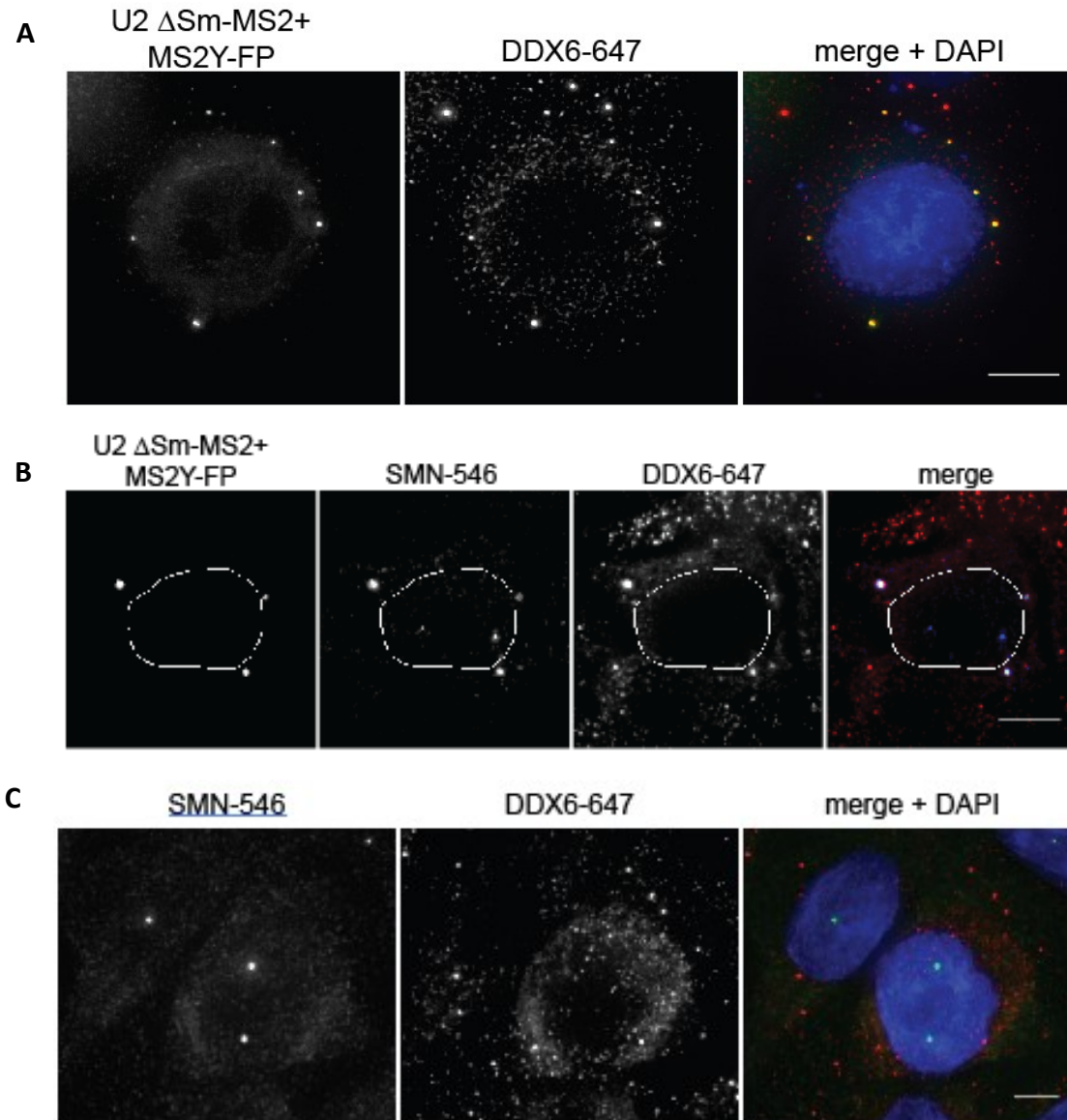


Figure 42: U2ΔSm MS2 is accumulated in the P bodies

(A,B) The cells were cotransfected with U2ΔSm MS2 and MS2-YFP (green). DDX6, marker of P bodies, were immunolabeled by Alexa-647 (red). DNA was labeled by DAPI (blue) (B,C) SMN was labeled by Alexa-546.. DNA was labeled by DAPI (blue). The scale bar represents 10 μm.

Our experiments showed that SMN complex is presented together with defective snRNAs in P bodies. Further, we wanted to test if the SMN complex plays a role in the targeting of the defective snRNAs into the P bodies. We prepared a double mutant lacking both Sm and SMN binding sites (U2 Δ SmSMN). We microinjected *in vitro* transcribed U2 Δ SmSMN snRNA into the cytoplasm of HeLa cells. The U2 Δ SmSMN mutant was accumulated in P bodies although the SMN binding site was disrupted (Fig. 43). This result indicates that SMN protein does not play a role in navigating defective snRNAs into the P bodies.

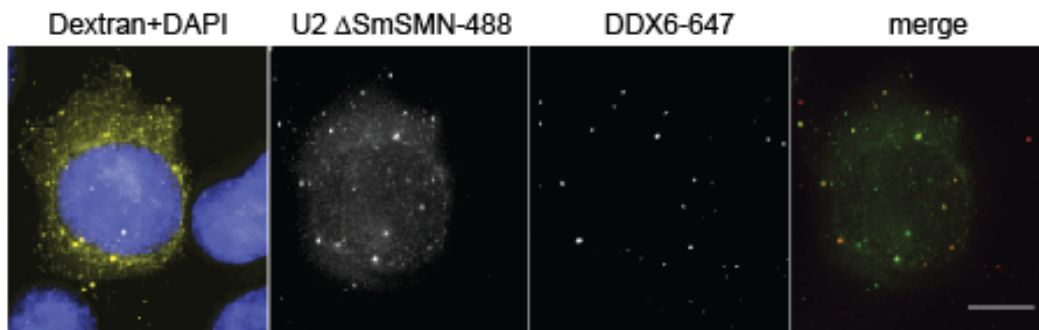


Figure 43: U2 Δ SmSMN mutant is localized in P bodies

In vitro transcribed U2 Δ SmSMN mutant was microinjected into the cytoplasm of HeLa cells. snRNA was labeled by Alexa-488 (green). DDX6, marker of P bodies, was immunostained by Alexa-647 (red). DNA was stained by DAPI (blue). The scale bar represents 10 μ m.

The LSM1 protein plays a role in the targeting of the immature snRNAs into the P bodies

The previous study showed that truncated form of U1 snRNA lacking the Sm site is degraded by the Xrn1 exonuclease (Shukla and Parker, 2014). Before Xrn1 can start degrading RNA, the 5' monomethyl guanosine cap has to be removed by the Dcp1/2 complex (STEIGER et al., 2003) The Dcp1/2 complex is stabilized by LSM1-7 proteins (Nissan et al., 2010). Therefore, we decided to test a potential role of LSM1-7 proteins in the targeting of immature snRNAs. We cloned the LSM1 protein into the GFP-N2 vector and transfected this plasmid into the SmB/B' depleted HeLa cells. After 24 h we analyzed snRNAs bound by LSM1-GFP by immunoprecipitation using anti-GFP antibody followed by RNA isolation and RT-qPCR. The proteins were analyzed by Western blot (Fig. 45B). Our

results showed an increased association of Sm-type snRNAs with LSm1-GFP after depletion of SmB/B' protein which prevents the Sm ring assembly (Fig. 45A). U6 snRNA, which does not contain the Sm site served as a negative control.

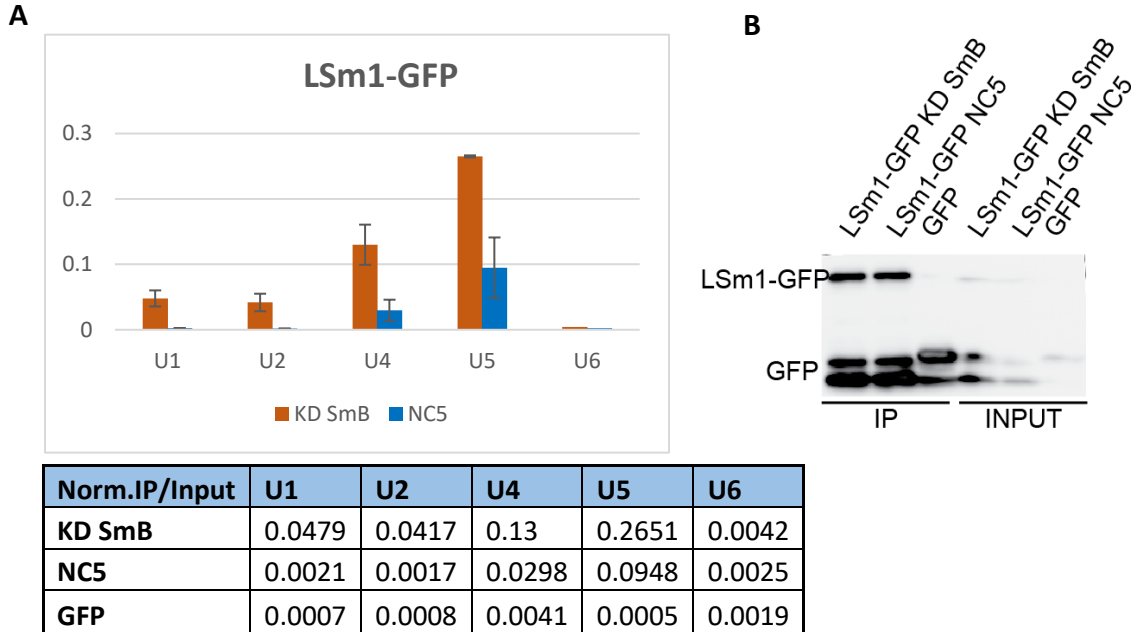


Figure 44: LSm1 protein binds the immature U snRNAs

(A) SmB/B' protein was depleted by RNAi, and the cells were transfected by LSm1-GFP protein. snRNAs in Inputs and immunoprecipitates were analyzed by RT followed by qPCR using primers shown in Table 1. qRT-PCR using the specific primers for detecting of U snRNAs (U1, U2, U4, and U5). U6 was used as a control. The average of three independent experiments is shown. Error bars represent standard deviation. (SD). (B) Proteins were analyzed by WB using the anti-GFP antibody.

Further, we tested a potential role of LSm1 protein in targeting of snRNAs to the P bodies. We depleted LSm1 protein by RNAi and co-transfected cells with U2ΔSm-MS2 and MS2-YFP plasmids. We detected P bodies by immunolabeling of DDX6 (Fig. 46 A). Our experiment showed that depletion of LSm1 protein inhibits localization of defective snRNAs to the P bodies. Therefore, we decided to test the localization of the endogenous snRNA after depletion of SmB/B' and LSm1. We depleted LSm1 together with SmB/B' protein by RNAi and detected U2 snRNA by fluorescence *in situ* hybridization. DDX6, a marker of P bodies, was detected by immunostaining (Fig. 46B). The LSm1 knockdown prevented localization of U2

snRNA into P bodies, which was observed in the cells treated with SmB siRNA only (Fig. 45B). These results indicate the role of LSm1 protein in snRNA degradation.

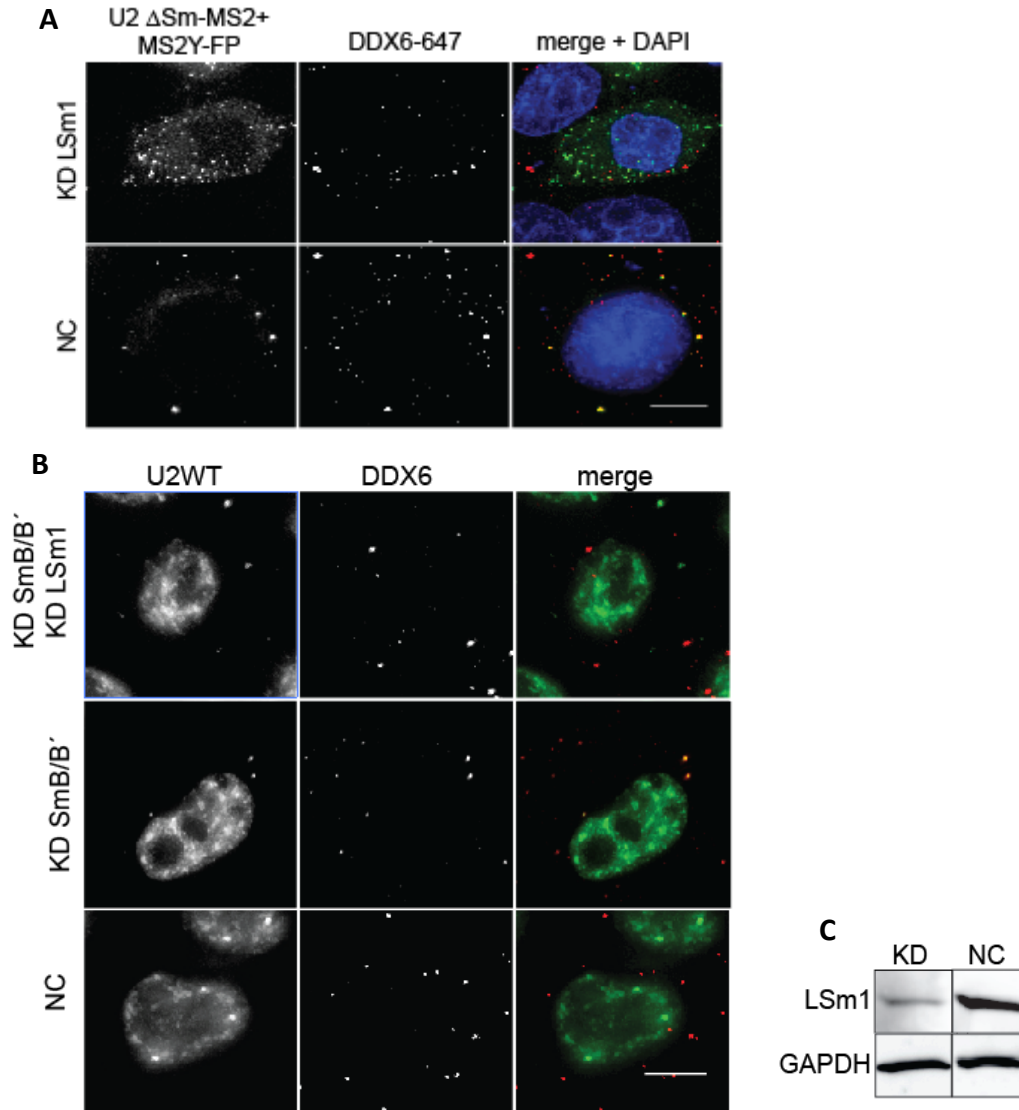


Figure 45: Depletion of LSm1 leads to disruption of snRNAs targeting into P bodies

(A) Lsm1 protein was depleted by RNAi and cells were cotransfected with U2 Δ Sm MS2 and MS2-YFP (green). As a control we treated cells by Negative control (NC). (B) SmB/B' and LSm1 proteins were depleted by RNAi. As a control we treated the cells with SmB/B' siRNA or Negative control (NC). U2 snRNA was detected by fluorescence *in situ* hybridization using the DNA probe labeled by Cy3. DDX6, marker of P bodies, was immunostained by Alexa-647. The scale bar represents 10 μ m. (C) Test of LSm1 siRNA efficiency. GAPDH was used as a loading control.

4. Discussion

Sm ring is essential for Cajal body targeting

snRNP biogenesis starts in the cell nucleus by snRNA transcription, continues in the cytoplasm, where snRNA acquires the ring of Sm proteins, and then the core snRNP returns to the nucleus, where snRNP matures and participates in splicing. After reimport to the nucleus snRNPs first appear in CBs where the final steps of their maturation occur (Nesic et al., 2004; Stanek et al., 2003). Despite an important role of CB in snRNP biogenesis, it has not been clear what targets the core snRNPs to CBs. In this project, we provide several lines of evidence, which Sm and SMN sites are necessary and sufficient to target snRNAs into CBs. At first, the Sm and SMN sites are both essential for CB targeting of microinjected snRNAs. Then we showed that the minimal U2 and U4 snRNA constructs which contain Sm and SMN sites are efficiently accumulated in the CBs when they are ectopically expressed in human cells. Finally, we were able to target RNAs, which are not normally present in the CBs (7SK, Alu from human and SRP from *E.coli*) to the CBs, when we added Sm and SMN binding sites on their 3' end.

Previously it has been shown that minimal sequence containing Sm and SMN binding sites is sufficient to bind the SMN complex (Golembe et al., 2005; Yong et al., 2004). It was proposed that this complex is able to facilitate nuclear import of newly assembled core snRNPs and target them via interaction with coilin to CBs (Narayanan et al., 2004). This indicates a role of SMN complex in the Cajal body targeting of core snRNPs. Our experiments showed that the depletion of Sm proteins blocks CB localization of U2 snRNA WT, which contains SMN binding sites (Fig. 28). Moreover, the core snRNPs (snRNA with Sm ring) were accumulated in CBs even if the SMN protein was depleted (Fig. 33). It should be noted that the depletion of SMN complex components (SMN, Gemin2, and Gemin5) reduces accumulation of snRNPs in the CBs (Fig. 29). However, in this case, the effect is likely indirect by disruption of Sm ring assembly. Thus, we conclude

that the SMN complex plays a role in Sm ring assembly but not in Cajal body targeting of snRNPs.

In the CBs the core snRNPs have to be captured. Coilin is the main CB protein and is the most prominent candidate, which can interact directly with the core snRNPs. The previous studies showed the direct interaction between coilin and Sm proteins via its Sm-fold (Toyota et al., 2010; Xu et al., 2005). However, the interaction is enhanced and stabilized by the C-terminal tails of Sm proteins (Xu et al., 2005). The C-terminus of coilin contains a Tudor domain (Shanbhag et al., 2010), which in other proteins interacts with methylated arginines (Pek et al., 2012). We have shown that deletion of GR repeats reduces Sm protein localization to CBs (Fig. 30 and 31). Thus, we can speculate that the coilin Tudor domain binds dimethylated arginines of GR repeats found in the C-termini of SmB/B', SmD1 and SmD3. However, it was shown before, that the isolated coilin Tudor domain did not exhibit any dimethylated arginine binding activity *in vitro* (Shanbhag et al., 2010) leaving the molecular mechanism of coilin-snRNP interaction resolved. The iCLIP data showed direct binding of coilin to snRNAs. This interaction likely provides an additional signal for CB localization of snRNAs (Machyna et al. 2014).

Recently, it has been proposed that the nuclear non-membrane organelles, such as Cajal bodies, are formed by phase-separation (reviewed in Courchaine et al., 2016). It was shown that the protein with low complexity domains (LCDs) are involved in this process (Berry et al., 2015; Lin et al., 2015; Patel et al., 2015). The GR repeats situated on C-termini of SmB/B', SmD1 and SmD3 proteins belong to LCDs. Here we demonstrate that deletion of GR repeats or replacement of these amino acids leads to decrease in CB accumulation of SmB/B', SmD1 and SmD3. Finding that, Sm proteins lacking all or a part of the GR repeats are still interacting with snRNAs strongly indicates that they are incorporated into the Sm ring with other endogenous Sm proteins (Fig. 30). It should be noted that SmB/B', D1 and D3 are found next to each other in the Sm ring. Thus we can speculate that GR domains of Sm proteins together create a larger low-complexity domain that phase separates with coilin and other CB proteins.

To our big surprise, snRNAs microinjected into the nucleus were accumulated in the CBs as well, which strongly indicates that they acquired the Sm ring, which is usually assembled only in the cytoplasm. We provide several lines of evidence that snRNAs microinjected into the nucleus are not transported into the cytoplasm. First, the snRNA mutants, which do not accumulate in CBs stayed in the nucleus in case of nuclear microinjection (Figs. 21D, E, F, 22 B, D, F and 23A, B). Second, when we microinjected WT U2 snRNA into one nucleus of heterokaryon, the snRNA was not observed in other nuclei (Fig. 23C). These findings indicate that the Sm ring was assembled in the nucleus. Previous studies showed that Sm ring formation is SMN complex dependent (Battle et al., 2006c; Yong et al., 2004). This indicates that the SMN complex can form the Sm ring also in the nucleus. This is consistent with the characterization of the nuclear 20S SMN complex containing Sm proteins (Meister et al., 2000). However, the mechanism of nuclear Sm ring formation is still unknown. Microinjected snRNAs resemble mature snRNAs rather than newly transcribed snRNAs that contain a monomethyl guanosine cap and extension at the 3' end. Therefore, we speculate that the nuclear Sm ring-assembly activity primary might rescue those snRNAs, which have lost the Sm ring. Taken together, our data reveal a new role for Sm proteins, and namely, GR repeats found at the C-terminus of several Sm proteins, as the CB targeting signal.

Sm ring and its role in the quality control in the CBs

The Cajal bodies are a place of the final maturation of the snRNPs (Stanek et al., 2003; Tanackovic and Kramer, 2005). Novotný and colleagues have previously shown that the inhibition of the final steps in the U4, U5 and U6 snRNP assembly pathway leads to sequestration of immature snRNPs in the CBs (Fig. 34)(Novotny et al., 2015). Similarly, the U2 snRNP lacking the SF3a complex has been shown to accumulate in CBs (Tanackovic and Kramer, 2005). Here, we observed the same phenotype in case of microinjection of U2 Δ SLI snRNA, which does not interact with the SF3a complex. In contrast, when we deleted Sm site in U2 Δ SLI mutant, this defective snRNA did not

accumulate in CBs. Based on our results that Sm ring serves as an essential CB targeting signal, we propose a model that the exposed Sm ring serves as a CB retention signal. We hypothesize that interactions of CB proteins with the unprotected Sm ring represent the molecular basis of the cellular mechanism controlling final steps of snRNP assembly. Core snRNPs are held in CBs until the specific proteins are bound, or the composite U4/U6•U5 tri-snRNP is formed. The binding of snRNP-specific proteins weakens the interaction between Sm proteins and CB factors allowing the mature snRNP to leave the CB. While the U2 snRNP structure is unknown, the atomic structure of yeast and mammalian tri-snRNP has already been determined. The structural data show Sm rings positioned on tri-snRNP edges with the side of the Sm ring that exposes C-terminal tails in close proximity of snRNP specific proteins (Nguyen et al., 2016; Agafonov et al., 2016). The U1 snRNP structure was also resolved. U1 specific proteins SNRNP70 and SNRNPC are in the positions covering some of the Sm proteins (Krummel et al., 2010), which might also explain why endogenous U1 snRNP is found at a lower concentration in CBs than other snRNPs.

Based on these data we proposed a model that the interaction of CB proteins with the Sm ring represents the molecular principle how cells discriminate between mature and immature snRNPs.

Gemin3 and its role in the Sm ring assembly

Previous studies showed that SMN complex plays an essential role in the snRNP biogenesis, especially in the Sm ring assembly but the molecular function of individual components is unclear (Massenet et al., 2002; Pellizzoni et al., 1998). Our data show that depletion of the Gemin3 decreases snRNA accumulation in the CBs (Fig. 35). However, the role of Gemin3 in the Sm ring assembly is still unknown. The Gemin3 was annotated as a DEAD-box RNA helicase. The main function of the RNA helicases is to unfold the secondary structures of RNAs using the energy from ATP. Our modeling of U2 snRNA secondary structure using RNAfold program predicts an extensive base-pairing of

sequences around the Sm site, which we called Near Sm site Stem (NSS) (Fig. 35). We provide several lines of evidence, that the NSS structure is involved in snRNP biogenesis and that Gemin3 plays a role in unfolding the NSS to release the Sm site for Sm ring formation. First, we prepared the mutant U2 stableNSS, where the NSS was more stable in comparison with WT was not efficiently transported to the nucleus, which indicates that the structure around the Sm site plays a role in Sm ring assembly (Figs 38A and 40B). Contrary, the U2 noNSS mutant, where the NSS was prevented and U2 snRNA forms the open structure, was transported and accumulated in the CBs when the Gemin3 was depleted (Fig. 38B and Fig. 40C). This finding strongly suggests that Gemin3 is important for unwinding the NSS and Sm ring formation. Consistently, we microinjected denaturated WT U2 snRNA into the Gemin3 depleted cells we observed that the WT U2 snRNA was also accumulated in CBs (Fig. 39). Finally, we showed that transiently expressed U2 stableNSS, and U2 noNSS mutants exhibit the same behavior as microinjected snRNAs (Fig. 37 and Fig. 39).

The newly transcribed snRNAs do not contain Sm proteins until they enter the cytoplasm. The NSS or other secondary structure can be formed around the Sm site immediately after transcription. For proper biogenesis the structure has to be opened for exposure of the Sm site. The Gemin3 has helicase activity and can execute this role during the Sm ring formation (Yan et al., 2003). Taken together we showed a new possible role for Gemin3 in the unfolding of the secondary structure of snRNAs for better accessibility of Sm site resulting in Sm ring formation.

Quality control of snRNPs in the cytoplasm

My previous data showed an important role for the Sm ring as a Cajal body targeting and retention signal. When the Sm ring is not assembled snRNAs stayed in the cytoplasm (Fischer et al., 1993). The previous study has shown that truncated form of U1 snRNA (without Sm site) was accumulated in the cytoplasmic foci, which were identified as P bodies (Ishikawa et al., 2014). We observed the same localization of endogenous or *in*

vitro transcribed snRNAs (U2, U4 and U5) after Sm proteins depletion or Sm site deletion resulting in disruption of Sm ring assembly. However, the recognition of immature snRNPs and their targeting into the P bodies is still unknown. P bodies are storage of many factors involved in RNA degradation such as exonuclease Xrn1 or decapping enzymes Dcp1 and Dcp2. Consistently it was proposed that U1 snRNA without Sm ring is degraded by 5'-3' exonuclease Xrn1 (Shukla and Parker, 2014). We found out that SMN protein is localized in the P bodies together with defective snRNAs (Fig. 43). However, microinjection of U2 Δ SmSMN snRNA lacking Sm and SMN binding sites also resulted in accumulation in the P bodies. This experiment suggests that transport of immature snRNAs into the P bodies is SMN independent.

The Lsm 1-7 ring is responsible for stabilizing of decapping enzymes on mRNA intended for degradation. We provide several lines of evidence that LSm1 play a role in the localization of snRNAs without Sm ring into the P bodies. First, we showed that LSm1 protein binds the snRNAs after the depletion of Sm protein (Fig. 45). Furthermore, the reduction of LSm1 prevents the localization of defective snRNAs into the P bodies (Fig. 46). After depletion of LSm1 protein defective snRNAs accumulated in the cytoplasm out of P bodies. Our data suggest that LSm1-7 ring play a role in snRNA degradation pathway. However, the binding mechanism of the LSm1-7 ring on snRNA is unknown. Previous studies have shown the interaction between LSm4 and SMN (Paushkin et al., 2002). It is possible that SMN interacts with the LSm4 when the Sm proteins are not available. This interaction can explain why the SMN is in P bodies together with defective snRNAs. However, the interaction between SMN and LSm4 is not essential for localization of defective snRNAs in P bodies (Fig. 43). The binding site for the LSm1-7 ring is composed of 8 uridines (Zhou et al., 2014a). A recent study has shown that truncated form of U2 snRNA (U2-tfs) are uridylated on 3' end by TUT4 and TUT7 (Ishikawa et al., 2018), which can create LSm binding site. Taken together our data suggest a new role for the LSm1-7 ring in snRNA degradation pathway. However, the precise mechanism remains unclear.

5. Summary

In this work I focused on the snRNP biogenesis. snRNPs are key components of spliceosome, which catalyzes the splicing of pre-mRNA. My project is separated into the three chapters undertake snRNP assembly and quality control of their maturation in the nucleus and the cytoplasm.

In the first project I studied the part of snRNP biogenesis, where the snRNPs are targeted into the nuclear structures called Cajal bodies, where the final maturation steps occur. We found out that Sm proteins navigate snRNPs into these structures and mapped this Cajal body targeting signal to GR repeats of SmB/B', SmD1 and SmD3. Our experiments also showed that incomplete snRNPs are more accumulated in the CBs comparing with the WT snRNAs. Based on my experiments we established a model, where Sm proteins are necessary for targeting of snRNP into CBs and also play a role in their quality control in the CBs.

In the second project I studied alternative snRNA secondary structure, which was observed during the mathematical modeling of U2 snRNA in my previous project. In this structure the Sm site is surrounded by an extensive stem, which we named Near Sm site Stem (NSS) and is not observed when the Sm ring is formed. It is known that SMN complex play a main role in Sm ring assembly. We depleted the component of the SMN complex, Gemin3, which is DEAD-box RNA helicase and observed less accumulation of snRNAs in CBs. We propose a model, where the Gemin3 can play a role in unwinding the NSS and opens this structure for Sm ring assembly.

In the third project I focused on the quality control of snRNA biogenesis in the cytoplasm. I showed that disruption of Sm ring assembly leads to accumulation of snRNAs in the cytoplasmic structures called P bodies and their consequential degradation by Xrn1. In this degradation pathway, many proteins are involved. I focused on LSm1-7 ring which play a role in stabilizing of decapping enzymes Dcp1 and Dcp2 and

identified a new role of Lsm1 protein in the navigating of defective snRNAs into the P bodies, where are decapped and degraded by Xrn1.

6. References

- Almeida, F., Saffrich, R., Ansorge, W. and Carmo-Fonseca, M.** (1998). Microinjection of Anti-coilin Antibodies Affects the Structure of Coiled Bodies. *J. Cell Biol.* **142**, 899 LP-912.
- Ast, G.** (2004). How did alternative splicing evolve? *Nat. Rev. Genet.* **5**, 773.
- Astuti, D., Morris, M. R., Cooper, W. N., Staals, R. H. J., Wake, N. C., Fews, G. A., Gill, H., Gentle, D., Shuib, S., Ricketts, C. J., et al.** (2012). Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nat. Genet.* **44**, 277–284.
- Baccon, J., Pellizzoni, L., Rappsilber, J., Mann, M. and Dreyfuss, G.** (2002). Identification and characterization of Gemin7, a novel component of the survival of motor neuron complex. *J. Biol. Chem.* **277**, 31957–31962.
- Bachand, F., Boisvert, F.-M., Cote, J., Richard, S. and Autexier, C.** (2002). The product of the survival of motor neuron (SMN) gene is a human telomerase-associated protein. *Mol. Biol. Cell* **13**, 3192–3202.
- Baillat, D., Hakimi, M.-A., Naar, A. M., Shilatifard, A., Cooch, N. and Shiekhattar, R.** (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276.
- Battle, D. J., Lau, C.-K., Wan, L., Deng, H., Lotti, F. and Dreyfuss, G.** (2006a). The Gemin5 protein of the SMN complex identifies snRNAs. *Mol. Cell* **23**, 273–279.
- Battle, D. J., Kasim, M., Yong, J., Lotti, F., Lau, C.-K., Mouaikel, J., Zhang, Z., Han, K., Wan, L. and Dreyfuss, G.** (2006b). The SMN complex: an assembly machine for RNPs. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 313–320.
- Battle, D. J., Lau, C.-K., Wan, L., Deng, H., Lotti, F. and Dreyfuss, G.** (2006c). The Gemin5 protein of the SMN complex identifies snRNAs. *Mol. Cell* **23**, 273–279.
- Berget, S. M.** (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414.
- Berry, J., Weber, S. C., Vaidya, N., Haataja, M. and Brangwynne, C. P.** (2015). RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5237-45.
- Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. and Luhrmann, R.** (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846–850.
- Bindereif, A. and Green, M. R.** (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.* **6**, 2415–2424.
- Bizarro, J., Dodré, M., Huttin, A., Charpentier, B., Schlotter, F., Branlant, C., Verheggen, C., Massenot, S. and Bertrand, E.** (2015). NUFIP and the HSP90/R2TP chaperone bind the SMN complex and facilitate assembly of U4-specific proteins. *Nucleic Acids Res.* **43**, 8973–89.
- Blencowe, B. J.** (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* **25**, 106–110.
- Boelens, W., Scherly, D., Beijer, R. P., Jansen, E. J., Dathan, N. A., Mattaj, I. W. and van Venrooij, W. J.** (1991). A weak interaction between the U2A' protein and U2 snRNA helps to stabilize their complex with the U2B'' protein. *Nucleic Acids Res.* **19**, 455–460.
- Bordonné, R.** (2000). Functional Characterization of Nuclear Localization Signals in Yeast Sm Proteins. *Mol. Cell. Biol.* **20**, 7943–7954.

- BRAHMS, H., MEHEUS, L., DE BRABANDERE, V., FISCHER, U. T. Z. and LÜHRMANN, R.** (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA* **7**, 1531–1542.
- Brody, Y. and Shav-Tal, Y.** (2011). Measuring the kinetics of mRNA transcription in single living cells. *J. Vis. Exp.* e2898.
- Brosi, R., Groning, K., Behrens, S. E., Luhrmann, R. and Kramer, A.** (1993). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* **262**, 102–105.
- Brunhilde, W.** (2000). An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). *Hum. Mutat.* **15**, 228–237.
- Busch, A. and Hertel, K. J.** (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* **3**, 1–12.
- Carissimi, C., Baccon, J., Straccia, M., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J. and Pellizzoni, L.** (2005). Unrip is a component of SMN complexes active in snRNP assembly. *FEBS Lett.* **579**, 2348–2354.
- Carissimi, C., Saieva, L., Baccon, J., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J. and Pellizzoni, L.** (2006). Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly. *J. Biol. Chem.* **281**, 8126–8134.
- Carrel, T. L., McWhorter, M. L., Workman, E., Zhang, H., Wolstencroft, E. C., Lorson, C., Bassell, G. J., Burghes, A. H. M. and Beattie, C. E.** (2006). Survival motor neuron function in motor axons is independent of functions required for small nuclear ribonucleoprotein biogenesis. *J. Neurosci.* **26**, 11014–11022.
- Cauchi, R. J., Davies, K. E. and Liu, J.-L.** (2008). A motor function for the DEAD-box RNA helicase, Gemin3, in *Drosophila*. *PLoS Genet.* **4**, e1000265.
- Chari, A., Golas, M. M., Klingenhöfer, M., Neuenkirchen, N., Sander, B., Englbrecht, C., Sickmann, A., Stark, H. and Fischer, U.** (2008). An Assembly Chaperone Collaborates with the SMN Complex to Generate Spliceosomal SnRNPs. *Cell* **135**, 497–509.
- Charroux, B., Pellizzoni, L., Perkinson, R. A., Shevchenko, A., Mann, M. and Dreyfuss, G.** (1999). Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. *J. Cell Biol.* **147**, 1181–1194.
- Charroux, B., Pellizzoni, L., Perkinson, R. A., Yong, J., Shevchenko, A., Mann, M. and Dreyfuss, G.** (2000). Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli. *J. Cell Biol.* **148**, 1177–1186.
- Chlebowski, A., Lubas, M., Jensen, T. H. and Dziembowski, A.** (2013). RNA decay machines: the exosome. *Biochim. Biophys. Acta* **1829**, 552–560.
- Chowdhury, A., Kalurupalle, S. and Tharun, S.** (2014). Pat1 contributes to the RNA binding activity of the Lsm1–7–Pat1 complex. *RNA* **20**, 1465–1475.
- Chowdhury, A., Kalurupalle, S. and Tharun, S.** (2016). Mutagenic Analysis of the C-Terminal Extension of Lsm1. *PLoS One* **11**, e0158876.
- Collier, S., Pendle, A., Boudonck, K., van Rij, T., Dolan, L. and Shaw, P.** (2006). A distant coilin homologue is required for the formation of cajal bodies in *Arabidopsis*. *Mol. Biol. Cell* **17**, 2942–2951.
- Cougot, N., Babajko, S. and Seraphin, B.** (2004). Cytoplasmic foci are sites of mRNA decay in

- human cells. *J. Cell Biol.* **165**, 31–40.
- Courchaine, E. M., Lu, A. and Neugebauer, K. M.** (2016). Droplet organelles? *EMBO J.* **35**, 1603–1612.
- Cuello, P., Boyd, D. C., Dye, M. J., Proudfoot, N. J. and Murphy, S.** (1999). Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *EMBO J.* **18**, 2867–2877.
- De Conti, L., Baralle, M. and Buratti, E.** (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **4**, 49–60.
- Eckwahl, M. J., Sim, S., Smith, D., Telesnitsky, A. and Wolin, S. L.** (2015). A retrovirus packages nascent host noncoding RNAs from a novel surveillance pathway. *Genes Dev.* **29**, 646–657.
- Eggert, C., Chari, A., Laggenbauer, B. and Fischer, U.** (2006). Spinal muscular atrophy: the RNP connection. *Trends Mol. Med.* **12**, 113–121.
- El Hage, A., Koper, M., Kufel, J. and Tollervey, D.** (2008). Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes Dev.* **22**, 1069–1081.
- Eulalio, A., Behm-Ansmant, I. and Izaurralde, E.** (2007). P bodies: at the crossroads of post-transcriptional pathways. *Nat. Rev. Mol. Cell Biol.* **8**, 9.
- Fernandez-Chamorro, J., Pineiro, D., Gordon, J. M. B., Ramajo, J., Francisco-Velilla, R., Macias, M. J. and Martinez-Salas, E.** (2014). Identification of novel non-canonical RNA-binding sites in Gemin5 involved in internal initiation of translation. *Nucleic Acids Res.* **42**, 5742–5754.
- Fischer, U., Sumpter, V., Sekine, M., Satoh, T. and Luhrmann, R.** (1993). Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *Embo J* **12**, 573–583.
- Ford, E., Strubin, M. and Hernandez, N.** (1998). The Oct-1 POU domain activates snRNA gene transcription by contacting a region in the SNAP(c) largest subunit that bears sequence similarities to the Oct-1 coactivator OBF-1. *Genes Dev.* **12**, 3528–3540.
- Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I. W.** (1997). CRM1 Is an Export Receptor for Leucine-Rich Nuclear Export Signals. *Cell* **90**, 1051–1060.
- Friesen, W. J. and Dreyfuss, G.** (2000). Specific Sequences of the Sm and Sm-like (Lsm) Proteins Mediate Their Interaction with the Spinal Muscular Atrophy Disease Gene Product (SMN). *J. Biol. Chem.* **275**, 26370–26375.
- Friesen, W. J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G. S., Van Duyne, G., Rappsilber, J., Mann, M. and Dreyfuss, G.** (2001a). The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol. Cell. Biol.* **21**, 8289–300.
- Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A. and Dreyfuss, G.** (2001b). SMN, the Product of the Spinal Muscular Atrophy Gene, Binds Preferentially to Dimethylarginine-Containing Protein Targets. *Mol. Cell* **7**, 1111–1117.
- Gillian, A. L. and Svaren, J.** (2004). The Ddx20/DP103 dead box protein represses transcriptional activation by Egr2/Krox-20. *J. Biol. Chem.* **279**, 9056–9063.
- Golembe, T. J., Yong, J. and Dreyfuss, G.** (2005). Specific sequence features, recognized by the SMN complex, identify snRNAs and determine their fate as snRNPs. *Mol. Cell. Biol.* **25**, 10989–11004.
- Goncalves, V., Matos, P. and Jordan, P.** (2009). Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. *Hum. Mol. Genet.* **18**, 3696–3707.

- Gornemann, J., Barrandon, C., Hujer, K., Rutz, B., Rigaut, G., Kotovic, K. M., Faux, C., Neugebauer, K. M. and Seraphin, B.** (2011). Cotranscriptional spliceosome assembly and splicing are independent of the Prp40p WW domain. *RNA* **17**, 2119–2129.
- Gozani, O., Feld, R. and 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. *Genes Dev.* **10**, 233–243.
- Grimm, C., Chari, A., Pelz, J.-P., Kuper, J., Kisker, C., Diederichs, K., Stark, H., Schindelin, H. and Fischer, U.** (2013). Structural basis of assembly chaperone-mediated snRNP formation. *Mol. Cell* **49**, 692–703.
- Grundhoff, A. T., Kremmer, E., Tureci, O., Glieden, A., Gindorf, C., Atz, J., Mueller-Lantsch, N., Schubach, W. H. and Grasser, F. A.** (1999). Characterization of DP103, a novel DEAD box protein that binds to the Epstein-Barr virus nuclear proteins EBNA2 and EBNA3C. *J. Biol. Chem.* **274**, 19136–19144.
- Gubitz, A. K., Mourelatos, Z., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G.** (2002). Gemin5, a novel WD repeat protein component of the SMN complex that binds Sm proteins. *J. Biol. Chem.* **277**, 5631–5636.
- Haas, G., Cetin, S., Messmer, M., Chane-Woon-Ming, B., Terenzi, O., Chicher, J., Kuhn, L., Hammann, P. and Pfeffer, S.** (2016). Identification of factors involved in target RNA-directed microRNA degradation. *Nucleic Acids Res.* **44**, 2873–2887.
- Hallais, M., Pontvianne, F., Andersen, P. R., Clerici, M., Lener, D., Benbahouche, N. E. H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al.** (2013a). CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* **20**, 1358–1366.
- Hallais, M., Pontvianne, F., Andersen, P. R., Clerici, M., Lener, D., Benbahouche, N. E. H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al.** (2013b). CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* **20**, 1358–1366.
- Heo, I., Joo, C., Cho, J., Ha, M., Han, J. and Kim, V. N.** (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* **32**, 276–284.
- Heo, I., Joo, C., Kim, Y.-K., Ha, M., Yoon, M.-J., Cho, J., Yeom, K.-H., Han, J. and Kim, V. N.** (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696–708.
- Heo, I., Ha, M., Lim, J., Yoon, M.-J., Park, J.-E., Kwon, S. C., Chang, H. and Kim, V. N.** (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* **151**, 521–532.
- Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H. and Lührmann, R.** (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J.* **14**, 2076–2088.
- Huang, Q., Jacobson, M. R. and Pederson, T.** (1997). 3' processing of human pre-U2 small nuclear RNA: a base-pairing interaction between the 3' extension of the precursor and an internal region. *Mol. Cell. Biol.* **17**, 7178–7185.
- Huranova, M., Hnilicova, J., Fleischer, B., Cvackova, Z. and Stanek, D.** (2009). A mutation linked to retinitis pigmentosa in HPRP31 causes protein instability and impairs its interactions with

- spliceosomal snRNPs. *Hum. Mol. Genet.* **18**, 2014–2023.
- Ilagan, J. O., Chalkley, R. J., Burlingame, A. L. and Jurica, M. S.** (2013). Rearrangements within human spliceosomes captured after exon ligation. *RNA* **19**, 400–412.
- Ingelfinger, D., Arndt-Jovin, D. J., Luhrmann, R. and Achsel, T.** (2002). The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. *RNA* **8**, 1489–1501.
- Ishikawa, H., Nobe, Y., Izumikawa, K., Yoshikawa, H., Miyazawa, N., Terukina, G., Kurokawa, N., Taoka, M., Yamauchi, Y., Nakayama, H., et al.** (2014). Identification of truncated forms of U1 snRNA reveals a novel RNA degradation pathway during snRNP biogenesis. *Nucleic Acids Res.* **42**, 2708–2724.
- Ishikawa, H., Nobe, Y., Izumikawa, K., Taoka, M., Yamauchi, Y., Nakayama, H., Simpson, R. J., Isobe, T. and Takahashi, N.** (2018). Truncated forms of U2 snRNA (U2-tfs) are shunted toward a novel uridylylation pathway that differs from the degradation pathway for U1-tfs. *RNA Biol.* **15**, 261–268.
- Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, a, McGuigan, C. and Mattaj, I. W.** (1995). A cap-binding protein complex mediating U snRNA export. *Nature* **376**, 709–712.
- Jablonka, S., Holtmann, B., Meister, G., Bandilla, M., Rossoll, W., Fischer, U. and Sendtner, M.** (2002). Gene targeting of Gemin2 in mice reveals a correlation between defects in the biogenesis of U snRNPs and motoneuron cell death. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10126–10131.
- Jady, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E. and Kiss, T.** (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J.* **22**, 1878–1888.
- Kitao, S., Segref, A., Kast, J., Wilm, M., Mattaj, I. W. and Ohno, M.** (2008). A compartmentalized phosphorylation/dephosphorylation system that regulates U snRNA export from the nucleus. *Mol. Cell. Biol.* **28**, 487–497.
- Klingauf, M., Stanek, D. and Neugebauer, K. M.** (2006). Enhancement of U4/U6 small nuclear ribonucleoprotein particle association in Cajal bodies predicted by mathematical modeling. *Mol. Biol. Cell* **17**, 4972–4981.
- Konarska, M. M., Vilardell, J. and Query, C. C.** (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol. Cell* **21**, 543–553.
- Kornblihtt, A. R., Schor, I. E., Allo, M., Dujardin, G., Petrillo, E. and Munoz, M. J.** (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–165.
- Krausova, M. and Stanek, D.** (2018). snRNP proteins in health and disease. *Semin. Cell Dev. Biol.* **79**, 92–102.
- Labno, A., Warkocki, Z., Kulinski, T., Krawczyk, P. S., Bijata, K., Tomecki, R. and Dziembowski, A.** (2016). Perlman syndrome nuclease DIS3L2 controls cytoplasmic non-coding RNAs and provides surveillance pathway for maturing snRNAs. *Nucleic Acids Res.* **44**, 10437–10453.
- Lanfranco, M., Vassallo, N. and Cauchi, R. J.** (2017). Spinal Muscular Atrophy: From Defective Chaperoning of snRNP Assembly to Neuromuscular Dysfunction. *Front. Mol. Biosci.* **4**, 41.
- Lee, K., Pisarska, M. D., Ko, J.-J., Kang, Y., Yoon, S., Ryou, S.-M., Cha, K.-Y. and Bae, J.** (2005). Transcriptional factor FOXL2 interacts with DP103 and induces apoptosis. *Biochem. Biophys.*

Res. Commun. **336**, 876–881.

- Li, W., Lin, W.-D., Ray, P., Lan, P. and Schmidt, W.** (2013). Genome-Wide Detection of Condition-Sensitive Alternative Splicing in Arabidopsis Roots. *Plant Physiol.* **162**, 1750 LP-1763.
- Lin, Y., Protter, D. S. W., Rosen, M. K. and Parker, R.** (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208–219.
- Liu, Q. and Dreyfuss, G.** (1996a). A novel nuclear structure containing the survival of motor neurons protein. *EMBO J.* **15**, 3555–3565.
- Liu, Q. and Dreyfuss, G.** (1996b). A novel nuclear structure containing the survival of motor neurons protein. *EMBO J.* **15**, 3555–3565.
- Liu, J.-L., Wu, Z., Nizami, Z., Deryusheva, S., Rajendra, T. K., Beumer, K. J., Gao, H., Matera, A. G., Carroll, D. and Gall, J. G.** (2009). Coilin is essential for Cajal body organization in *Drosophila melanogaster*. *Mol. Biol. Cell* **20**, 1661–1670.
- Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z. and Lingner, J.** (2008). The Rat1p 5' to 3' Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in *Saccharomyces cerevisiae*. *Mol. Cell* **32**, 465–477.
- Lund, E. and Dahlberg, J. E.** (1992). Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science* **255**, 327–330.
- Madore, S. J., Wieben, E. D. and Pederson, T.** (1984). Intracellular site of U1 small nuclear RNA processing and ribonucleoprotein assembly. *J. Cell Biol.* **98**, 188–192.
- Malinova, A., Cvackova, Z., Mateju, D., Horejsi, Z., Abeza, C., Vandermoere, F., Bertrand, E., Stanek, D. and Verheggen, C.** (2017). Assembly of the U5 snRNP component PRPF8 is controlled by the HSP90/R2TP chaperones. *J. Cell Biol.* **216**, 1579–1596.
- Maraia, R. J., Driscoll, C. T., Bilyeu, T., Hsu, K. and Darlington, G. J.** (1993). Multiple dispersed loci produce small cytoplasmic Alu RNA. *Mol. Cell. Biol.* **13**, 4233–4241.
- Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I. W. and Dreyfuss, G.** (2002). The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Mol. Cell. Biol.* **22**, 6533–6541.
- Masuyama, K., Taniguchi, I., Kataoka, N. and Ohno, M.** (2004). RNA length defines RNA export pathway. *Genes Dev.* **18**, 2074–2085.
- Matera, A. G.** (1999). RNA splicing: more clues from spinal muscular atrophy. *Curr. Biol.* **9**, R140-2.
- Matera, A. G. and Wang, Z.** (2014a). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–21.
- Matera, A. G. and Wang, Z.** (2014b). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121.
- Matera, A. G. and Ward, D. C.** (1993). Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. *J. Cell Biol.* **121**, 715–727.
- Meister, G., Bühler, D., Lagerbauer, B., Zobawa, M., Lottspeich, F. and Fischer, U.** (2000). Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins. *Hum. Mol. Genet.* **9**, 1977–86.
- Meister, G., Eggert, C., Bühler, D., Brahms, H., Kambach, C. and Fischer, U.** (2001). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.* **11**, 1990–1994.

- Minasaki, R., Puoti, A. and Streit, A.** (2009). The DEAD-box protein MEL-46 is required in the germ line of the nematode *Caenorhabditis elegans*. *BMC Dev. Biol.* **9**, 35.
- Morris, D. P. and Greenleaf, A. L.** (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **275**, 39935–39943.
- Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A. G., Bertrand, E., Tazi, J. and Bordonne, R.** (2003). Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep.* **4**, 616–622.
- Mouillet, J.-F., Yan, X., Ou, Q., Jin, L., Muglia, L. J., Crawford, P. A. and Sadovsky, Y.** (2008). DEAD-box protein-103 (DP103, Ddx20) is essential for early embryonic development and modulates ovarian morphology and function. *Endocrinology* **149**, 2168–2175.
- Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J. A.** (1983). The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell* **33**, 509–518.
- Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G.** (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728.
- Mullen, T. E. and Marzluff, W. F.** (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev.* **22**, 50–65.
- Narayanan, U., Ospina, J. K., Frey, M. R., Hebert, M. D. and Matera, A. G.** (2002). SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. *Hum. Mol. Genet.* **11**, 1785–1795.
- Narayanan, U., Achsel, T., L??hrmann, R. and Matera, A. G.** (2004). Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol. Cell* **16**, 223–234.
- Natalizio, A. H. and Matera, A. G.** (2013). Identification and characterization of *Drosophila* Snurportin reveals a role for the import receptor Moleskin/importin-7 in snRNP biogenesis. *Mol. Biol. Cell* **24**, 2932–2942.
- Nesic, D., Tanackovic, G. and Kramer, A.** (2004). A role for Cajal bodies in the final steps of U2 snRNP biogenesis. *J. Cell Sci.* **117**, 4423–4433.
- Nissan, T., Rajyaguru, P., She, M., Song, H. and Parker, R.** (2010). Decapping Activators in *Saccharomyces cerevisiae* Act by Multiple Mechanisms. *Mol. Cell* **39**, 773–783.
- Novotny, I., Blazikova, M., Stanek, D., Herman, P. and Malinsky, J.** (2011). In vivo kinetics of U4/U6.U5 tri-snRNP formation in Cajal bodies. *Mol. Biol. Cell* **22**, 513–523.
- Novotny, I., Malinova, A., Stejskalova, E., Mateju, D., Klimesova, K., Roithova, A., Sveda, M., Knejzlik, Z. and Stanek, D.** (2015). SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. *Cell Rep.*
- Ohno, M.** (2012). Size matters in RNA export. *RNA Biol.* **9**, 1413–1417.
- Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I. W.** (2000a). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187–198.
- Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I. W.** (2000b). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187–198.
- Ou, Q., Mouillet, J. F., Yan, X., Dorn, C., Crawford, P. A. and Sadovsky, Y.** (2001). The DEAD box protein DP103 is a regulator of steroidogenic factor-1. *Mol. Endocrinol.* **15**, 69–79.
- Pacheco, A., Lopez de Quinto, S., Ramajo, J., Fernandez, N. and Martinez-Salas, E.** (2009). A

- novel role for Gemin5 in mRNA translation. *Nucleic Acids Res.* **37**, 582–590.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. and Blencowe, B. J.** (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415.
- Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoykov, S., Mahamid, J., Saha, S., Franzmann, T. M., et al.** (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077.
- Paushkin, S., Charroux, B., Abel, L., Perkinson, R. A., Pellizzoni, L. and Dreyfuss, G.** (2000). The survival motor neuron protein of *Schizosaccharomyces pombe*. Conservation of survival motor neuron interaction domains in divergent organisms. *J. Biol. Chem.* **275**, 23841–23846.
- Paushkin, S., Gubitz, A. K., Massenet, S. and Dreyfuss, G.** (2002). The SMN complex, an assemblysome of ribonucleoproteins. *Curr. Opin. Cell Biol.* **14**, 305–312.
- Pek, J. W., Anand, A. and Kai, T.** (2012). Tudor domain proteins in development. *Development* **139**, 2255–2266.
- Pellizzoni, L.** (2007). Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Rep.* **8**, 340 LP-345.
- Pellizzoni, L., Kataoka, N., Charroux, B. and Dreyfuss, G.** (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* **95**, 615–624.
- Pellizzoni, L., Charroux, B. and Dreyfuss, G.** (1999). SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11167–11172.
- Pellizzoni, L., Baccon, J., Rappsilber, J., Mann, M. and Dreyfuss, G.** (2002). Purification of native survival of motor neurons complexes and identification of Gemin6 as a novel component. *J. Biol. Chem.* **277**, 7540–7545.
- Peterlin, B. M., Brogie, J. E. and Price, D. H.** (2012). 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdiscip. Rev. RNA* **3**, 92–103.
- Pillai, R. S., Grimmier, M., Meister, G., Will, C. L., Luhrmann, R., Fischer, U. and Schumperli, D.** (2003). Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes Dev.* **17**, 2321–2333.
- Plessel, G., Fischer, U. and Luhrmann, R.** (1994). m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol. Cell. Biol.* **14**, 4160–4172.
- Prusty, A. B., Meduri, R., Prusty, B. K., Vanselow, J., Schlosser, A. and Fischer, U.** (2017). Impaired spliceosomal UsnRNP assembly leads to Sm mRNA down-regulation and Sm protein degradation. *J. Cell Biol.* **216**, 2391–2407.
- Reed, R. and Maniatis, T.** (1986). A role for exon sequences and splice-site proximity in splice-site selection. *Cell* **46**, 681–690.
- Robberson, B. L., Cote, G. J. and Berget, S. M.** (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94.
- Sadowski, C. L., Henry, R. W., Lobo, S. M. and Hernandez, N.** (1993). Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. *Genes Dev.* **7**, 1535–1548.

- Sadowski, C. L., Henry, R. W., Kobayashi, R. and Hernandez, N.** (1996). The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 4289–4293.
- Sanchez, G., Bondy-Chorney, E., Laframboise, J., Paris, G., Didillon, A., Jasmin, B. J. and Cote, J.** (2016). A novel role for CARM1 in promoting nonsense-mediated mRNA decay: potential implications for spinal muscular atrophy. *Nucleic Acids Res.* **44**, 2661–2676.
- Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T. and Lührmann, R.** (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J.* **23**, 3000–3009.
- Schmidt, M.-J., West, S. and Norbury, C. J.** (2011). The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA* **17**, 39–44.
- Schwer, B. and Gross, C. H.** (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* **17**, 2086–2094.
- Scott, D. D. and Norbury, C. J.** (2013). RNA decay via 3' uridylation. *Biochim. Biophys. Acta* **1829**, 654–665.
- Shanbhag, R., Kurabi, A., Kwan, J. J. and Donaldson, L. W.** (2010). Solution structure of the carboxy-terminal Tudor domain from human Coilin. *FEBS Lett.* **584**, 4351–4356.
- Shao, W., Kim, H.-S., Cao, Y., Xu, Y.-Z. and Query, C. C.** (2012). A U1-U2 snRNP interaction network during intron definition. *Mol. Cell. Biol.* **32**, 470–478.
- Sheth, U. and Parker, R.** (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**, 805–808.
- Shukla, S. and Parker, R.** (2014). Quality control of assembly-defective U1 snRNAs by decapping and 5'-to-3' exonucleolytic digestion. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3277–E3286.
- Sleeman, J. E. and Lamond, A. I.** (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.* **9**, 1065–1074.
- Spellman, R. and Smith, C. W. J.** (2006). Novel modes of splicing repression by PTB. *Trends Biochem. Sci.* **31**, 73–76.
- Stanek, D., Rader, S. D., Klingauf, M. and Neugebauer, K. M.** (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J. Cell Biol.* **160**, 505–516.
- STEIGER, M., CARR-SCHMID, A., SCHWARTZ, D. C., KILEDJIAN, M. and PARKER, R. O. Y.** (2003). Analysis of recombinant yeast decapping enzyme. *RNA* **9**, 231–238.
- Stejskalova, E. and Stanek, D.** (2014). The splicing factor U1-70K interacts with the SMN complex and is required for nuclear gem integrity. *J. Cell Sci.* **127**, 3909–3915.
- Strzelecka, M., Oates, A. C. and Neugebauer, K. M.** (2010). Dynamic control of Cajal body number during zebrafish embryogenesis. *Nucleus* **1**, 96–108.
- Sugnet, C. W., Kent, W. J., Ares, M. J. and Haussler, D.** (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.* 66–77.
- Sun, J. S. and Manley, J. L.** (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* **9**, 843–854.
- Sun, J., Xu, H., Subramony, S. H. and Hebert, M. D.** (2005). Interactions between Coilin and PIASy partially link Cajal bodies to PML bodies. *J. Cell Sci.* **118**, 4995 LP-5003.

- Suzuki, T., Izumi, H. and Ohno, M.** (2010). Cajal body surveillance of U snRNA export complex assembly. *J. Cell Biol.* **190**, 603–612.
- Takizawa, Y., Qing, Y., Takaku, M., Ishida, T., Morozumi, Y., Tsujita, T., Kogame, T., Hirota, K., Takahashi, M., Shibata, T., et al.** (2010). GEMIN2 promotes accumulation of RAD51 at double-strand breaks in homologous recombination. *Nucleic Acids Res.* **38**, 5059–5074.
- Tanackovic, G. and Kramer, A.** (2005). Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. *Mol. Biol. Cell* **16**, 1366–1377.
- Tang, X., Bharath, S. R., Piao, S., Tan, V. Q., Bowler, M. W. and Song, H.** (2016). Structural basis for specific recognition of pre-snRNA by Gemin5. *Cell Res.* **26**, 1353–1356.
- Tange, T. Ø., Damgaard, C. K., Guth, S., Valcárcel, J. and Kjems, J.** (2001). The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* **20**, 5748 LP-5758.
- Tazi, J., Forne, T., Jeanteur, P., Cathala, G. and Brunel, C.** (1993). Mammalian U6 small nuclear RNA undergoes 3' end modifications within the spliceosome. *Mol. Cell. Biol.* **13**, 1641–1650.
- Teixeira, D., Sheth, U., Valencia-Sanchez, M. A., Brengues, M. and Parker, R.** (2005). Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA* **11**, 371–382.
- Totaro, A., Renzi, F., La Fata, G., Mattioli, C., Raabe, M., Urlaub, H. and Achsel, T.** (2011). The human Pat1b protein: a novel mRNA deadenylation factor identified by a new immunoprecipitation technique. *Nucleic Acids Res.* **39**, 635–647.
- Toyota, C. G., Davis, M. D., Cosman, A. M. and Hebert, M. D.** (2010). Coilin phosphorylation mediates interaction with SMN and SmB'. *Chromosoma* **119**, 205–215.
- Tritschler, F., Huntzinger, E. and Izaurralde, E.** (2010). Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of déjà vu. *Nat. Rev. Mol. Cell Biol.* **11**, 379–384.
- Tucker, K. E., Berciano, M. T., Jacobs, E. Y., LePage, D. F., Shpargel, K. B., Rossire, J. J., Chan, E. K., Lafarga, M., Conlon, R. A. and Matera, A. G.** (2001). Residual Cajal bodies in coilin knockout mice fail to recruit Sm snRNPs and SMN, the spinal muscular atrophy gene product. *J. Cell Biol.* **154**, 293–307.
- Urlaub, H., Raker, V. A., Kostka, S. and Lührmann, R.** (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* **20**, 187–196.
- Ustianenko, D., Pasulka, J., Feketova, Z., Bednarik, L., Zigackova, D., Fortova, A., Zavolan, M. and Vanacova, S.** (2016). TUT-DIS3L2 is a mammalian surveillance pathway for aberrant structured non-coding RNAs. *EMBO J.* **35**, 2179–2191.
- Valcarcel, J., Gaur, R. K., Singh, R. and Green, M. R.** (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273**, 1706–1709.
- Vincent, H. A. and Deutscher, M. P.** (2009). Insights Into How RNase R Degrades Structured RNA: Analysis of the Nuclease Domain. *J. Mol. Biol.* **387**, 570–583.
- Wahl, M. C. and Fischer, U.** (2016). The right pick: structural basis of snRNA selection by Gemin5. *Genes Dev.* **30**, 2341–2344.
- Walker, M. P., Tian, L. and Matera, A. G.** (2009). Reduced viability, fertility and fecundity in mice lacking the cajal body marker protein, coilin. *PLoS One* **4**, e6171.
- Wang, M. and Pestov, D. G.** (2011). 5'-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay. *Nucleic Acids Res.* **39**, 1811–1822.
- Will, C. L. and Lührmann, R.** (2011). Spliceosome structure and function. *Cold Spring Harb.*

Perspect. Biol. **3**,

- Xu, H., Pillai, R. S., Azzouz, T. N., Shpargel, K. B., Kambach, C., Hebert, M. D., Sch??mperli, D. and Matera, A. G.** (2005). The C-terminal domain of coilin interacts with Sm proteins and U snRNPs. *Chromosoma* **114**, 155–166.
- Xu, C., Ishikawa, H., Izumikawa, K., Li, L., He, H., Nobe, Y., Yamauchi, Y., Shahjee, H. M., Wu, X.-H., Yu, Y., et al.** (2016). Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly. *Genes Dev.* **30**, 2376–2390.
- Yan, X., Mouillet, J.-F., Ou, Q. and Sadovsky, Y.** (2003). A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity. *Mol. Cell. Biol.* **23**, 414–423.
- Yang, Z., Jakymiw, A., Wood, M. R., Eystathioy, T., Rubin, R. L., Fritzler, M. J. and Chan, E. K. L.** (2004). GW182 is critical for the stability of GW bodies expressed during the cell cycle and cell proliferation. *J. Cell Sci.* **117**, 5567–5578.
- Yong, J., Pellizzoni, L. and Dreyfuss, G.** (2002a). Sequence-specific interaction of U1 snRNA with the SMN complex. *EMBO J.* **21**, 1188–1196.
- Yong, J., Pellizzoni, L. and Dreyfuss, G.** (2002b). Sequence-specific interaction of U1 snRNA with the SMN complex. *EMBO J.* **21**, 1188–1196.
- Yong, J., Golembe, T. J., Battle, D. J., Pellizzoni, L. and Dreyfuss, G.** (2004). snRNAs contain specific SMN-binding domains that are essential for snRNP assembly. *Mol. Cell. Biol.* **24**, 2747–2756.
- Yuo, C. Y., Ares, M. J. and Weiner, A. M.** (1985). Sequences required for 3' end formation of human U2 small nuclear RNA. *Cell* **42**, 193–202.
- Zhang, D., Abovich, N. and Rosbash, M.** (2001). A biochemical function for the Sm complex. *Mol. Cell* **7**, 319–329.
- Zhang, R., So, B. R., Li, P., Yong, J., Glisovic, T., Wan, L. and Dreyfuss, G.** (2011). Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* **146**, 384–395.
- Zhong, X.-Y., Wang, P., Han, J., Rosenfeld, M. G. and Fu, X.-D.** (2009). SR Proteins in Vertical Integration of Gene Expression from Transcription to RNA Processing to Translation. *Mol. Cell* **35**, 1–10.
- Zhou, L., Zhou, Y., Hang, J., Wan, R., Lu, G., Yan, C. and Shi, Y.** (2014a). Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex. *Cell Res.* **24**, 497–500.
- Zhou, L., Zhou, Y., Hang, J., Wan, R., Lu, G., Yan, C. and Shi, Y.** (2014b). Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex. *Cell Res.* **24**, 497–500.