



# **Insights into the regulation of RNA helicases by protein cofactors**

Dissertation

for the award of the degree

“Doctor of Philosophy”

Division of Mathematics and Natural Sciences

of the Georg-August-Universität Göttingen

within the doctoral program Molecular Biology

of the Georg-August University School of Science (GAUSS)

submitted by

**Indira Memet**

from Medgidia, Romania

Göttingen 2018

## **Thesis Committee**

Prof. Dr. Markus Bohnsack	Department of Molecular Biology University Medical Center Göttingen
Prof. Dr. Marina Rodnina	Department of Physical Biochemistry Max Planck Institute for Biophysical Chemistry Göttingen
Prof. Dr. Peter Rehling	Department of Cellular Biochemistry University Medical Center Göttingen

## **Members of the Examination Board**

Referee: Prof. Dr. Markus Bohnsack	Department of Molecular Biology University Medical Center Göttingen
2 <sup>nd</sup> Referee: Prof. Dr. Marina Rodnina	Department of Physical Biochemistry Max Planck Institute for Biophysical Chemistry Göttingen

## **Further members of the Examination Board**

Prof. Dr. Peter Rehling	Department of Cellular Biochemistry University Medical Center Göttingen
Dr. Ricarda Richter-Dennerlein	Department of Cellular Biochemistry University Medical Center Göttingen
Prof. Dr. Silvio Rizzoli	Department of Neuro- and Sensory Physiology University Medical Center Göttingen
Prof. Dr. Jörg Stülke	Department of General Microbiology University of Göttingen

Date of oral examination: 05.02.2019

## **Affidavit**

Herewith I declare that I prepared the Ph.D. dissertation "Insights into the regulation of RNA helicases by protein cofactors" on my own and with no other sources and aids than quoted.

Göttingen, 30.11.2018

Indira Memet

---

## **Publications associated with this dissertation**

Memet, I., Doebele, C., Sloan, K.E., and Bohnsack, M.T. (2017). The G-patch protein NF-kappaB-repressing factor mediates the recruitment of the exonuclease XRN2 and activation of the RNA helicase DHX15 in human ribosome biogenesis. *Nucleic Acids Res.* 45, 5359–5374.

## Table of contents

Summary .....	viii
<b>1. INTRODUCTION</b>	
1.1 General aspects of RNA and RNA-binding proteins .....	1
1.2 RNA helicases: characteristics and mode of action .....	2
1.2.1 Classification and conserved sequence elements .....	2
1.2.2 Structure and mechanism of action .....	4
1.2.3 Biochemical activities .....	6
1.3 Cellular pathways involving RNA helicases .....	7
1.3.1 Ribosome biogenesis .....	7
1.3.2 Pre-mRNA splicing .....	11
1.3.3 Other pathways .....	13
1.4 Regulation of RNA helicases .....	14
1.4.1 General regulatory mechanisms .....	14
1.4.2 Protein cofactor-independent regulation .....	14
1.4.3 Regulation by protein cofactors .....	15
1.5 G-patch proteins as regulators of DEAH/RHA helicases .....	16
1.5.1 General characteristics and regulatory mechanisms .....	16
1.5.2 Human G-patch proteins .....	18
1.5.3 G-patch proteins and RNA helicases in disease .....	21
1.6 Scope of the thesis .....	23
<b>2. MATERIALS AND METHODS</b>	
2.1 Materials .....	24
2.1.1 Chemicals and buffers .....	24
2.1.2 Plasmids .....	26
2.1.3 Small interfering (si)RNAs .....	27
2.1.4 RNA oligonucleotides .....	28
2.1.5 Northern blotting probes .....	28
2.1.6 Antibodies .....	29



2.2 Standard molecular biology methods .....	30
2.2.1 Molecular cloning .....	30
2.2.2 Site-directed mutagenesis .....	31
2.2.3 Transformation of <i>E. coli</i> .....	31
2.2.4 Plasmid DNA extraction .....	32
2.3 Cell culture-based methods .....	32
2.3.1 Cell culture growth conditions .....	32
2.3.2 Generation of inducible stable cell lines .....	32
2.3.3 Gene knockdown with siRNAs .....	33
2.3.4 Immunofluorescence microscopy .....	33
2.4 Protein analysis methods .....	34
2.4.1 SDS-PAGE and western blotting .....	34
2.4.2 Immunoprecipitation (IP) of protein complexes .....	34
2.4.3 Purification of nucleoli and preparation of nucleolar lysates .....	35
2.4.4 Liquid chromatography tandem mass spectrometry (LC-MS/MS) .....	36
2.4.5 Separation of (pre)-ribosomal complexes on sucrose gradients .....	36
2.5 RNA analysis methods .....	37
2.5.1 Total RNA extraction .....	37
2.5.2 Northern blotting .....	37
2.5.3 Pulse-chase analysis of RNA .....	38
2.5.4 Quantitative and gel-based reverse transcription PCR (RT-PCR) .....	38
2.5.5 Site-specific RNase H cleavage assay .....	39
2.6 Next-generation sequencing-based methods and data analysis .....	40
2.6.1 RNA-seq sample preparation .....	40
2.6.2 Crosslinked RNA immunoprecipitation (RIP)-seq .....	40
2.6.3 Crosslinking and analysis of cDNA (CRAC) .....	41
2.6.4 Next-generation sequencing data analysis .....	43
2.7 <i>In vitro</i> methods .....	44
2.7.1 Recombinant protein expression and purification .....	44

2.7.2 Steady-state ATPase assay .....	45
2.7.3 Fluorescence anisotropy .....	46
2.7.4 Unwinding assay .....	47
<b>3. RESULTS</b>	
3.1 Identification and characterization of G-patch protein-RNA helicase interactions.....	48
3.1.1 Human G-patch proteins interact with a subset of DEAH/RHA helicases.....	48
3.1.2 The G-patch domain is sufficient in most cases for the regulation of the cognate RNA helicase.....	52
3.1.3 The RNA binding affinity of DHX15 is enhanced by the G-patch domains of its cofactors .....	55
3.2 Functional characterization of DHX15 and its G-patch cofactors .....	57
3.2.1 The majority of human G-patch proteins are localized in the nucleoplasm.....	57
3.2.2 Genome-wide analyses reveal that DHX15 and most of its G-patch cofactors regulate alternative splicing.....	59
3.2.3 DHX15 and its G-patch cofactors regulate splicing either directly or indirectly ...	67
3.2.4 DHX15 and the G-patch protein NKRF function together in ribosome biogenesis .....	72
3.2.4.1 DHX15 and NKRF form a nucleolar complex with XRN2.....	72
3.2.4.2 NKRF crosslinks to the pre-rRNA transcript at several sites.....	77
3.2.4.3 DHX15, NKRF and XRN2 are required for efficient pre-rRNA processing..	79
3.2.4.4 NKRF regulates the function of XRN2 in the degradation of pre-rRNA processing by-products .....	81
3.2.4.5 NKRF stimulates the activity of DHX15 for efficient processing of the pre-rRNA transcript at a specific site.....	82
<b>4. DISCUSSION</b>	
4.1 G-patch protein-RNA helicase interactions in human cells .....	85
4.2 Mechanism of RNA helicase regulation by G-patch proteins .....	89
4.3 Regulation of alternative splicing by DHX15 and G-patch proteins.....	93
4.4 The function of DHX15 and NKRF in ribosome biogenesis .....	98
4.5 Conclusions and perspectives.....	101

References .....	103
Supplementary information .....	124
List of Figures .....	129
List of Tables .....	131
Abbreviations.....	132
Acknowledgements .....	134
Curriculum vitae .....	135

## Summary

RNA helicases comprise a large family of ubiquitously expressed enzymes that remodel RNA structures and RNA-protein complexes in an NTP-dependent manner. These proteins are essential regulators of every RNA-related process, including pre-mRNA splicing and ribosome biogenesis, where they are suggested to perform various activities, such as unwinding RNA duplexes and displacing proteins from RNA. RNA helicases require a complex regulation due to the lack of specificity of their conserved helicase core, their generally low intrinsic activity and the involvement of individual helicases in multiple cellular functions. The activity of these enzymes can be modulated in diverse ways, including through interactions with effector proteins termed helicase cofactors. Several helicase cofactors identified so far share a conserved glycine-rich domain known as a G-patch domain. In yeast, the role of these G-patch proteins as RNA helicase regulators has been established. However, a larger number of G-patch proteins are expressed in human cells and much less is known about their functions and interactions with RNA helicases.

In this study, a comprehensive analysis of the human G-patch protein family was performed in order to assess their functions and potential role as cofactors of RNA helicases. Our results show that all 22 human G-patch proteins interact with an RNA helicase and, in most cases, they enhance the RNA binding affinity and/or the ATPase activity of their helicase partner. Only three human DEAH/RHA helicases associate with G-patch proteins, among which DHX15 is regulated by approximately 20 G-patch cofactors, suggesting that this helicase could have a central role in the cell. Our transcriptome analyses indicate that DHX15 and the majority of G-patch proteins are involved in alternative splicing, where they regulate specific genes but also have common targets. We confirm the role in alternative splicing for several proteins and propose that DHX15 functions together with its G-patch cofactors in this pathway. Our results further suggest that G-patch proteins and DHX15 could potentially regulate alternative splicing by direct binding to pre-mRNAs as well as in an indirect manner. In addition to its role in splicing, DHX15 also functions in ribosome biogenesis together with its G-patch cofactor NKRF and with the exonuclease XRN2. These proteins associate into a nucleolar complex that is involved in the efficient processing of the pre-rRNA transcript at a specific site. The catalytic activity of DHX15 is required for this function, implying that the helicase might remodel structures in this region to facilitate the cleavage event. Taken together, our data provide essential insights into the role of human G-patch proteins as cofactors of RNA helicases and also reveal functions for DHX15 in alternative splicing and ribosome biogenesis together with its G-patch cofactors. Therefore, the findings of this study provide the basis for further understanding the function and regulation of RNA helicases.

## 1. INTRODUCTION

### 1.1 General aspects of RNA and RNA-binding proteins

RNA is a highly versatile molecule that plays a central role in almost every cellular process and is capable of a wide range of functions, such as mediating the transfer of genetic information from DNA to proteins, catalyzing biochemical reactions and regulating gene expression at the transcriptional, post-transcriptional and translational levels (Cech, 2012). A typical cell contains a multitude of RNA molecules that can be broadly classified into protein-coding RNA (or messenger RNA) and non-coding RNA, with the latter including, among others, ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and the related small Cajal body-associated RNA (scaRNA) (Cech and Steitz, 2014; Morris and Mattick, 2014). Messenger RNAs (mRNAs) are intermediaries in the gene expression pathway that transmit information from genes to proteins and their splicing requires the action of snRNAs, which are essential components of the spliceosome. Protein synthesis is catalyzed by rRNA in the context of the ribosome and involves decoding of the three-nucleotide code of the mRNA sequence into the corresponding amino acids, which is accomplished with the help of tRNAs. Ribose methylation at 2'-OH groups and the isomerization of uridine into pseudouridine in rRNAs and snRNAs are guided by snoRNAs and scaRNAs, which act together with proteins that install these modifications.

The functions of RNAs are brought about by their ability to fold into unique and complex structures, which involve local secondary structure elements that contribute to higher-order tertiary arrangements (Russell, 2008). In addition, many non-coding RNAs establish base-pairing interactions with other RNAs to carry out their activity. This includes, for example, the association of snRNAs with pre-mRNAs in splicing and of snoRNAs and scaRNAs with their target RNAs for directing modifications. However, RNA molecules are rarely found alone in the cell but instead they are bound by proteins to form RNA-protein (RNP) complexes. These RNA-binding proteins (RBPs) are suggested to regulate the fate or function of RNAs by stabilizing or remodeling their structure, mediating interactions with other macromolecules, assisting in their transport or installing modifications (Cusack, 1999; Hentze et al., 2018). In general, RBPs recognize either sequence motifs, structural elements or both and bind short stretches (3-8 nucleotides) of RNA that have a low sequence complexity (Jankowsky and Harris, 2015; Dominguez et al., 2018). Classical RBPs interact with RNA through established RNA-binding domains, such as the RNA recognition motif (RRM) or the helicase core. In addition, a multitude of RBPs that use intrinsically disordered regions, protein-protein interaction interfaces and other unconventional modes to bind RNA have been recently described (Castello et al., 2016).

The structure and composition of RNPs are dynamically regulated throughout their lifetime and this has been suggested to take place mainly through the action of RNA helicases (Jarmoskaite and Russell, 2014; Hentze et al., 2018). The role of RNA helicases in RNP remodeling as well as their other biochemical activities, mode of action and regulation are discussed in detail below.

## **1.2 RNA helicases: characteristics and mode of action**

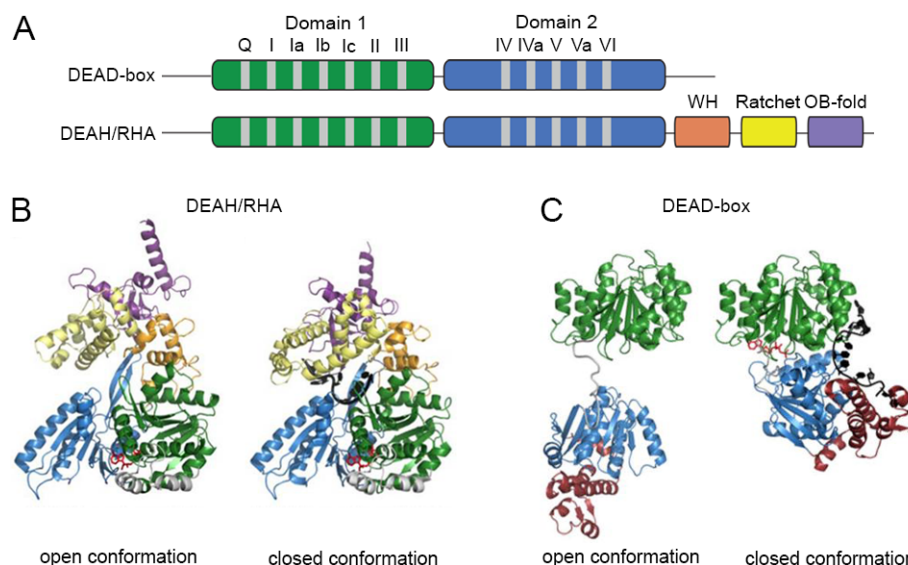
RNA helicases comprise a large family of ubiquitously expressed RBPs that are involved in every aspect of RNA metabolism through their function in remodeling RNA-RNA and RNA-protein interactions using the energy of nucleoside triphosphate (NTP) hydrolysis. Originally designated as helicases based on the ability of some members of the family to unwind duplex structures, it is now widely accepted that RNA helicases display a broad range of activities and mechanisms of action, with their common characteristic being that they possess RNA-dependent NTPase activity (Jarmoskaite and Russell, 2014; Ozgur et al., 2015).

### **1.2.1 Classification and conserved sequence elements**

RNA helicases and the related DNA helicases have been classified into six superfamilies (SF1-SF6) based on primary sequence and structural and functional analyses (Gorbalenya and Koonin, 1993; Singleton et al., 2007). Most helicases of SF1 and SF2 act as monomers, while SF3-SF6 helicases associate into hexameric ring-like structures. RNA helicases belong almost exclusively to SF1 and SF2 and are further subdivided into five SF2 families (DEAD-box, DEAH/RHA, RIG-I-like, Ski2-like, NS3/NPH-II) and the Upf1-like family that is part of SF1 (Fairman-Williams et al., 2010; Byrd and Raney, 2012). More than 70 RNA helicases have been identified in human cells and the majority are either DEAD-box or DEAH/RHA proteins. Consequently, much of our mechanistic understanding of RNA helicases stems from studies of these two families.

The hallmark of SF1 and SF2 helicases is the presence of a conserved helicase core consisting of two globular domains connected by a flexible linker, which are designated as RecA-like domains based on homology to bacterial RecA (Bleichert and Baserga, 2007; Jankowsky and Fairman, 2007). Within the two RecA-like domains, a series of conserved sequence motifs have been described, with roles in RNA and NTP substrate binding, NTP hydrolysis and its coordination with remodeling events (Figure 1.1A). The Walker A motif (motif I) containing the characteristic glycine-lysine-threonine (GKT) sequence, the Walker B motif (motif II) with the aspartate-glutamate-alanine-glutamate/histidine (DEAD/H) signature that gives the name of the respective families, and motif VI, which contains an

essential ‘arginine finger’, are highly conserved and participate in NTP binding and hydrolysis (Walker et al., 1982; Caruthers and McKay, 2002). The Q motif confers specificity for ATP through interactions between the glutamine and the adenine base and is absent from certain families, such as the DEAH/RHA helicases, that are able to bind and hydrolyze other NTPs (Cordin et al., 2004). Other conserved regions include motifs implicated in the coordination of the RNA and NTP binding sites (III and Va) and motifs required for binding of the RNA substrate (Ia, Ib, Ic, IV, IVa, V) (Cordin et al., 2006; Banroques et al., 2010; Fitzgerald et al., 2017).



**Figure 1.1. Structural comparison of DEAD-box and DEAH/RHA RNA helicases.** (A) Schematic representation of the primary sequence of DEAD-box and DEAH/RHA helicases depicting the two domains of the conserved core as well as the winged helix (WH), ratchet-like (Ratchet) and OB-fold domains that are specific for the DEAH/RHA family. Each core domain contains several conserved motifs that are colored in gray and labelled above. (B) Crystal structures of the DEAH/RHA helicase Prp43 in the open (left) and closed (right) conformations based on the PDB entries 5LTK and 5LTA. The different domains are colored as in (A) and the bound nucleotide is shown in red. The closed conformation also contains the RNA substrate, which is depicted in black. (C) Structures of the DEAD-box helicase Mss116 in the open (left) and closed (right) conformations. The open conformation is based on small-angle X-ray scattering data, while the closed conformation is obtained from the PDB entry 3I5X. The two helicase domains are colored as in (A) and the C-terminal extension that is found only in some DEAD-box proteins is shown in magenta. The nucleotide and the RNA substrate, which are only present in the closed conformation, are colored in red and black respectively. This figure was adapted from Gilman et al., 2017.

The conserved core of RNA helicases is flanked in most cases by N-terminal and C-terminal auxiliary regions, which are in general highly variable between, and also within, the different families and can encompass specific domains or folds (Fairman-Williams et al., 2010). In DEAD-box proteins, the N- and C-terminal extensions adopt a variety of conformations and are not conserved between the different helicases. For example, the bacterial helicase DbpA contains an RRM domain at its C-terminus, while the yeast helicase Mss116 has a

positively charged C-tail, and both features are important for their regulation. In some cases, DEAD-box proteins can be represented exclusively by the helicase core as is the case for eIF4A, which is defined as a minimal helicase (Rudolph and Klostermeier, 2015). On the other hand, while the N-terminal extension of DEAH/RHA helicases is specific for each individual protein, this family is characterized by the presence of a highly-conserved C-terminus that has an essential role in their function and regulation and is composed of a winged helix (WH) domain, a ratchet-like domain and an oligonucleotide/oligosaccharide-binding (OB)-fold (He et al., 2010; Walbott et al., 2010). Interestingly, OB-folds are versatile domains composed of a five-stranded  $\beta$ -barrel structure that are found in a wide range of proteins where they are involved in binding nucleic acids, proteins or other molecules (Arcus, 2002; Theobald et al., 2003). In general, it has been suggested that the auxiliary domains regulate the activity of RNA helicases by different mechanisms that involve interactions with proteins or RNA and some of these regulatory mechanisms are discussed in the following sections.

### 1.2.2 Structure and mechanism of action

In three-dimensional structure, the two helicase domains form a cleft where the motifs involved in NTP binding and hydrolysis cluster, while the RNA-binding motifs are located on the surface of the two domains, opposite the NTP binding site (Jankowsky and Fairman, 2007; Pyle, 2008). Structural and biochemical studies of RNA helicases in different functional conformations have provided insight into how these enzymes couple NTP binding and hydrolysis to RNA unwinding and have revealed major differences in the mode of action of DEAD-box and DEAH/RHA helicases (Yang et al., 2007; Walbott et al., 2010; Mallam et al., 2012; He et al., 2017; Tauchert et al., 2017).

DEAH/RHA helicases are generally defined as processive enzymes that unwind duplexes by directional translocation powered by cycles of NTP hydrolysis. These helicases load onto single-stranded RNA overhangs adjacent to a duplex region and have been suggested to translocate preferentially in the 3'-5' direction, leading to the displacement of the complementary strand (Pyle, 2008). An important feature of DEAH/RHA helicases is the presence of a highly-conserved C-terminal region consisting of a WH domain, a ratchet-like domain and an OB-fold (He et al., 2010; Walbott et al., 2010). These C-terminal domains establish essential interactions with both RecA-like domains that contribute to the formation of an RNA-binding channel located inside the core. Binding of NTP triggers rearrangements of the C-terminal region, which leads to an open conformation that allows access of the RNA substrate to the binding channel (Tauchert et al., 2017; Figure 1.1B). The RNA is bound in a specific configuration with its 5' end located in domain 2 and its 3' end contacting domain 1, thus providing a basis for the 3'-5' polarity exhibited by these enzymes (Pyle,



2008; Tauchert et al., 2017). Only four RNA bases are stacked in the helicase binding channel and relatively few contacts are formed, involving almost exclusively the sugar-phosphate backbone of the nucleic acid, which suggests a transient and unspecific interaction of these helicases with their substrate. RNA binding triggers further rearrangements of the helicase core, mostly in domain 2, and this leads to a closed conformation that is favorable for NTP hydrolysis (He et al., 2017; Tauchert et al., 2017).

Translocation along the RNA substrate is driven by changes in the interactions between the RNA and the two helicase core domains during the NTP hydrolysis cycle. More specifically, two  $\beta$ -hairpin regions in each helicase domain contact the RNA, acting as 'bookends' for a region of four nucleotides (nt), and their movement relative to each other by a stepping mechanism leads to translocation by one nucleotide (He et al., 2017). How the actual unwinding takes place is still not fully understood, but it has been suggested to involve a combination of active disruption of the duplex during translocation and passive, spontaneous dissociation of base-pairs at the duplex termini followed by translocation to prevent re-annealing (Pyle, 2008). In the cell, helicases act on highly structured substrates, raising the question of how DEAH/RHA proteins access and remodel their RNA targets, which are generally buried inside large RNPs. For this, a winching mechanism was proposed that involves loading of the helicase onto an exposed single-stranded RNA region and, if its translocation is physically blocked, pulling onto the complementary RNA strand, thereby disrupting the base-pairing interactions (Gilman et al., 2017).

In contrast to DEAH/RHA helicases, DEAD-box proteins are generally considered to be non-processive enzymes that unwind short duplex regions by a local strand separation mechanism (Yang et al., 2007). Due to the absence of the specific C-terminal region found in DEAH/RHA helicases, the two RecA-like domains in DEAD-box proteins are more flexible and unwinding takes place through switching of the helicase core between an open and a closed conformation during one cycle of ATP hydrolysis (Ozgur et al., 2015; Figure 1.1C). In the unbound state, the two RecA-like domains are spatially separated and inter-domain contacts are lacking. The cooperative binding of ATP and RNA to conserved sites located in both domains leads to the formation of an intricate network of interactions and triggers the conversion to a closed conformation (Hilbert et al., 2009; Mallam et al., 2012). Similar to DEAH/RHA helicases, DEAD-box proteins contact exclusively the sugar-phosphate backbone of RNA, demonstrating the intrinsic lack of specificity of the helicase core. Interestingly, while DEAD-box proteins can dock directly onto double-stranded RNA, their binding can be stabilized in some cases by the presence of nearby single-stranded RNA regions that interact with auxiliary domains of the helicase (Yang et al., 2007; Russell et al., 2013).

The closed conformation with ATP and RNA bound induces a kink in the RNA that distorts the duplex structure and further stochastic dissociation of base-pairs next to the unwound region leads to the release of one strand (Yang et al., 2007; Hilbert et al., 2009; Gilman et al., 2017). ATP hydrolysis and inorganic phosphate ( $P_i$ ) release induce the transition to an open conformation, dissociating the second RNA strand and recycling the helicase for more catalytic cycles (Liu et al., 2008). In the cellular context, most RNAs and RNPs likely contain relatively short duplex regions and would therefore be appropriate targets for unwinding by a local strand separation mechanism (Linder and Jankowsky, 2011). Putative physiological unwinding substrates that have been suggested to require the action of DEAD-box helicases include snoRNA-rRNA interactions and the base-pairing of U1 snRNA to the pre-mRNA 5' splice site (Staley and Guthrie, 1999; Kos and Tollervey, 2005; Srivastava et al., 2010).

### 1.2.3 Biochemical activities

In addition to the NTP-dependent unwinding of RNA duplexes described above, RNA helicases can perform other activities, such as acting as RNA chaperones to facilitate RNA folding, displacing proteins from RNA, nucleating the assembly of RNP complexes (clamping) or annealing RNA strands (Pyle, 2008; Linder and Jankowsky, 2011). RNA structures have a strong tendency to adopt misfolded or non-functional conformations and were suggested to require the action of RNA chaperones, such as RNA helicases, to achieve their native configuration (Herschlag, 1995). This RNA chaperoning activity has been shown for a few DEAD-box proteins, which bind RNA non-specifically and either assist directly in their folding or resolve non-native structures that would then facilitate their proper folding (Russell, 2008; Pan and Russell, 2010). On the other hand, protein displacement has been described both for translocating and non-translocating helicases and, although it was shown to be independent of duplex unwinding for specific helicases, the exact mechanism is not known (Fairman et al., 2004; Jankowsky and Bowers, 2006; Putnam and Jankowsky, 2013). Interestingly, in some cases, the diverse activities carried out by helicases are a result of their regulation by interacting proteins. For example, the exon junction complex RNA helicase eIF4A-III binds RNA in an ATP-dependent manner and arrest of its ATP hydrolysis cycle by the MAGOH-Y14 heterodimer leads to clamping of the helicase onto RNA and the formation of a stable RNP complex that serves as an assembly platform for other factors (Ballut et al., 2005; Nielsen et al., 2009). Furthermore, the RNA helicase Rok1 was shown to undergo conformational changes in the presence of Rrp5 that stimulate its ability to anneal RNA strands (Young et al., 2013). RNA annealing activity was also observed for Ded1 and Mss116 in addition to their unwinding function and, interestingly, it was found to be ATP-independent, suggesting that these enzymes can

catalyze complex rearrangements of RNA structures depending on their nucleotide status (Linder and Jankowsky, 2011).

### **1.3 Cellular pathways involving RNA helicases**

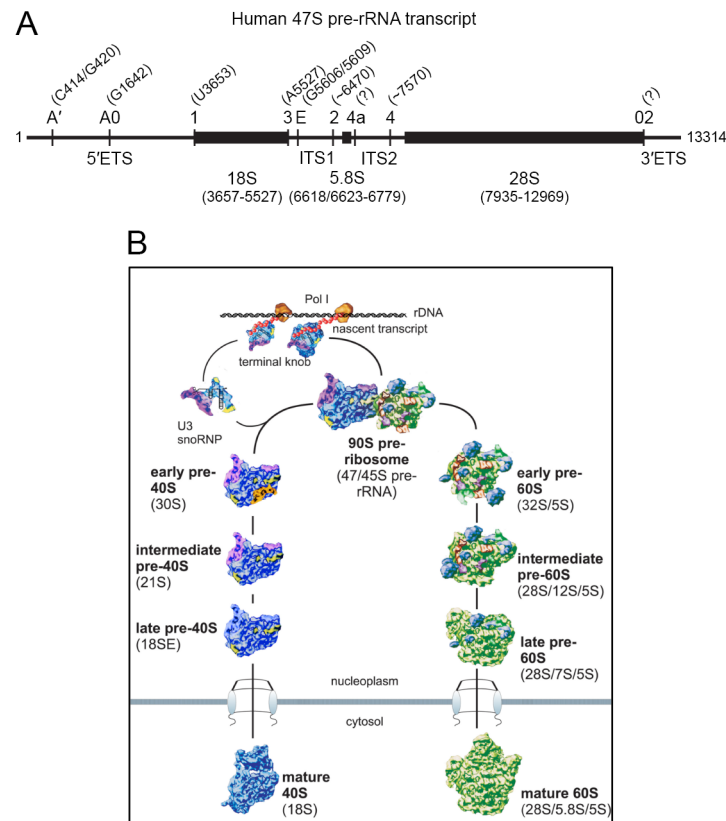
The diverse ways in which RNA helicases express their function in remodeling RNAs and RNPs is also reflected in the wide range of activities that they perform in the cell. These enzymes are essential effectors of all RNA-related processes and are involved in pathways such as ribosome biogenesis, splicing, transcription, mRNA export, translation, mRNA decay and innate immunity (Jankowsky, 2011). The function of RNA helicases has been described mostly in the context of two large and highly complex RNPs, the ribosome and the spliceosome, which undergo extensive remodeling during their assembly and functional cycle (Jarmoskaite and Russell, 2014).

#### **1.3.1 Ribosome biogenesis**

Ribosomes are essential ribonucleoprotein complexes that are responsible for protein synthesis in all three domains of life. The eukaryotic ribosome sediments at 80S and is composed of four rRNAs and approximately 80 ribosomal proteins (RPs) that are assembled into two asymmetric subunits. In human cells, the 40S small subunit (SSU) contains the 18S rRNA and 33 RPs, while the 60S large subunit (LSU) is composed of the 5S, 5.8S and 28S rRNAs together with 47 RPs. During translation, decoding of the mRNA sequence takes place in the SSU and peptide bond formation is accomplished in a catalytic center consisting of rRNA that is located in the LSU. The functional core of the ribosome is highly conserved in all organisms, but eukaryotic ribosomes have an increased size and complexity compared to their bacterial counterparts due to the presence of rRNA expansion segments, additional RPs and RP extensions (Melnikov et al., 2012; Wilson and Doudna, 2012; Yusupova and Yusupov, 2014).

The production of eukaryotic ribosomes starts with the transcription of ribosomal RNA precursors (pre-rRNAs), which undergo processing, folding and modification and are concurrently assembled with RPs to generate the mature ribosomal subunits. This requires the assistance of hundreds of assembly factors, which bind transiently and in a defined order and generally perform irreversible reactions that drive the process forward (Strunk and Karbstein, 2009; Kressler et al., 2010). Thus, ribosome biogenesis is a highly regulated and energy-consuming process. The pathway of ribosome assembly is best studied in *Saccharomyces cerevisiae* (*S. cerevisiae*) and, while most features are conserved across eukaryotes, in human cells there are specific differences in pre-rRNA processing as well as

a larger number of assembly factors (Tafforeau et al., 2013; Henras et al., 2015; Tomecki et al., 2017).



**Figure 1.2. Overview of ribosome biogenesis in human cells. (A)** Schematic representation of the 47S pre-rRNA transcript, which contains the sequences of the mature 18S, 5.8S and 28S rRNAs flanked by external transcribed spacer (5'ETS and 3'ETS) and separated by internal transcribed spacer (ITS1 and ITS2) regions. This precursor is processed by endonucleolytic cleavage at sites that are marked above. The position of the first and last nucleotide of the mature rRNAs within the precursor are indicated below. This panel is based on Mullineux and Lafontaine, 2012 and Henras et al., 2015. **(B)** During ribosome assembly, a multitude of factors associate with the nascent transcript to generate the 90S pre-ribosome, which undergoes a pre-rRNA cleavage event that separates the precursors of the two ribosomal subunits. These pre-ribosomal complexes are further processed in the nucleus and cytoplasm to produce the mature 40S and 60S subunits. This panel was adapted from Martin, 2014.

In human cells, the 18S, 5.8S and 28S rRNAs are co-transcribed in the nucleolus by RNA polymerase I to generate the 47S pre-rRNA transcript, in which the sequences of the mature rRNAs are interspersed with external transcribed spacer (5'ETS and 3'ETS) and internal transcribed spacer (ITS1 and ITS2) regions (Figure 1.2A). Multiple assembly factors and several RPs are recruited co-transcriptionally to the nascent pre-rRNA, leading to the formation of the earliest biogenesis precursor, the 90S pre-ribosome, which contains the full-length transcript and predominantly proteins required for SSU maturation (Grandi et al., 2002; Phipps et al., 2011; Figure 1.2B). Structures of 90S pre-ribosomes from *S. cerevisiae* and *Chaetomium thermophilum* have revealed that this macromolecular complex assembles as a scaffold around the pre-rRNA and ensures that its maturation takes place

in a coordinated and sequential manner (Kornprobst et al., 2016; Kressler et al., 2017; Sun et al., 2017a). The processing pathways of the SSU and LSU diverge after an endonucleolytic cleavage in ITS1. The precursors of the two subunits undergo additional maturation steps that involve the dynamic association and release of assembly factors, the incorporation of RPs and further pre-rRNA processing and structural rearrangement events. These pre-ribosomal complexes transition through the nucleolus and nucleoplasm and are then exported to the cytoplasm, where final maturation occurs and the two subunits associate for translation (Kressler et al., 2017; Pena et al., 2017; Chaker-Margot, 2018). The 5S rRNA precursor is transcribed separately by RNA polymerase III at sites adjacent to the nucleolus and joins the assembly pathway at an early stage (Ciganda and Williams, 2011).

The processing of the 47S pre-rRNA transcript into the mature 18S, 5.8S and 28S rRNAs involves sequential endonucleolytic cleavages that take place at defined sites in the spacer regions and are followed by exonucleolytic trimming (Henras et al., 2015; Aubert et al., 2018; Figure 1.2A). The initial precursor is processed first at sites A' in the 5'ETS and 02 in the 3'ETS to produce the 45S pre-rRNA. The cleavage at site A' was shown to not be required for the downstream steps and, while its role is not known, it is interesting to note that this site is only present in metazoans (Sloan et al., 2014). Two parallel pathways exist for processing of the 45S precursor that differ in the relative order of the 5'ETS removal and ITS1 cleavage events and give rise to different pre-rRNA species. Cleavage of the 45S pre-rRNA at site 2 in ITS1 prior to 5'ETS excision generates the 30S and 32.5S precursors. The 5'ETS region of the 30S pre-rRNA is subsequently removed by coordinated cleavages at sites A0 and 1, giving rise to the 21S intermediate, which is then processed at its 3' terminus through the combined action of endo- and exonucleases to produce the 18SE pre-rRNA. This precursor is exported to the cytoplasm where a final cleavage at site 3 in ITS1 generates the mature 18S rRNA (Henras et al., 2015; Aubert et al., 2018).

Alternatively, excision of the 5'ETS region in the 45S pre-rRNA leads to the formation of the 41S intermediate, which can be further processed via two pathways. In the major pathway, cleavage takes place at site 2 in ITS1 and creates the 21S and 32.5S precursors, while in the minor pathway processing occurs instead at site E in ITS1 and produces the 18SE and 36S pre-rRNAs (Prete et al., 2013; Sloan et al., 2013). The 21S and 18SE precursors of the small ribosomal subunit are matured as described above. The 36S pre-rRNA is trimmed at its 5' end by the 5'-3' exonuclease XRN2 to produce the 32.5S intermediate, which is the common LSU biogenesis precursor for all the alternative pathways. The 5' end of the 32.5S pre-rRNA is further digested by XRN2 to generate the abundant 32S intermediate, which contains the sequences of the 5.8S and 28S rRNAs. Cleavage at site 4 in ITS2 followed by

exonucleolytic digestion releases the mature forms of these rRNAs (Henras et al., 2015; Aubert et al., 2018). Interestingly, a second ITS2 cleavage has been reported at site 4a, which leads to the excision of a fragment corresponding to the 4a-4 region that is degraded by XRN2 (Schillewaert et al., 2012). Other pre-rRNA spacer regions are released during processing and XRN2 has also been linked to the turnover of the 5'-A', A0-1 and E-2 fragments (Wang and Pestov, 2011; Sloan et al., 2013; Sloan et al., 2014).

In addition to nucleases that participate directly in pre-rRNA processing, the ribosome assembly pathway requires the action of a multitude of other factors, such as RNA helicases, GTPases, kinases, structural proteins and snoRNAs that associate with proteins into snoRNPs. These assembly factors are essential for a wide range of processes, which include, among others, folding and modification of pre-rRNAs, remodeling and export of pre-ribosomal complexes, acting as structural scaffolds within pre-ribosomes or chaperoning and assisting the integration of RPs (Strunk and Karbstein, 2009; Kressler et al., 2010; Watkins and Bohnsack, 2012; Konikkat and Woolford, 2017; Pillet et al., 2017).

The role of RNA helicases in ribosome biogenesis has been mainly characterized in yeast, where 21 helicases participate in this process. These enzymes were suggested to remodel RNA-RNA and protein-RNA interactions within pre-ribosomes and were recently shown to also mediate the export of pre-ribosomal complexes and the acetylation of pre-rRNA (Martin et al., 2013; Rodriguez-Galan et al., 2013; Neumann et al., 2016; Sharma et al., 2017). The RNA/RNP remodeling function of RNA helicases is exerted in diverse ways during ribosome biogenesis. For example, multiple RNA helicases were suggested to mediate the release of snoRNPs from pre-ribosomes by unwinding snoRNA-rRNA interactions. This includes Dbp4, Rok1, Has1, Dhr1 and Prp43, whose depletion or inactivation led to the accumulation of specific snoRNPs in pre-ribosomal particles (Kos and Tollervey, 2005; Liang and Fournier, 2006; Bohnsack et al., 2008; Bohnsack et al., 2009; Sardana et al., 2015). RNA helicases can also unwind secondary structures in pre-rRNAs that facilitate the binding of snoRNPs to their target sites as has been suggested for Prp43. Another role proposed for Prp43 is the remodeling of late pre-ribosomal complexes to enable access of the endonuclease Nob1 to its cleavage site (Lebaron et al., 2009; Pertschy et al., 2009). Therefore, in addition to snoRNPs, RNA helicases might also regulate the binding or dissociation of ribosome assembly proteins either in a direct or indirect manner. A remodeling function was also described for the RNA helicase Mtr4, which is required to unfold structured pre-rRNA substrates and facilitate their processing or degradation by the nuclear exosome (Thoms et al., 2015; Schuller et al., 2018; Weick et al., 2018).

RNA helicases in higher eukaryotes are expected to perform similar functions in ribosome biogenesis as their yeast counterparts (Martin et al., 2013; Rodriguez-Galan et al., 2013).

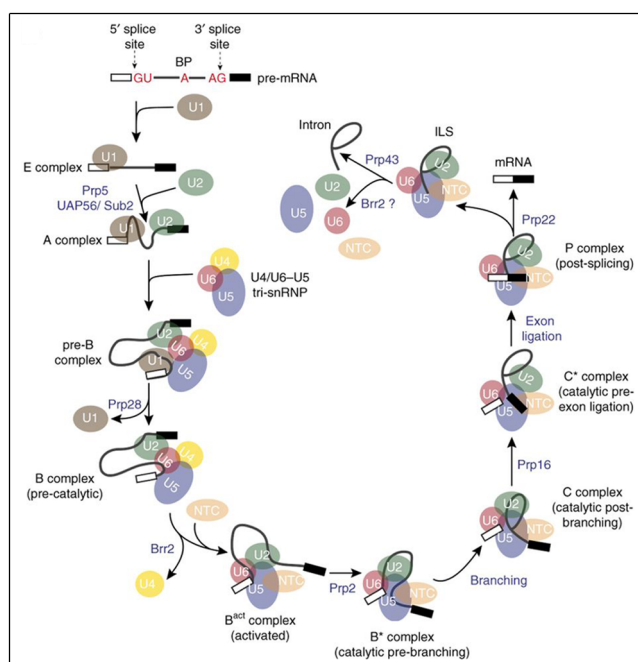
Consistent with this, some mammalian helicases have already been implicated in snoRNA release/association within pre-ribosomes (Srivastava et al., 2010; Sloan et al., 2015). However, the function of most RNA helicases in human ribosome biogenesis is poorly characterized and, given the increased complexity of this pathway compared to yeast, additional roles might be revealed.

### 1.3.2 Pre-mRNA splicing

The transcription of protein-coding genes yields precursor mRNAs (pre-mRNAs) that require processing by splicing to remove the non-coding segments (introns) and join the coding sequences (exons). Pre-mRNA splicing takes place within the spliceosome, a large RNP complex that assembles *de novo* on each intron from five snRNPs, composed of the U1, U2, U4, U5 and U6 snRNAs and their associated proteins, and numerous additional non-snRNP proteins. Introns are defined by conserved sequence elements, namely the 5' splice site, the branch point and the 3' splice site, and are removed in two sequential transesterification reactions. First, the 2'-OH group of a conserved adenosine in the branch point initiates a nucleophilic attack on the phosphodiester bond at the 5' splice site that releases the 5' exon and generates an intron lariat-3' exon intermediate. In the next step, the phosphodiester bond at the 3' splice site is attacked by the 3'-OH group of the first exon, leading to ligation of the exons and release of the excised intron lariat (Wahl et al., 2009; Will and Lührmann, 2011). Both reactions take place in an RNA-based catalytic core, in which the splice sites are brought into proximity by a network of base-pairing interactions and the U6 snRNA coordinates metal ions that are directly involved in catalysis (Fica et al., 2013; Fica and Nagai, 2017; Shi, 2017).

During its assembly and functional cycle, the spliceosome undergoes extensive structural and compositional rearrangements that involve the dynamic exchange of proteins as well as restructuring of RNA-RNA interactions (Figure 1.3). In *S. cerevisiae*, these remodeling events are driven by eight conserved RNA helicases that belong to the DEAD-box, DEAH/RHA and Ski2-like families (Cordin and Beggs, 2013). While DEAD-box proteins act in the early steps of spliceosome assembly, DEAH/RHA helicases are implicated at the later stages, possibly reflecting the requirement for different types of helicase activity at the pre-catalytic and catalytic phases (Gilman et al., 2017). Following the association of U1 snRNP with the 5' splice site, the U2 snRNP is stably recruited at the branch point in a process that requires the action of the RNA helicases Sub2 and Prp5. Next, joining of the pre-assembled U4/U6.U5 tri-snRNP complex generates the pre-B complex, which undergoes major rearrangements driven by Prp28 and Brr2 to form the activated spliceosome complex ( $B^{act}$  complex). Prp28 removes the U1 snRNP from the 5' splice site, while Brr2 unwinds the U4/U6 duplex, enabling the base-pairing of U6 snRNA to the 5' splice site and the formation

of U6-U2 snRNA interactions. The final transition step to the catalytic spliceosome (B\* complex) requires the action of Prp2, which remodels the branch point region, exposing the adenosine for the nucleophilic attack. After the first transesterification reaction, the resulting C complex is remodeled by Prp16 to position the reactive groups for the second catalytic step. Following exon ligation, Prp22 releases the mRNA, while Prp43 disassembles the spliceosome, recycling the snRNPs and releasing the intron lariat (Cordin and Beggs, 2013; Fica and Nagai, 2017; Shi, 2017). In addition to their remodeling function, several RNA helicases ensure the fidelity of the splicing process by discriminating against suboptimal splice sites. At the assembly stage, Prp5 was suggested to check the accuracy of the branch point-U2 snRNA pairing, while Prp28 proofreads the 5' splice site. During the catalytic steps, proofreading at the 5' and 3' splice sites is carried out by Prp16 and Prp22 respectively. Subsequently, the suboptimal spliceosomal complexes are directed to Prp43 for disassembly (Koodathingal et al., 2010; Koodathingal and Staley, 2013; Semlow et al., 2016).



**Figure 1.3. Overview of the splicing cycle.**

A model pre-mRNA containing the 5' splice site, the branch point (BP) and the 3' splice site conserved sequences is depicted with exons as rectangles and the intron as a line. The spliceosome assembles on this substrate in a step-wise manner with the U1 and U2 snRNPs binding first, followed by the U4/U6.U5 tri-snRNP, the NineTeen Complex (NTC) and additional splicing factors. During assembly, the spliceosome undergoes structural and compositional changes that are necessary for the two transesterification reactions (branching and exon ligation), which lead to joining of the two exons and the release of the intron lariat. These remodeling events are driven by eight conserved RNA helicases that are indicated in blue. This figure was originally published in Fica and Nagai, 2017 and is reprinted here with permission from Springer Nature.

Although the core aspects of the splicing pathway are conserved between yeast and human, specific differences and the presence of additional factors lead to a more complex spliceosome machinery in human cells. In yeast, the splice site sequences are highly conserved and this correlates with the prevalence of constitutive splicing compared to alternative splicing (Will and Lührmann, 2011). In contrast, these sites are more degenerate in humans, where it is estimated that 95-100% of genes undergo alternative splicing, which



expands the proteome significantly and contributes to the complexity of higher organisms (Lee and Rio, 2015). The regulation of alternative splicing is enabled by the presence of *cis*-acting regulatory elements in metazoan pre-mRNAs, which serve as binding sites for proteins that can either promote or inhibit spliceosome assembly. It has been suggested that the combined action of multiple such *trans*-acting splicing factors determines the choice of splice sites both at the early and late stages of the splicing cycle (Will and Lührmann, 2011; Fu and Ares, 2014). Furthermore, the regulation of alternative splicing is also linked to transcription, chromatin organization and signal transduction mechanisms (Wahl and Lührmann, 2015).

### 1.3.3. Other pathways

Splicing is tightly connected to mRNA export and several RNA helicases are implicated in this process as well. For example, Sub2/UAP56 mediates the recruitment of specific adaptor proteins to mRNAs to form export-competent complexes, while on the cytoplasmic side of the nuclear pore, Dbp5/DDX19 displaces export factors from mRNPs to ensure directionality (Tieg and Krebber, 2013; Jarmoskaite and Russell, 2014; Bourgeois et al., 2016). Similarly, multiple RNA helicases are involved at different stages during translation. In the early steps of cap-dependent translation initiation, eIF4A-I/DDX2A, eIF4A-II/DDX2B and Ded1/DDX3 are required to unwind structures in the 5' untranslated regions (UTR) of the mRNA, thereby facilitating the loading of the pre-initiation complex and scanning for the start codon. These helicases have complementary but distinct activities, with eIF4A-I/II being suggested to disrupt weak structures and promote translation of all mRNAs, while mRNAs that contain strong secondary structures in the 5' UTR are more dependent on Ded1 (Sen et al., 2015; Yourik et al., 2017; Gupta et al., 2018). The translation of mRNAs with highly structured 5' UTRs is also assisted by DHX29, which interacts with the 40S ribosomal subunit near the mRNA entry channel and was proposed to act by remodeling ribosomal complexes (Dhote et al., 2012). Other RNA helicases function at later stages of translation. This includes DHX33, which has a role in the formation of elongation-competent 80S ribosomes, and Dbp5/DDX19, which is involved in translation termination (Gross et al., 2007; Tieg and Krebber, 2013; Zhang et al., 2015; Mikhailova et al., 2017). Multiple other cellular processes involving RNA helicases have been described, including mRNA storage and decay, miRNA-induced silencing and viral RNA recognition in the immune response (Jankowsky, 2011; Bourgeois et al., 2016).

## 1.4 Regulation of RNA helicases

### 1.4.1 General regulatory mechanisms

RNA helicases are essential for all RNA-related processes through their function in remodeling RNA and RNA-protein complexes. Furthermore, many *multifunctional* RNA helicases that regulate several aspects of RNA metabolism have been identified, indicating that these enzymes have a central role in the coordination of different cellular events (Jankowsky, 2011; Bourgeois et al., 2016). On the other hand, structural studies have revealed that the conserved core of RNA helicases binds the RNA substrate in a non-sequence specific manner and biochemical analyses have shown that some of these enzymes have a low activity *in vitro* (Ozgun et al., 2015; Gilman et al., 2017).

These features indicate the need for a complex regulation of RNA helicases, which can be achieved *in cis* through the influence of auxiliary domains flanking the helicase core, as well as *in trans* through the action of accessory proteins termed RNA helicase cofactors (Ozgun et al., 2015). Furthermore, other regulatory mechanisms, such as post-translational modifications and interactions with long non-coding RNAs (lncRNAs), have been suggested (Sloan and Bohnsack, 2018). These diverse modes of RNA helicase regulation can result in the stimulation or inhibition of their activity, enable their recruitment and/or confer specificity for certain RNA targets, as well as expand the repertoire of the biochemical activities that they perform.

### 1.4.2 Protein cofactor-independent regulation

The N-terminal and C-terminal auxiliary regions present in the majority of RNA helicases generally establish interactions with RNA and/or proteins and can influence helicase function in multiple ways. For example, DHX36 contains a characteristic motif in its N-terminal extension that contributes to specific recognition of G-quadruplex structures (Lattmann et al., 2010). Some DEAD-box proteins such as Mss116 possess a positively-charged C-tail region that binds adjacent to the target site and anchors the helicase core to the substrate in a non-sequence specific manner (Russell et al., 2013). On the other hand, the RRM domain present in the C-terminus of the *Escherichia coli* (*E. coli*) RNA helicase DbpA and of its *Bacillus subtilis* orthologue YxiN was found to recruit the helicase to rRNA by specifically binding to a region of 23S rRNA (Diges and Uhlenbeck, 2001; Kossen et al., 2002). Interestingly, this is also linked to the stimulation of the ATPase and unwinding activities, at least in the case of YxiN (Samatanga et al., 2017). A different role has been attributed to the C-terminal region of Ded1, which is involved both in the oligomerization of the helicase and in the interaction with the translation initiation factor eIF4G. Due to the mutually exclusive nature of these processes, it has been suggested that the helicase might

be targeted to distinct functions depending on the interactions that its C-terminus establishes (Putnam et al., 2015). Furthermore, the C-terminal OB-fold domain that is specific for DEAH/RHA helicases contributes to RNA binding and also serves as a platform for interaction with protein cofactors, indicating that it represents an essential hub for the regulation of helicase activity (He et al., 2010; Walbott et al., 2010). Auxiliary domains can also function to keep the helicase inactive in the absence of the RNA substrate, preventing futile ATP hydrolysis. For example, DDX19 contains an N-terminal segment wedged between its two RecA-like domains that is displaced upon RNA binding to allow formation of the active, closed conformation (Collins et al., 2009). In addition to the effects exerted by auxiliary domains, RNA helicases can be regulated by post-translational modifications that either directly influence their catalytic activity or presumably modulate their interactions with other factors. Furthermore, several lncRNAs were found to impact the function of RNA helicases and the mechanisms proposed include competition with the cognate RNA substrate or acting as acceptors for the unwound products (Sloan and Bohnsack, 2018). These examples of cofactor-independent regulatory mechanisms highlight the complex control of RNA helicase function. An additional layer of regulation is provided by RNA helicase protein cofactors, which are discussed in the next section.

#### 1.4.3 Regulation by protein cofactors

In addition to the mechanisms described above, a growing number of RNA helicases whose functions are modulated by *trans*-acting proteins, termed cofactors, have been identified. Overall, the helicase cofactors characterized so far comprise a heterogeneous group of proteins that interact with RNA helicases either in the conserved core or in the auxiliary domains and can influence every aspect of their catalytic cycle, including substrate binding and release, ATP hydrolysis and unwinding activity (Young and Karbstein, 2012; Sloan and Bohnsack, 2018). Interestingly, a significant number of these effector proteins don't share any obvious similarity with other cofactors, implying that they represent individual proteins that have evolved to regulate RNA helicases. This includes, for example, the ribosome biogenesis factors Utp14 and Esf2, which were shown to stimulate the activity of Dhr1 and Dbp8 respectively (Granneman et al., 2006; Zhu et al., 2016). Some cofactors can have other defined functions in the cell that are independent of their role as RNA helicase regulators as is the case for NUP98, which is a component of the nuclear pore complex but also interacts in the nucleoplasm with DHX9 and was suggested to activate this helicase for its functions in transcription and splicing (Capitanio et al., 2017).

On the other hand, the identification of effector proteins that modulate the activity of RNA helicases through a common domain has revealed the existence of dedicated families of helicase cofactors. This includes proteins that contain an MIF4G (middle domain of eIF4G)

or a G-patch domain, several of which have been shown to act as helicase regulators. Interestingly, MIF4G domain-containing proteins regulate DEAD-box helicases, while G-patch proteins are cofactors of DEAH/RHA helicases (Ozgur et al., 2015; Robert-Paganin et al., 2015; Sloan and Bohnsack, 2018).

Structural studies of various MIF4G cofactors in complex with their cognate helicase have revealed that they act mainly by modulating the transition between the open and closed conformation of the helicase core and can either stimulate or repress the activity of the RNA helicase. The MIF4G domain consists of five antiparallel  $\alpha$ -helices termed HEAT repeats that are assembled into a crescent-shaped configuration, with the N-terminal and C-terminal sides contacting the helicase domains 2 and 1 respectively. In the case of activating cofactors, such as eIF4G, Gle1 and CNOT1, these interactions bring the RecA-like domains together, leading to a 'half-open' conformation that enhances activity (Schutz et al., 2008; Montpetit et al., 2011; Mathys et al., 2014; Ozgur et al., 2015). Intriguingly, the MIF4G domain of CWC22 does not have a stimulatory role but instead inhibits the activity of eIF4A-III. This is due to a different binding mode of the C-terminus of the MIF4G module to domain 1 of the helicase core, which leads to an inactive conformation in which the ATP and RNA binding sites located in the two RecA-like domains of the helicase are distant (Buchwald et al., 2013). Therefore, despite having a similar architecture, MIF4G domains can induce opposing effects depending on a few key intermolecular contacts that stabilize the helicase core either in an active or inactive conformation. Similarly, subtle variations in the MIF4G domains have been suggested to control their binding specificity, allowing the discrimination of the cognate DEAD-box helicase based on a few favorable or unfavorable interactions (Buchwald et al., 2013; Ozgur et al., 2015). The regulation of DEAH/RHA helicases by G-patch proteins is discussed separately in the next section.

## **1.5 G-patch proteins as regulators of DEAH/RHA helicases**

### **1.5.1 General characteristics and regulatory mechanisms**

In addition to the MIF4G domain proteins and the other cofactors described above, several yeast and human proteins containing a G-patch domain have been shown to regulate the activity of RNA helicases (Robert-Paganin et al., 2015; Sloan and Bohnsack, 2018). G-patch proteins are found in eukaryotes as well as in some viruses and are characterized by a common glycine-rich region of approximately 50 amino acids that constitutes the G-patch domain. The consensus sequence of the G-patch domain was defined as HHX<sub>3</sub>GAX<sub>2</sub>GXGHGX<sub>4</sub>G (H - hydrophobic, A - aromatic, X - non-conserved amino acid) and includes five highly conserved glycine residues, an invariant aromatic amino acid following

the first conserved glycine and several conserved hydrophobic residues. In addition, a sixth glycine can be found downstream in most proteins (Aravind and Koonin, 1999; Robert-Paganin et al., 2015).

Most of our knowledge of the regulation of DEAH/RHA helicases by G-patch cofactors is based on studies in *S. cerevisiae*, where five G-patch proteins have been identified. These include Spp382 (Ntr1), Pxr1 (Gno1), Sqs1 (Pfa1) and Cmg1, which were shown to act as positive regulators of the multifunctional RNA helicase Prp43, as well as Spp2, which is a cofactor of Prp2 (Robert-Paganin et al., 2015; Sloan and Bohnsack, 2018). The best characterized G-patch cofactor is Spp382, which is required together with Prp43 for the disassembly of aberrant spliceosomes and of late-stage, intron lariat spliceosomes. The N-terminal region of Spp382 containing the G-patch domain enhances the activity of Prp43 and enables coupling of ATP hydrolysis to its remodeling function, while the C-terminal region controls the timing of Prp43 activation through interactions with other proteins (Fourmann et al., 2016; Fourmann et al., 2017). On the other hand, Pxr1 and Sqs1 were suggested to stimulate the activity of Prp43 during ribosome biogenesis for its function in the release of a subset of snoRNAs from pre-rRNA and in the remodeling of pre-ribosomal complexes to facilitate final maturation of 18S rRNA respectively (Lebaron et al., 2009; Pertschy et al., 2009; Robert-Paganin et al., 2017). For the Cmg1-Prp43 complex, a role in the remodeling or disassembly of cytoplasmic RNPs has been proposed (Heininger et al., 2016). Furthermore, Spp2 is essential for the remodeling activity of Prp2, which is required for the transition to a catalytically active spliceosome (Silverman et al., 2004; Warkocki et al., 2015).

The above-mentioned interactions have provided valuable insight into the regulation of RNA helicases by G-patch cofactors. In all cases, G-patch proteins were shown to have a stimulatory role and enhance the ATPase and/or unwinding activity of their interacting helicase. In contrast to MIF4G cofactors, which bind in the helicase core and control its conformational transitions, G-patch proteins interact with the C-terminal auxiliary region of the RNA helicase. Protein-protein crosslinking experiments and interaction studies with helicase mutants have pinpointed the OB-fold domain as the main contact site for the G-patch motif (Silverman et al., 2004; Walbott et al., 2010; Christian et al., 2014; Heininger et al., 2016). Furthermore, the G-patch domain is the essential module for activating RNA helicases as no stimulation is observed in its absence, although in some cases other regions of the G-patch protein can bind the helicase (Lebaron et al., 2009; Christian et al., 2014). In addition to mediating the interaction with the OB-fold region, the G-patch domain was shown to bind RNA in specific cases (Svec et al., 2004; Lebaron et al., 2009). However,

this characteristic is not shared by all G-patch domains, for example Spp382 being suggested to interact with RNA only in the presence of the helicase (Christian et al., 2014).

Elucidating the mechanistic details of the G-patch protein-mediated regulation of RNA helicases would require structural information of the G-patch domain alone or in the presence of the helicase, which is currently limited. Circular dichroism spectroscopy studies have determined that the G-patch motif is unstructured in solution but has the capacity to adopt secondary structure elements and it was proposed that this would take place upon binding to the helicase (Frenal et al., 2006; Christian et al., 2014). In addition, based on structure probing and protein-RNA crosslinking experiments, it was suggested that binding of the G-patch protein induces structural rearrangements in the C-terminal region of the helicase that increase the availability of the RNA-binding channel (Christian et al., 2014). On the other hand, a recent study proposed that activation by G-patch proteins is required to disrupt a non-productive, auto-inhibited state of the RNA helicase, which is imposed by stacking of the nucleotide base between two residues found in each of the helicase domains (Robert-Paganin et al., 2017).

An important aspect that remains to be elucidated is whether G-patch proteins act solely as enhancers of helicase activity or if they have other regulatory effects. Both Prp43 and Prp2 were found to associate with their target RNPs in the absence of their cofactors, indicating that they are recruited to the spliceosome and pre-ribosome independently of G-patch proteins (Lebaron et al., 2009; Warkocki et al., 2015; Fourmann et al., 2016). In addition, swapping experiments have shown that in some cases G-patch domains can substitute for each other in activating the helicase (Banerjee et al., 2015; Fourmann et al., 2016). On the other hand, the effects exerted by G-patch proteins are likely to be more complex than simply acting as activators of RNA helicases, as demonstrated for Spp382, whose C-terminal region lacking the G-patch domain is also essential for the regulation of Prp43 (Fourmann et al., 2017). This is further supported by the finding that overexpression of certain G-patch cofactors of Prp43 leads to a relocalization of the helicase and to its withdrawal from specific functions, indicating that G-patch proteins control the distribution of the RNA helicase between different pathways (Heininger et al., 2016).

### **1.5.2 Human G-patch proteins**

The human genome encodes 22 proteins that have a G-patch domain in their sequence as indicated by the UniProt database (Apweiler et al., 2004). In addition, G-patch domains are found as part of endogenous retroviral elements, which are remnants of ancient retroviral infections that were preserved in the genome. However, the expression of these human endogenous retroviruses is suppressed in most cases and there is limited knowledge of their function (Hanke et al., 2016; Garcia-Montojo et al., 2018). The 22 human G-patch

proteins comprise a heterogeneous group of proteins, some of which contain additional defined domains, such as the RRM and R3H domains, which are involved in RNA binding (Figure 1.4A). Sequence alignment of the human G-patch domains revealed the presence of most of the conserved elements described originally for the G-patch motif, with the first and fourth glycine, as well as the aromatic amino acid and one hydrophobic residue, being found in all proteins (Figure 1.4B).

A

G-patch protein	Molecular weight (kDa)	Other domains
AGGF1	81	FHA
CHERP	104	SURP, CID
CMTR1	95	Methyltransferase, WW
GPANK1	39	ANK
GPATCH1	103	-
GPATCH2	59	-
GPATCH3	59	-
GPATCH4	50	-
GPATCH8	164	-
GPATCH11	33	-
GPKOW	52	KOW
NKRF	78	R3H
PINX1	37	-
RBM5	92	RRM, ZNF
RBM6	129	RRM
RBM10	104	RRM, ZNF
RBM17	45	RRM
SUGP1	72	SURP
SUGP2	120	SURP
SON	264	DRBM
TFIP11	97	-
ZGPAT	57	ZNF

B

GPATCH1_152-198	KLSVGFELLRKMGWKEGGVGVPRVKRRPRRQ---KPDPGVKI---YGCALPPG
RBM17_235-283	GGTVAHKIMQKYGFREGGGLGKHEQGLSTALSVEKTSKRGG---KIIVGDAT
CMTR1_87-133	YNSVSQKLMAKMGFREGEGLGKYSGRDKDIV---EASSQKGR---RGLGLTLR
GPATCH4_11-57	GMKFAEEQLKHGWTQKGLGRKENGITQAL---RVTLKQDT---HGVGHDP
GPATCH3_410-458	TKGIGRKVMERQGWAEQGLGCRCSVPEALDSGQHPRCK---RGLGYHGE
RBM6_1051-1097	TSSKGGCVQATGWKGTGLGYGHPGLASSE---EAEGMRMG---PSVGASGR
GPANK1_255-301	ISSPGFKLLRGGWEPGMGLGPRGEGRANPI---PTVLKRDQ---EGLGYRSA
GPATCH11_69-115	CENKGFALLQKMGYKSGQALGKSGGGIVEPI---PLNIKTGK---SGIGHEAS
GPATCH8_40-86	SDNIGHRLQKHGWLGGGLGKSLQGRTPDI---PIVVKYDV---MGMRMEM
ZGPAT_333-379	TRGIGSRLLTKMGYEFKGLGRHAEGRVEPI---HAVVLPRG---KSLDQCVE
GPKOW_164-210	VEAYGLAMLRGMGWKPGEIGRTFNQVVKPR---VNSLRPKG---LGLGANLT
CHERP_841-891	EENKGHOMLVKMGWSSGGLGAKEQGIQDPI---KGGDVDRKWDQYKGVGVALD
SON_2305-2351	TGGMGAVLMRKMGWREGEGLGKNEKNEKEPI---LVDFKTRD---KGLVAVGE
TFIP11_149-195	TKGIGQKLLQKMGYVPGRGLGKNAOGIINPI---EAKQRKKG---GAVGAYGS
GPATCH2_467-513	ENNIGNRMLQNMGWTPGSLGRDGKISEPI---QAMQRPKG---LGLGFPLP
RBM5_743-789	HSNIGNKMLQAMGWREGEGLGRKCOGITAPI---EAQVRLKG---AGLGAKGS
RBM10_858-904	SDNIGSRMLQAMGWKEGSLGRKKOGIVTPI---EAQTRVRG---SGLGARGS
NKRF_551-596	EDNIGNQLLRKMGWT---GGGLGKSGEGIREPI---SVKEQHKK---EGLGLDVE
SUGP1_562-609	VENIGYQMLMKMGWKEGEGLGSEGGIKNPVN---KGTITVDG---AGFGIDRP
SUGP2_1011-1057	DKNLGFQMLQKMGWKEGHGLGSLGKGIREFV---SVGTPESE---EGLGADGQ
PINX1_26-72	DSKFGQRMLEKMGWSKGKGLGAQEQGATDHI---KVQVKNNH---LGLGATIN
AGGF1_619-665	DSNKGKRMLEKMGWKKGEGLGKDGGMKTPI---QLQLRRTH---AGLGTGKP

**Figure 1.4. Characteristics and sequence alignment of human G-patch proteins. (A)** The human genome encodes 22 G-patch proteins that differ in size and, in some cases, contain additional domains apart from the G-patch domain. Abbreviations: FHA - forkhead-associated domain; SURP (SWAP) - suppressor-of-white-apricot and PRP21 motif; CID - CTD (C-terminal domain of RNA polymerase II)-interaction domain; ANK - ankyrin repeat domain; KOW - Kyprides, Ouzounis, Woese motif; RRM - RNA recognition motif; ZNF - zinc finger motif; DRBM - double-stranded RNA-binding motif. The information in this panel is based on the UniProt database. **(B)** Primary sequence alignment of the G-patch domain regions corresponding to the 22 human G-patch proteins indicated in (A). The conserved amino acids are marked below the panel and the residues that correspond to the consensus sequence of the G-patch domain are highlighted in blue. The symbols indicate fully conserved residues (\*), conservation between residues of strongly similar properties (:), and conservation between residues of weakly similar properties (.). The alignment was generated using MUSCLE (Edgar, 2004).

Compared to yeast G-patch proteins, much less is known about the role of human G-patch proteins as regulators of RNA helicases. Only CMTR1, GPATCH2, RBM5 and ZGPAT have been described as cofactors that stimulate the activity of the RNA helicase DHX15, which is the homologue of yeast Prp43 (Lin et al., 2009; Niu et al., 2012; Chen et al., 2017; Inesta-Vaquera et al., 2018; Toczydlowska-Socha et al., 2018). Furthermore, the G-patch proteins TFIP11 and PINX1 were also shown to bind DHX15, whereas GPKOW associates with DHX16 (Yoshimoto et al., 2009; Chen et al., 2014; Zang et al., 2014). These interactions are conserved from yeast where their homologues Spp382, Pxr1 and Spp2 are known to bind and activate Prp43 and Prp2 respectively, implying that the stimulatory role of these cofactors is preserved as well. For the other human G-patch proteins, there is no information

regarding their action as helicase cofactors, and in most cases, there is also limited knowledge of their function.

For AGGF1, multiple cellular functions have been proposed, including as an angiogenesis factor and as a transcriptional regulator (Tian et al., 2004; Major et al., 2008). The G-patch proteins CHERP and RBM17 were suggested to form a subcomplex with U2SURP that regulates the alternative splicing of a subset of genes, most of which encode RNA-processing factors (De Maio et al., 2018). Interestingly, CMTR1 is the only protein that contains a catalytic domain in combination with a G-patch domain and was shown to methylate the first transcribed nucleotide of mRNAs at the ribose 2'-OH position, thus contributing to the formation of the mRNA cap (Smietanski et al., 2014). CMTR1 was also described as a cofactor that stimulates the activity of the RNA helicase DHX15. Although the exact role of this complex was not determined, expression of a CMTR1 mutant that does not interact with DHX15 led to a more efficient translation of a subset of mRNAs, which in turn impaired cell growth (Inesta-Vaquera et al., 2018). Conversely, DHX15 might be required to facilitate the methylation activity of CMTR1 on highly structured RNA substrates (Toczydlowska-Socha et al., 2018). GPATCH1 was detected in the spliceosomal C complex and was also found to interact with centrosomal proteins, but it is not clear if these are related or separate functions (Agafonov et al., 2011; Gupta et al., 2015). GPATCH2 was shown to stimulate the ATPase activity of DHX15 and was suggested to participate in ribosome biogenesis similar to the related yeast G-patch protein Sqs1 (Lin et al., 2009; Robert-Paganin et al., 2015). GPATCH3 was recently described as a negative regulator of the antiviral immune response (Nie et al., 2017). GPKOW associates with the RNA helicase DHX16 and probably regulates its function in splicing similar to the interaction of their yeast homologues (Zang et al., 2014). Interestingly, it was suggested that the RNA binding affinity of GPKOW is regulated by phosphorylation (Aksaas et al., 2011). NKRF was originally described as a transcriptional repressor of NF- $\kappa$ B target genes (Nourbakhsh and Hauser, 1999). During the course of this work, an independent study reported that NKRF acts as a regulator of nucleolar homeostasis and pre-rRNA processing in heat shock conditions (Coccia et al., 2017). These results are in accordance with some of the findings of the present study, which are presented in the following sections. PINX1 was shown to interact with DHX15 and to substitute the function of its yeast homologue Pxr1 in ribosome biogenesis, raising the possibility that it performs a similar role in human cells together with the RNA helicase (Chen et al., 2014). Furthermore, PINX1 was described as a telomerase inhibitor and was also suggested to regulate chromosome segregation, but it is not known if these functions involve DHX15 as well (Zhou and Lu, 2001; Yuan et al., 2009). The highly similar proteins RBM5, RBM6 and RBM10 were suggested to regulate the alternative



splicing of distinct sets of pre-mRNAs (Bechara et al., 2013). While RBM5 and RBM10 have been detected in spliceosome preparations, RBM6 was not found in any spliceosomal complex so far (Agafonov et al., 2011). Furthermore, RBM5 was shown to stimulate the activity of DHX15, leading to the suggestion that these two proteins function together in splicing (Niu et al., 2012). SON is a splicing factor that regulates both intron removal in constitutively spliced transcripts and the alternative splicing of specific genes involved in essential processes such as cell cycle regulation, apoptosis or pluripotency maintenance (Ahn et al., 2011; Lu et al., 2013; Hickey et al., 2014). In addition, SON was linked to transcriptional regulation at specific promoters. The two related proteins SUGP1 and SUGP2 are putative splicing factors (Utans and Kramer, 1990; Sampson and Hewitt, 2003; Agafonov et al., 2011). Likewise, TFIP11 is involved in splicing where it is suggested to act together with DHX15 in the disassembly of intron lariat spliceosomes similar to their yeast counterparts (Yoshimoto et al., 2009). ZGPAT was recently described as a component of a U4/U6.U5 tri-snRNP complex intermediate that also contains DHX15 and was shown to stimulate the activity of the helicase, which might be required for maturation of the tri-snRNP (Chen et al., 2017). For GPANK1, GPATCH4, GPATCH8 and GPATCH11 there is no information available about their cellular functions.

### **1.5.3 G-patch proteins and RNA helicases in disease**

Due to their central role in fundamental cellular pathways, dysregulation of the activity of RNA helicases often leads to disease, and several RNA helicases have already been implicated in cancer as a result of altered expression levels or mutations (Fuller-Pace, 2013; Robert and Pelletier, 2013). In addition, the function of RNA helicases can be disrupted by alterations of their protein cofactors. For example, GPATCH2 was found to be overexpressed in breast cancer and, based on its role in stimulating the activity of DHX15, it was suggested that this would lead to hyperactivation of the helicase and promote cancer cell growth (Lin et al., 2009). Furthermore, a recurrent mutation in DHX15, which involves substitution of an arginine by a glycine residue (R222G), was detected in several cases of acute myeloid leukaemia. This mutant helicase showed reduced binding to the G-patch protein TFIP11 and its overexpression impaired splicing, implying that disruption of the DHX15-TFIP11 interaction might contribute to disease (Faber et al., 2016). Other diseases caused by mutations of RNA helicases or G-patch proteins were also identified (Tian et al., 2004; Kaneko et al., 2011; Steimer and Klostermeier, 2012). For example, nonsense and frame-shift mutations in the RBM10 gene were shown to cause TARP syndrome (Talipes equinovarus, atrial septal defect, Robin sequence and persistent left superior vena cava) (Gripp et al., 2011). One such mutation in a patient led to the deletion of a segment of the

G-patch protein and induced changes in alternative splicing that might explain the molecular basis of the disorder (Wang et al., 2013).

In addition to their implication in cancer and other diseases, RNA helicases play a role in viral infection, as most viruses do not encode their own RNA helicase but instead rely on host helicases for replication. Multiple cellular RNA helicases were found to be recruited by viruses to support various stages of their replication cycle, yet the mechanisms involved remain elusive (Ranji and Boris-Lawrie, 2010; Steimer and Klostermeier, 2012). Interestingly, some retroviruses encode a G-patch domain, raising the possibility that they use it to co-opt host RNA helicases for their purposes (Gifford et al., 2005; Jern et al., 2005). Furthermore, the G-patch domain expressed by the Mason-Pfizer monkey virus was found to be required for the activity of the reverse transcriptase and for viral infectivity, but it is not known if any RNA helicase is implicated in this function (Bauerova-Zabranska et al., 2005; Krizova et al., 2012). Overall, these findings highlight the importance of understanding the interactions between RNA helicases and G-patch proteins both in normal conditions and in disease.

## 1.6 Scope of the thesis

RNA helicases are essential regulators of the structure and function of RNA and RNA-protein complexes and participate in fundamental pathways such as ribosome biogenesis, splicing and translation. These enzymes share a conserved helicase core that binds RNA in a non-sequence specific manner, indicating that substrate recognition is regulated through other means. Furthermore, the majority of RNA helicases in higher eukaryotes are suggested to have more than one cellular role, but the mechanisms that determine their distribution between different functions are not known.

In recent years, regulation by *trans*-acting proteins, termed cofactors, has emerged as a powerful means to control the activity of RNA helicases. Among the diverse helicase cofactors that have been identified so far, several of them belong to the G-patch protein family. These proteins contain a characteristic glycine-rich motif known as a G-patch domain that is essential for their function as helicase regulators.

In order to understand the regulation of RNA helicases by G-patch protein cofactors, the complete inventory of such complexes needs to be identified. Furthermore, to decipher the functions of RNA helicases whose activities are modulated by interacting proteins, these enzymes have to be studied together with their cofactors. However, despite the large number of G-patch proteins expressed in human cells, only few of them have been functionally characterized and their role as RNA helicase cofactors has remained largely unexplored.

In this context, the goals of the present study were to:

- Identify the interacting RNA helicase(s) of the 22 human G-patch proteins;
- Characterize the identified G-patch protein-RNA helicase interactions *in vitro* to assess the effects that the cofactors exert on the activity of the helicase;
- Determine the subcellular localization of all human G-patch proteins;
- Functionally characterize the G-patch protein-RNA helicase complexes in a global manner using genome-wide approaches;
- Explore the function of selected G-patch protein-RNA helicase complexes in depth.

Together, these complementary strategies aimed to establish the human G-patch protein family as regulators of RNA helicases and to provide key insights into the mode of interaction and function of G-patch cofactor-RNA helicase complexes.

## 2. MATERIALS AND METHODS

### 2.1 Materials

#### 2.1.1 Chemicals and buffers

Standard laboratory reagents were purchased from Sigma-Aldrich, Carl Roth and Applichem. All restriction enzymes used in this study were obtained from Thermo Fisher. Other specific reagents used in this study together with their supplier are listed in Table 2.1 and the composition of standard solutions and buffers is given in Table 2.2.

**Table 2.1. Specific reagents used in this study**

Reagent	Supplier
[ $\gamma$ - $^{32}$ P]-ATP and [ $^{32}$ P]-orthophosphate	PerkinElmer
Anti-FLAG M2 Magnetic Beads	Sigma-Aldrich
Blasticidin S Hydrochloride	Applichem
cOmplete His-Tag Purification Resin	Roche
cOmplete Mini Protease Inhibitor Cocktail	Roche
Coomassie Plus (Bradford) assay kit	Thermo Fisher
Dulbecco's Modified Eagle Medium	Thermo Fisher
Fetal Bovine Serum Superior	Merck
FLAG Peptide	Sigma-Aldrich
Hygromycin B	Applichem
Immobilon Western Chemiluminescent HRP Substrate	Merck
Isopropyl- $\beta$ -D-thiogalactoside (IPTG)	Sigma-Aldrich
LightCycler 480 SYBR Green I Master kit	Roche
Lipofectamine RNAiMAX	Thermo Fisher
$\beta$ -Nicotinamide adenine dinucleotide reduced (NADH)	Sigma-Aldrich
Ni-NTA Agarose Resin	Qiagen
Nuclease-free H <sub>2</sub> O	Qiagen
NuPAGE LDS Sample Buffer (4X)	Thermo Fisher
Opti-MEM	Thermo Fisher
Penicillin-Streptomycin	Thermo Fisher
PfuTurbo DNA polymerase	Bohnsack lab
Phosphoenolpyruvate	Sigma-Aldrich
Phusion High-Fidelity DNA polymerase	Thermo Fisher
Poly-L-lysine solution (0.1%)	Sigma-Aldrich
Protein G Sepharose	GE Healthcare
Proteinase K	Roche
Pyruvate kinase/lactic dehydrogenase	Sigma-Aldrich
RiboLock Ribonuclease Inhibitor	Thermo Fisher

(continued on next page)

(Table 2.1 continued)

RNase-It Ribonuclease Cocktail	Agilent
RNase H	NEB
RNasin Ribonuclease Inhibitor	Promega
SuperScript III Reverse Transcriptase	Thermo Fisher
T4 DNA Ligase	Thermo Fisher
T4 Polynucleotide Kinase	Thermo Fisher / NEB
T4 RNA Ligase 1	NEB
T4 RNA Ligase 2, Deletion Mutant	Epicentre
TaKaRa LA Taq DNA polymerase	Clontech
4-Thiouridine	Sigma-Aldrich
TRI Reagent	Sigma-Aldrich
Trypsin-EDTA (0.25%)	Thermo Fisher
TSAP Thermosensitive Alkaline Phosphatase	Promega
TURBO DNase	Thermo Fisher
VECTASHIELD Antifade Mounting Medium with DAPI	Vector Labs
X-tremeGENE 9 DNA transfection reagent	Roche

**Table 2.2. Composition of standard solutions and buffers**

<b>Solution / buffer</b>	<b>Composition</b>
PBS	137 mM NaCl, 2.7 mM KCl, 8 mM Na <sub>2</sub> HPO <sub>4</sub> , 1.5 mM KH <sub>2</sub> PO <sub>4</sub>
TBS	25 mM Tris-HCl pH 7.4, 137 mM NaCl, 2.7 mM KCl
TBE	90 mM Tris, 90 mM boric acid, 2.55 mM EDTA
SDS-PAGE resolving gel buffer	375 mM Tris-HCl pH 8.8, 3.5 mM SDS
SDS-PAGE stacking gel buffer	125 mM Tris-HCl pH 6.8, 3.5 mM SDS
SDS-PAGE running buffer	25 mM Tris, 0.19 M glycine, 0.05% SDS
Western blotting transfer buffer	25 mM Tris, 0.19 M glycine, 20% methanol
NuPAGE MES running buffer	50 mM MES, 50 mM Tris, 0.1% SDS, 1 mM EDTA
NuPAGE western blotting transfer buffer	25 mM bicine, 25 mM Bis-Tris, 1 mM EDTA
BTPE	10 mM PIPES, 30 mM Bis-Tris, 1 mM EDTA
SSC	150 mM NaCl, 15 mM sodium citrate
SES1	0.25 M sodium phosphate pH 7.0, 7 % SDS, 1 mM EDTA
DNA loading dye (6X)	60% glycerol, 60 mM EDTA, 0.2% xylene cyanol, 0.2% bromophenol blue
SDS loading dye (4X)	200 mM Tris-HCl pH 6.8, 8% SDS, 50 mM EDTA, 4% 2-mercaptoethanol, 0.08% bromophenol blue, 40% glycerol
Formamide loading dye (2X)	80% formamide, 1 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol
Glyoxal loading dye	72 ml DMSO, 24 ml glyoxal, 14.4 ml BTPE (10X), 7.2 ml 80% glycerol, 20 µg/ml ethidium bromide

## 2.1.2 Plasmids

**Table 2.3. Mammalian expression plasmids used in this study**

Construct	CDS	Vector	Source	ID
AGGF1-His <sub>6</sub> -2xFLAG	NM_018046	pcDNA5/FRT/TO	This study	p1339
CHERP-His <sub>6</sub> -2xFLAG	NM_006387	pcDNA5/FRT/TO	This study	p1338
CMTR1-His <sub>6</sub> -2xFLAG	NM_015050	pcDNA5/FRT/TO	This study	p1302
His <sub>6</sub> -2xFLAG-DHX15	NM_001358	pcDNA5/FRT/TO	Bohnsack lab	p420
His <sub>6</sub> -2xFLAG-DHX15 <sub>1-698</sub>	NM_001358	pcDNA5/FRT/TO	This study	p1253
His <sub>6</sub> -2xFLAG-DHX15-siRNA resistant	NM_001358	pcDNA5/FRT/TO	This study	p1038
His <sub>6</sub> -2xFLAG-DHX15 <sub>E261Q</sub> -siRNA resistant	NM_001358	pcDNA5/FRT/TO	This study	p1309
DHX35-His <sub>6</sub> -2xFLAG	NM_021931	pcDNA5/FRT/TO	This study	p1472
DHX35 <sub>1-597</sub> -His <sub>6</sub> -2xFLAG	NM_021931	pcDNA5/FRT/TO	This study	p1536
His <sub>6</sub> -2xFLAG	-	pcDNA5/FRT/TO	Bohnsack lab	p187
GPANK1-His <sub>6</sub> -2xFLAG	NM_001199237	pcDNA5/FRT/TO	This study	p1326
GPATCH1-His <sub>6</sub> -2xFLAG	NM_018025	pcDNA5/FRT/TO	This study	p1327
GPATCH1 <sub>199-931</sub> -His <sub>6</sub> -2xFLAG	NM_018025	pcDNA5/FRT/TO	This study	p1593
GPATCH2-His <sub>6</sub> -2xFLAG	NM_018040	pcDNA5/FRT/TO	This study	p1328
GPATCH3-His <sub>6</sub> -2xFLAG	NM_022078	pcDNA5/FRT/TO	This study	p1329
GPATCH4-His <sub>6</sub> -2xFLAG	KJ902705	pcDNA5/FRT/TO	This study	p1370
GPATCH8-His <sub>6</sub> -2xFLAG	NM_001002909	pcDNA5/FRT/TO	This study	p1487
GPATCH11-His <sub>6</sub> -2xFLAG	AK294697	pcDNA5/FRT/TO	This study	p1330
GPKOW-His <sub>6</sub> -2xFLAG	NM_015698	pcDNA5/FRT/TO	This study	p1332
His <sub>6</sub> -2xFLAG-NKRF	NM_001173487	pcDNA5/FRT/TO	This study	p1036
His <sub>6</sub> -2xFLAG-NKRF <sub>110-705</sub>	NM_001173487	pcDNA5/FRT/TO	This study	p1169
His <sub>6</sub> -2xFLAG-NKRF <sub>G1-6A</sub>	NM_001173487	pcDNA5/FRT/TO	This study	p1248
PINX1-His <sub>6</sub> -2xFLAG	NM_017884	pcDNA5/FRT/TO	This study	p1258
RBM5-His <sub>6</sub> -2xFLAG	NM_005778	pcDNA5/FRT/TO	This study	p1340
His <sub>6</sub> -2xFLAG-RBM6	NM_005777	pcDNA5/FRT/TO	This study	p1603
RBM10-His <sub>6</sub> -2xFLAG	NM_005676	pcDNA5/FRT/TO	This study	p1369
RBM17-His <sub>6</sub> -2xFLAG	NM_001145547	pcDNA5/FRT/TO	This study	p1334
SON-His <sub>6</sub> -2xFLAG	NM_138927	pcDNA5/FRT/TO	This study	p1371
His <sub>6</sub> -2xFLAG-SUGP1	NM_172231	pcDNA5/FRT/TO	This study	p1372
SUGP2-His <sub>6</sub> -2xFLAG	NM_001017392	pcDNA5/FRT/TO	This study	p1337
TFIP11-His <sub>6</sub> -2xFLAG	NM_001008697	pcDNA5/FRT/TO	This study	p1335
His <sub>6</sub> -2xFLAG-XRN2	NM_012255	pcDNA5/FRT/TO	Bohnsack lab	p810
GFP-XRN2	NM_012255	pcDNA5/FRT/TO	Bohnsack lab	p811
ZGPAT-His <sub>6</sub> -2xFLAG	NM_001195653	pcDNA5/FRT/TO	This study	p1333

**Table 2.4. Plasmids for recombinant protein expression used in this study**

Construct	CDS	Vector	Source	ID
ZZ-AGGF1_619-665-His <sub>7</sub>	NM_018046	pQE-80 (A15)	This study	p1501
ZZ-CHERP_841-891-His <sub>7</sub>	NM_006387	pQE-80 (A15)	This study	p1506
ZZ-CMTR1_87-133-His <sub>7</sub>	NM_015050	pQE-80 (A15)	This study	p1505
MBP-DHX15-His <sub>10</sub>	NM_001358	pQE-80 (A102)	Bohnsack lab	p1102
MBP-DHX15 <sub>E261Q</sub> -His <sub>10</sub>	NM_001358	pQE-80 (A102)	This study	p1310
MBP-DHX16-His <sub>10</sub>	NM_003587	pQE-80 (A102)	This study	p1481
ZZ-GPANK1_255-301-His <sub>7</sub>	NM_001199237	pQE-80 (A15)	This study	p1513
ZZ-GPATCH1_152-198-His <sub>7</sub>	NM_018025	pQE-80 (A15)	This study	p1491
ZZ-GPATCH2_467-513-His <sub>7</sub>	NM_018040	pQE-80 (A15)	This study	p1496
ZZ-GPATCH3_410-458-His <sub>7</sub>	NM_022078	pQE-80 (A15)	This study	p1516
ZZ-GPATCH4_11-57-His <sub>7</sub>	KJ902705	pQE-80 (A15)	This study	p1492
ZZ-GPATCH8_40-86-His <sub>7</sub>	NM_001002909	pQE-80 (A15)	This study	p1503
ZZ-GPATCH11_69-115-His <sub>7</sub>	AK294697	pQE-80 (A15)	This study	p1502
ZZ-GPKOW_164-210-His <sub>7</sub>	NM_015698	pQE-80 (A15)	This study	p1495
ZZ-GPKOW-His <sub>7</sub>	NM_015698	pQE-80 (A15)	This study	p1417
ZZ-NKRF_551-596-His <sub>7</sub>	NM_001173487	pQE-80 (A15)	This study	p1517
MBP-NKRF-His <sub>10</sub>	NM_001173487	pQE-80 (A102)	This study	p1040
ZZ-PINX1_26-72-His <sub>7</sub>	NM_017884	pQE-80 (A15)	This study	p1489
ZZ-RBM5_743-789-His <sub>7</sub>	NM_005778	pQE-80 (A15)	This study	p1499
ZZ-RBM6_1051-1097-His <sub>7</sub>	NM_005777	pQE-80 (A15)	This study	p1497
ZZ-RBM10_858-904-His <sub>7</sub>	NM_005676	pQE-80 (A15)	This study	p1514
ZZ-RBM17_235-283-His <sub>7</sub>	NM_001145547	pQE-80 (A15)	This study	p1498
ZZ-SON_2305-2351-His <sub>7</sub>	NM_138927	pQE-80 (A15)	This study	p1512
ZZ-SUGP1_562-609-His <sub>7</sub>	NM_172231	pQE-80 (A15)	This study	p1490
ZZ-SUGP2_1011-1057-His <sub>7</sub>	NM_001017392	pQE-80 (A15)	This study	p1504
ZZ-TFIP11_149-195-His <sub>7</sub>	NM_001008697	pQE-80 (A15)	This study	p1515
ZZ-ZGPAT_313-359-His <sub>7</sub>	NM_001195653	pQE-80 (A15)	This study	p1500

### 2.1.3 Small interfering (si)RNAs

**Table 2.5. siRNAs used in this study**

Name / target gene	Sense sequence (5'-3')	Reference
siAGGF1	GUCGGAAGAUGUUGGAGAA(dTdT)	Major et al., 2008
siCHERP	GGUUUAGGUCUAGAAAGAAGAAU(dAdC)	Sasaki-Osugi et al., 2013
siCMTR1	GUGAAGGAUUGGGUAAAUA(dTdT)	Schuberth-Wagner et al., 2015
siDHX15_1	GGUUAUAGUUAUGAGCGCUACUCUA(dTdT)	Mosallanejad et al., 2014
siDHX15_2	GAGAAGGAGUUGCGAGCUU(dTdT)	Dharmacon

(continued on next page)

siGPANK1	GGACUUACAUGAACCCUCGA(dTdT)	Qiagen
siGPATCH3	GCAUGAAGUUUCGGACAGA(dTdT)	Qiagen
siGPATCH8	CCGUGUCCUAGAAGUAGAA(dTdT)	Qiagen
siGPATCH11	GCUAUUUUGUGAUACGUGA(dTdT)	Qiagen
siGPKOW	CGGCCGCACCUUCAUCAA(dTdT)	Qiagen
siNKRF_1	GGCUAUGC UUGUGAAGUUUA(dTdT)	Tafforeau et al., 2013
siNKRF_2	GUAUUGAAGUUAGAGUUGU(dTdT)	Tafforeau et al., 2013
siNT (non-target)	UCGUAAAGUAAGCGCAACCC(dTdT)	Elbashir et al., 2001
siRBM5	AACUCGCAAUACUACUAUA(dTdT)	Qiagen
siRBM6	GAGUCAUGUUUGCUCGAUA(dTdT)	Qiagen
siRBM10	CAACGUGCGCGUCAUAAAG(dTdT)	This study
siRBM17	GCGUAAAGACAGACAUGAA(dTdT)	Qiagen
siSON	GCAUUUGGCCCAUCUGAGA(dTdT)	Ahn et al., 2011
siSUGP1	CGAUAAAGAAUAGCAGGGAA(dTdT)	Qiagen
siSUGP2	GCGAGCAGAUACAGGGUA(dTdT)	Qiagen
siTFIP11	GGAUUAGCAAGAAGCUCAC(dTdT)	Stanek et al., 2008
siXRN2_1	GGGAAGAAAUUUUGGCAAA(dTdT)	West et al., 2004
siXRN2_2	AAGAGUACAGAUGAUGAUG(dTdT)	West et al., 2004
siZGPAT	CGUUCUUCUGGAGGGAAA(dTdT)	Qiagen

[illegible]

Probe	Sequence (5'-3')
Actin	AGGGATAGCACAGCCTGGATAGCAAC
ETS1	CGGAGGCCCAACCTCTCCGACGACAGGTCGCCAGAGGACAGCGTGTCCAGC
ETS3	ACCGGTCACGACTCGGCA



(Table 2.7 continued)

5'ITS1	CCTCGCCCTCCGGGCTCCGTTAATGATC
ITS1	AGGGGTCTTTAAACCTCCGCGCCGGAACGCGCTAGGTAC
ITS2	GCTCTCTCTTTCCCTCTCCGTCTTCC
SCARNA2	CCTCGTCTATCTGATCAATTCATCAC
SCARNA5	CAGCTGCTCCATGATCCCATACACA
SCARNA10	CCTTGGCCCTGATACCCTGAACAT
SCARNA12	TCTGGCCTTACAGTGGGGAGTCATG
SCARNA13	GTCATTATCACCGTGGCAAC
U1 snRNA	GGTCAGCACATCCGGAGTGC
U2 snRNA	CATTTAATATATTGTCCTCGG
U4 snRNA	CCAGTGCCGACTATATTGC
U5 snRNA	GACTCAGAGTTGTTCTCTCC
U6 snRNA	GAACGCTTACGAATTTGCGTGTC

## 2.1.6 Antibodies

Table 2.8. Primary antibodies used in this study

Target	Supplier
DHX9	Bethyl (A300-854A)
DHX15	Bethyl (A300-390A)
DHX16	Bethyl (A301-537A)
DDX21	Bethyl (A300-628A)
DHX35	Abcam (ab179442)
Fibrillarin	Santa Cruz Biotechnology (sc-25397)
FLAG M2	Sigma-Aldrich (F3165)
NKRF	Bethyl (A304-016A)
NSUN5	Santa Cruz Biotechnology (sc-376147)
Nucleophosmin	Sigma-Aldrich (B0556)
PCNA	Santa Cruz Biotechnology (sc-56)
PES1	Bethyl (A300-903A)
PWP2	GeneTex (GTX105344)
RPL15	Aviva Systems Bio (ARP65141)
RPS3A	ProteinTech (14123-1-AP)
$\alpha$ -Tubulin	Sigma-Aldrich (T6199)
UTP14A	ProteinTech (11474-1-AP)
WBSCR22	Abgent (AP20254b)
XRN2	Bethyl (A301-103A)

**Table 2.9. Secondary antibodies used in this study**

Secondary antibody	Supplier
Goat anti-mouse - Alexa Fluor 488-conjugated	Jackson ImmunoResearch
Goat anti-rabbit - Alexa Fluor 488-conjugated	Jackson ImmunoResearch
Goat anti-mouse - Alexa Fluor 594-conjugated	Jackson ImmunoResearch
Goat anti-rabbit - Alexa Fluor 594-conjugated	Jackson ImmunoResearch
Goat anti-mouse - HRP-conjugated	Jackson ImmunoResearch
Goat anti-rabbit - HRP-conjugated	Jackson ImmunoResearch

## 2.2 Standard molecular biology methods

### 2.2.1 Molecular cloning

Cloning of the vector constructs used in this study was performed using standard molecular biology procedures (Sambrook and Russell, 2001). First, the coding sequences of the proteins of interest were amplified by PCR with Phusion polymerase (Thermo Fisher) either from existing plasmids or from HEK293 or HeLa cDNA. The general PCR reaction conditions and the primers used are listed in Table 2.10 and in Supplementary Table S1.

**Table 2.10. PCR parameters for cloning**

PCR reaction components (cloning)		
5X buffer	10 µl	
dNTP mix (10 mM)	1 µl	
Forward primer (10 µM)	1 µl	
Reverse primer (10 µM)	1 µl	
Template (cDNA or plasmid DNA)	2 µl	
H <sub>2</sub> O	34.5 µl	
Phusion polymerase	0.5 µl	
PCR reaction conditions (cloning)		
98°C (initial denaturation)	2 min	x1
98°C (denaturation)	30 sec	x35
50-58°C (annealing)	40 sec	
72°C (elongation)	1 min/kb	
72°C (final elongation)	10 min	x1

Ten percent of the PCR reaction was checked by agarose gel electrophoresis for the correct product size and the remainder was purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) following manufacturer's instructions. The vector backbones and the purified PCR products were digested with the corresponding restriction enzymes (Thermo Fisher) for 2 h at 37°C and gel-purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). Ligation reactions containing 50 ng of digested vector and five-fold molar excess of the PCR insert were carried out with 5 U of T4 DNA Ligase (Thermo Fisher)

either at 22°C for 1 h or at 16°C overnight. Afterwards, the ligation mix was transformed into chemically competent *E. coli* cells as described in 2.2.3 and plasmid DNA was extracted from single colonies and analyzed by Sanger sequencing to confirm cloning of the correct construct (2.2.4).

### 2.2.2 Site-directed mutagenesis

Primers for site-directed mutagenesis were designed with approximately 20 bases surrounding the region containing the desired mutations on each side (Supplementary Table S2). PCR reactions consisting of 50 ng plasmid DNA template and 125 ng each of the forward and reverse primers were carried out using the parameters listed in Table 2.11. Afterwards, the parental wild-type methylated plasmids were digested with 10 U of DpnI for 2 h at 37°C and the reactions were precipitated for 1 h at -80°C with 0.3 M sodium acetate pH 5.3 and 3 volumes of 100% ethanol. After centrifugation for 20 min at 20,000 g, the pellets were washed with 70% ethanol, resuspended in H<sub>2</sub>O and transformed in DH5α cells (2.2.3). Plasmid DNA extraction and verification of the mutagenesis by Sanger sequencing are described in 2.2.4.

**Table 2.11. PCR parameters for site-directed mutagenesis**

PCR reaction components (mutagenesis)		
10X buffer	5 μl	
dNTP mix (10 mM)	1 μl	
Forward primer	125 ng	
Reverse primer	125 ng	
Template (50 ng)	1 μl	
H <sub>2</sub> O	up to 50 μl	
PfuTurbo polymerase	0.5 μl	
PCR reaction conditions (mutagenesis)		
95°C (initial denaturation)	5 min	x1
95°C (denaturation)	30 sec	x12-18
55°C (annealing)	1 min	
68°C (elongation)	1 min/kb	

### 2.2.3 Transformation of *E. coli*

Chemically competent *E. coli* strains were thawed on ice, incubated for 20 min on ice with the DNA sample and transformed by heat-shock for 1 min at 42°C. Afterwards, the cells were briefly chilled on ice, LB media was added and samples were incubated for 1 h at 37°C with shaking. The bacteria were plated on agar plates containing the appropriate selection antibiotics (100 µg/ml ampicillin with or without 34 µg/ml chloramphenicol) and grown overnight at 37°C. *E. coli* strains used in this study were DH5α (Thermo Fisher) for cloning and plasmid propagation, and BL21 (DE3) pLysS (Novagen) and BL21 (DE3)

CodonPlus-RIL (Agilent) for recombinant protein expression. Cloning of GPATCH8-His<sub>6</sub>-2xFLAG was done using CopyCutter EPI400 cells (Epicentre) due to the instability of the full-length GPATCH8 insert.

#### **2.2.4 Plasmid DNA extraction**

Plasmid DNA was extracted using the NucleoSpin Plasmid kit (Macherey-Nagel) following manufacturer's protocol. Single colonies obtained after bacterial transformation were used to inoculate LB cultures containing selection antibiotics, which were grown overnight at 37°C. The bacterial cultures were centrifuged at 20,000 g for 1 min and the cell pellet was resuspended in 250 µl of Solution A1. For alkaline cell lysis, 250 µl of Solution A2 were added, the samples were gently mixed and incubated for 5 min at room temperature. The reactions were neutralized with 300 µl Solution A3 and centrifuged at 20,000 g for 10 min. The soluble fraction containing plasmid DNA was added to a silica spin column and centrifuged for 1 min at 20,000 g. The column was washed with 700 µl Solution A4 and centrifuged as before, followed by another centrifugation at 20,000 g for 2 min to remove the residual ethanol. To elute the plasmid DNA, 30 µl H<sub>2</sub>O were added and the columns were incubated for 1 min at room temperature followed by centrifugation at 20,000 g for 1 min. The purified plasmids were verified for the presence of the correct construct by Sanger sequencing (Eurofins Genomics) and the results were analyzed with the DNASTAR Lasergene 10 SeqMan software.

### **2.3 Cell culture-based methods**

#### **2.3.1 Cell culture growth conditions**

HeLa CCL-2 (ATCC) and HEK293 Flp-In T-REx (Thermo Fisher) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM; Thermo Fisher) supplemented with 10% fetal bovine serum (FBS; Merck) and penicillin-streptomycin (1:100 dilution; Thermo Fisher) at 37°C in a humidified atmosphere with 5% CO<sub>2</sub>. Cells were passaged every 2-3 days at a ratio of 1:5 or 1:10 using 0.25% Trypsin-EDTA (Thermo Fisher). Cell lines used in this study were checked for mycoplasma contamination with the *Mycoplasma*check service (Eurofins Genomics).

#### **2.3.2 Generation of inducible stable cell lines**

For generation of stable cell lines, HEK293 Flp-In T-REx cells (Thermo Fisher), which have an FRT site integrated into their genome, were plated in antibiotic-free media at a concentration of 350,000 cells/well in a 6-well plate. The following day, two wells for each construct were transfected with the pOG44 plasmid encoding the Flp recombinase and with

the appropriate pcDNA5/FRT/TO-derived plasmid (Table 2.3) to enable site-specific integration of the transgene at the FRT locus. Transfection was carried out with X-tremeGENE 9 DNA transfection reagent (Roche) and 100  $\mu$ l of transfection mix containing 91  $\mu$ l Opti-MEM (Thermo Fisher), 9  $\mu$ l transfection reagent, 0.6  $\mu$ g pcDNA5 plasmid and 1.8  $\mu$ g pOG44 plasmid were added to each well. The selection was started 48 h after transfection with 82.4  $\mu$ g/ml Hygromycin B (Applichem) and 10  $\mu$ g/ml Blasticidin S Hydrochloride (Applichem), and was maintained for 2 weeks. Afterwards, cells were generally cultured in medium without hygromycin and blasticidin, and tested regularly for expression of the transgene. To confirm expression of the encoded proteins, cells were induced for 24 h with 1  $\mu$ g/ml tetracycline, harvested with trypsin, lysed in 4X SDS loading dye and analyzed by SDS-PAGE and western blotting (2.4.1).

### **2.3.3 Gene knockdown with siRNAs**

Knockdowns were carried out either in HeLa or HEK293 cells with Lipofectamine RNAiMAX (Thermo Fisher) following manufacturer's protocol for reverse transfections. In general, 20-50 nM siRNAs (Table 2.5) were mixed with 5  $\mu$ l transfection reagent and 500  $\mu$ l Opti-MEM (Thermo Fisher) in the well of a 6-well plate and incubated for 15 min. Afterwards, the appropriate number of cells diluted in antibiotic-free media (e.g. approximately 120,000 HeLa cells or 275,000 HEK293 cells for a 3-day knockdown) were added to each well and knockdowns were carried out for 3-5 days as specified. The knockdowns were scaled up accordingly when required.

### **2.3.4 Immunofluorescence microscopy**

Glass coverslips were coated with 0.01% poly-L-lysine (Sigma-Aldrich) for 15 min at room temperature, washed with H<sub>2</sub>O and sterilized with UV light. The coverslips were placed in wells of 24-well plates and 25,000-50,000 cells were plated per well depending on the experiment. Cells were generally grown on coverslips for 2 days, and in the case of HEK293 stable cell lines, expression of the transgene was induced for 24 h with 1  $\mu$ g/ml tetracycline. All sample preparation steps for microscopy were carried out at room temperature. First, cells were fixed with 2.5% paraformaldehyde in PBS for 15 min and then washed two times with PBS for 5 min. Permeabilization was done for 20 min with 0.1% Triton X-100 in PBS and afterwards, cells were blocked for 1.5 h with PBS containing 1% FBS (Merck) and 0.1% Triton X-100. Protein localization was determined by staining with antibodies against the respective target or against the FLAG tag. Antibodies against markers for different cellular structures (nucleoli - UTP14A/NSUN5, centrosome - PCM1) were used for counterstaining. After blocking, the cells were incubated with the primary antibodies for 2 h, washed three times for 10 min each with PBS and stained for 1.5 h with Alexa Fluor 488 and 594-conjugated secondary antibodies (Table 2.9). The cells were then washed with PBS as

before and mounted on microscope slides using VECTASHIELD Antifade Mounting Medium with DAPI (Vector Labs) for visualization of nuclei. The slides were imaged using a ZEISS LSM 510 META laser scanning microscope. For the GFP-XRN2 localization studies, the seeding of the cells was coupled with setting up siRNA-mediated knockdowns and the cells were fixed and subjected to immunofluorescence detection with an UTP14A antibody as above, while the GFP signal was visualized directly.

## **2.4 Protein analysis methods**

### **2.4.1 SDS-PAGE and western blotting**

For SDS-polyacrylamide gel electrophoresis (SDS-PAGE), protein samples were mixed with 4X SDS loading dye and denatured at 95°C for 10 min. Electrophoresis was carried out using 25 mA/gel in a discontinuous system with different composition and pH of the resolving and stacking gel buffers (Table 2.2). After the run, the gels were stained for 1 h with a Coomassie solution (0.1% Coomassie R-250, 10% acetic acid, 40% methanol) followed by destaining with a solution of 10% acetic acid and 20% methanol to enable direct protein visualization.

Alternatively, the samples were transferred onto a Hybond P 0.45 PVDF blotting membrane (GE Healthcare) in a wet-transfer system. The membrane was first activated in 100% methanol for 2 min and the transfer was carried out for 75 min at 100 V in western blotting transfer buffer. For experiments involving detection of the G-patch protein SON, which has a molecular weight of more than 260 kDa, the buffer was supplemented with 0.05% SDS and the transfer was done for 16 h at 25 V. After transfer, the membrane was blocked for 1 h at room temperature in 5% milk in TBS with 0.1% Tween-20 (TBS-T), followed by incubation with primary antibodies (Table 2.8) overnight at 4°C. Next, the membrane was washed three times for 10 min each in TBS-T and incubated with HRP-coupled secondary antibodies (Table 2.9) for 1 h at room temperature. After removal of the secondary antibody, washing steps with TBS-T were carried out as before and detection was done by exposure to X-ray films using Immobilon Western Chemiluminescent HRP Substrate (Merck).

### **2.4.2 Immunoprecipitation (IP) of protein complexes**

For each sample, approximately  $7 \times 10^6$  cells were plated in a 15 cm dish and the next day, expression of the FLAG-tagged proteins was induced with 1 µg/ml tetracycline. After 24 h, the cells were washed with PBS, harvested with 0.25% Trypsin-EDTA (Thermo Fisher) and centrifuged for 3 min at 1000 g. The cell pellet was resuspended in 1 ml IP buffer (20 mM HEPES-NaOH pH 7.5, 250 mM NaCl, 0.5% Triton X-100, 10% glycerol, 0.5 mM EDTA) supplemented with cOmplete Mini Protease Inhibitor Cocktail (Roche). Cells were lysed by

sonication (3 cycles of 15 sec with 0.3 sec on/0.7 off) using a Branson Digital Sonifier set to 20% amplitude, followed by centrifugation for 10 min at 20,000 g and 4°C to remove the insoluble material. The lysate was added to 30 µl slurry of pre-equilibrated anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) and binding was carried out for 2 h at 4°C in the presence of 50 µg/ml RNase A (Applichem). Afterwards, the beads were washed five times with the IP buffer and bound complexes were eluted with 250 µg/ml FLAG Peptide (Sigma-Aldrich) diluted in IP buffer for 30 min at 4°C. The eluates were precipitated with 20% trichloroacetic acid for 20 min on ice and centrifuged for 20 min at 20,000 g and 4°C. The pellets were then washed with ice-cold acetone and air-dried briefly. For mass spectrometry analysis (2.4.4), the samples were resuspended in 1X NuPAGE LDS Sample Buffer (Thermo Fisher) supplemented with 50 mM DTT and denatured at 70°C for 10 min. For SDS-PAGE and western blotting analysis (2.4.1), the pellets were resuspended first in 100 mM Tris-HCl pH 8.4, mixed with 4X SDS loading dye and boiled at 95°C for 10 min.

To study the interactions between DHX15, NKRF and XRN2 (3.2.4), IPs using antibodies against the FLAG tag or against NKRF were performed as above with some modifications. The buffer for cell lysis consisted of 20 mM HEPES pH 8.0, 150 mM KCl and 0.5 mM EDTA. The cells were lysed by sonication and centrifuged to remove cell debris as described above. Afterwards, the lysates were supplemented with 10% glycerol, 1.5 mM MgCl<sub>2</sub> and 0.2% Triton X-100, and incubated with the pre-equilibrated beads for 2 h at 4°C. For IP with an anti-NKRF antibody, Protein G Sepharose beads (GE Healthcare) were prepared beforehand by an overnight incubation at 4°C with 5 µl antibody followed by washing with IP buffer (20 mM HEPES pH 8.0, 150 mM KCl, 10% glycerol, 1.5 mM MgCl<sub>2</sub> and 0.2% Triton X-100). After binding, the beads were washed with the IP buffer and in the case of the anti-NKRF IP, the co-precipitated proteins were eluted at 95°C with 4X SDS loading dye. For the anti-FLAG IP, the bound proteins were eluted with 250 µg/ml FLAG Peptide (Sigma-Aldrich) and precipitated with 20% trichloroacetic acid as above. The samples were then analyzed by SDS-PAGE and western blotting (2.4.1). In the case of the IP experiments coupled to siRNA-mediated knockdowns, the cells were treated for 96 h with 50 nM siRNAs and expression of the FLAG-tagged proteins was induced 24 h prior to harvesting.

#### **2.4.3 Purification of nucleoli and preparation of nucleolar lysates**

HEK293 cell lines expressing FLAG-tagged DHX15 or the FLAG tag only were induced for 24 h with 1 µg/ml tetracycline and nucleoli were isolated based on Chamousset et al., 2010 with a few changes. Cells (4x15 cm plates) were harvested with trypsin, washed twice with ice-cold PBS and lysed for 10 min on ice in 5 ml lysis buffer containing 20 mM Tris-HCl pH 7.4, 10 mM KCl, 3 mM MgCl<sub>2</sub>, 0.1% NP-40 and 10% glycerol. The released nuclei were

pelleted by centrifugation at 1350 g for 10 min, resuspended in 3 ml of solution S1 (0.25 M sucrose, 10 mM  $\text{MgCl}_2$ ) and layered on top of 3 ml solution S2 (0.35 M sucrose, 0.5 mM  $\text{MgCl}_2$ ). After centrifugation at 1430 g for 5 min, the nuclear pellet was resuspended in 3 ml solution S2 and sonicated six times for 10 sec each at 20% amplitude using a Branson Digital Sonifier. The lysed sample containing intact nucleoli was layered on top of 3 ml solution S3 (0.88 M sucrose, 0.5 mM  $\text{MgCl}_2$ ) and centrifuged at 3000 g for 10 min. The nucleolar pellet was resuspended in 500  $\mu\text{l}$  solution S2 and centrifuged again at 1430 g for 5 min to remove contaminants. The isolated nucleoli were disrupted by incubation for 30 min on ice in 400  $\mu\text{l}$  high-salt buffer (20 mM HEPES pH 8.0, 500 mM KCl, 0.5 mM EDTA, 0.8% Triton X-100, 0.4% CHAPS) supplemented with 16 U TURBO DNase (Thermo Fisher), followed by sonication as described for the IP protocol. The nucleolar lysate was diluted 1:4 with a buffer containing 20 mM HEPES pH 8.0, 2 mM  $\text{MgCl}_2$  and 13.3% glycerol, and immunoprecipitation with anti-FLAG M2 beads (Sigma-Aldrich) was performed as detailed above.

#### **2.4.4 Liquid chromatography tandem mass spectrometry (LC-MS/MS)**

Protein IP samples obtained as described in 2.4.2 were separated on 4-12% NuPAGE Bis-Tris gels (Thermo Fisher) and entire lanes were excised, divided into 12 fragments and digested with trypsin. The samples were analyzed on a mass spectrometer with two technical replicates for each. Proteins were identified with Mascot (Matrix Science) by searching against the UniProt human protein database. The results were further processed with Scaffold (Proteome Software) and proteins containing at least two detected peptides were identified at a false discovery rate of less than 1%. These steps were performed by the Proteomics Service Facility (University Medical Center Göttingen). For further data analysis, the proteins were ranked based on the total spectral counts and the fold change for each protein was calculated relative to the control sample. To enable calculations, a spectral count of 1 was added to proteins that had zero counts. The final results were expressed as the mean of two technical replicates or, when available, two biological replicates.

#### **2.4.5 Separation of (pre)-ribosomal complexes on sucrose gradients**

HeLa cells grown to 80% confluency in a 10 cm plate were detached with trypsin, centrifuged for 3 min at 1000 g and resuspended in 0.5 ml of lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 5 mM  $\text{MgCl}_2$ , 1 mM DTT). The cells were lysed by sonication as described for the IP method (2.4.2) and the lysates were cleared by centrifugation for 10 min at 20,000 g and 4°C. The obtained whole cell extracts (400  $\mu\text{l}$ ) were layered on top of 6 ml sucrose gradients (10-45%), which were prepared using a Gradient Master (BioComp). The separation of complexes was carried out for 16 h at 23,500 rpm and 4°C



in a SW40 Ti rotor (Beckman Coulter). After centrifugation, the gradient was fractionated into 530 µl samples that were precipitated with 20% trichloroacetic acid (2.4.2) and prepared for SDS-PAGE and western blotting analysis (2.4.1). When indicated, cells were treated with siRNAs for 96 h prior to the experiment.

## **2.5 RNA analysis methods**

### **2.5.1 Total RNA extraction**

Total RNA was extracted using TRI Reagent (Sigma-Aldrich) according to manufacturer's recommendations. Briefly, the media was removed, cells were washed once with PBS and the appropriate amount of TRI Reagent was added directly to the cells. The lysate was pipetted up and down several times to disrupt RNA-protein complexes and the samples were incubated for 5 min at room temperature. Next, chloroform (1/5 volume) was added and the samples were thoroughly mixed and incubated for 2-3 min at room temperature before centrifuging for 20 min at 20,000 g and 4°C. The upper aqueous phase containing RNA was transferred to a fresh tube and 1 volume isopropanol was added. Samples were incubated for 5 min and then centrifuged for 15 min at 20,000 g and 4°C to precipitate the RNA. The RNA pellet was washed with 75% ethanol, air-dried briefly and resuspended in H<sub>2</sub>O at a concentration of 0.5-2 µg/µl.

### **2.5.2 Northern blotting**

To analyze small RNA species, total RNA samples or elution samples from RNA IP experiments (2.6.2) were mixed with 2X formamide loading dye and denatured at 90°C for 3 min. The samples were separated on 8% or 12% 7M urea-polyacrylamide gels in 1X TBE and transferred to a Hybond-N membrane (GE Healthcare) for 2 h at 60 V and 4°C in the presence of 0.5X TBE. Alternatively, for northern blotting analysis of pre-ribosomal RNA processing, 5 µg total RNA were mixed with 5 volumes of glyoxal loading dye and incubated at 55°C for 1 h. The samples were separated for 16 h at 60 V on a 1.2% agarose gel in 1X BTPE. Prior to transfer, the gel was incubated for 20 min in 0.1 M NaOH followed by two washes of 15 min each in 0.5 M Tris-HCl pH 7.4 and 0.3 M NaCl, and a final equilibration step of 20 min in 6X SSC. The samples were then transferred to a Hybond-N membrane (GE Healthcare) by vacuum blotting for 2 h at 300 mbar in the presence of 6X SSC. After transfer, the RNA was crosslinked to the membrane at 254 nm with 240 mJ/cm<sup>2</sup> in a Stratalinker UV Crosslinker (Stratagene). The membrane was pre-hybridized in SES1 buffer for 40 min at 37°C and <sup>32</sup>P-labelled oligonucleotide probes were used to detect the RNA species of interest.

To generate labelled probes, 20 µl reactions containing 20 pmol DNA oligonucleotide (Table 2.7), 20 µCi [ $\gamma$ - $^{32}$ P]-ATP (PerkinElmer) and 10 U T4 Polynucleotide Kinase (Thermo Fisher) were set up and incubated at 37°C for 40 min. The labelled oligonucleotides were diluted in SES1 buffer and added to the membrane for overnight incubation at 37°C. The next day, the membrane was washed with 6X SSC and with 2X SSC + 0.1% SDS for 30 min each, and exposed to a phosphor screen (GE Healthcare). Detection of the radioactive signals was done with a Typhoon FLA 9500 scanner (GE Healthcare).

### 2.5.3 Pulse-chase analysis of RNA

HeLa cells were treated with siRNAs for 84 h as described in 2.3.3 and, prior to metabolic labelling, the cells were grown for 1 h at 37°C in phosphate-free DMEM. Labelling was carried out for 1 h at 37°C in phosphate-free DMEM supplemented with 10 µCi/ml [ $^{32}$ P]-orthophosphate (PerkinElmer). Afterwards, the labelling media was replaced with complete DMEM and the cells were harvested at specific time points for extraction of total RNA (2.5.1). The RNA samples were analyzed by agarose gel electrophoresis and northern blotting, followed by detection of the radioactive signals corresponding to the newly-synthesized (pre)-rRNAs by phosphorimaging (2.5.2). In addition, the levels of actin mRNA were determined using a specific northern blotting probe (Table 2.7).

### 2.5.4 Quantitative and gel-based reverse transcription PCR (RT-PCR)

First-strand cDNA synthesis was performed using the SuperScript III Reverse Transcriptase kit (Thermo Fisher) according to manufacturer's instructions. Two µg of template RNA were denatured for 5 min at 65°C in a volume of 13 µl with 10 nmol dNTP mix (Roche) and 50 pmol anchored oligo(dT)<sub>24</sub>VN primer or 75 pmol random hexamer primer. The samples were briefly placed on ice and 7 µl of reverse transcription mix containing 4 µl 5X first-strand buffer, 1 µl 100 mM DTT, 1 µl RiboLock RNase Inhibitor (Thermo Fisher) and 1 µl SuperScript III Reverse Transcriptase were added. The reactions were incubated for 1 h at 50°C and subsequently the reverse transcriptase was inactivated at 70°C for 15 min. The obtained cDNA was used for quantitative RT-PCR (qRT-PCR) or for PCR coupled to gel-based analysis depending on the experiment.

Quantitative RT-PCR was done using the LightCycler 480 SYBR Green I Master kit (Roche). Ten µl reactions containing 0.65X SYBR Green mix, 0.33 pmol forward and reverse primer (Supplementary Table S3) and 3 µl cDNA diluted as necessary (generally 1:10 dilution) were pipetted in each well in triplicates. Samples were amplified in a LightCycler 480 machine (Roche) using the following program: 5 min pre-incubation at 95°C and 50 cycles of 10 sec denaturation at 95°C, 20 sec annealing at 58°C and 15 sec amplification at 72°C. Melting curve analysis was subsequently performed and involved

incubation for 10 sec at 95°C and for 1 min at 55°C followed by continuous acquisition of fluorescence until 97°C. The presence of a single melting peak indicated specific amplification of one product. Crossing point (Cp) values were determined with the second derivative maximum method and relative quantification was done based on the  $\Delta\Delta C_t$  method. To account for differences in the input material, the results were normalized to the expression level of housekeeping genes (GAPDH or EMC7). All analysis steps were performed using the LightCycler 480 software.

For alternative splicing analysis, PCR reactions were carried out as specified in Table 2.12 using the primers listed in Supplementary Table S3. The samples (1/5 of the PCR reactions) were mixed with 6X DNA loading dye and separated on a 4% polyacrylamide gel in 1X TBE. The PCR products were stained with 0.5 µg/ml ethidium bromide for 10 min and visualized using a Typhoon FLA 9500 scanner (GE Healthcare). Quantification was done with ImageQuant TL 8.1 and the *percentage spliced in* (PSI) values were calculated as the ratio of the inclusion isoform relative to the sum of the inclusion and skipping products. The results were expressed as the mean of two biological replicates  $\pm$  standard error of the mean (s.e.m), unless otherwise stated.

**Table 2.12. PCR parameters for alternative splicing analysis**

PCR reaction components (alternative splicing)		
5X buffer	5 µl	
dNTP mix (10 mM)	0.5 µl	
Forward primer (10 µM)	0.5 µl	
Reverse primer (10 µM)	0.5 µl	
cDNA template (1:2 dilution)	1 µl	
H <sub>2</sub> O	17.25 µl	
Phusion polymerase	0.25 µl	
PCR reaction conditions (alternative splicing)		
98°C (initial denaturation)	2 min	x1
98°C (denaturation)	30 sec	x26-30
55°C (annealing)	40 sec	
72°C (elongation)	25 sec	
72°C (final elongation)	10 min	x1

### 2.5.5 Site-specific RNase H cleavage assay

Total RNA (250 ng) was mixed with 10 pmol of 2'-O-methyl RNA-DNA chimeric oligonucleotides (Table 2.6) and denatured at 85°C for 3 min, followed by incubation at 37°C for 10 min. The reactions were supplemented with 40 U RiboLock RNase Inhibitor (Thermo Fisher) and 5 U RNase H (NEB) in the presence of 1X RNase H buffer (NEB) and were incubated for 1 h at 37°C. Afterwards, the enzyme was inactivated with 4 volumes of

0.3 M sodium acetate pH 5.3 and 1 mM EDTA. The RNA was extracted with phenol-chloroform-isoamyl alcohol (25:24:1) and precipitated overnight at -20°C with 3 volumes of 100% ethanol and 1 µl glycogen. The samples were analyzed by urea-polyacrylamide gel electrophoresis and northern blotting as described in 2.5.2.

## **2.6 Next-generation sequencing-based methods and data analysis**

### **2.6.1 RNA-seq sample preparation**

HEK293 cells were treated with the indicated siRNAs and total RNA was extracted using TRI Reagent (Sigma-Aldrich) as described in 2.5.1. Approximately 10 µg of RNA were incubated with 2 U TURBO DNase (Thermo Fisher) for 15 min at 37°C, in a reaction supplemented with 40 U RNasin Ribonuclease Inhibitor (Promega). The DNase-treated RNA was purified using the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. Briefly, the reactions were brought to 100 µl with nuclease-free H<sub>2</sub>O (Qiagen), and 350 µl buffer RLT and 250 µl 100% ethanol were added. The samples were mixed well and the RNA was bound to the columns by centrifugation at 11,000 g for 15 sec. The columns were washed first with 500 µl RPE buffer and then with 500 µl of 80% ethanol. The residual ethanol was removed by centrifugation at 20,000 g for 5 min and the RNA was eluted in 14 µl nuclease-free H<sub>2</sub>O (Qiagen). Two µg RNA for each sample were used for rRNA depletion and library preparation with the Illumina TruSeq Stranded Total RNA Kit followed by 50 bp single-end sequencing on an Illumina HiSeq 4000 sequencer, which yielded approximately 20-50 million reads per sample. Library preparation and next-generation sequencing were carried out by the Transcriptome and Genome Analysis Laboratory (University Medical Center Göttingen).

### **2.6.2 Crosslinked RNA immunoprecipitation (RIP)-seq**

HEK293 stable cell lines expressing FLAG-tagged proteins were grown to confluency in 15 cm plates (six plates for each sample) and prior to crosslinking protein expression was induced for 24 h with 1 µg/ml tetracycline. The cells were washed once with PBS and then crosslinked in the presence of 6 ml PBS at 254 nm with 400 mJ/cm<sup>2</sup> in a Stratalinker UV Crosslinker (Stratagene). The crosslinked cells were collected by scraping and cells from three plates were pooled together, centrifuged at 800 g for 3 min and resuspended in 1 ml RIP buffer (20 mM HEPES-NaOH pH 7.5, 250 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.5% Triton X-100, 10% glycerol) containing cOmplete Mini Protease Inhibitor Cocktail (Roche) and 80 U RNasin Ribonuclease Inhibitor (Promega). The cells were lysed by sonication with a Branson Digital Sonifier at 25% amplitude in 3 cycles of 15 sec (0.3 sec on/0.7 sec off) with 30 sec pause in between. The cell lysate was cleared by centrifugation for 15 min at 20,000

g and 4°C, and incubated with 75 µl slurry of pre-equilibrated anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) for 2-3 h. The beads were then washed five times with RIP buffer and bound complexes were eluted for 2 h at 4°C with 250 µg/ml FLAG Peptide (Sigma-Aldrich) diluted in RIP buffer. Afterwards, the eluates were incubated with 2 U TURBO DNase (Thermo Fisher) for 20 min at 37°C in the presence of 40 U RNasin Ribonuclease Inhibitor (Promega). Eluates corresponding to the same sample were pooled together and treated with 275 µg/ml Proteinase K (Roche) for 2 h at 42°C in a reaction containing 1% SDS and 0.5 mM EDTA. The released RNA was extracted with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) in the presence of 0.3 M sodium acetate pH 5.2. After centrifugation for 20 min at 20,000 g and 4°C, the upper phase was transferred to a new tube and the RNA was precipitated overnight at -20°C with 2.5 volumes 100% ethanol and 1 µl glycogen. The RNA pellet was washed twice with 75% ethanol and resuspended in nuclease-free H<sub>2</sub>O (Qiagen). The RNA eluates were used for rRNA depletion and library preparation with the Illumina TruSeq Stranded Total RNA Kit. Sequencing of the libraries was done in single-end mode with a read length of 50 bp on an Illumina HiSeq 4000 sequencer and generated approximately 50 million reads per sample. Library preparation and sequencing were performed by the Transcriptome and Genome Analysis Laboratory (University Medical Center Göttingen). Alternatively, the RNA eluates were analyzed by northern blotting as described in 2.5.2.

### **2.6.3 Crosslinking and analysis of cDNA (CRAC)**

Stable cell lines expressing the FLAG tag or FLAG-tagged NKRF were induced for 24 h with 1 µg/ml tetracycline and approximately  $1.2 \times 10^8$  cells (3x15 cm plates) per sample were used. When specified, the cells were treated for 6 h with 100 µM of the photoactivatable ribonucleoside 4-thiouridine (Sigma-Aldrich) prior to crosslinking. The protein-RNA crosslinking and anti-FLAG immunoprecipitation steps were performed as in 2.6.2 with the following modifications. Crosslinking at 254 nm (UV-CRAC) was done with 2400 mJ/cm<sup>2</sup>, while 360 mJ/cm<sup>2</sup> were used for crosslinking at 365 nm in the presence of 4-thiouridine (photoactivatable ribonucleoside-enhanced-CRAC; PAR-CRAC). Cells were harvested by scraping in 1 ml of TNM150 buffer (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.1% NP-40, 5 mM 2-mercaptoethanol) and lysates were prepared as described above except that sonication was done at 40% amplitude. The cleared lysates were incubated with 100 µl slurry of pre-equilibrated anti-FLAG M2 beads (Sigma-Aldrich) for 2-3 h at 4°C, followed by washes with TNM150 and TNM1000 (50 mM Tris-HCl pH 7.8, 1 M NaCl, 1.5 mM MgCl<sub>2</sub>, 0.1% NP-40, 5 mM 2-mercaptoethanol). Elution was carried out overnight at 4°C with 250 µg/ml FLAG Peptide (Sigma-Aldrich) diluted in TNM150.

Next, the eluates were treated with 0.1 U RNase-It Ribonuclease Cocktail (Agilent) for 30 sec at 37°C. The samples were then supplemented with 6 M guanidine hydrochloride, 300 mM NaCl and 10 mM imidazole, and incubated with 50 µl Ni-NTA beads (Qiagen) that had been pre-equilibrated with WB1 (50 mM Tris-HCl pH 7.8, 300 mM NaCl, 0.1% NP-40, 6 M guanidine hydrochloride, 10 mM imidazole, 5 mM 2-mercaptoethanol). After binding for 2 h at 4°C, the beads were washed with WB1 and PNK buffers (50 mM Tris-HCl pH 7.8, 10 mM MgCl<sub>2</sub>, 0.5% NP-40, 5 mM 2-mercaptoethanol). The next steps involved dephosphorylation, ligation of 3' and 5' linkers and radioactive labelling by 5' phosphorylation of the RNA. These reactions were carried out in the presence of 1X PNK buffer and 80 U RNasin Ribonuclease Inhibitor (Promega) as detailed below. Dephosphorylation was performed for 30 min at 37°C with 8 U TSAP Thermosensitive Alkaline Phosphatase (Promega) and afterwards the beads were washed with WB1 and PNK buffers. Ligation reactions containing 1 µM 3' linker, 10% PEG8000 (Sigma-Aldrich) and 800 U T4 RNA Ligase 2, Deletion Mutant (Epicentre) were set up and incubated overnight at 16°C. The beads were washed as before and phosphorylation with 80 U T4 Polynucleotide Kinase (NEB) and 40 µCi [ $\gamma$ -<sup>32</sup>P]-ATP (PerkinElmer) was done at 37°C for 40 min, followed by a 20 min incubation at 37°C with 1.25 mM ATP-lithium salt (Roche). Next, ligation of the 5' linker (1.25 µM) was carried out overnight at 16°C with 40 U T4 RNA Ligase 1 (NEB) and 1 mM ATP. The beads were washed with WB2 (50 mM Tris-HCl pH 7.8, 50 mM NaCl, 10 mM imidazole, 0.1% NP-40, 5 mM 2-mercaptoethanol) and the crosslinked protein-RNA complexes were eluted with the same buffer containing 150 mM imidazole instead. The eluates were concentrated by trichloroacetic acid precipitation and denatured in 1X NuPAGE LDS Sample Buffer (Thermo Fisher) as described in 2.4.2. This was followed by separation on 4-12% NuPAGE Bis-Tris gels (Thermo Fisher) using MES buffer and western blotting transfer onto a Hybond-C membrane (GE Healthcare). The membrane was exposed to X-ray films to detect the radioactive signals and the regions containing crosslinked complexes were excised and incubated overnight at 42°C with WB2 supplemented with 1% SDS, 5 mM EDTA and 275 µg/ml Proteinase K (Roche). The RNA was recovered by phenol-chloroform-isoamyl alcohol extraction as described in 2.6.2 and used for reverse transcription with SuperScript III Reverse Transcriptase (Thermo Fisher), which was performed as in 2.5.4 with a primer specific for the 3' linker.

PCR amplification of the resulting cDNA was done with the TaKaRa LA Taq DNA polymerase kit (Clontech) using the conditions shown in Table 2.13. The PCR products were separated on 3% MetaPhor agarose (Lonza) gels in 1X TBE and extracted using the MinElute Gel Extraction Kit (Qiagen) according to manufacturer's recommendations. The concentration of the cDNA libraries was determined using a Qubit fluorometer (Thermo

Fisher) and next-generation sequencing was performed by the Transcriptome and Genome Analysis Laboratory (University Medical Center Göttingen) on a HiSeq 2500 system (Illumina).

**Table 2.13. PCR parameters for CRAC**

PCR reaction components (CRAC)		
10X buffer	5 µl	
dNTP mix (2.5 mM)	2.5 µl	
Forward primer (10 µM)	1 µl	
Reverse primer (10 µM)	1 µl	
cDNA template	1 µl	
H <sub>2</sub> O	39 µl	
TaKaRa LA Taq polymerase	0.5 µl	
PCR reaction conditions (CRAC)		
98°C (initial denaturation)	2 min	x1
98°C (denaturation)	30 sec	x40
60°C (annealing)	40 sec	
68°C (elongation)	40 sec	
72°C (final elongation)	5 min	x1

## 2.6.4 Next-generation sequencing data analysis

Raw sequencing data obtained from the RNA-seq and RIP-seq experiments were processed with Flexbar 2.7 (Dodt et al., 2012) to discard low-quality reads and to remove contaminating adapter sequences. In addition, the RNA-seq knockdown reads were trimmed to 45 nt and shorter reads were discarded to fit the requirements of the downstream analyses. After quality-control, the reads were mapped to the human GRCh38.p12 genome assembly using the STAR RNA-seq aligner (Dobin et al., 2013). For spliced alignments, only reads crossing annotated splice junctions were selected. The mapped reads were then analyzed with featureCounts (Liao et al., 2014) to obtain cumulated read counts for all annotated genes, which were expressed as counts per million (CPM) relative to the total number of mapped reads. For the RIP-seq samples, a highly variable accumulation of reads mapping antisense to annotated genomic features was observed, which were considered to be an artefact introduced during library preparation and were therefore excluded from the analysis. To determine changes in gene expression levels, a pseudocount of 0.5 was assigned to all genes to enable calculation of the log<sub>2</sub> fold change between the sample of interest and the control.

Alternative splicing analysis was done with MISO (Katz et al., 2010) based on a set of precomputed splicing events derived from annotated exon-intron boundaries, which are classified into *cassette exon* (CE), *alternative 5' and 3' splice site* (A5'SS and A3'SS), *mutually exclusive exons* (MXE) and *retained intron* (RI) events (Figure 3.11). Each splicing

event can lead to two outcomes that can be described in general as an inclusion and an exclusion isoform, and MISO calculates a *percentage spliced in* (PSI) value to denote the expression of the inclusion isoform. For an alternatively spliced exon, PSI is estimated based on the reads that map to that exon and to its junctions with the adjacent constitutive exons relative to the reads that span the junction between the two constitutive exons. Changes in alternative splicing were detected using a threshold of at least 20% difference in PSI values between the knockdown sample and the control, and a Bayes factor greater than or equal to 10, which is a measure of the likelihood that the event is indeed differentially expressed. For visualization of RNA-seq and RIP-seq data, the read coverage for specific genes was analyzed using Integrative Genomics Viewer (IGV; Robinson et al., 2011) and scatterplots and Venn diagrams were generated using packages available in RStudio. The analysis of CRAC sequencing data was done using a bioinformatics pipeline that had already been established in the lab. Briefly, raw sequencing reads were processed with Flexbar 2.7 (Dodt et al., 2012) and the pyCRAC software (Webb et al., 2014) to remove low-quality sequences, contaminating adapters and duplicated reads, and to discard reads shorter than 21 nt. The remaining reads were mapped with Bowtie (Langmead et al., 2009) to the GRCh37.75 human genome version and further analysis steps were done using tools from the pyCRAC software. The workflow for RNA-seq and RIP-seq data analysis was established in collaboration with Jens Kretschmer, who performed all the bioinformatics analysis steps described above.

## **2.7 In vitro methods**

### **2.7.1 Recombinant protein expression and purification**

Recombinant protein expression plasmids (Table 2.4) were transformed in BL21 (DE3) CodonPlus-RIL cells or, in the case of the DHX15 constructs, in BL21 (DE3) pLysS (2.2.3). Single colonies were used to inoculate LB cultures, which were grown overnight at 37°C. The next day, 1 l cultures were initiated from the overnight cultures at a 1:1000 dilution and grown at 37°C until an O.D. of 0.6. Induction of protein expression was done either with 500 mM IPTG for 3 h at 37°C for the G-patch domains or with 250 mM IPTG at 18°C overnight in the case of full-length DHX15, DHX16, GPKOW and NKRF. The cultures were harvested by centrifugation at 5000 g for 10 min and after a wash with PBS, the cell pellet was resuspended in lysis buffer and purification was carried out as described below. For the G-patch domains, 1 l of bacterial culture was used, while purification of full-length proteins required approximately 6-10 l of culture. All purification steps were done at 4°C.



*Purification of full-length DHX15, DHX16, GPKOW and NKRF*

The cells were resuspended in 16 ml of lysis buffer (50 mM Tris-HCl pH 7.4, 600 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.5% Triton X-100, 10% glycerol, 25 mM imidazole, 2 mM 2-mercaptoethanol) per liter of culture and lysed with EmulsiFlex-C3 (Avestin) by three passes at 10,000 psi. The lysate was cleared by centrifugation at 50,000 g for 30 min and incubated with pre-equilibrated Ni-NTA beads (Qiagen) for 1.5 h. Approximately 125 µl of slurry were used per liter of culture. After binding, the beads were washed with 20 ml of wash buffer 1 (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 10% glycerol, 40 mM imidazole) followed by a high-salt wash with wash buffer 2 (same as wash buffer 1 but with 1 M NaCl instead) and a final washing step with wash buffer 1. The bound proteins were eluted with elution buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 10% glycerol, 250 mM imidazole) and the fractions were checked by SDS-PAGE. The fractions containing the highest amount of protein were pooled together and buffer exchange was done on PD-10 columns (GE Healthcare) against a buffer with 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1.5 mM MgCl<sub>2</sub> and 10% glycerol.

*Purification of G-patch domains*

The cell pellets were resuspended in 18 ml of lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 10% glycerol, 5 mM imidazole) and sonicated on ice using a Branson Digital Sonifier for 4 cycles at 45% amplitude with 0.5 sec on/0.5 sec off pulses and with 30 sec pause between the cycles. The lysate was cleared by centrifugation at 20,000 g for 20 min and bound to 1.5 ml Ni-NTA resin (Roche) for 1 h. The beads were then washed with 10 ml wash buffer 1 (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 10 mM imidazole), 10 ml wash buffer 2 (50 mM Tris-HCl pH 7.4, 1 M NaCl, 1 mM MgCl<sub>2</sub>, 10 mM imidazole) and again with 10 ml wash buffer 1. Elution was carried out with a buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub> and 250 mM imidazole. The fractions containing the highest amount of protein were pooled together and dialyzed against a buffer with 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub> and 10% glycerol using Spectra/Por Dialysis Tubing (Spectrum Labs) with a cutoff of 6-8 kDa. All recombinant proteins purified in this study were stored at -80°C and protein concentration was determined using Coomassie Plus (Bradford) assay (Thermo Fisher).

**2.7.2 Steady-state ATPase assay**

The ATPase activity was determined using an NADH-based enzymatic assay, in which the hydrolysis of ATP to ADP leads, in several steps, to the oxidation of NADH to NAD<sup>+</sup> and to a concomitant decrease in absorbance at 340 nm (Kiianitsa et al., 2003). Reactions containing 45 mM Tris-HCl pH 7.4, 25 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 300 µM NADH, 1 mM

phosphoenolpyruvate, 20 U/ml pyruvate kinase/lactic dehydrogenase, 4 mM ATP and 2  $\mu$ M polyU<sub>32</sub> RNA oligonucleotide (IDT) were set up, and recombinantly purified MBP-DHX15-His<sub>10</sub> or MBP-DHX16-His<sub>10</sub> were added to a final concentration of 250 nM. The reactions were further supplemented with 1.5  $\mu$ M ZZ-tagged G-patch domains as indicated. The experiments were carried out at 30°C and the absorbance at 340 nm was measured every min for 50 min using a BioTEK Synergy plate reader. The amount of ATP hydrolyzed, which is equimolar to the amount of NADH oxidized, was determined from the slope of the linear absorbance decrease and results from three experiments were expressed as mean  $\pm$  s.e.m.

### 2.7.3 Fluorescence anisotropy

To determine RNA binding affinities, the anisotropy of an ATTO647N-labelled RNA substrate (Table 2.6) obtained from IDT was followed in the presence of increasing amounts of protein. Reactions contained 50 nM RNA and binding was analyzed in a buffer with 50 mM Tris-HCl pH 7.4, 90 mM NaCl, 1 mM MgCl<sub>2</sub> and 4% glycerol. For the DHX15 binding curves, the helicase was titrated from 0 to 600 nM and the reactions were supplemented when indicated with the purified ZZ-tagged G-patch domains or the ZZ tag only at a concentration of 1.2  $\mu$ M for all data points. Alternatively, binding curves in which the G-patch domains of GPATCH1, GPATCH2, GPATCH8 and GPKOW were titrated from 0 to 10  $\mu$ M were generated in the same conditions as above.

After mixing all the components, the samples were incubated for 2 min at 25°C before being transferred to a Quartz SUPRASIL 10x2 mm High Precision cuvette (Hellma-Analytics). The steady-state fluorescence anisotropy was measured in a FluoroMax-4 spectrofluorometer (Horiba) using the following settings: excitation wavelength - 644 nm, emission wavelength - 661 nm, excitation and emission slits - 8, integration time - 0.5 sec, temperature - 25°C. Measurements were performed with a target error of less than 2% up to a maximum of six trials. The averaged data from two or three biological replicates were plotted and fitted with Origin 8.2 using the equation below (eq.1) to obtain dissociation constant (K<sub>d</sub>) values  $\pm$  s.e.m. Please note that for reactions containing both DHX15 and G-patch cofactors, only the concentration of the helicase was taken into account for calculations due to the insignificant level of RNA binding displayed by the G-patch domains themselves.

$$r = r_0 + \frac{\Delta r_{max}}{[RNA]} \times \left( \frac{[protein] + [RNA] + K_d}{2} - \sqrt{\left( \frac{[protein] + [RNA] + K_d}{2} \right)^2 - [protein] \times [RNA]} \right) \quad (\text{eq. 1})$$

( $r_0$  – anisotropy of unbound RNA;  $\Delta r_{max}$  – amplitude; [protein], [RNA] – protein, RNA concentration)

#### 2.7.4 Unwinding assay

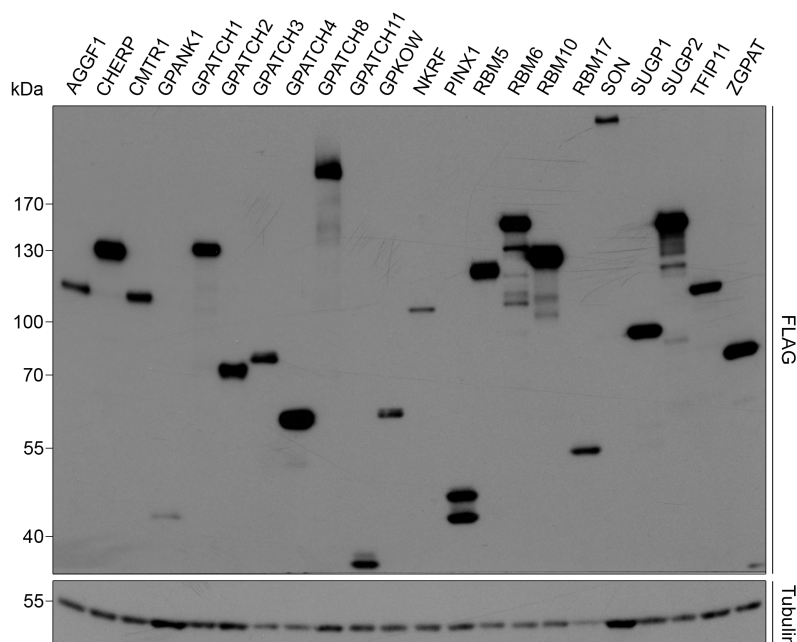
Unwinding experiments were performed using an RNA-DNA duplex as a substrate, which consisted of a  $^{32}\text{P}$ -labelled DNA oligonucleotide (GCTGATCATCTCTGTATTG) complementary to an *in vitro* transcribed RNA of 118 nt. For annealing, the RNA and DNA were mixed at a 1:1 molar ratio in annealing buffer (30 mM HEPES pH 7.5 and 100 mM potassium acetate) and heated at 95°C for 3 min, followed by incubation at 65°C for 5 min and cooling down to room temperature. Unwinding reactions consisted of 1 nM duplex and 2.5 nM DHX15 with or without 2.5 nM NKRF, and were carried out in a buffer containing 50 mM HEPES pH 7.5, 50 mM KCl, 0.5 mM  $\text{MgCl}_2$ , 100  $\mu\text{g/ml}$  BSA, 5% glycerol, 2 mM DTT and 20 U RiboLock RNase Inhibitor (Thermo Fisher). To ensure single-turnover conditions, the samples were supplemented with 50-fold excess of unlabelled complementary DNA oligonucleotide. The reactions were initiated with 2 mM ATP/ $\text{MgCl}_2$  and incubated for up to 20 min at 30°C, followed by quenching with 4X quenching buffer (50 mM Tris-HCl pH 8.0, 2.5% SDS, 50 mM EDTA, 25% glycerol). The samples were separated on 10% polyacrylamide gels in 0.5X TBE that were afterwards exposed to phosphor screens and imaged using a Typhoon FLA 9500 scanner (GE Healthcare).

### 3. RESULTS

#### 3.1 Identification and characterization of G-patch protein-RNA helicase interactions

##### 3.1.1 Human G-patch proteins interact with a subset of DEAH/RHA RNA helicases

RNA helicases have essential roles in the cell and their activity is tightly controlled by a variety of mechanisms, including through interactions with protein cofactors. The inventory of helicase cofactors identified so far includes several proteins that contain a G-patch domain. However, to date, a systematic analysis of the role of the human G-patch protein family as cofactors of RNA helicases is lacking. In order to identify the entire network of G-patch protein-RNA helicase interactions in human cells, we set out to study all proteins that are annotated as containing a G-patch domain in the UniProt database, excluding only those that belong to the endogenous retroviral elements family for which there is little evidence of expression (Hanke et al., 2016). First, the coding sequences of 22 human G-patch proteins were cloned into mammalian expression vectors (Table 2.3) and tetracycline-inducible stable cell lines for expressing these proteins with a FLAG tag were generated. Induction for 24 h followed by western blotting analysis with an anti-FLAG antibody confirmed synthesis of the FLAG-tagged proteins and, using  $\alpha$ -Tubulin as a loading control, their relative expression levels could be compared (Figure 3.1).



**Figure 3.1. Confirmation of the expression of the FLAG-tagged variants of human G-patch proteins from inducible stable cell lines.** HEK293 stable cell lines encoding the indicated human G-patch proteins were induced for 24 h with tetracycline and cell extracts were analyzed by SDS-PAGE and western blotting. The expression of the FLAG-tagged proteins was detected with an anti-FLAG antibody and  $\alpha$ -Tubulin was used as a loading control.

These cell lines were then used together with control cells encoding only the FLAG tag in IP experiments with an anti-FLAG antibody, followed by identification and quantification of the co-precipitated proteins by LC-MS/MS analysis and spectral counting. The experiments were performed in the presence of RNase to avoid bridging of interactions by RNA and to retain only protein-protein interactions. The proteins identified by MS analysis of the IP eluates were ranked based on the total spectral counts, which are an indicator of protein abundance (Old et al., 2005) and their fold enrichment in the samples expressing FLAG-tagged G-patch proteins relative to the FLAG control was calculated.

**Table 3.1. MS analysis results of IP experiments with FLAG-tagged G-patch proteins.** The fold enrichment for the indicated DEAH/RHA helicases was calculated based on the spectral counts measured in the samples expressing FLAG-tagged G-patch proteins relative to the FLAG tag control. The asterisk (\*) marks the samples for which two biological replicates were performed, while the results for the other samples are based on two technical replicates.

G-patch protein bait	Top DEAH/RHA helicase	Fold enrichment helicase
AGGF1	DHX15	4.8
CHERP	DHX15	5.6
CMTR1*	DHX15	42.9
GPANK1	DHX15	7.8
GPATCH1	DHX35	236
GPATCH2	DHX15	21.9
GPATCH3	DHX15	214.9
GPATCH4	DHX15	9
GPATCH8	DHX15	14.6
GPATCH11	DHX15	46.8
GPKOW	DHX16	1720.5
NKRF	DHX15	9
PINX1	DHX15	57
RBM5	DHX15	11.9
RBM6	DHX15	1.6
RBM10	DHX15	13.6
RBM17	DHX15	6.6
SON	DHX15	21.4
SUGP1*	-	-
SUGP2	DHX15	3.7
TFIP11	DHX15	22
ZGPAT	DHX15	47

Among the proteins that were co-precipitated with the G-patch protein baits, we focused our attention on RNA helicases that belong to the DEAH/RHA family, since this subset of helicases contain an OB-fold domain, which has previously been shown to be essential for the interaction with several G-patch cofactors (Robert-Paganin et al., 2015; Sloan and

Bohnsack, 2018). Interestingly, the MS data revealed that DHX15 was the top co-precipitated DEAH/RHA helicase for the majority (19/22) of G-patch proteins, being recovered between 1.6 and 215-fold more compared to its levels in the FLAG control (Table 3.1). Thus, in addition to confirming the known interactions with CMTR1, GPATCH2, PINX1, RBM5, TFIP11 and ZGPAT, these results uncovered a plethora of previously unknown G-patch protein-DHX15 complexes, suggesting that DHX15 might be regulated by an extensive network of cofactors (Lin et al., 2009; Yoshimoto et al., 2009; Niu et al., 2012; Chen et al., 2014; Chen et al., 2017; Inesta-Vaquera et al., 2018). Apart from the DHX15-interacting G-patch proteins, we identified a novel interaction between GPATCH1 and the RNA helicase DHX35, which have previously been detected together in spliceosomal and centrosomal complexes (Agafonov et al., 2011; Gupta et al., 2015). Furthermore, these data confirmed the known association of GPKOW with DHX16, which is conserved from yeast (Silverman et al., 2004; Zang et al., 2014). Lastly, in the case of SUGP1, no DEAH/RHA helicase was reproducibly enriched relative to the FLAG control in the two experiments performed.

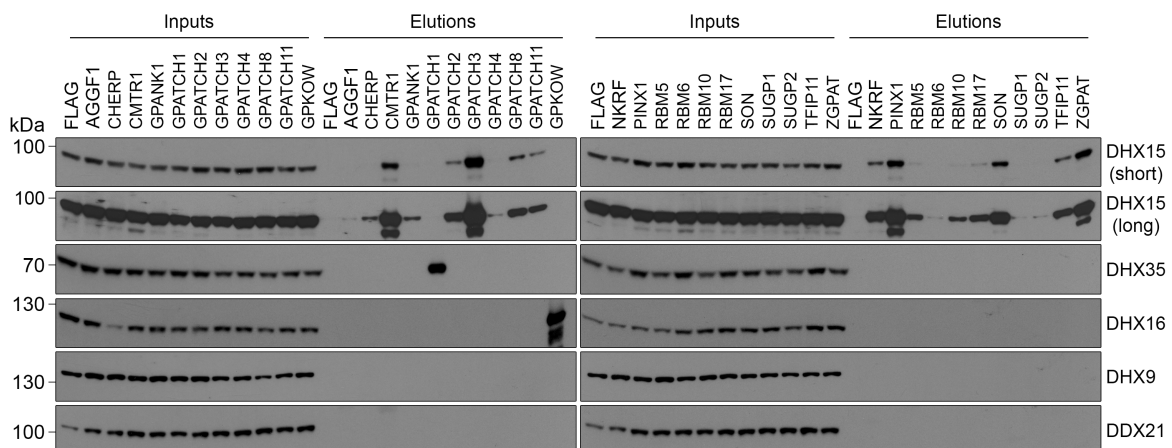
**Table 3.2. MS analysis results of the FLAG-DHX15 IP.** The spectral counts (SC) measured for the G-patch proteins that were enriched in the FLAG-DHX15 IP sample compared to the FLAG tag control are shown together with the fold enrichment.

G-patch protein	SC FLAG-DHX15	Fold enrichment
NKRF	613	94.3
CMTR1	324.5	72.1
TFIP11	176.5	58.8
PINX1	165.5	165.5
GPATCH8	116	116
ZGPAT	88.5	16.1
GPATCH3	40	40
RBM6	35	1.8
RBM17	28	28
CHERP	27.5	1.9
GPATCH11	19.5	19.5
GPATCH2	16.5	16.5
SUGP2	6.5	6.5

To test some of the identified interactions in a reciprocal manner, IPs with lysates from cells expressing FLAG-tagged DHX15 were also carried out and analyzed by MS. Consistent with our previous findings, 13 G-patch proteins were enriched in the FLAG-DHX15 IP compared to the FLAG control, including known interactors as well as novel binding partners (Table 3.2). Interestingly, the mammalian-specific G-patch protein NKRF was the most highly enriched protein in the DHX15 IP eluate, indicating a strong association of the

helicase with this putative cofactor. Furthermore, RBM6 and SUGP2, which showed a weak binding to DHX15 in the reciprocal analysis, were also co-precipitated with the helicase, implying that they are indeed DHX15 interactors. Importantly, no interaction between GPKOW and FLAG-DHX15 was observed despite both proteins being well-established spliceosome components and the specificity of this method is further supported by the lack of GPATCH1 in the DHX15 pulldown.

Next, to confirm the interactions identified by MS, IPs with the G-patch protein-expressing cell lines were performed and this time the eluates were analyzed by western blotting using antibodies against several RNA helicases. Strong binding of DHX16 to GPKOW-FLAG and of DHX35 to GPATCH1-FLAG was observed, suggesting that a significant proportion of these helicases is found in complex with their interacting G-patch protein in cells (Figure 3.2). On the other hand, as our previous results indicated, DHX15 is distributed among a multitude of G-patch interactors and would probably interact to different extents with each of these proteins. In line with this, the anti-FLAG IP results showed that for all G-patch proteins except GPATCH1 and GPKOW, DHX15 was enriched in the eluates to variable levels compared to the FLAG tag only control. In addition to the 19 G-patch proteins that were already indicated by MS as DHX15-interactors, this included SUGP1, which showed a weak but reproducible binding to DHX15. Importantly, the RNA helicases DHX9 and DDX21 were not recovered in any of the IP samples, further supporting the specificity of the interactions detected.

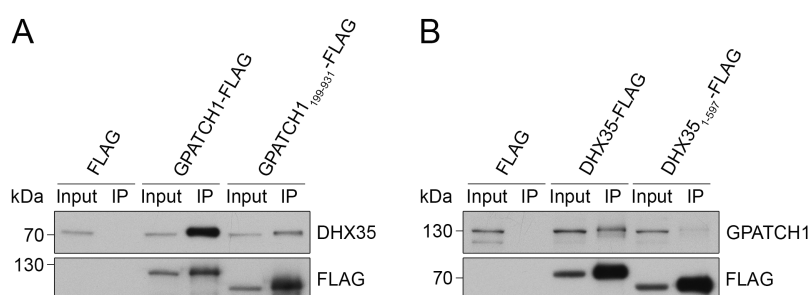


**Figure 3.2. Human G-patch proteins interact specifically with DHX15, DHX16 or DHX35 *in vivo*.** The expression of the FLAG tag or of FLAG-tagged G-patch proteins from HEK293 stable cell lines was induced for 24 h with tetracycline. Cell lysates were used for immunoprecipitation with an anti-FLAG antibody and the co-precipitated proteins were detected by western blotting using antibodies against the indicated RNA helicases. Input samples representing 1% of the lysates for DHX15, DHX9 and DDX21, and 2% for DHX35 and DHX16 were processed alongside the elution fractions. Two different exposures are shown for DHX15 (short and long).

Overall, these results revealed that human G-patch proteins interact with a small subset of RNA helicases: DHX15 has a network of 20 G-patch proteins, while DHX35 and DHX16

bind GPATCH1 and GPKOW respectively. Excitingly, 14 novel G-patch protein interactors were discovered for DHX15 in addition to the 6 cofactors that had been described by others either prior to or during this work.

The identification of DHX35 as a binding partner of GPATCH1 suggested the existence of an additional RNA helicase regulated by G-patch cofactors apart from DHX15 and DHX16 (Robert-Paganin et al., 2015). Several studies have shown that the interaction between G-patch proteins and RNA helicases involves the G-patch domain of the cofactor and the OB-fold region of the helicase (Walbott et al., 2010; Christian et al., 2014; Heininger et al., 2016; Inesta-Vaquera et al., 2018). Therefore, to assess if GPATCH1 and DHX35 form a *bona fide* G-patch cofactor-RNA helicase complex, the requirement for these domains for their interaction was analyzed by deleting the N-terminal region of GPATCH1, which includes the G-patch domain (GPATCH1<sub>199-931</sub>), as well as the OB-fold domain of DHX35 (DHX35<sub>1-597</sub>). These truncated proteins were expressed with a FLAG tag from stable cell lines and used together with the wild-type proteins and with a FLAG only control in IP experiments. Western blotting analysis of the IP eluates showed that full-length GPATCH1 and DHX35 interacted in a reciprocal manner, however, binding was strongly reduced when either the G-patch domain of the cofactor or the OB-fold of the helicase were lacking (Figure 3.3). Thus, these results indicate that GPATCH1 and DHX35 constitute a novel G-patch cofactor-RNA helicase complex as their binding follows the same mode of interaction that has been described for other such complexes.



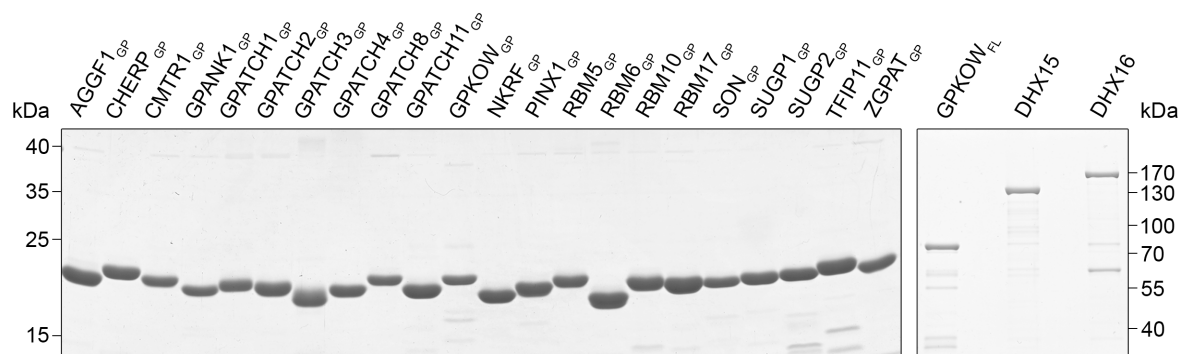
**Figure 3.3. GPATCH1 and DHX35 interact through conserved domains. (A)** IP experiments with anti-FLAG beads were performed using extracts prepared from cells induced for 24 h with tetracycline to express the FLAG tag, FLAG-GPATCH1 or FLAG-GPATCH1<sub>199-931</sub>. The levels of DHX35 and of the bait proteins (FLAG) in the input (1%) and elution samples were detected by western blotting. **(B)** Stable cell lines expressing the FLAG tag and FLAG-tagged DHX35 and DHX35<sub>1-597</sub> were used as in (A) and the samples were analyzed with the indicated antibodies.

### 3.1.2 The G-patch domain is sufficient in most cases for the regulation of the cognate RNA helicase

The G-patch domain has been suggested to mediate protein-protein interactions and in specific cases protein-RNA interactions, and has previously been shown to be essential for activating RNA helicases in yeast and human cells (Lebaron et al., 2009; Niu et al., 2012).



To assess whether the G-patch protein-RNA helicase complexes identified by IPs involve direct protein-protein binding and to test the potential of the G-patch domains to regulate their cognate helicase, we studied ATP hydrolysis and RNA binding as parameters of RNA helicase activity *in vitro*, in the presence or absence of the putative G-patch cofactors. For this, the fragments corresponding to the 22 G-patch domains were recombinantly expressed and purified with a ZZ-His<sub>7</sub> tag together with full-length MBP-DHX15-His<sub>10</sub>, MBP-DHX16-His<sub>10</sub> and ZZ-GPKOW-His<sub>7</sub> (Figure 3.4).

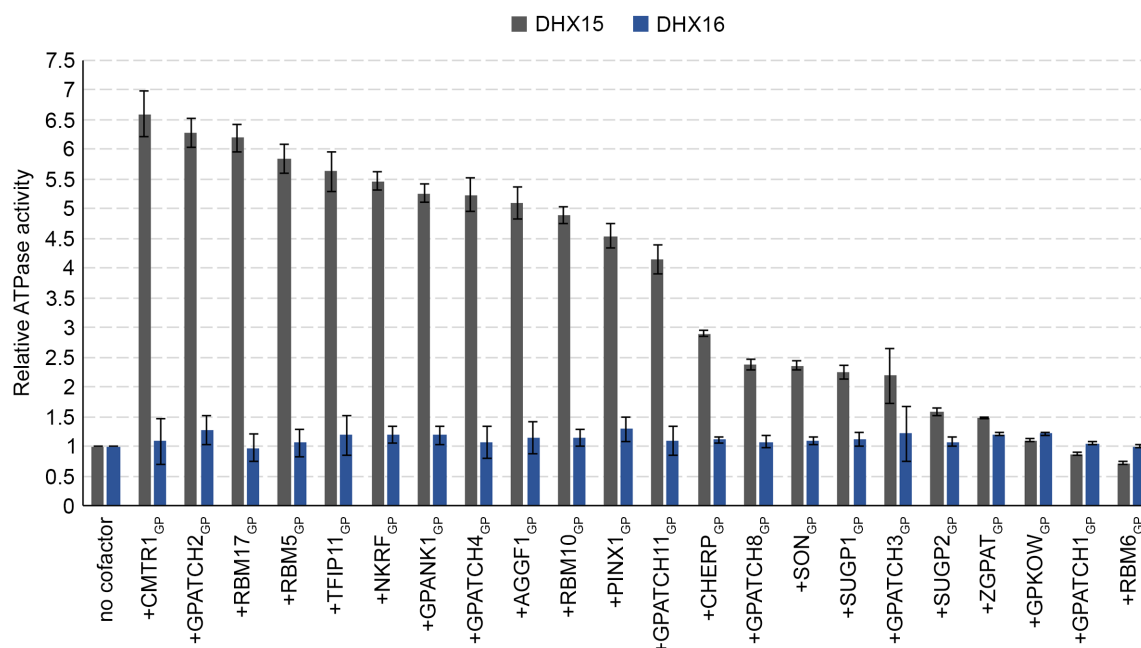


**Figure 3.4. Overview of the recombinantly purified G-patch domains and full-length GPKOW, DHX15 and DHX16.** The G-patch domains (GP) of the indicated proteins were expressed in *E. coli* with a ZZ-His<sub>7</sub> tag and purified on Ni-NTA. The resulting protein preparations were analyzed by SDS-PAGE and Coomassie staining (left panel). Similarly, full-length GPKOW containing a ZZ-His<sub>7</sub> tag (GPKOW<sub>FL</sub>) and DHX15 and DHX16 carrying an MBP-His<sub>10</sub> tag were purified on Ni-NTA, separated by SDS-PAGE and visualized by Coomassie staining (right panel).

The steady-state ATP hydrolysis activity of the RNA helicase DHX15 was analyzed in the presence or absence of excess G-patch domains using an NADH-coupled enzymatic assay (Kiianitsa et al., 2003). Since RNA helicases are RNA-dependent ATPases, all samples were supplemented with *in vitro* synthesized polyU<sub>32</sub> RNA. Compared to the activity of the helicase alone, addition of the G-patch domains to DHX15 led to an increase of ATP hydrolysis between 1.5 and 6.6-fold for 19 out of the 22 samples, indicating that these proteins act as cofactors of the helicase and have a stimulatory effect (Figure 3.5). Importantly, the G-patch domains of the non-DHX15 interactors, GPATCH1 and GPKOW, did not influence the activity of the helicase.

In order to confirm the specificity of the effects observed, the ATPase activity of DHX16 with or without addition of the G-patch domains was also determined (Figure 3.5). In contrast to DHX15, ATP hydrolysis by DHX16 was not significantly stimulated by the presence of any of the G-patch domains compared to its activity in the absence of cofactor (maximum fold increase 1.3). The small changes observed are likely caused by unspecific binding of the G-patch domains to DHX16 due to their presence in excess. However, we failed to observe a clear stimulation of DHX16 by the G-patch domain of GPKOW (GPKOW<sub>GP</sub>) despite the

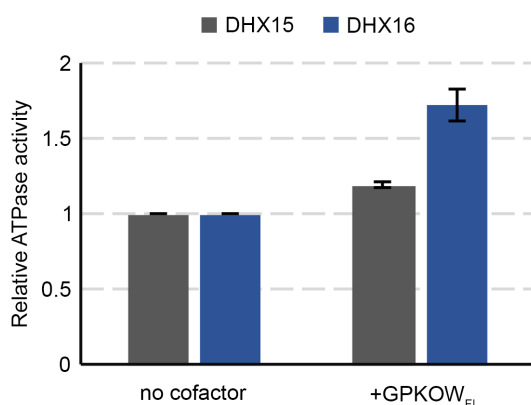
well-established interaction between these proteins. Similarly, addition of RBM6<sub>GP</sub> did not enhance the ATPase activity of DHX15, although these proteins interacted weakly *in vivo*.



**Figure 3.5. The G-patch domain is generally sufficient to stimulate the ATPase activity of the interacting RNA helicase.** An NADH-coupled assay was used to measure ATP hydrolysis in reactions containing 250 nM of recombinantly purified DHX15 and DHX16 together with 2  $\mu$ M RNA and the indicated G-patch domains (GP) at a concentration of 1.5  $\mu$ M. The amount of ATP hydrolyzed in each sample was expressed relative to the samples containing only the RNA helicase and RNA (no cofactor) and results from three independent experiments are shown as mean  $\pm$  s.e.m.

This suggested that the G-patch domain is not always sufficient to regulate the activity of RNA helicases and that other regions of the G-patch protein might contribute to binding and activation of the helicase. Therefore, the ATP hydrolysis activity of DHX15 and DHX16 was also analyzed in the presence of full-length GPKOW (GPKOW<sub>FL</sub>). Upon addition of GPKOW<sub>FL</sub> to the reaction, a higher increase in ATPase activity was observed for DHX16 compared to DHX15 relative to their basal activities, which have similar values (Figure 3.6). This supports the specificity of the interaction between GPKOW and DHX16 as well as the stimulatory role of the cofactor. Since full-length RBM6 could not be purified due to its large size, it remains to be determined whether this G-patch protein regulates the ATPase activity of DHX15 as well.

Taken together, these results demonstrate that the interactions identified *in vivo* represent *bona fide* helicase-cofactor complexes and that the G-patch domain is sufficient in most cases to confer specificity for interaction with the cognate helicase and to regulate its activity. Furthermore, these data show that G-patch proteins have primarily a stimulatory role and are capable of promoting ATP hydrolysis by the helicase.



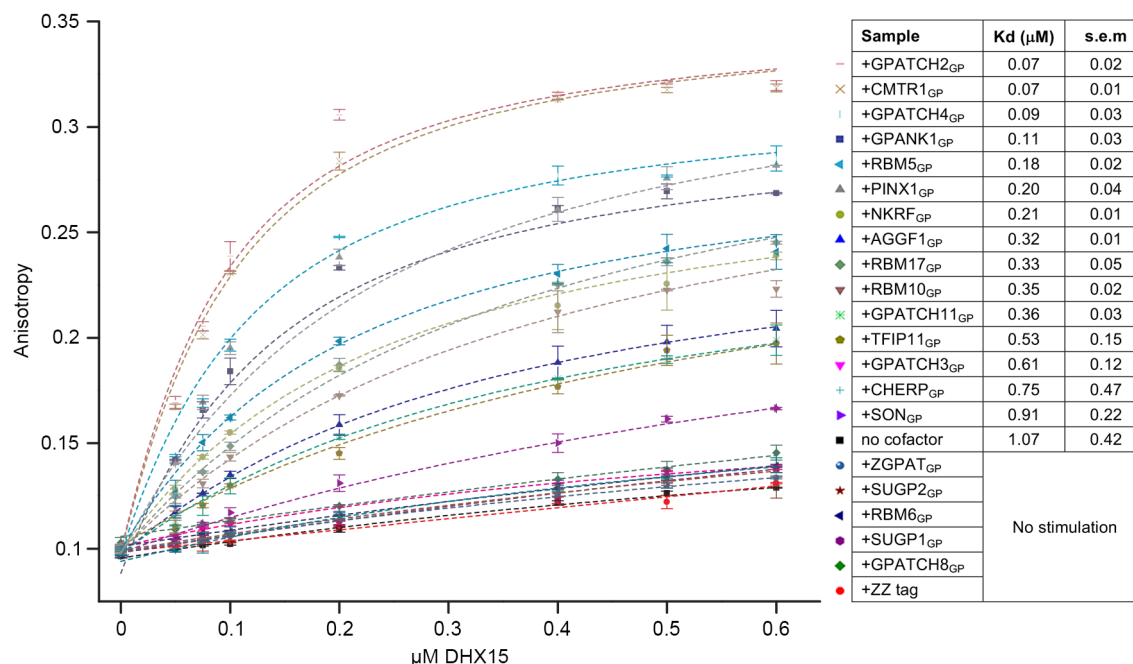
**Figure 3.6. Full-length GPKOW stimulates the ATPase activity of DHX16.** The ATPase activity of DHX15 and DHX16 (250 nM) in the presence of RNA (2  $\mu$ M) was measured with or without the addition of 1.5  $\mu$ M full-length GPKOW (GPKOW<sub>FL</sub>) using an NADH-coupled assay. The level of ATP hydrolysis in the presence of GPKOW was calculated relative to the samples containing only DHX15 or DHX16 and was represented as mean  $\pm$  s.e.m of three experiments.

### 3.1.3 The RNA binding affinity of DHX15 is enhanced by the G-patch domains of its cofactors

In the functional cycle described for DEAH/RHA helicases, binding of the RNA substrate leads to conformational changes that favor ATP hydrolysis, thereby providing the basis for the RNA-dependent ATPase activity exhibited by these enzymes (He et al., 2017; Tauchert et al., 2017). To gain more insight into the mechanisms involved in the stimulation of the RNA helicase ATPase activity by G-patch proteins, the effect of the G-patch domains on the RNA binding affinity of DHX15 was determined *in vitro* by fluorescence anisotropy using an 11 nt RNA substrate labelled with ATTO647N. Binding curves were generated by titration of the helicase to the RNA substrate, with or without the addition of a constant concentration of the DHX15-interacting G-patch domains at all data points. In the absence of cofactor, DHX15 bound RNA with an estimated  $K_d$  of 1.07  $\mu$ M (Figure 3.7). Upon addition of the G-patch domains, the RNA binding activity of DHX15 was stimulated in most cases, up to a  $K_d$  value of 70 nM, with the highest effects being observed in the presence of GPATCH2<sub>GP</sub> and CMTR1<sub>GP</sub>, consistent with the results of the ATPase assay (Figure 3.5). Importantly, no stimulation was observed when the purified ZZ tag was added to DHX15, demonstrating that the G-patch moiety is indeed responsible for the observed increase in RNA binding.

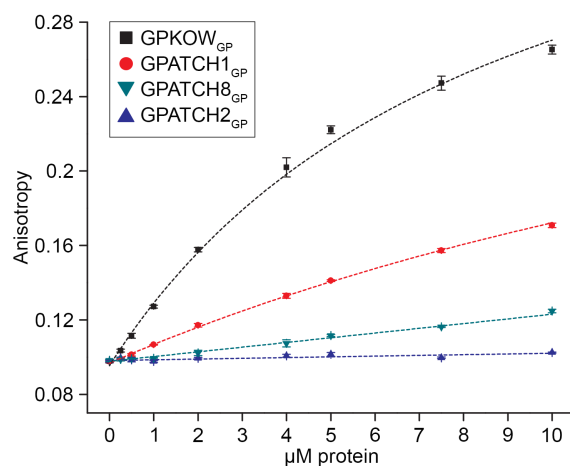
On the other hand, the presence of the G-patch domains of GPATCH8, RBM6, SUGP1, SUGP2 and ZGPAT did not stimulate RNA binding by DHX15, which suggested, in line with earlier results (Figure 3.5), that these proteins might be weaker cofactors. Of note, neither of the DHX15-interacting G-patch domains showed any significant RNA binding at the selected concentration (Figure 3.7, first data point). Interestingly, however, in the course of these experiments we observed that GPKOW<sub>GP</sub> and GPATCH1<sub>GP</sub>, which interact with DHX16 and DHX35 respectively, had a considerably higher RNA affinity than the other G-patch domains. Titration of these G-patch domains with the RNA substrate revealed a  $K_d$  of 8.7  $\mu$ M for GPKOW<sub>GP</sub> and of 26.2  $\mu$ M for GPATCH1<sub>GP</sub> (Figure 3.8). These binding affinities were several orders of magnitude higher than the affinity of the other G-patch

domains, with GPATCH2<sub>GP</sub> and GPATCH8<sub>GP</sub> being shown here as representative examples.



**Figure 3.7. The RNA binding affinity of DHX15 is enhanced by the G-patch domains of its cofactors.** The fluorescence anisotropy of an 11 nt ATTO647N-labelled RNA was determined in the presence of increasing amounts of DHX15 (0-0.6 μM) either without cofactor or with the addition of 1.2 μM of the indicated ZZ-tagged G-patch domains or the ZZ tag only. Data from two independent experiments were averaged and fitted in Origin 8.2 to calculate Kd values ± s.e.m, which are shown alongside the binding curves.

Taken together, these results demonstrate that most G-patch domains of the DHX15-interacting proteins can enhance the RNA binding affinity of the helicase, indicating that this might be the major mechanism contributing to the stimulation of its ATPase activity observed in the presence of cofactors. Furthermore, these data show that G-patch domains themselves can have significantly different RNA binding affinities, and in some cases, can mediate protein-RNA interactions in addition to protein-protein interactions.



**Figure 3.8. The G-patch domains of GPKOW and GPATCH1 bind RNA.** An 11 nt RNA substrate labelled with ATTO647N was incubated with increasing concentrations (0-10 μM) of the indicated G-patch domains (GP) and the steady-state fluorescence anisotropy was measured. Results from three experiments were analyzed in Origin 8.2 to determine Kd values ± s.e.m. The obtained Kd values were  $8.7 \pm 1.3$  μM for GPKOW<sub>GP</sub> and  $26.2 \pm 4.8$  μM for GPATCH1<sub>GP</sub>, while the RNA affinities of GPATCH2<sub>GP</sub> and GPATCH8<sub>GP</sub> were in the mM range.

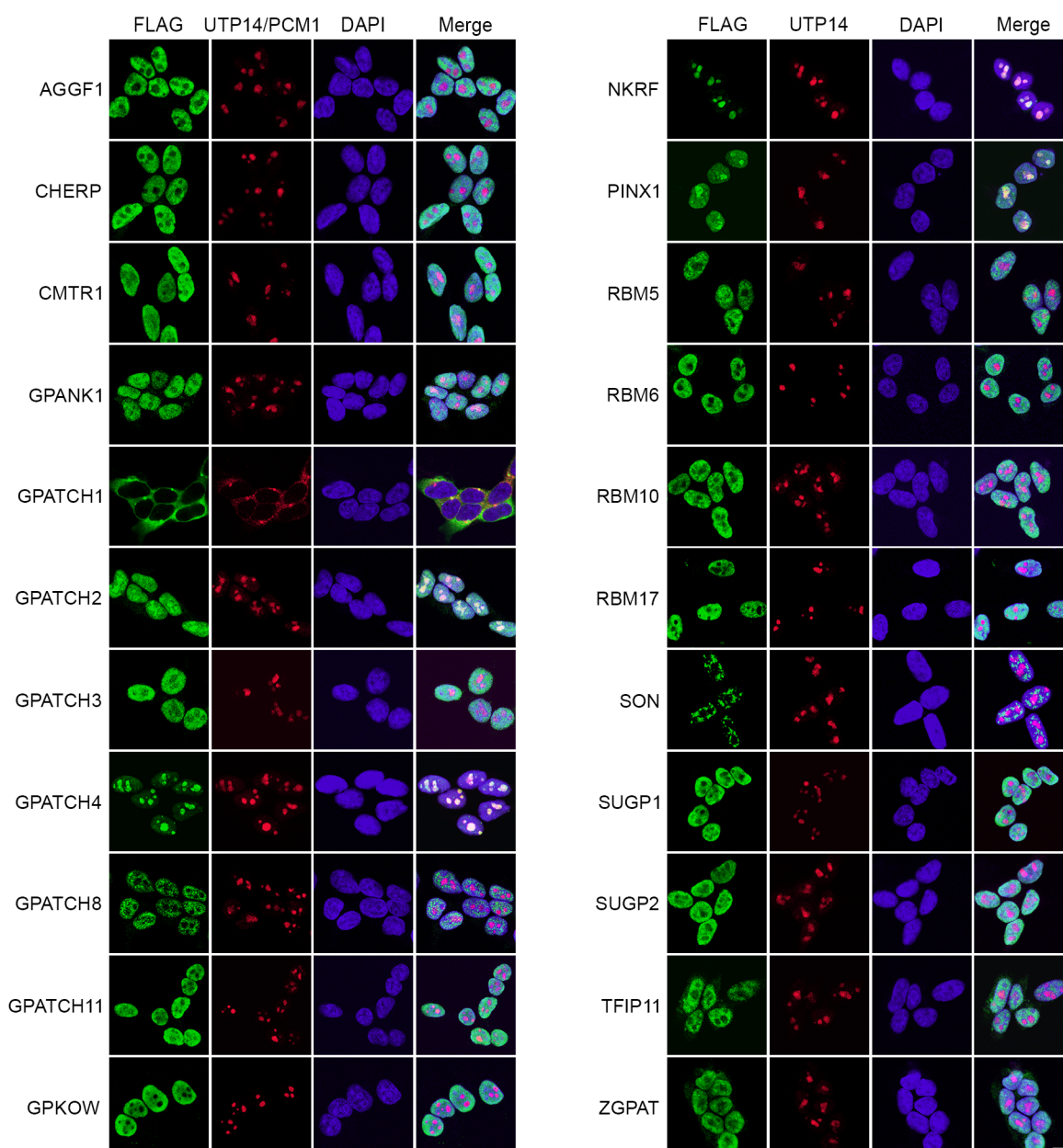
## **3.2 Functional characterization of DHX15 and its G-patch cofactors**

### **3.2.1 The majority of human G-patch proteins are localized in the nucleoplasm**

Having identified and characterized the interactions that human G-patch proteins establish with their cognate RNA helicases, the next aim was to study their target pathways and cellular functions. First, immunofluorescence microscopy was used to investigate the localization of the FLAG-tagged G-patch proteins expressed from the HEK293 stable cell lines. Staining of cells with an anti-FLAG antibody, with antibodies against markers for different cellular structures (UTP14A - nucleoli, PCM1 - centrosome) and with DAPI for visualization of the nuclear material revealed that all G-patch proteins were localized in the nucleus with the exception of GPATCH1, which was found in the cytoplasm and at the centrosome (Figure 3.9). Interestingly, the presence of GPATCH1 at the centrosome in close proximity to PCM1 is in line with the previously described association of this protein with centrosomal complexes (Gupta et al., 2015).

Within the nucleus, most G-patch proteins had a nucleoplasmic localization without any nucleolar staining, as demonstrated by the lack of colocalization with UTP14A. A more particular localization was observed for GPATCH8, which showed a punctate staining in the nucleoplasm. Similarly, SON was detected in specific foci that correspond to splicing speckles as has been reported in other studies (Huen et al., 2010; Ahn et al., 2011). Furthermore, in addition to its nucleoplasmic localization, ZGPAT was enriched in structures adjacent to nucleoli, which based on results from another study represent Cajal bodies (Chen et al., 2017).

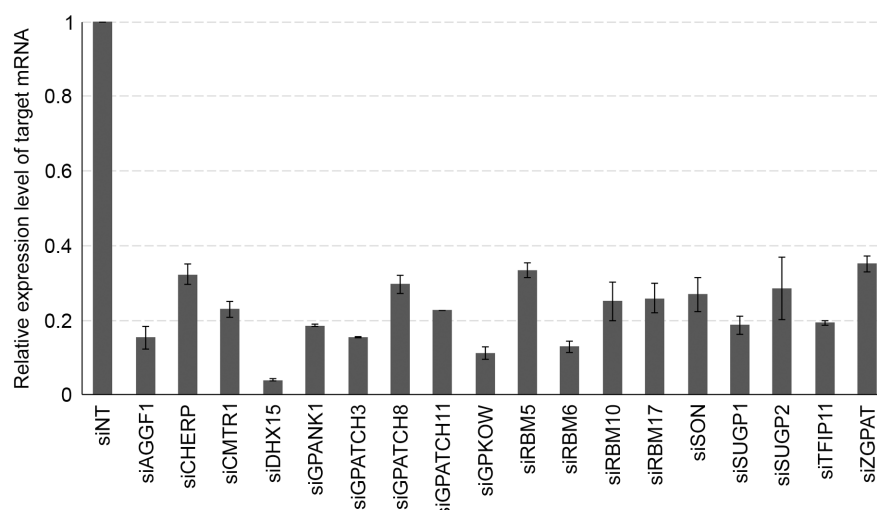
On the other hand, GPATCH2 and PINX1 showed both a nucleoplasmic and a nucleolar staining, while GPATCH4 and NKRF were strongly enriched in nucleoli. The localization of these proteins in nucleoli suggested that they might be involved in ribosome biogenesis, which is initiated at these sites. In line with this, it has been proposed that GPATCH2 and PINX1 are implicated in ribosome production similar to their yeast homologs Sqs1 and Pxr1 (Chen et al., 2014; Robert-Paganin et al., 2015). Taken together, these results provide an overview of the subcellular localization of all human G-patch proteins and represent a starting point for further functional studies.



**Figure 3.9. Human G-patch proteins are localized almost exclusively in the nucleus.** FLAG-tagged G-patch proteins were expressed from the HEK293 stable cell lines after induction with tetracycline for 24 h and their localization was followed by immunofluorescence microscopy using an antibody against the FLAG tag (green). Antibodies against the centrosomal protein PCM1 (only for the GPATCH1 sample) and the nucleolar marker UTP14A were used for counterstaining (red) and nuclei were visualized with DAPI (blue). Overlays of the signals from the three channels are provided (Merge) and the scale bar represents 10  $\mu$ m.

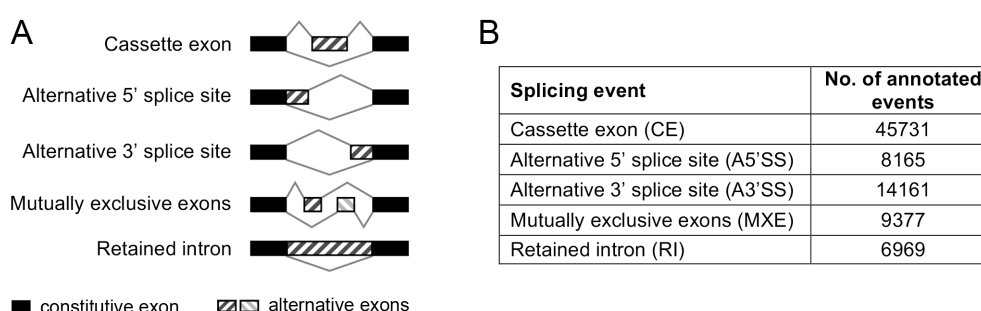
### 3.2.2 Genome-wide analyses reveal that DHX15 and most of its G-patch cofactors regulate alternative splicing

The nucleoplasmic localization observed for the majority of human G-patch proteins suggested that they might impact RNA metabolism at an early stage in the gene expression pathway. Thus, in order to gain more insight into the function of G-patch proteins and RNA helicases, we investigated in a genome-wide manner changes in alternative splicing and gene expression upon their siRNA-mediated knockdown. Given the intriguing finding that DHX15 interacts with such a large number of G-patch proteins, our analyses focused on this helicase and its cofactors with the aim to understand this complex network of interactions and the interplay between its components. However, the G-patch proteins that are likely involved in ribosome biogenesis (GPATCH2, GPATCH4, NKRF, PINX1) as indicated by their nucleolar localization, were not included in the analysis since they were not expected to have a direct effect on the transcriptome. To enable comparison with the DHX15 cofactors, GPKOW, which interacts with DHX16, was also studied. Therefore, knockdowns for DHX15 and for 17 human G-patch protein genes together with a control knockdown against the firefly luciferase gene (non-target, NT) were established in HEK293 cells using one siRNA for each, and the efficiency of the depletion was tested by qRT-PCR. This showed a decrease in the expression of the target mRNAs compared to the NT control ranging between 65% and 96% relative to the levels of the reference housekeeping gene (Figure 3.10).



**Figure 3.10. The expression levels of the G-patch protein-encoding genes are efficiently reduced by siRNA treatment.** HEK293 cells were treated for 72 h with a non-target siRNA (siNT) or with siRNAs against the indicated genes and the levels of the target mRNAs were determined by qRT-PCR. The housekeeping genes GAPDH and EMC7 were used for normalization and for each sample the expression level of the targeted mRNA is shown relative to the siNT control.

Next, total RNA extracted from siRNA-treated cells was subjected to rRNA depletion and library preparation for RNA-seq analysis, and the obtained sequencing reads were mapped onto the human genome and analyzed for changes in alternative splicing and gene expression levels. Alternative splicing analysis was performed with the MISO software (Katz et al., 2010), which uses the genome annotation to derive all known splicing events and determines changes in the extent of these events between the knockdown samples and the control based on the difference in PSI values (see pages 43-44 for a detailed explanation). The most abundant type of alternative splicing is represented by *cassette exons*, which are exons that can be either included or skipped during the splicing process (Figure 3.11). The other alternative splicing events are classified into *alternative 5' and 3' splice sites*, *mutually exclusive exons* and *retained intron*.



**Figure 3.11. Classification of alternative splicing events.** (A) Schematic representation of the five types of alternative splicing events. Modified from Park et al., 2013. (B) The number of annotated splicing events for each category of alternative splicing based on the GRCh38 human genome annotation as determined using the MISO software.

Interestingly, the results of the MISO analysis, which are summarized in Table 3.3, showed that depletion of DHX15 and of all the G-patch proteins tested affected a multitude of splicing events belonging to all categories of alternative splicing. The highest numbers of differentially expressed splicing events were observed upon knockdown of SON (2450 events), ZGPAT (2202 events), DHX15 (737 events), AGGF1 (460 events), TFIP11 (336 events) and SUGP2 (325 events), indicating that these factors have important roles in splicing regulation. In some cases, multiple splicing changes were detected for the same gene and the number of target genes is also shown in Table 3.3.

For all proteins analyzed, the majority of regulated events belonged to the CE category, consistent with this being the most abundant type of alternative splicing (Figure 3.11B). In general, DHX15 and G-patch proteins promoted both exon skipping and inclusion to similar levels, however, in some cases a stronger effect was observed for the former or the latter (Figure 3.12). For example, depletion of AGGF1 led to three times more exon skipping events than inclusion, while knockdown of RBM10 regulated four times more exon inclusion compared to skipping events. Despite the large number of differentially expressed cassette



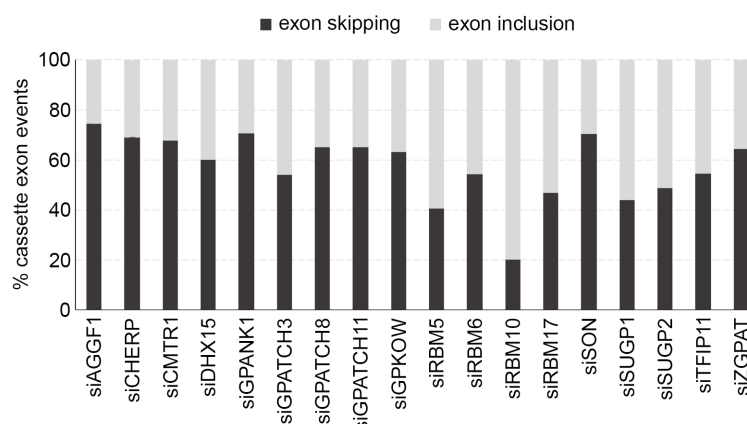
exons that were detected, these represented a small fraction of all known CE events, with only about 2% of the total being affected by depletion of SON and ZGPAT. Interestingly, analysis of the other types of alternative splicing revealed a high level of intron retention following SON knockdown (12.5% of total RI events; Figure 3.11B; Table 3.3), which is in agreement with previous reports (Lu et al., 2013). Furthermore, depletion of ZGPAT preferentially regulated the choice between two adjacent alternative exons, impacting 6.5% of the total number of annotated MXE events.

**Table 3.3. Summary of the RNA-seq and alternative splicing analysis results.** The number of differentially expressed alternative splicing events for the indicated samples relative to the control is shown for each category (CE - cassette exon, A5'SS and A3'SS - alternative 5' and 3' splice sites, MXE - mutually exclusive exons, RI - retained intron). The number of genes showing changes in alternative splicing and the percentage of unique events for each sample are also indicated.

Sample	CE events	A5'SS events	A3'SS events	MXE events	RI events	Total events (genes)	Unique events (%)
siAGGF1	223	40	53	63	81	460 (337)	46.3
siCHERP	145	34	22	26	36	263 (205)	35.7
siCMTR1	68	12	20	16	19	135 (122)	37.8
siDHX15	427	38	90	61	121	737 (542)	41.7
siGPANK1	92	20	15	37	28	192 (147)	35.4
siGPATCH3	63	12	18	18	13	124 (107)	47.6
siGPATCH8	103	23	30	44	27	227 (192)	11.0
siGPATCH11	97	21	32	50	28	228 (192)	13.6
siGPKOW	84	20	11	25	21	161 (137)	35.4
siRBM5	52	7	15	14	6	94 (84)	42.6
siRBM6	59	8	7	24	15	113 (92)	46.0
siRBM10	70	14	16	18	11	129 (113)	38.8
siRBM17	120	16	26	54	19	235 (180)	50.6
siSON	935	162	301	168	884	2450 (1230)	69.7
siSUGP1	89	18	35	34	25	201 (161)	41.8
siSUGP2	193	25	18	51	38	325 (258)	35.7
siTFIP11	180	35	28	53	40	336 (262)	44.6
siZGPAT	1009	191	228	608	166	2202 (1341)	74.3

Comparison of the differentially expressed splicing events revealed that for most proteins analyzed approximately 35-75% of the detected changes were unique for that sample, however, there was also overlap between the targets identified for G-patch proteins and for DHX15. For example, SON regulated 256 splicing events that overlapped with the DHX15 targets, while ZGPAT and SUGP2 had 78 and 48 events in common with the helicase. Common events were also detected between DHX15 and GPKOW, suggesting that multiple factors and/or mechanisms might be responsible for alternative splicing regulation at the same sites. A notable result was obtained in the case of GPATCH8 and GPATCH11, for

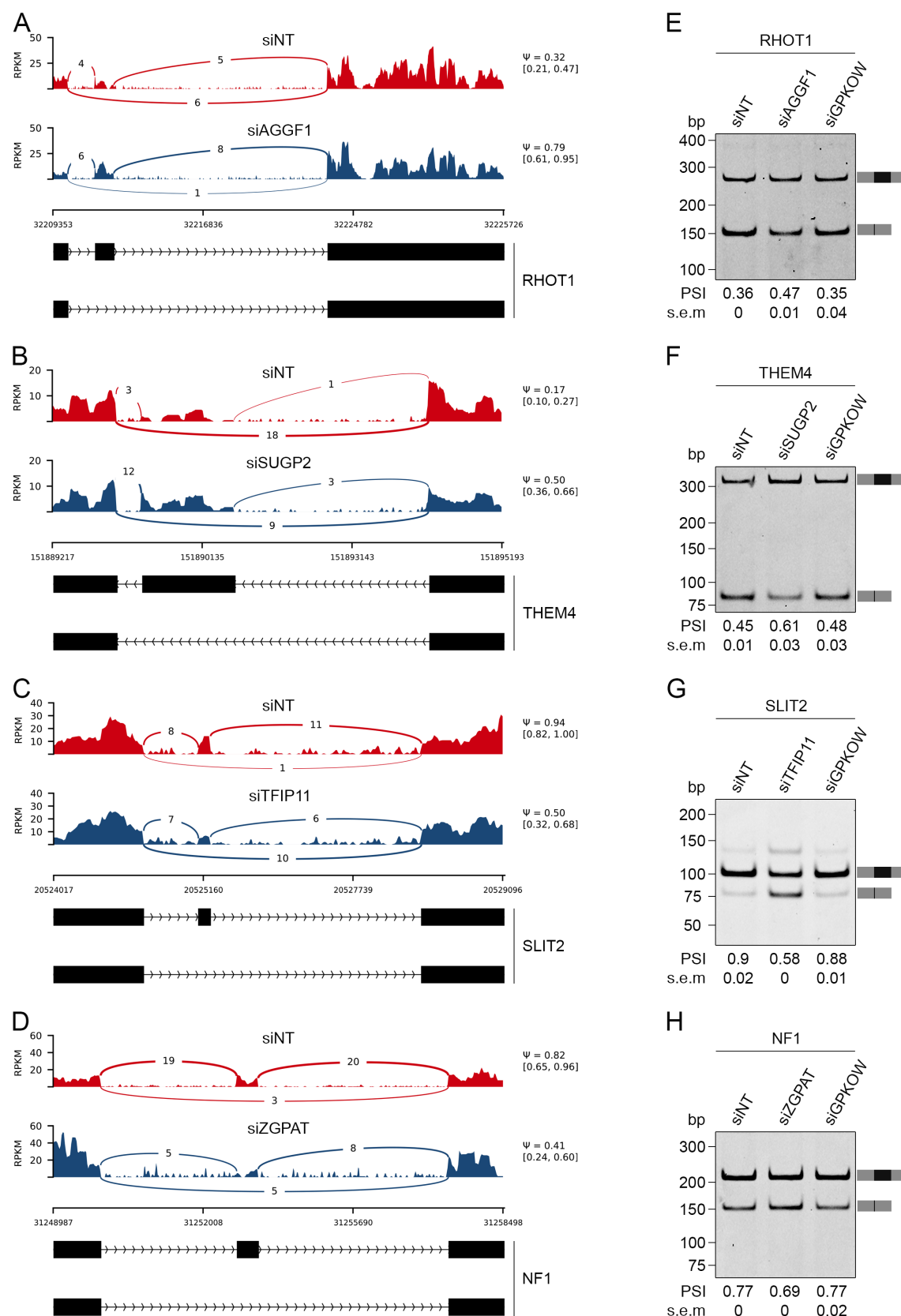
which 80% of the splicing changes identified were identical, leading to 180 common target events and suggesting that G-patch proteins might also act synergistically or redundantly.



**Figure 3.12. Depletion of G-patch proteins and DHX15 affects both exon skipping and inclusion.** Total RNA extracted from cells treated for 72 h with a non-target siRNA (NT) or with siRNAs against the indicated genes was analyzed by RNA-seq. Changes in alternative splicing between the samples and the NT control were determined using MISO and the differentially expressed cassette exon events were further divided into exon skipping ( $\Delta\text{PSI} \geq 0.2$ ) and inclusion ( $\Delta\text{PSI} \leq -0.2$ ) events.

To confirm their role in alternative splicing regulation, four G-patch proteins were examined in more detail and their effect on the splicing of a few targets was tested experimentally. The selected proteins included factors that had not been linked to splicing until now (AGGF1) and proteins previously associated with the spliceosome but whose functions or targets were not known (SUGP2, TFIP11 and ZGPAT). First, Sashimi plots were generated from the MISO analysis for one CE event for each of the four G-patch proteins in order to enable visualization of alternative splicing. The Sashimi plots, which display the number of reads spanning splice junctions and the read coverage for the alternative and constitutive exons alongside the calculated PSI values, clearly indicated that, compared to the NT control, depletion of AGGF1 and SUGP2 promoted exon inclusion for RHOT1 and THEM4 respectively, while knockdown of TFIP11 and ZGPAT led to exon skipping in the case of SLIT2 and NF1 (Figure 3.13A-D).

These changes were also analyzed experimentally by RT-PCR using primers that bind in the flanking constitutive exons, which would amplify different amplicon sizes depending on whether the alternative exon is included or skipped. For this, total RNA extracted from cells treated with a NT siRNA or with siRNAs against the genes of interest was reverse-transcribed into cDNA, which was used as a template for PCR with primers designed as described above. The PCR products were separated on polyacrylamide gels, visualized by ethidium bromide staining and the PSI value for each sample was calculated as the ratio of the inclusion isoform relative to the total level of the inclusion and skipping isoforms.

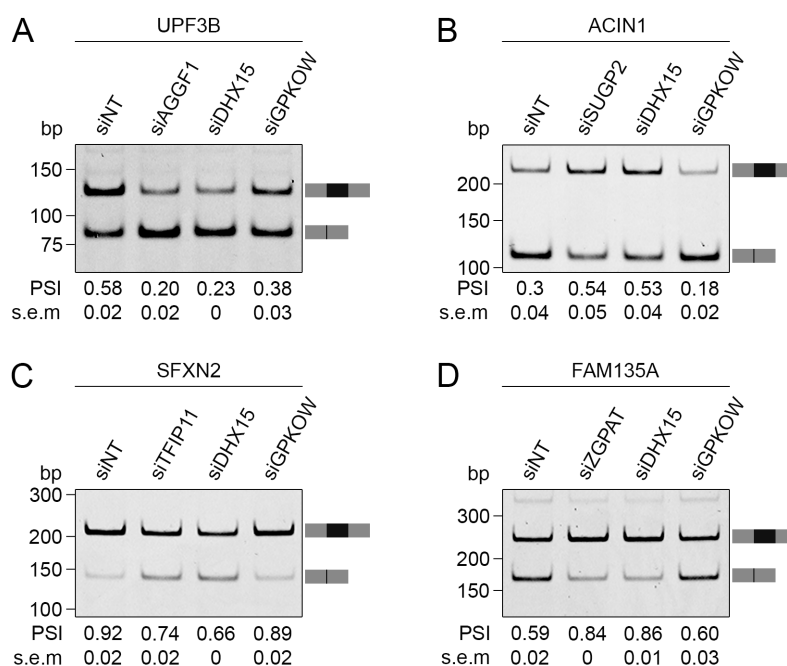


**Figure 3.13. Knockdown of G-patch proteins leads to changes in alternative splicing for specific genes.** (A-D) RNA extracted from cells treated for 72 h with siRNAs against AGGF1 (A), SUGP2 (B), TFIP11 (C), ZGPAT (D) and with a non-target siRNA (siNT) was subjected to RNA-seq followed by alternative splicing (legend continued on next page)

(legend continued)

analysis using MISO. Sashimi plots were generated for one splicing event/gene for each sample to display the read coverage as well as the number of reads spanning each splice junction. The alternative exon and the flanking constitutive exons are shown as rectangles, while the introns are represented as lines. The PSI ( $\Psi$ ) values are shown on the right. **(E-H)** Knockdown samples were obtained as in (A-D) and the same splicing events for RHOT1 (E), THEM4 (F), SLIT2 (G) and NF1 (H) were analyzed by RT-PCR and PAGE. The position of the exon inclusion and skipping isoforms is marked on the right. The PSI values were determined as the ratio of the inclusion isoform relative to the level of both products and the mean  $\pm$  s.e.m of two experiments is shown.

The results in Figure 3.13E-H showed that knockdown of AGGF1, SUGP2, TFIP11 and ZGPAT induced changes in the splicing of their target genes compared to the NT control sample, thus confirming the observations from the RNA-seq data analysis. However, these effects were weaker than those determined computationally, which might be due to the semi-quantitative nature of the RT-PCR and PAGE methods. GPKOW was used as an additional control and its depletion did not affect the splicing of the selected cassette exons.

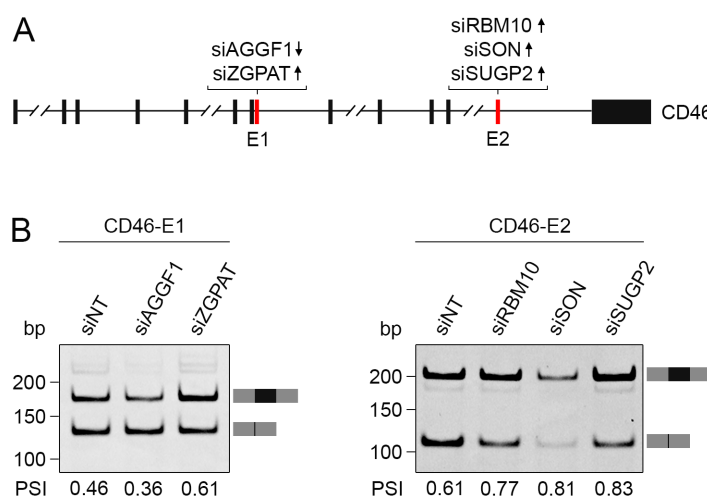


**Figure 3.14. Specific splicing events are regulated by both DHX15 and its G-patch cofactors. (A-D)** Total RNA extracted from cells treated for 72 h with a non-target siRNA (NT) or with siRNAs against the indicated genes was analyzed by RT-PCR and PAGE to determine the splicing of an alternative exon for UPF3B (A), ACIN1 (B), SFXN2 (C) and FAM135A (D). The inclusion and skipping products are indicated on the right and the PSI values were calculated as the ratio of the inclusion isoform relative to the total level of both isoforms. Results are shown as mean  $\pm$  s.e.m of two experiments.

Next, splicing events that were common between these G-patch proteins and DHX15 were analyzed using the same approach and changes in the alternative splicing of the target genes were observed upon depletion of both the helicase and of its G-patch cofactors compared to the control. Knockdown of DHX15 and AGGF1 and of DHX15 and TFIP11 promoted exon skipping for UPF3B and SFXN2 respectively and similarly, the level of exon

inclusion for ACIN1 and FAM135A was increased upon depletion of both the helicase and of SUGP2 or ZGPAT (Figure 3.14A-D). Again, knockdown of GPKOW did not show the same effects on the splicing of the genes analyzed. Thus, these results suggest that DHX15 regulates alternative splicing at these sites together with its G-patch cofactors.

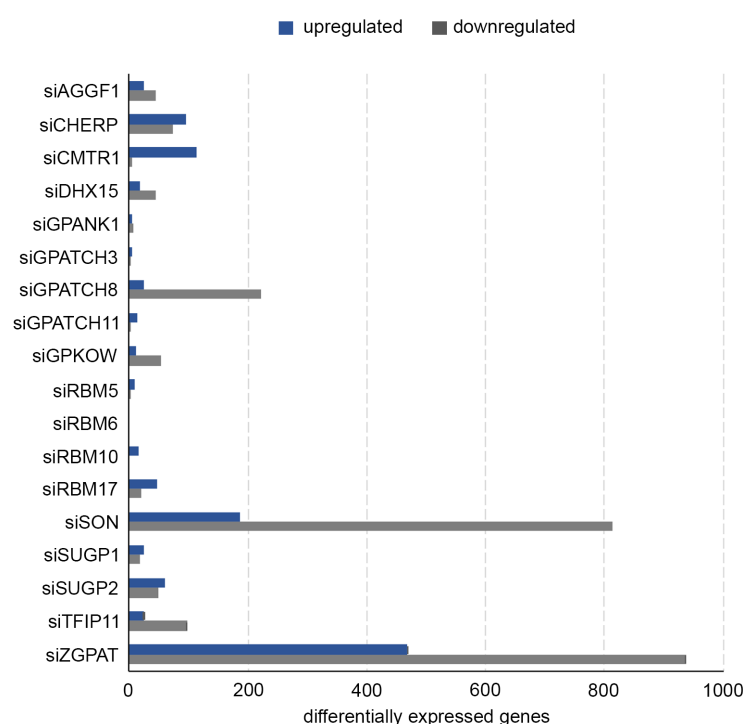
The finding that such a large number of the G-patch protein cofactors of DHX15 are involved in alternative splicing raised the possibility that their action is coordinated. Interestingly, genes targeted by multiple G-patch proteins at the same splice sites or at different locations were also identified in our analysis. One example of such a gene is CD46, for which splicing of two alternative exons was found to be impacted by depletion of several G-patch proteins. More specifically, knockdown of AGGF1 and ZGPAT had an antagonistic effect on one cassette exon event, while depletion of RBM10, SON and SUGP2 promoted exon inclusion at a separate site (Figure 3.15A). These splicing changes were tested by RT-PCR, which confirmed that multiple G-patch proteins can target a gene at different splice sites and that they can regulate the same splicing event either cooperatively or antagonistically (Figure 3.15B).



**Figure 3.15. Alternative splicing of the CD46 gene is regulated by multiple G-patch proteins. (A)** Schematic representation of the structure of the CD46 gene with exons and introns depicted by rectangles and lines respectively. The two alternatively spliced exons targeted by several G-patch proteins are colored in red (E1 and E2) and the proteins whose depletion affects their splicing are shown above, with the arrows indicating the effects that they exert (↓ exon skipping, ↑ exon inclusion). **(B)** The splicing of the two alternative exons depicted in (A) was analyzed by RT-PCR in cells depleted of the indicated G-patch proteins or in cells treated with a control siRNA (siNT). The products corresponding to the inclusion and skipping isoforms are indicated to the right of each panel and the PSI values were calculated as the ratio of the inclusion isoform relative to the total level of both isoforms.

The DHX15 and G-patch protein knockdown RNA-seq samples were also analyzed for changes in gene expression levels in order to determine correlations with the alternatively spliced target genes or other effects on the transcriptome. For this, the normalized read count for each gene was expressed as counts per million and the fold change relative to

the control sample was calculated. The output of the analysis showed that depletion of DHX15 and of the majority of G-patch proteins tested did not significantly impact gene expression, with only a small number of genes (<100) being upregulated or downregulated with a log2 fold change of more than 1.5 (Figure 3.16). Exceptions to this were GPATCH8, SON and ZGPAT for which 222, 813 and 937 genes were downregulated, and SON and ZGPAT for which 186 and 469 genes were upregulated using the same threshold. Comparison of the differentially expressed genes with the alternative splicing targets identified by MISO revealed that only few of the genes whose splicing was regulated by DHX15 and G-patch proteins also showed changes in expression levels. The highest overlap was observed in the case of the SON knockdown sample, for which 167 out of the 1230 splicing targets had a decreased expression, while only one gene was expressed at a higher level.



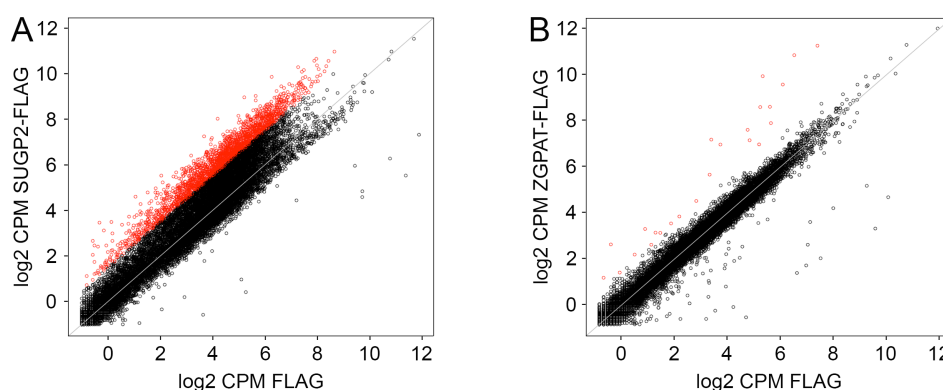
**Figure 3.16. Knockdown of DHX15 and of most G-patch proteins does not lead to significant changes in gene expression levels.** Total RNA extracted from cells treated with the indicated siRNAs was used for RNA-seq analysis and the obtained reads were mapped onto the human genome. The read count for each gene was normalized to the total number of mapped reads and the log2 fold change was calculated in each knockdown sample relative to the control sample. Genes were considered differentially expressed if the log2 fold change values were  $\geq 1.5$  (upregulated) or  $\leq -1.5$  (downregulated).

Taken together, the transcriptome analyses following knockdown of DHX15 and of its G-patch cofactors revealed that in most cases these proteins regulate alternative splicing without influencing gene expression levels. The identification of genes targeted by multiple G-patch proteins at the same splice sites or at different locations indicates that the action of these cofactors might be coordinated or redundant in some cases. Furthermore, the detection of splicing events regulated by both DHX15 and its G-patch cofactors suggests that they function together in this pathway. These data also propose novel factors involved in alternative splicing such as AGGF1 and provide an overview of the splicing targets for the analyzed proteins that represents a resource for more in-depth functional studies.

### 3.2.3 DHX15 and its G-patch cofactors regulate splicing either directly or indirectly

The finding that a large number of human G-patch proteins are involved in alternative splicing raised the question of whether they regulate the splicing of their target mRNAs by direct binding or in an indirect manner. To address this, the HEK293 stable cell lines expressing FLAG-tagged G-patch proteins were used in RNA IP experiments to identify binding to cellular RNAs. Cells were crosslinked *in vivo* with UV light of 254 nm, which introduced covalent bonds between proteins and RNA, and the tagged G-patch proteins were recovered using anti-FLAG beads. The co-precipitated RNA was extracted and analyzed by next-generation sequencing followed by mapping of the sequencing reads onto the human genome and their assignment to the corresponding genomic features.

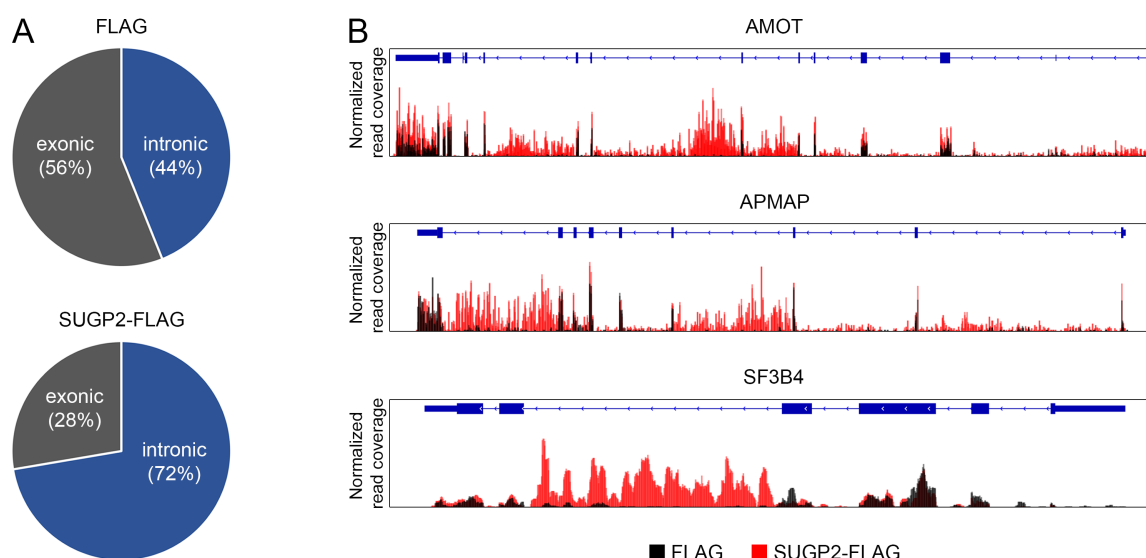
For this analysis, we focused on SUGP2 and ZGPAT, which were among the G-patch proteins with the highest number of regulated splicing events (Table 3.3). First, the reads mapping to protein-coding genes were examined and, interestingly, for SUGP2 more than 2000 mRNAs were found enriched compared to the FLAG control with a log<sub>2</sub> fold change greater than 1.5 (2.25-fold enrichment; Figure 3.17A). Approximately 40% of these target mRNAs were enriched at least four-fold and 6% accumulated more than six-fold. In contrast, for ZGPAT only about 20 mRNAs were enriched with a log<sub>2</sub> fold change greater than 1.5 relative to the control, suggesting that this protein does not crosslink to mRNAs (Figure 3.17B).



**Figure 3.17. SUGP2 crosslinks to thousands of mRNAs. (A-B)** Stable cell lines expressing SUGP2-FLAG or ZGPAT-FLAG and cells encoding the FLAG tag only were crosslinked at 254 nm and protein-RNA complexes were immunoprecipitated. The isolated RNA was analyzed by next-generation sequencing and the obtained reads were mapped to the corresponding genomic features. Scatterplots displaying the log<sub>2</sub> values of the normalized read counts (counts per million, CPM) for protein-coding genes in these samples were generated. The mRNAs that were enriched with a log<sub>2</sub> fold change  $\geq 1.5$  in the SUGP2-FLAG (A) or ZGPAT-FLAG (B) sample compared to the FLAG control are colored in red.

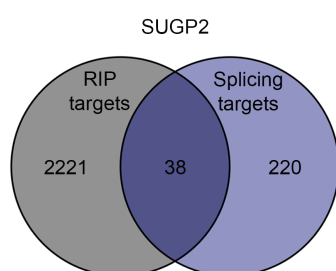
Closer inspection of the distribution of reads mapping to protein-coding genes in the SUGP2 IP sample revealed a higher percentage of intronic reads compared to reads mapping to exons, which was different than the distribution observed for the FLAG control (Figure

3.18A). Representative views of the read coverage for three of the SUGP2 target genes (AMOT, APMAP and SF3B4) confirmed the accumulation of reads mapping to introns (Figure 3.18B).



**Figure 3.18. SUGP2 binds preferentially pre-mRNAs.** (A) The reads mapping to protein-coding genes in the FLAG and SUGP2-FLAG RIP-seq samples were analyzed to obtain their distribution into exonic and intronic regions. (B) The normalized read coverage in the FLAG (black) and SUGP2-FLAG (red) samples for three representative genes (AMOT, APMAP and SF3B4) were illustrated using IGV. The exons and introns are depicted in blue as rectangles and lines respectively.

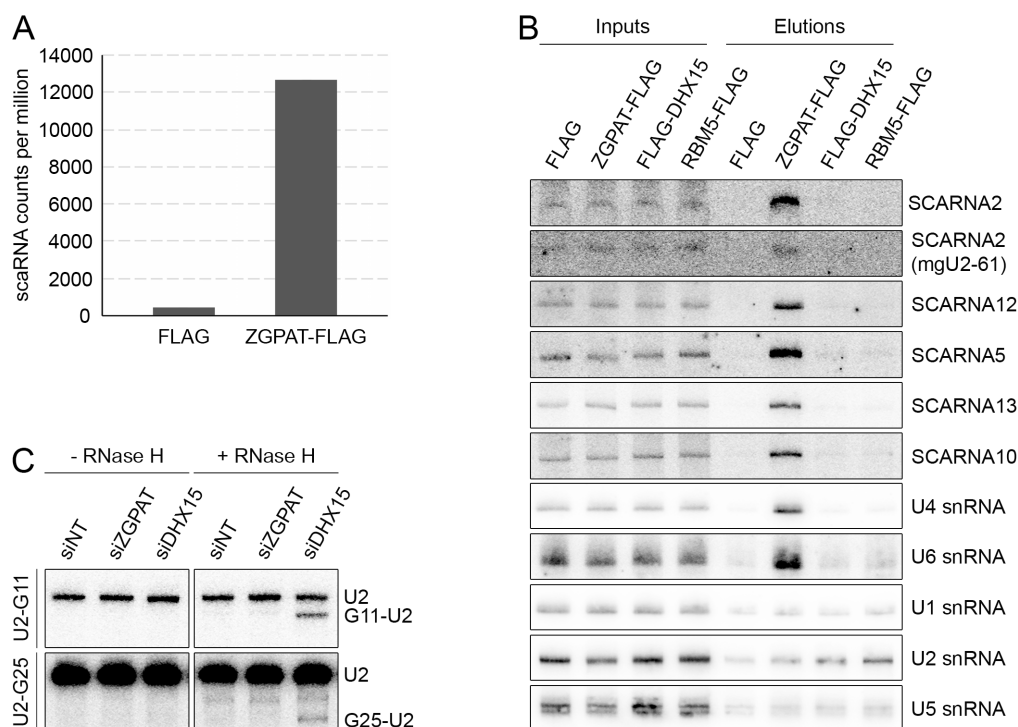
Overall, these results strongly suggested that SUGP2 binds pre-mRNAs at an early stage during their processing, which would be consistent with a role in splicing regulation. Therefore, the mRNAs bound by SUGP2 in the RIP experiment were compared with the mRNAs whose splicing was affected by depletion of this G-patch protein and 38 common targets were found (Figure 3.19), including THEM4 for which the role of SUGP2 in its alternative splicing had been confirmed by RT-PCR (Figure 3.13F). These results indicate that SUGP2 might regulate alternative splicing by direct binding to pre-mRNAs and that the two complementary approaches, namely the RIP-seq and the knockdown coupled to RNA-seq analysis, could together reveal the full spectrum of RNA targets for SUGP2 and for other G-patch proteins.



**Figure 3.19. Overlap between the SUGP2 bound mRNAs and its alternative splicing targets.** The protein-coding genes enriched in the SUGP2-FLAG RIP-seq analysis with a  $\log_2$  fold change  $\geq 1.5$  relative to the FLAG control (RIP targets) were compared with the genes whose splicing was affected by knockdown of SUGP2 based on the RNA-seq and MISO analysis (Splicing targets) and the overlap is shown.



Apart from mRNAs, binding to other RNA types was analyzed and, excitingly, for ZGPAT a 30-fold increase in the number of reads mapping to scaRNAs was observed compared to the FLAG control (Figure 3.20A). These small RNA molecules guide the 2'-O-methylation and pseudouridylation of snRNAs that take place in Cajal bodies where ZGPAT was also detected by our localization experiments and by prior studies (Figure 3.9; Chen et al., 2017). Analogous to scaRNAs, snoRNAs direct the methylation and pseudouridylation of rRNAs in the nucleolus. However, a few snoRNAs are involved in the modification of U6 snRNA, which follows a different biogenesis pathway than the other spliceosomal snRNAs.



**Figure 3.20. ZGPAT crosslinks to multiple scaRNAs/snRNAs and depletion of DHX15 affects snRNA modification.** (A) Stable cell lines expressing the FLAG tag or ZGPAT-FLAG were induced with tetracycline and crosslinked at 254 nm. The bait proteins were immunoprecipitated and the co-eluted RNA was extracted and analyzed by next-generation sequencing. The obtained reads were mapped onto the human genome and the normalized read count for scaRNA genes in the two samples is shown. (B) Samples were prepared as in (A) except that the isolated RNA was analyzed by northern blotting using the probes indicated on the right. Input samples representing 0.5% of the lysates were also included. (C) Total RNA extracted from cells treated with a non-target siRNA (siNT) or with siRNAs against ZGPAT and DHX15 was used for site-specific RNase H cleavage with oligonucleotides targeting the G11 and G25 residues of U2 snRNA as indicated on the left. The reactions were analyzed by northern blotting with a probe against the U2 snRNA and the position of the uncleaved and cleaved products is indicated on the right.

Analysis of the RIP-seq read distribution for individual scaRNA genes revealed that multiple scaRNAs were significantly enriched in the ZGPAT eluate, with SCARNA2 accumulating approximately 80 times more compared to the control sample (Table 3.4). Interestingly, SCARNA2 guides the 2'-O-methylation of U2 snRNA G25 and C61 and is also processed into a shorter fragment (mgU2-61) that likewise directs methylation of U2 snRNA C61 and

also of U2 snRNA G11 (Lestrade and Weber, 2006). Other ZGPAT targets included scaRNAs involved in the methylation and/or pseudouridylation of U1, U2, U4 and U5 snRNAs (Table 3.4), as well as the snoRNAs SNORA79B and SNORD10 that act on U6 snRNA.

**Table 3.4. Overview of the scaRNAs and snRNAs enriched in the ZGPAT RIP-seq analysis.** The normalized read count for scaRNA and snRNA genes expressed as counts per million (CPM) is shown for the ZGPAT-FLAG RIP-seq sample together with the log2 fold change relative to the FLAG control sample.

scaRNA/snRNA	CPM ZGPAT	log2 CPM (ZGPAT/FLAG)
SCARNA2	5298.8	6.3
SCARNA12	1783.1	5.9
SCARNA17	54.2	5.4
SCARNA9	162.4	5.1
SCARNA6	418.9	5.1
SCARNA5	562.1	5.0
SCARNA13	963.6	4.9
SCARNA10	2425.7	4.8
SCARNA7	752.6	4.0
SCARNA21	43.3	3.5
SCARNA22	37.7	2.1
SCARNA4	3.9	1.9
SCARNA8	2.5	1.2
SCARNA14	0.8	1.1
SCARNA11	1.2	1.1
SCARNA3	1.2	0.9
SCARNA21B	0.3	0.7
SCARNA1	10.4	0.6
SCARNA16	140.2	0.4
SCARNA23	4.1	0.4
SCARNA20	0.5	0.3
SCARNA15	0.1	0.3
SCARNA18B	0.1	0.0
SCARNA18	0.6	-0.1
U4 snRNA	10527	4
U6 snRNA	254	3.9
U1 snRNA	11473.4	0.6
U2 snRNA	1667	0.2
U5 snRNA	203.9	0.2

In addition to scaRNAs and snoRNAs, an accumulation of U4 and U6 snRNAs was observed for the ZGPAT IP sample compared to the FLAG control, which is in agreement with results from another study (Chen et al., 2017). However, our data did not show binding of the G-patch protein to U5 snRNA as reported in the same study, suggesting that ZGPAT

does not directly contact U5 snRNA and that the previously described interaction was most likely indirect.

To confirm the RNA interactions detected for ZGPAT, the RNA IP was repeated and this time the samples were analyzed by northern blotting using probes complementary to a few selected scaRNA and snRNA targets. The experiment was done with cells expressing only the FLAG tag or FLAG-tagged ZGPAT, DHX15 and RBM5, with the latter serving as an additional control for specificity. The results of the northern blotting analysis confirmed the binding of ZGPAT to multiple scaRNAs (SCARNA2 full-length and mgU2-61 fragment, SCARNA12, SCARNA5, SCARNA13 and SCARNA10) as well as to U4 and U6 snRNAs (Figure 3.20B). At the same time, the levels of U1, U2 and U5 snRNAs were not increased in the ZGPAT eluate relative to the FLAG control, indicating specific binding of the G-patch protein to its scaRNA and snRNA targets. For DHX15, despite its strong interaction with ZGPAT (Figure 3.2; Table 3.2), none of the analyzed scaRNAs were enriched in the elution fraction compared to the FLAG and RBM5 IP samples, suggesting that the helicase might be associated with these small RNAs indirectly or transiently and that only the G-patch protein contacts them directly. RBM5 did not interact with any of the scaRNAs and snRNAs tested except for U2 snRNA, which is in line with its detection in U2 snRNP-related spliceosomal complexes (Agafonov et al., 2011).

The direct crosslinking of ZGPAT to U4 and U6 snRNAs is consistent with the identification of these components in a 35S splicing complex that also contains U5 snRNA, DHX15 and other splicing factors and that represents an intermediate in the assembly of the U4/U6.U5 tri-snRNP (Chen et al., 2017). Interestingly, our data revealed that ZGPAT binds scaRNAs that guide modification of all spliceosomal snRNAs and not only of U4 and U6 snRNAs, indicating that this protein might have several independent functions in splicing. Therefore, given that its association with U4 and U6 has been described elsewhere and has been suggested to involve DHX15, the next aim was to understand the functional basis of the interactions that ZGPAT establishes with scaRNAs and to determine whether the helicase is implicated in this as well.

To check if ZGPAT and DHX15 play a role in the scaRNA-guided modification of snRNAs, the modification status of U2 snRNA at two sites of 2'-O-methylation (G11 and G25) that are targeted by the SCARNA2 fragment mgU2-61 or by full-length SCARNA2 respectively was analyzed using a site-specific RNase H cleavage assay. In this method, RNase H is guided by a 2'-O-methyl RNA-DNA chimeric oligonucleotide and cleaves the snRNA-DNA duplex region at a specific residue unless the target base is 2'-O-methylated. Total RNA extracted from cells treated with NT siRNA or with siRNAs against ZGPAT or DHX15 was

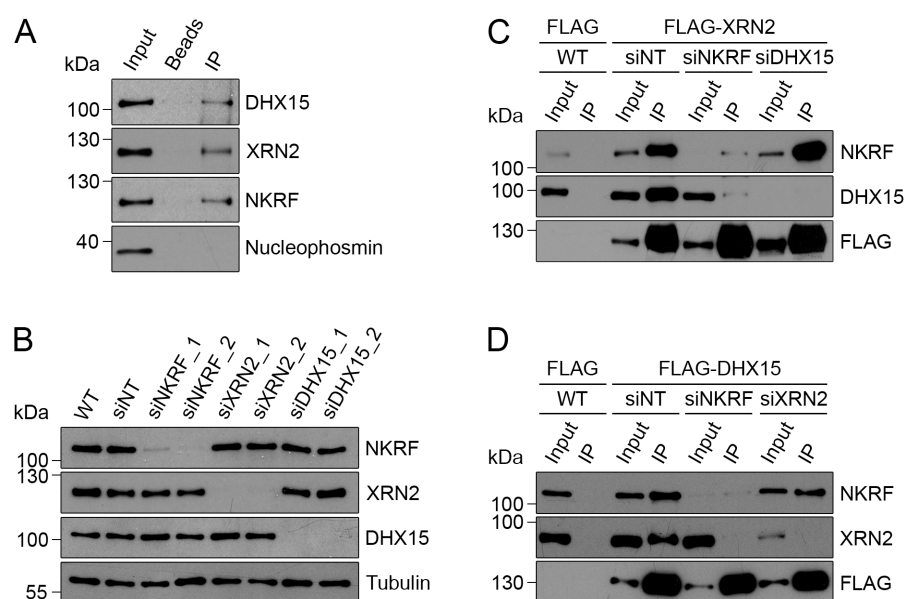
incubated with or without RNase H in the presence of the chimeric oligonucleotides and the reaction products were analyzed by northern blotting using a probe to detect U2 snRNA. In the control samples treated with NT siRNA, U2 snRNA was not cleaved by RNase H (Figure 3.20C), which is consistent with the almost complete 2'-O-methylation of the G11 and G25 residues in normal conditions (Krogh et al., 2017). On the other hand, in the DHX15 knockdown sample, U2 snRNA was cleaved at both G11 and G25, albeit weakly at the latter site, indicating a decrease in the methylation level of these residues. Considering that ZGPAT was found to crosslink to the scaRNAs responsible for these modifications, this suggested that DHX15 might have a direct role in the scaRNA-guided methylation of U2 snRNA together with the G-patch protein. Contrary to this hypothesis, knockdown of ZGPAT did not affect the extent of 2'-O-methylation at the sites tested (Figure 3.20C). This might be explained by the less efficient knockdown of ZGPAT, implying that the residual protein is sufficient for carrying out this function (Figure 3.10). Alternatively, ZGPAT might not be required for snRNA modification, but instead might bind scaRNAs for another purpose, such as assisting in their biogenesis. Furthermore, it is also possible that the role of DHX15 in snRNA modification is not linked to scaRNAs and is instead mediated through the protein components of the 2'-O-methylation machinery (FBL, NOP56, NOP58, SNU13), but inspection of the RNA-seq data did not reveal any changes in the expression level or alternative splicing of these factors upon knockdown of DHX15. Taken together, these data show that ZGPAT interacts directly with multiple scaRNAs and suggest that this might be linked to the role of DHX15 in snRNA modification, which in turn could influence pre-mRNA splicing.

### **3.2.4 DHX15 and the G-patch protein NKRF function together in ribosome biogenesis**

#### **3.2.4.1 DHX15 and NKRF form a nucleolar complex with XRN2**

The results so far showed that DHX15 and a large number of its G-patch cofactors regulate splicing, either by probably binding directly to target mRNAs as in the case of SUGP2 or by influencing snRNA modification patterns as suggested for DHX15 possibly together with ZGPAT. Thus, DHX15 likely impacts splicing at multiple levels through its G-patch cofactors and acts as a master regulator of this process. In higher eukaryotes, the majority of RNA helicases are suggested to be multifunctional and to regulate different steps of gene expression in a coordinated manner (Jankowsky, 2011; Bourgeois et al., 2016). Therefore, we explored additional functions of DHX15 and, since four of its G-patch cofactors are localized in nucleoli (Figure 3.9), the attention was focused on ribosome biogenesis, which is initiated at these sites. IP experiments followed by MS analysis identified the nucleolar protein NKRF as the top interacting G-patch protein for FLAG-DHX15 based on the total spectral counts and the enrichment compared to the control (Table 3.2). Apart from NKRF,

another protein highly abundant in the DHX15 IP sample was the 5'-3' exonuclease XRN2 (124-fold enrichment relative to the FLAG control), which is involved in multiple aspects of RNA metabolism, including ribosome biogenesis (Miki and Grosshans, 2013). Interestingly, XRN2 was found to interact in mammals with proteins containing a DUF3469 domain (also called XRN-Two-Binding-Domain, XTBD) and sequence analysis revealed that the G-patch protein NKRF is among the three human XTBD-containing proteins (Miki et al., 2014; Richter et al., 2016). This indicated that NKRF might also bind XRN2, and given the strong association between NKRF and DHX15, we hypothesized that these interactions might be related. Therefore, to study whether DHX15, NKRF and XRN2 are present together in endogenous complexes, immunoprecipitation with an anti-NKRF antibody was performed and the co-precipitated proteins were analyzed by western blotting. This showed that together with NKRF, both DHX15 and XRN2 were recovered in the anti-NKRF IP eluate but not in the control sample (Figure 3.21A). At the same time, the nucleolar protein Nucleophosmin was not co-precipitated with NKRF, demonstrating the specificity of the interactions with DHX15 and XRN2.

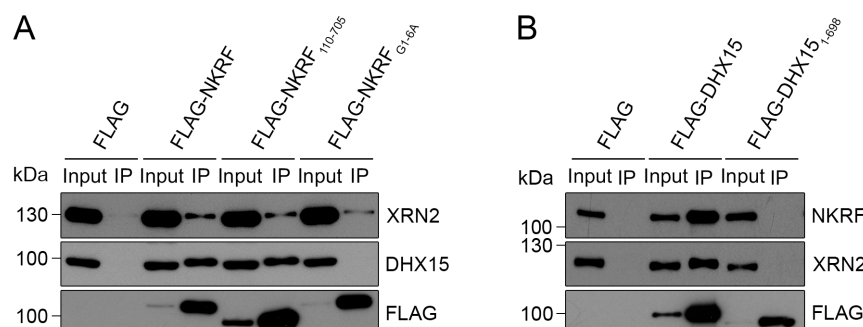


**Figure 3.21. DHX15, NKRF and XRN2 are associated into a complex *in vivo*.** (A) HeLa cells were used for immunoprecipitation with an anti-NKRF antibody and the eluates (IP) were analyzed by western blotting with the indicated antibodies. A control IP using beads lacking the antibody (Beads) and an input sample representing 1% of the lysate were also analyzed. (B) Extracts from HeLa cells that were either untreated (WT) or treated for 96 h with a control siRNA (siNT) or with siRNAs against the indicated targets were analyzed by western blotting to determine the levels of the proteins shown on the right. (C-D) HEK293 cells were treated with the indicated siRNAs for 96 h and the expression of FLAG-tagged XRN2 (C) or DHX15 (D) was induced in the last 24 h with tetracycline. Lysates prepared from these cells were used for anti-FLAG IPs and the co-precipitated proteins were detected by western blotting using the indicated antibodies. Untreated cells (WT) expressing the FLAG tag were also analyzed. The panels in this figure were originally published in Memet et al., 2017.

To gain insight into how the DHX15-NKRF-XRN2 complex is organized, the interdependency of these interactions was examined by depleting each protein and analyzing the association between the other two factors. First, siRNA-mediated knockdowns of DHX15, NKRF and XRN2 were established and efficient depletion of the target proteins was confirmed by western blotting (Figure 3.21B). Stable cell lines encoding FLAG-tagged XRN2 or DHX15 were treated with these siRNAs or with a non-target siRNA, followed by induction of transgene expression and IP with an anti-FLAG antibody. The results of the western blotting analysis showed that FLAG-XRN2 interacted efficiently with both DHX15 and NKRF when treated with non-target siRNA, but upon depletion of NKRF, its interaction with DHX15 was strongly reduced (Figure 3.21C). In contrast, knockdown of DHX15 did not influence binding of FLAG-XRN2 to NKRF. Similarly, both NKRF and XRN2 were recovered with FLAG-DHX15 in cells treated with siNT and depletion of NKRF affected the interaction between FLAG-DHX15 and XRN2 (Figure 3.21D). However, knockdown of XRN2 did not influence binding of FLAG-DHX15 to NKRF. In all cases, cells expressing the FLAG tag were used as a control and none of the analyzed proteins were co-precipitated with this sample. Overall, these results demonstrate that DHX15, NKRF and XRN2 associate in a complex *in vivo* and that NKRF mediates the interaction between DHX15 and XRN2.

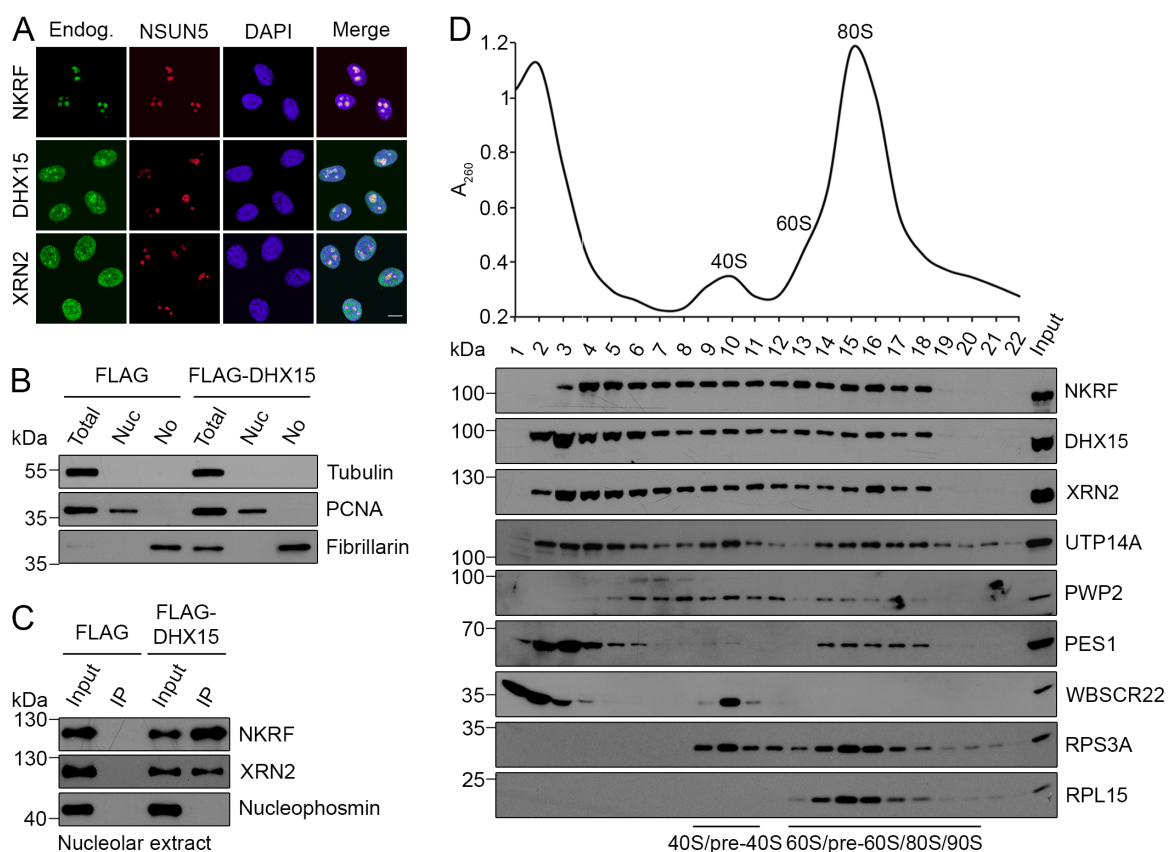
Several studies have shown that the regulation of RNA helicases by G-patch proteins requires the G-patch domain of the cofactor and the OB-fold region of the helicase (Lebaron et al., 2009; Christian et al., 2014; Heininger et al., 2016; Inesta-Vaquera et al., 2018). In addition to its G-patch domain, NKRF contains a domain (XTBD) that has been described in other proteins to mediate binding to XRN2 (Miki et al., 2014; Richter et al., 2016). Based on these observations, the contribution of the different regions of NKRF and of the DHX15 OB-fold domain for their interactions within the complex with XRN2 was analyzed. For NKRF, a truncated form lacking the XTBD (NKRF<sub>110-705</sub>) and a mutant protein containing six glycine to alanine substitutions in the G-patch domain (NKRF<sub>G1-6A</sub>) were generated, while for DHX15 the OB-fold domain was deleted to produce a truncated protein (DHX15<sub>1-698</sub>). These constructs were expressed with a FLAG tag from HEK293 stable cell lines together with the wild-type proteins and the FLAG tag only control, and immunoprecipitation experiments were carried out. Western blotting analysis of the IP samples revealed, as expected, that wild-type FLAG-NKRF interacted with XRN2 and DHX15 (Figure 3.22A). Similarly, FLAG-NKRF<sub>110-705</sub> bound both proteins, indicating that for NKRF the XTBD is not required for association with XRN2 as in other proteins. On the other hand, mutations in the G-patch domain of NKRF or deletion of the OB-fold domain of DHX15 led to a loss of the NKRF-DHX15 interaction, demonstrating that these proteins associate in a similar mode to

other G-patch protein-RNA helicase complexes (Figure 3.22A-B). In addition, while binding of NKRF<sub>G1-6A</sub> to XRN2 was only mildly affected, DHX15<sub>1-698</sub> was not able to interact with XRN2, supporting the model that NKRF is the core component of the complex with DHX15 and XRN2.



**Figure 3.22. The NKRF-DHX15 interaction is mediated by the G-patch domain of NKRF and the OB-fold of the helicase. (A-B)** HEK293 stable cell lines were induced for 24 h with tetracycline to express the indicated NKRF (A) and DHX15 (B) variants or the FLAG tag. Cell extracts were used in IP experiments with an anti-FLAG antibody followed by western blotting detection of the co-precipitated proteins using the indicated antibodies. The panels in this figure were originally published in Memet et al., 2017.

Having established that DHX15, NKRF and XRN2 associate in a complex *in vivo*, the next aim was to determine in which cellular compartment these interactions occur. All three factors have been identified in proteomic screens of isolated nucleoli and the localization studies for FLAG-tagged NKRF confirmed that the G-patch protein is present at these sites (Andersen et al., 2005; Figure 3.9). To assess if endogenous NKRF is also found in nucleoli and to check for the localization of DHX15 and XRN2, immunofluorescence microscopy was performed in HeLa cells using antibodies against these proteins together with an antibody against the rRNA methyltransferase NSUN5 as a nucleolar marker. The results in Figure 3.23A showed that DHX15, NKRF and XRN2 were present in nucleoli based on colocalization with NSUN5, suggesting that they interact at these sites. DHX15 and XRN2 also displayed a nucleoplasmic staining consistent with their known functions in processes such as pre-mRNA splicing and RNA polymerase II transcription respectively. To assess if the DHX15-NKRF-XRN2 complex is assembled in the nucleolus, subcellular fractionation of cells expressing the FLAG tag or FLAG-DHX15 was performed to isolate nucleoli, whose purity was confirmed by western blotting using antibodies against  $\alpha$ -Tubulin, PCNA and Fibrillarin as markers for the cytoplasm, nucleoplasm and nucleoli respectively (Figure 3.23B). Lysates prepared from the purified nucleoli were used in anti-FLAG IP experiments, which showed that FLAG-DHX15 interacted with both NKRF and XRN2 in this subcellular compartment (Figure 3.23C). In contrast, the nucleolar protein Nucleophosmin was not co-precipitated with the helicase, implying that DHX15 binds specifically to NKRF and XRN2.



**Figure 3.23. DHX15, NKRF and XRN2 interact in the nucleolus and co-migrate with pre-ribosomal complexes.** (A) HeLa cells were analyzed by immunofluorescence microscopy using antibodies against endogenous NKRF, DHX15 or XRN2 (Endog., green) and the nucleolar protein NSUN5 (red). Nuclei were visualized by DAPI staining (blue) and overlays of the three channels are shown (Merge). The scale bar represents 10  $\mu$ m. (B) Subcellular fractionation of cells expressing the FLAG tag or FLAG-DHX15 was performed and samples corresponding to whole lysates (Total) and the nucleoplasmic (Nuc) and nucleolar (No) fractions were analyzed by western blotting using antibodies against  $\alpha$ -Tubulin, PCNA and Fibrillarin as markers for the cytoplasm, nucleoplasm and nucleoli respectively. (C) Lysates prepared from nucleoli isolated as in (B) were used in IP experiments with an anti-FLAG antibody and the co-eluted proteins were analyzed by western blotting together with input samples representing 1% of the nucleolar lysates. (D) HeLa cell extracts were separated on sucrose gradients and the obtained fractions were analyzed by western blotting using antibodies against the indicated proteins. The absorbance at 260 nm was measured for each fraction to determine the position of the 40S and 60S ribosomal subunits and of the 80S monosome. This figure was originally published in Memet et al., 2017.

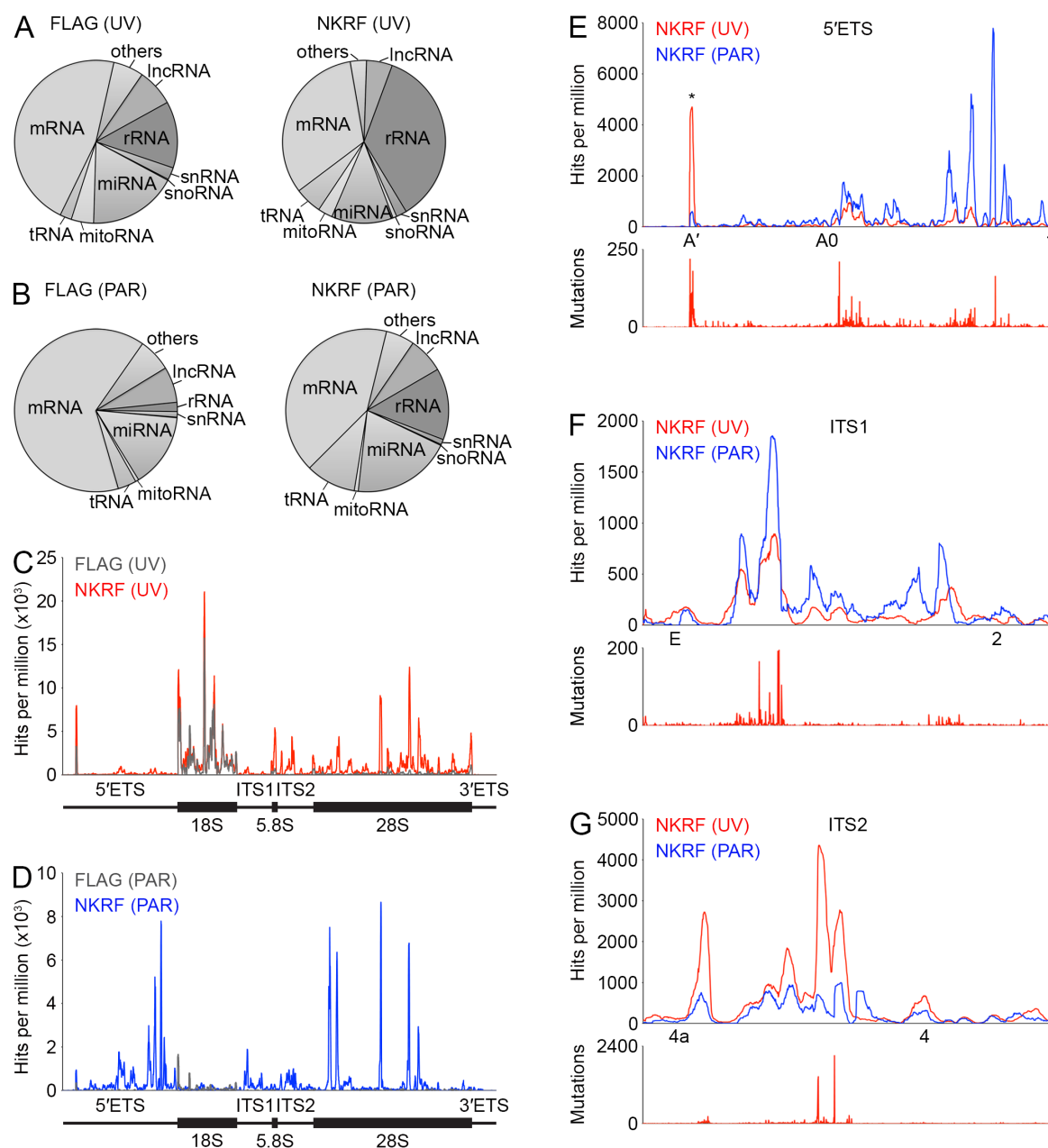
The nucleolus represents the site at which ribosome production is initiated and contains a high concentration of pre-ribosomal complexes that are formed through the recruitment of ribosomal proteins and ribosome assembly factors to sites of rDNA transcription. Therefore, to check whether DHX15, NKRF and XRN2 associate with pre-ribosomes, whole cell extracts were separated on sucrose gradients and the distribution of these proteins in fractions of different sedimentation coefficients was followed by western blotting. The fractions containing the mature 40S small ribosomal subunit, the 60S large ribosomal subunit and the 80S ribosome were determined by measuring the RNA content (absorbance



at 260 nm, A260) and by the presence of the ribosomal proteins RPS3A and RPL15 (Figure 3.23D). The position of the different pre-ribosomal complexes was established based on the migration patterns of previously described ribosome biogenesis factors: the 90S/early pre-40S components UTP14A and PWP2, and the pre-60S and late pre-40S factors PES1 and WBSR22 respectively (Dosil and Bustelo, 2004; Rohrmoser et al., 2007; Hu et al., 2011; Ounap et al., 2013; Haag et al., 2015; Sloan et al., 2015). Analysis of the distribution of DHX15, NKRF and XRN2 in the sucrose gradient fractions showed that they co-migrated both in the top part of the gradient that contains free proteins and small complexes as well as in fractions containing (pre)-ribosomal particles (Figure 3.23D). Together with the nucleolar localization of these proteins, these results indicate that DHX15, NKRF and XRN2 associate with pre-ribosomes during their maturation.

#### **3.2.4.2 NKRF crosslinks to the pre-rRNA transcript at several sites**

To decipher the function of the DHX15-NKRF-XRN2 complex, the RNA interactions that the core component NKRF forms in cells were studied, similar to the genome-wide analyses employed to identify the RNA targets of other G-patch proteins described in 3.2.2. However, due to its association with pre-ribosomes, a different approach was required for NKRF that would also enable the identification of its binding site(s) on (pre)-rRNA. Thus, the crosslinking and analysis of cDNA (CRAC) method was used, in which cells expressing FLAG-NKRF or the FLAG tag only were crosslinked with UV at 254 nm to introduce covalent bonds between proteins and their interacting RNAs (UV-CRAC). Alternatively, cells were treated for 6 h with the photoactivatable ribonucleoside 4-thiouridine, which when incorporated into RNA can be specifically crosslinked to proteins using UV light of 365 nm (PAR-CRAC; Hafner et al., 2010). After crosslinking, the FLAG-tagged proteins, which also contain a His<sub>6</sub> tag in their sequence (Table 2.3), were recovered in a two-step purification using anti-FLAG and Ni-NTA beads. A partial RNase digestion step was performed to remove any RNAs that were not protected by the protein and the trimmed RNA was ligated to sequencing adapters, reverse-transcribed and amplified by PCR to generate a library for next-generation sequencing. Mapping of the sequencing reads onto the human genome showed an accumulation of reads corresponding to (pre)-ribosomal RNA in the FLAG-NKRF sample compared to the FLAG control (36% compared to 13% for UV-CRAC and 14% compared to 2% for PAR-CRAC of the total number of mapped reads; Figure 3.24A-B). Visualization of the CRAC read coverage at the rDNA locus encoding the 47S pre-rRNA transcript revealed several binding sites for NKRF, in the spacer regions as well as within the 28S rRNA sequence (Figure 3.24C-D).



**Figure 3.24. NKRF crosslinks to the pre-rRNA transcript in the spacer regions and within the 28S rRNA.**

(A-B) HEK293 stable cells expressing the FLAG tag or FLAG-NKRF were crosslinked at 254 nm (UV; A) or were treated with 4-thiouridine for 6 h and crosslinked at 365 nm (PAR; B). Protein-RNA complexes were isolated and the RNA fragments protected by the protein were extracted and used to generate a cDNA library for next-generation sequencing. The sequencing reads were mapped onto the human genome and the distribution of reads into the different classes of RNA is shown. (C-D) UV-CRAC (C) or PAR-CRAC (D) sequencing reads were analyzed as in (A-B) and the read coverage for the rDNA locus encoding the 47S pre-rRNA transcript is shown for the FLAG (gray) and FLAG-NKRF (red/blue) samples. A schematic representation of the 47S transcript with the mature rRNAs and the spacer regions drawn as rectangles and lines respectively is shown below each panel. (E-G) Magnified views of the read coverage corresponding to the 5'ETS (E), ITS1 (F) and ITS2 (G) regions of the pre-rRNA with the cleavage sites marked below each panel. The lower panels show the number of reads containing deletions or substitutions (Mutations) in the UV-CRAC experiment, which are indicative of crosslinked nucleotides. This figure was originally published in Memet et al., 2017.

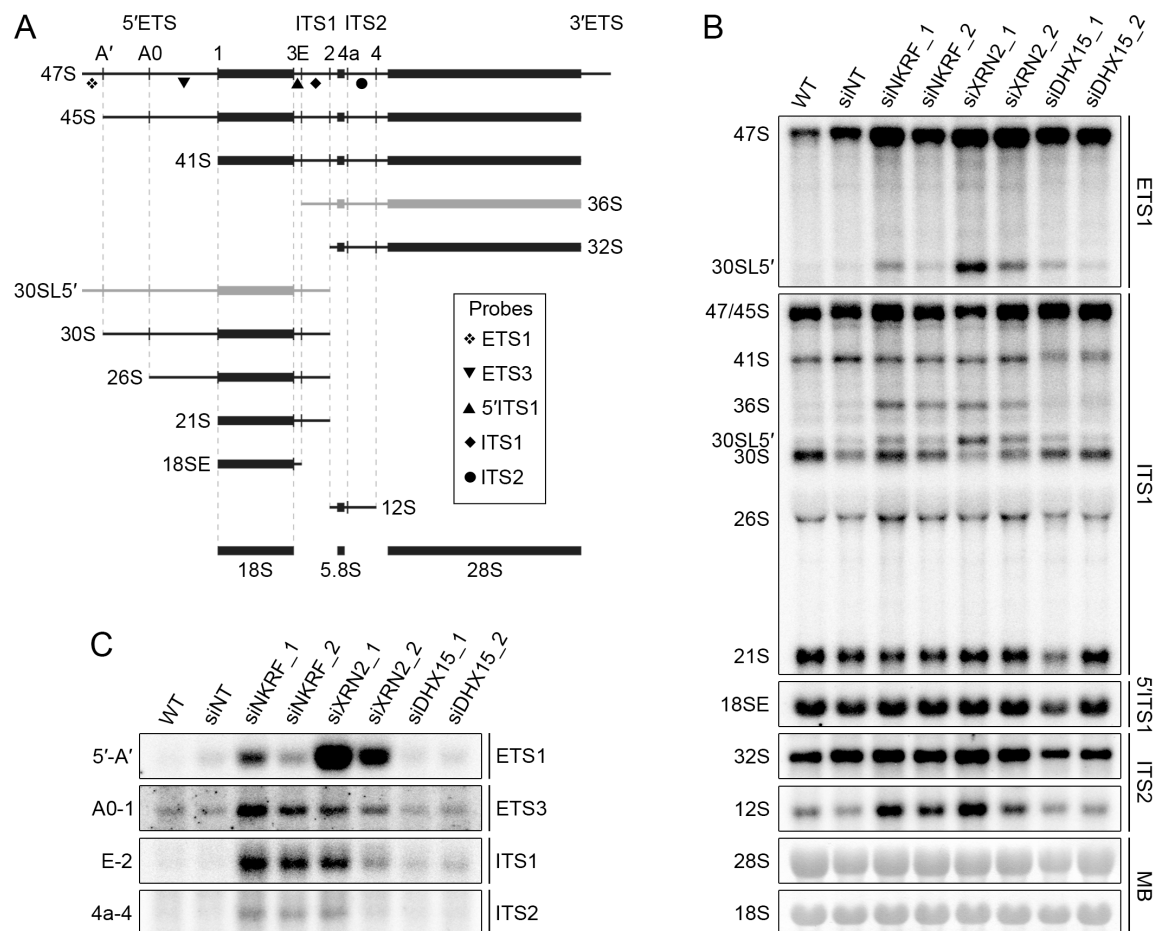
Closer examination of the reads mapping to pre-rRNA spacers showed NKRF crosslinking sites in the 5'ETS, ITS1 and ITS2 regions that mostly overlapped in the two methods used (Figure 3.24E-G). For UV-CRAC, these binding sites were supported by the detection of mutations at these loci, which are generally introduced at the reverse transcription step when a crosslinked nucleotide is encountered. Thus, these data imply that NKRF binds the pre-rRNA transcript at an early stage during its maturation and taken together with the association of DHX15, NKRF and XRN2 with pre-ribosomal particles suggest that this nucleolar complex might have a role in ribosome production.

#### **3.2.4.3 DHX15, NKRF and XRN2 are required for efficient pre-rRNA processing**

To gain further insight into the function of the complex involving DHX15, NKRF and XRN2, knockdowns of all three factors were performed and the processing of pre-rRNA was analyzed by northern blotting using probes hybridizing at specific sites to detect the different precursors of the mature rRNAs (Figure 3.25A). This showed that, compared to untreated cells (WT) or cells treated with NT siRNA, depletion of DHX15, NKRF and XRN2 led to an increase in the levels of the 47S and 30SL5' intermediates, which accumulate when processing at site A' in 5'ETS is impaired and is preceded by cleavage in ITS1 (Figure 3.25B). Nevertheless, knockdown of these factors did not impact the production of mature ribosomal subunits as revealed by pulse-chase labelling and detection of newly-synthesized 18S and 28S rRNA, as well as analysis of the steady-state levels of the mature rRNAs by methylene blue staining, indicating that only the order of the processing events is affected (Figure 3.25B; Figure 3.26). Consistent with these results, changes in the kinetics of the different pre-rRNA cleavage steps have been previously observed after knockdown of XRN2 and it has been suggested that the existence of several processing pathways might ensure that the production of the mature rRNAs is not blocked (Sloan et al., 2013). In addition, despite the fact that A' cleavage is suggested to take place co-transcriptionally, the pulse-chase analysis did not show any significant changes in pre-rRNA transcription after knockdown of these factors (Figure 3.26; Lazdins et al., 1997).

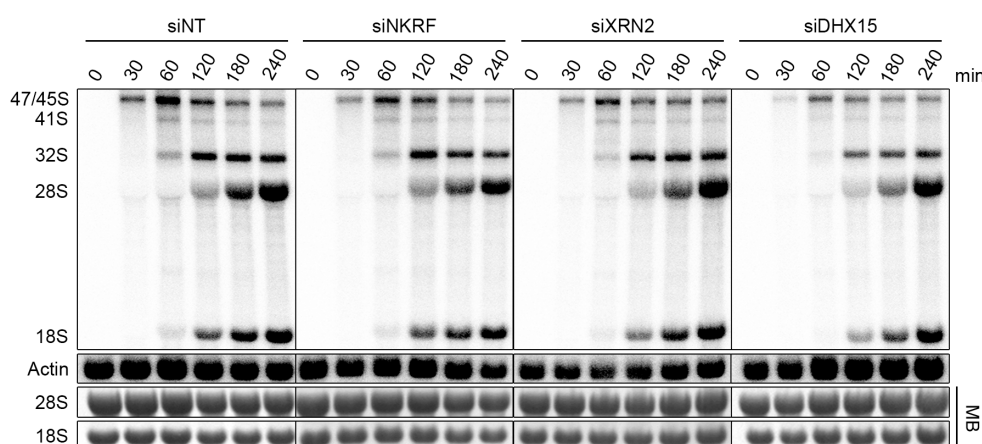
Depletion of NKRF and XRN2, but not of DHX15, also caused an accumulation of the 36S precursor, which forms when cleavage in ITS1 takes place at site E before site 2, and of the 12S intermediate, which is generated by processing of 32S and is the precursor to the 5.8S mature rRNA (Figure 3.25B). This indicated that NKRF and XRN2 might have additional roles in pre-rRNA processing that are independent of the helicase. In line with this, knockdown of NKRF and XRN2 also led to an increase in the levels of the pre-rRNA spacer fragments corresponding to the 5'-A' and A0-1 regions in 5'ETS, and the E-2 and 4a-4 regions in ITS1 and ITS2 respectively, which are normally degraded by the exonuclease after being excised (Figure 3.25C; Wang and Pestov, 2011; Schillewaert et

al., 2012; Sloan et al., 2013; Sloan et al., 2014). Considering that NKRF was found to crosslink to these pre-rRNA spacers (Figure 3.24), these results imply that the G-patch protein is required for the function of XRN2 in their turnover. Taken together, these data suggest that DHX15, NKRF and XRN2 are required for efficient pre-rRNA cleavage at site A' in the 5'ETS, while NKRF and XRN2 also function together in the degradation of pre-rRNA spacer sequences that are released during processing.



**Figure 3.25. DHX15, NKRF and XRN2 are required for pre-rRNA processing and turnover of by-products.**

**(A)** Schematic representation of the pre-rRNA processing pathway in human cells. The mature rRNAs are shown as rectangles and the spacer regions are drawn as lines. The cleavage sites are marked above the 47S transcript and the hybridization position of the probes indicated in the box is shown below. The aberrant intermediates that are not part of the normal processing pathway are colored in gray. **(B)** HeLa cells were either untreated (WT) or treated for 96 h with a control siRNA (siRNA) or with siRNAs against the indicated genes. Total RNA extracted from these cells was analyzed by northern blotting and the different pre-rRNA intermediates, which are marked on the left, were detected using the probes indicated on the right side. The 28S and 18S mature rRNAs were visualized by methylene blue (MB) staining. **(C)** Samples were prepared as in (B) and the levels of the pre-rRNA fragments indicated on the left were analyzed by northern blotting using the corresponding probes. This figure was originally published in Memet et al., 2017.



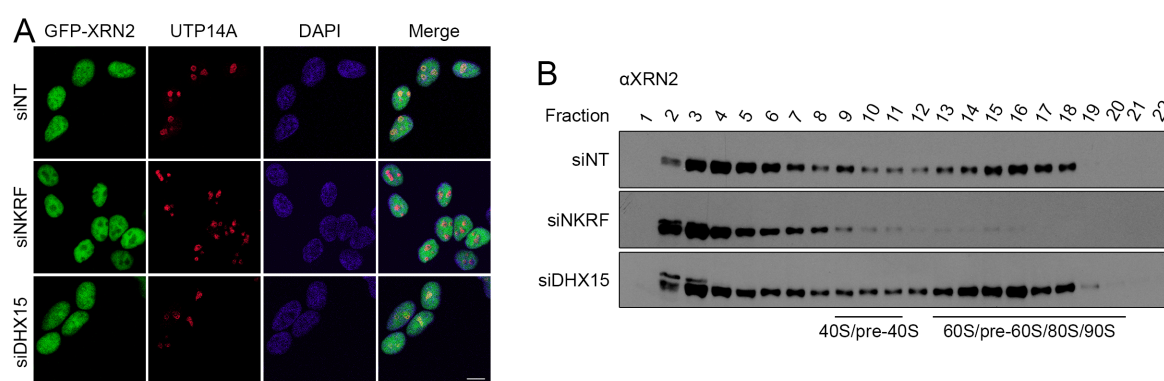
**Figure 3.26. Knockdown of DHX15, NKRF or XRN2 does not affect pre-rRNA transcription and the production of mature rRNAs.** HeLa cells were treated for 84 h with a non-target siRNA (siNT) or with siRNAs against NKRF, XRN2 or DHX15 and nascent RNAs were labelled with [ $^{32}$ P]-orthophosphate. After labelling, cells were grown in unlabelled media and harvested at the indicated time points. Total RNA was extracted and analyzed by northern blotting followed by detection of the labelled nascent RNAs by phosphorimaging. In addition, actin mRNA, which served as a loading control, was detected using a specific probe and the steady-state levels of the mature 28S and 18S rRNA were visualized by methylene blue staining (MB). This figure was originally published in Memet et al., 2017.

### 3.2.4.4 NKRF regulates the function of XRN2 in the degradation of pre-rRNA processing by-products

The common defect in the turnover of fragments excised during pre-rRNA processing observed in cells lacking either NKRF or XRN2 indicated that depletion of the G-patch protein impaired the activity of the 5'-3' exonuclease at these sites. Interestingly, this function of XRN2 was previously found to be regulated by CARF, whose overexpression led to a relocalization of the exonuclease into the nucleoplasm and, consequently, to an accumulation of these pre-rRNA fragments (Sato et al., 2015). Similar to NKRF, CARF contains an XTBD and interacts with XRN2, raising the possibility that the nucleolar localization and function of the exonuclease in the turnover of pre-rRNA processing by-products is dynamically regulated through interactions with both NKRF and CARF (Miki et al., 2014; Richter et al., 2016).

Therefore, to assess if NKRF also influences the localization of XRN2, the levels of the G-patch protein were depleted in cells stably expressing GFP-tagged XRN2 from a tetracycline-inducible promoter and the subcellular distribution of the exonuclease was followed by fluorescence microscopy. Compared to cells treated with NT siRNA in which GFP-XRN2 was detected in the nucleoplasm and nucleoli, knockdown of NKRF led to a strong reduction of the nucleolar fraction of XRN2, indicating that the G-patch protein is required for its recruitment to these sites (Figure 3.27A). In contrast, depletion of DHX15 did not influence the localization of the exonuclease. To further demonstrate that in the

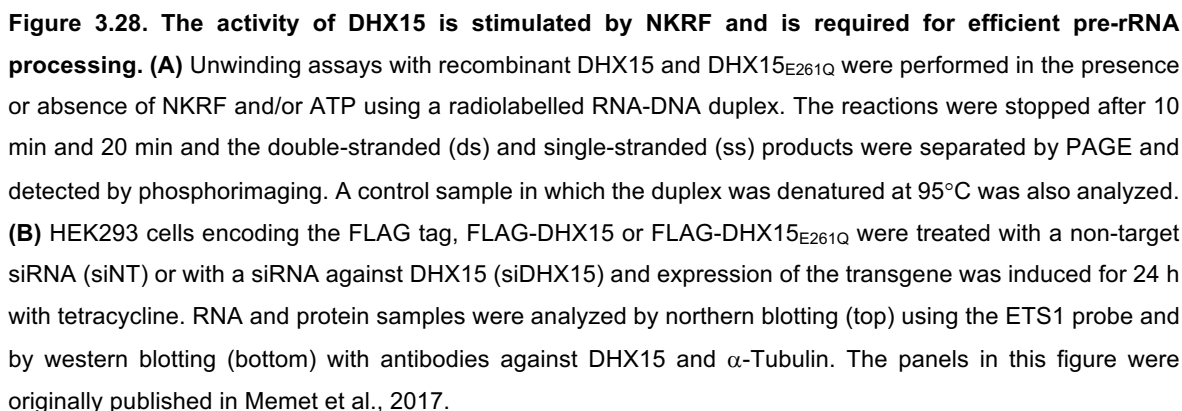
absence of NKRF, XRN2 is not able to associate with nucleolar pre-ribosomal particles, lysates prepared from cells treated with a control siRNA or with siRNAs against NKRF or DHX15 were separated on sucrose gradients and the distribution of the 5'-3' exonuclease in the different fractions was determined by western blotting. In line with our previous findings, XRN2 was found to co-migrate with pre-ribosomal complexes in cells treated with NT siRNA and depletion of DHX15 did not influence the levels of XRN2 in these fractions (Figure 3.27B). In contrast, knockdown of NKRF led to a strong reduction of the amount of XRN2 associated with pre-ribosomal particles, demonstrating that NKRF is indeed required for the nucleolar recruitment of XRN2 and consequently for its activity in the turnover of pre-rRNA substrates.



**Figure 3.27. NKRF is required for the nucleolar localization of XRN2.** (A) Stable cells encoding GFP-XRN2 were treated for 96 h with the siRNAs indicated on the left and the expression of the transgene was induced in the last 24 h with tetracycline. The GFP-XRN2 signal was visualized directly (green), while the localization of the nucleolar protein UTP14A was detected using a specific antibody (red). Nuclei were visualized by DAPI staining (blue). Overlays of the three channels are shown on the right (Merge) and the scale bar represents 10  $\mu$ m. (B) Lysates prepared from HeLa cells treated with a non-target siRNAs or with siRNAs against NKRF or DHX15 were subjected to sucrose gradient centrifugation and the distribution of XRN2 in the different fractions was analyzed by western blotting. The position of the co-migrating ribosomal/pre-ribosomal complexes is indicated below the panel. The panels in this figure were originally published in Memet et al., 2017.

### 3.2.4.5 NKRF stimulates the activity of DHX15 for efficient processing of the pre-rRNA transcript at a specific site

After establishing that the common pre-rRNA processing defects observed after depletion of NKRF and XRN2 likely arise due to a failure to recruit the exonuclease to its nucleolar substrates, the next focus was to elucidate the molecular basis of the impaired A' cleavage that was detected upon knockdown of DHX15, NKRF and XRN2 (Figure 3.25B). Our previous *in vitro* data showed that the G-patch domain of NKRF enhanced the activity of DHX15 both at the level of RNA binding and ATP hydrolysis (Figure 3.5; Figure 3.7). To gain more insight into the role of this G-patch protein as an RNA helicase cofactor, the effects of full-length NKRF on the ATPase and unwinding activities of DHX15 were also tested. After confirming that full-length recombinant MBP-NKRF-His<sub>10</sub> was capable of



83

and only the FLAG tag was expressed, compared to the control sample treated with NT siRNA (Figure 3.28B). Expression of FLAG-tagged DHX15 from a transgene rescued these defects, however, upon expression of DHX15<sub>E261Q</sub> to a similar level, the 47S and 30SL5' precursors still accumulated. Unwinding assays with recombinant DHX15<sub>E261Q</sub> confirmed, as expected, that this protein is catalytically inactive and, moreover, anti-FLAG IP experiments showed that the mutant helicase was able to associate with NKRF and XRN2 *in vivo* (Figure 3.28A and C). Together, these results indicate that the catalytic activity of DHX15, which is stimulated by the G-patch cofactor NKRF, is the main determinant for the function of the complex in pre-rRNA processing at site A', suggesting that structural remodeling of RNA or protein-RNA interactions by the helicase at this site might facilitate efficient cleavage.



## 4. DISCUSSION

### 4.1 G-patch protein-RNA helicase interactions in human cells

RNA helicases are ubiquitous enzymes that catalyze the NTP-dependent remodeling of RNA and RNA-protein complexes. Individually, most RNA helicases have the ability to perform more than one cellular activity and, altogether, they accomplish a wide range of functions and participate in every RNA-related process (Jankowsky, 2011). In recent years, a multitude of interacting proteins that modulate different aspects of the function of RNA helicases have been identified (Young and Karbstein, 2012; Sloan and Bohnsack, 2018). These helicase cofactors provide several layers of regulation that are necessary due to the characteristics of this family of enzymes. First, RNA helicases share a highly similar helicase core that lacks the ability to specifically recognize RNA substrates. Therefore, it has been suggested that in order to act on diverse substrates and carry out such broad activities, the specificity of RNA helicases for their RNA targets and/or functions is controlled through other mechanisms. Furthermore, several RNA helicases were shown to have a low activity *in vitro* and it was proposed that they are found in an auto-inhibited conformation in the cell, from which they would be released by the action of interacting proteins (Ozgun et al., 2015; Robert-Paganin et al., 2017). The default inactive state might be required to prevent unwanted cellular activities and enable the activation of RNA helicases at specific stages or subcellular locations. Similarly, the distinct activities of multifunctional RNA helicases could be regulated in a spatial and temporal manner by interactions with multiple cofactors. The requirement for helicase cofactors can be extended to other functions as well. For example, the MIF4G domain cofactor CWC22 chaperones the RNA helicase eIF4A-III to its target site by preventing unspecific RNA binding (Barbosa et al., 2012). At a subsequent stage, the interaction of eIF4A-III with other cofactors, namely MAGOH and Y14, inhibits the release of  $P_i$  after ATP hydrolysis and clamps the RNA helicase on the substrate RNA to promote the assembly of the exon-junction complex (Ballut et al., 2005; Nielsen et al., 2009). Thus, RNA helicases associate with protein cofactors that modulate their activity in diverse ways and increase the complexity of their function, which is necessary for the wide range of specific roles that these enzymes have in the cell.

The inventory of RNA helicase cofactors described so far includes several yeast and human proteins that contain a G-patch domain. The potential role of G-patch proteins as a dedicated family of RNA helicase cofactors has recently become apparent after the characterization of a growing number of such helicase-cofactor complexes (Robert-Paganin et al., 2015; Sloan and Bohnsack, 2018). The G-patch domain was found to be essential for the regulation of RNA helicases, raising the possibility that every protein containing this motif might act as a helicase cofactor. Indeed, all five G-patch proteins identified in yeast

were shown to modulate the activity of an RNA helicase. However, much less was known about the role of human G-patch proteins as cofactors of RNA helicases. Interestingly, the human genome encodes 22 proteins that contain a G-patch domain compared to five proteins in yeast. This might be linked to the larger number of RNA helicases expressed in human cells as well as to the higher complexity needed for their regulation, as most helicases in higher eukaryotes are probably multifunctional and coordinate several cellular events. At the start of this project, only four human G-patch proteins were known to associate with an RNA helicase and three of these interactions were conserved from yeast (Lin et al., 2009; Yoshimoto et al., 2009; Niu et al., 2012; Chen et al., 2014). Two more G-patch cofactor-RNA helicase complexes have been described recently (Chen et al., 2017; Inesta-Vaquera et al., 2018), but numerous additional human G-patch proteins remained to be characterized.

In this work, we first sought to establish if all human G-patch proteins interact with an RNA helicase and determine the effects that they have on the activity of their helicase partner. Immunoprecipitation experiments of FLAG-tagged G-patch proteins expressed from HEK293 stable cell lines revealed that GPKOW and GPATCH1 associate with DHX16 and DHX35 respectively, whereas the remaining 20 G-patch proteins interact with DHX15. Thus, these results confirmed that every human G-patch protein can be found in complex with an RNA helicase. In addition, *in vitro* experiments validated the role as genuine helicase cofactors for 20 out of the 22 G-patch proteins based on their influence on the ATP hydrolysis and/or RNA binding activities of their cognate RNA helicase. On the other hand, the effects induced by GPATCH1 and RBM6 remain to be further characterized experimentally and are discussed in more detail in the next section.

The intriguing finding that DHX15 might be regulated by approximately 20 cofactors raised the question of whether these interactions take place in every cell or if they are modulated by variable expression of G-patch proteins depending on the cell and tissue type or developmental stage. Despite the fact that our results are based on the exogenous expression of G-patch proteins, the reciprocal experiment with FLAG-tagged DHX15 confirmed the interaction with 13 out of the 20 G-patch cofactors. This suggests that the majority of the identified G-patch protein-DHX15 complexes are present in normal conditions in HEK293 cells and are not a result of overexpression. The failure to identify the remaining G-patch proteins could be due to their low expression level or their weak association with the helicase, which might be below the detection sensitivity of the method. Interestingly, a complementary study using HA-tagged DHX15 transiently expressed in HeLa cells identified 11 G-patch proteins that were co-precipitated with the helicase, six of which overlapped with our data (Inesta-Vaquera et al., 2018). Together, these two studies

validated the interactions of DHX15 with approximately 18 G-patch proteins in a reciprocal manner and imply that the helicase might be differentially associated with its extensive network of cofactors depending on the cell type or conditions. The dynamic nature of the G-patch protein-DHX15 interactions is further supported by the tissue-dependent expression levels and the upregulation/downregulation in cancer cells that have been reported for multiple G-patch proteins (Laplante et al., 2000; Jianfeng et al., 2003; Sampath et al., 2003; Sampson and Hewitt, 2003; Lin et al., 2009; Xu et al., 2014; Wang et al., 2015). Furthermore, post-translational modifications have been detected for some G-patch proteins, but it remains to be determined if these can modulate the helicase-cofactor interaction (Aksaas et al., 2011; Liu et al., 2013; Rother et al., 2016).

The OB-fold domain of RNA helicases was defined as the binding site for G-patch proteins based on protein-protein crosslinking experiments and interaction studies with truncated helicases (Silverman et al., 2004; Walbott et al., 2010; Christian et al., 2014; Inesta-Vaquera et al., 2018). Our results showed that deletion of the OB-fold region of DHX15 and DHX35 strongly diminished their interactions with NKRF and GPATCH1, indicating a similar binding mode. These findings also imply that the interactions established by DHX15 with its G-patch cofactors are mutually exclusive and that the helicase associates with each cofactor in distinct subcomplexes that probably carry out separate functions. This model is supported by the finding that in yeast four G-patch cofactors can compete for binding to the multifunctional RNA helicase Prp43 *in vitro*. In addition, overexpression of certain Prp43 cofactors *in vivo* led to a relocalization of the helicase and to its withdrawal from specific functions (Heininger et al., 2016). Thus, changes in the expression level and the dynamic interplay of cofactors might serve as a mechanism to regulate multifunctional RNA helicases. Since DHX15 is the human homologue of Prp43, the ability of these related helicases to interact with a multitude of G-patch proteins seems to be conserved, but the larger network of cofactors for DHX15 suggests that it performs more complex or additional functions compared to its yeast counterpart. The association of GPKOW with DHX16 is also conserved from yeast, where their homologues Spp2 and Prp2 function together in splicing (Silverman et al., 2004; Warkocki et al., 2015). On the other hand, DHX35 does not have a defined yeast homologue and its interaction with GPATCH1 is mammalian-specific, thus making this novel helicase-cofactor complex interesting to explore in future studies.

Although the OB-fold of RNA helicases was identified as the binding platform for G-patch proteins, this domain is present in a large number of other proteins, where it is mostly involved in binding nucleic acids, but can also mediate interactions with proteins or other molecules (Arcus, 2002; Theobald et al., 2003). In RNA helicases, OB-folds are specific for the DEAH/RHA family and are always found in conjunction with a winged helix and a

ratchet-like domain. Together, these domains constitute the C-terminal auxiliary region, which contributes to RNA binding and is essential for the function and regulation of the helicase (He et al., 2010; Walbott et al., 2010). There are approximately 15 DEAH/RHA helicases in human cells that contain an OB-fold domain, but our data indicate that only three of them interact with G-patch proteins. It is currently unknown which features of DHX15, DHX16 and DHX35 enable these helicases to associate specifically with G-patch cofactors. However, homology modelling studies of the C-terminal domains of DHX8, DHX15, DHX16 and DHX38 revealed the existence of conserved regions, but also of divergent surface patches that were suggested to contribute to interactions with different proteins (Kudlinzki et al., 2012). In the absence of structural information of a G-patch cofactor-RNA helicase complex, the exact determinants of specificity for G-patch protein binding remain to be identified. Interestingly, for some RNA helicases that are not regulated by G-patch proteins, other functions have been described for the OB-fold domain apart from its common role in substrate binding and ensuring the overall architecture of helicases. For example, the OB-fold domain of DHX36 is implicated in the specific recognition of G-quadruplexes together with the N-terminal region of the helicase, while the OB-fold of DHX9 was suggested to modulate the splicing of proviral transcripts (Xing et al., 2014; Chen et al., 2018).

Despite the fact that only three human DEAH/RHA helicases were found to interact with G-patch cofactors, it is likely that the remaining helicases are also regulated by accessory proteins. Consistent with this, the function of DHX37 and DHX9 was shown to be modulated by the cofactors UTP14A and NUP98 respectively (Capitanio et al., 2017; Choudhury et al., *submitted*). The modularity achieved by having protein cofactors that influence the activity of RNA helicases *in trans* would enable the independent regulation of both components and might be particularly well-suited for complex systems such as those in higher eukaryotes. In this way, post-translational modifications, changes in expression levels or interactions with competing proteins for both the helicase and cofactor can modulate the function of the complex, thus providing an additional layer of regulation. This might also be advantageous for the regulation of multifunctional RNA helicases, whose distinct functions could be accomplished by having dedicated cofactors for each activity. In line with this, the four G-patch cofactors of Prp43 were suggested to target the helicase to specific functions (Tanaka et al., 2007; Lebaron et al., 2009; Chen et al., 2014; Fourmann et al., 2016; Heininger et al., 2016). As a comparison, the prokaryotic RNA helicase DbpA and its other bacterial orthologues are recruited to their RNA substrate through interactions established by the C-terminal RRM domain (Diges and Uhlenbeck, 2001; Kossen et al., 2002). In this case, the helicase activity and the specificity for the target RNA are encoded in the same

polypeptide, which might reflect the fact that these enzymes have a single cellular function. However, helicase cofactors have also been described in bacteria and in viruses, indicating that the regulation of RNA helicases by accessory proteins is a widespread mechanism (Silverman et al., 2003; Redder et al., 2015). This might also be linked to the high level of conservation between RNA helicases across all domains of life (Fairman-Williams et al., 2010). Whether or not every RNA helicase is regulated by a protein cofactor is currently not established. Nevertheless, other means of regulation have been described and given their characteristic features and their essential cellular functions, it is highly likely that the activity of RNA helicases is modulated at multiple levels.

## 4.2 Mechanism of RNA helicase regulation by G-patch proteins

Although the mechanistic details of the G-patch protein-mediated regulation of RNA helicases have not been elucidated, several common characteristics can be derived from the existing studies. G-patch proteins were shown so far to have an exclusively stimulatory role and to enhance the RNA binding affinity, the ATPase and/or unwinding activities of their interacting helicase. Furthermore, the G-patch domain was found to be essential for the regulation since no stimulation of activity was observed in its absence (Lebaron et al., 2009). This domain was also suggested to be sufficient for modulating the helicase activity, although most studies used slightly longer fragments than the G-patch motif only (Christian et al., 2014; Heininger et al., 2016).

Here, ATPase assays of DHX15 and DHX16 in the presence of the G-patch domains confirmed that this region is sufficient in most cases to stimulate the activity of the helicase partner. Despite the weak effects induced by some cofactors, the increase in ATPase activity was observed only for the specific helicase, indicating that the G-patch domain alone is also able to distinguish the cognate RNA helicase. On the other hand, the G-patch domain of GPKOW did not enhance the activity of DHX16, despite a mild stimulation being observed in the presence of the full-length cofactor. This is consistent with the finding that, in some cases, other regions of G-patch proteins contribute to binding to the helicase, suggesting that for GPKOW sequences outside of the G-patch domain are also involved in the interaction with DHX16 (Lebaron et al., 2009). The G-patch domain of RBM6 did not stimulate the ATPase activity of DHX15 either, but instead caused a mild reduction. In the absence of experiments with the full-length protein, it is not clear what effects this cofactor induces. Similarly, no *in vitro* data is available for the GPATCH1-DHX35 interaction.

Therefore, even though our results showed that 20 out of the 22 human G-patch cofactors enhanced the ATPase activity of their interacting helicase, it is not yet known if G-patch

proteins function exclusively as activators. Considering that the MIF4G domain cofactors of DEAD-box helicases were shown to have both stimulatory and inhibitory effects, it is possible that G-patch proteins could influence the activity of their helicase partner in other ways as well (Ozgun et al., 2015; Sloan and Bohnsack, 2018). The mild decrease of the DHX15 ATPase activity induced by the G-patch domain of RBM6 suggests that this cofactor might act as an inhibitor. In the cellular environment, this putative inhibitory role of RBM6 could be manifested in a similar way to the function of the MIF4G cofactor CWC22, which holds the RNA helicase eIF4A-III in an inactive conformation to prevent unspecific RNA binding (Barbosa et al., 2012; Buchwald et al., 2013). Alternatively, RBM6 could also act as a placeholder that blocks the interaction of DHX15 with other G-patch cofactors, sequestering the helicase in a non-productive state.

The results described above were based on ATP hydrolysis assays performed in the presence of RNA. Since RNA helicases are RNA-dependent NTPases, the higher ATPase activity observed upon addition of the G-patch cofactors could be due to an increased binding of the helicase to the RNA substrate. This is supported by anisotropy measurements that showed most G-patch domains can also enhance the RNA affinity of DHX15. In addition, other G-patch cofactors were previously found to stimulate both RNA binding and ATP hydrolysis by their helicase (Christian et al., 2014; Heininger et al., 2016). However, this mechanism cannot exclusively account for the stimulation of ATP hydrolysis because not all G-patch domains increased the RNA binding affinity of DHX15 despite their effects on its ATPase activity. Consistent with this, Spp2 was found to enhance the ATPase activity of Prp2, but not its affinity for the RNA substrate (Warkocki et al., 2015). It is also possible that the effects induced by these cofactors on the helicase RNA affinity are weak and below the detection limit.

Several studies have shown that G-patch proteins can also stimulate the ATPase activity of the helicase in the absence of the RNA substrate (Lebaron et al., 2009; Chen et al., 2014; Christian et al., 2014). In the cellular context, this mechanism might not be significant since RNA helicases need to couple ATP hydrolysis to mechanical work on the RNA substrate. Based on the available information, a model can be proposed in which binding of the G-patch domain to the OB-fold domain of the RNA helicase leads to conformational changes both in the C-terminal region and in the helicase core. This enhances binding of the RNA substrate in some cases and also brings the helicase in a conformation that is more favorable for ATP hydrolysis. In turn, the bound RNA triggers further rearrangements that stimulate ATP hydrolysis, which then leads to translocation along the substrate (He et al., 2017; Tauchert et al., 2017). Thus, the cooperative action of the RNA and the G-patch protein is required for the remodeling function. Indeed, Spp2 was shown to be essential for

coupling the ATPase and remodeling activities of Prp2 (Warkocki et al., 2015). A similar role was suggested for Pxr1, which was proposed to act by disrupting stacking interactions between the nucleotide base and specific residues located in the two helicase domains (Robert-Paganin et al., 2017). Other G-patch cofactors were shown to stimulate the ATPase and unwinding activities of their interacting helicase, including NKRF in the present study, providing further support that G-patch proteins are essential for both functions (Lebaron et al., 2009; Chen et al., 2017; Inesta-Vaquera et al., 2018).

It is currently not known if the G-patch domain can contact the RNA directly or if it enhances RNA binding by the helicase indirectly by promoting conformational changes. Our anisotropy studies showed that the DHX15-interacting G-patch domains don't bind RNA on their own. Other G-patch cofactors were found to lack the ability to bind RNA, while Spp382 was suggested to contact RNA only in complex with the helicase (Christian et al., 2014; Warkocki et al., 2015; Heininger et al., 2016). On the other hand, some virus-encoded G-patch domains were shown to associate with RNA with different substrate specificities (Svec et al., 2004). Our results also revealed that the G-patch domains of GPKOW and GPATCH1 bind RNA with  $K_d$  values of 8.7  $\mu$ M and 26.2  $\mu$ M respectively, providing one of the first quantitative reports of binding affinities for isolated G-patch domains. Strikingly, there is a clear separation between the DHX15-interacting G-patch domains, which don't bind RNA, and those belonging to GPKOW and GPATCH1, which are cofactors of DHX16 and DHX35 and are able to associate with RNA. This might indicate distinct mechanisms for the regulation of these three DEAH/RHA helicases by G-patch cofactors, with the contribution of the G-patch domain to RNA binding depending on the identity of the helicase-cofactor complex. The G-patch domain is often found in combination with canonical RNA-binding domains such as the RRM or the R3H domains (Aravind and Koonin, 1999; Figure 1.4A). Thus, some G-patch proteins can bind the RNA substrate through these additional domains. Similar to the C-terminal tail of certain DEAD-box helicases that anchors the helicase core to the substrate, it is possible that these RNA-binding domains help to strengthen the interaction of the G-patch cofactor-helicase complex with the target RNA by binding to adjacent regions.

The results of the ATPase and anisotropy assays also uncovered a putative hierarchy of the G-patch cofactors of DHX15 based on the strength of their effect, with GPATCH2 and CMTR1 being strong cofactors and proteins such as GPATCH3, SUGP2 or ZGPAT having weak effects. However, only the G-patch domains were used in these experiments and it is possible that the full-length cofactors might show different effects. Furthermore, in the cellular environment, other factors, such as the expression levels of G-patch proteins, their

subcellular localization or the presence of competing proteins are likely to influence the interaction of each helicase-cofactor complex.

It is currently unknown which residues of the G-patch domain are directly involved in the binding and regulation of RNA helicases, and consequently, the features that would make a cofactor stronger or weaker are elusive. The consensus sequence that was originally defined for the G-patch domain includes five glycine residues at conserved positions, an aromatic amino acid following the first glycine residue and a few conserved hydrophobic amino acids. However, these residues are not found in all G-patch proteins, for example, yeast Spp2 containing only two of the five conserved glycines (Aravind and Koonin, 1999; Robert-Paganin et al., 2015). Several mutational studies have attempted to define the amino acids of the G-patch domain that are essential for RNA helicase binding. For example, in the case of RBM5, mutation of conserved glycine residues in different combinations reduced the interaction with DHX15 (Niu et al., 2012). In this work, a similar result was obtained for NKRF, for which substituting six conserved glycine residues with alanine abolished binding to DHX15. Other studies have shown that several conserved leucine residues in the G-patch domain of specific proteins are also important for the interaction with RNA helicases (Tanaka et al., 2007; Inesta-Vaquera et al., 2018). Intriguingly, the first four glycine residues of the consensus sequence are highly conserved in human G-patch proteins, but the fifth glycine is less conserved and at this position all the DHX15 cofactors have a glycine residue, whereas the non-DHX15 interactors contain an arginine (for GPATCH1) or a glutamine (for GPKOW) (Figure 1.4B). Further experimental work is required to determine if the amino acid at this position is important for discriminating between different RNA helicases. Small differences at the level of the primary sequence have already been suggested to enable MIF4G domains to distinguish their cognate DEAD-box helicases due to a few steric clashes or unfavorable electrostatic interactions and a similar mechanism is possible for G-patch proteins (Buchwald et al., 2013; Ozgur et al., 2015).

Interestingly, MIF4G cofactors that influence the activity of more than one helicase have been identified. For example, Gle1 activates the RNA helicase Dbp5 in the mRNA export pathway and also inhibits the function of Ded1 in translation, but it remains to be determined if the MIF4G domain is the regulatory module in both cases (Bolger and Wentz, 2011; Montpetit et al., 2011). Although our data did not find any evidence for G-patch proteins acting as cofactors of multiple RNA helicases, this possibility cannot be excluded and requires further investigation.



### 4.3 Regulation of alternative splicing by DHX15 and G-patch proteins

Alternative splicing is a key step of gene expression regulation that involves the differential usage of splice sites within a primary transcript derived from a single gene locus to produce multiple transcript variants. Almost every human gene undergoes alternative splicing, which generates diverse transcripts and protein isoforms that are essential determinants of cell identity and function (Lee and Rio, 2015). Alternative splicing relies on regulatory *cis* elements that are located either in introns or in exons and can be functionally classified into splicing enhancers and silencers. These regulatory elements are bound by *trans*-acting factors, including SR proteins and hnRNP proteins, that promote or inhibit splicing through various mechanisms, such as the recruitment of spliceosome complexes or steric hindrance of splice sites. Both the *cis* elements and the *trans*-acting factors can exhibit dual roles that are dependent on the sequence context. In addition, other features modulate the choice of splice sites, including pre-mRNA secondary structures or multiple proteins binding to the same regulatory elements (Wang and Burge, 2008; Chen and Manley, 2009; Fu and Ares, 2014). Overall, the combination of these positive and negative effects determines splice site selection, but the underlying mechanisms are not yet elucidated.

The majority of alternative splicing events are a result of exon skipping or inclusion, intron retention, the usage of alternative 5' and 3' splice sites or the choice between mutually exclusive exons. These patterns of alternative splicing occur independently or in different combinations, thus producing a multitude of diverse transcripts (Wang et al., 2015b). Alternative splicing can be investigated in a global manner using methods such as RNA-seq or, the more recently developed, long-read sequencing. Most commonly, alternative splicing analysis based on RNA-seq data is performed using libraries of known splicing events derived from annotated genomes. However, methods that enable the identification of novel splice junctions have also been developed to overcome the limitations of incompletely annotated or unannotated reference genomes (Conesa et al., 2016).

In this study, changes in alternative splicing patterns at annotated splice sites were examined using RNA-seq in cells lacking either DHX15 or G-patch proteins. The results revealed differentially expressed splicing events upon knockdown of each of the 18 factors tested, with some proteins regulating only a small number of events (~100) and others, such as SON and ZGPAT, influencing thousands of sites. The changes in alternative splicing were not linked to altered gene expression levels, indicating that these proteins regulate splice site selection without affecting the transcription or stability of the target mRNAs. Several G-patch proteins have been previously implicated in splicing, including CHERP, RBM5, RBM6, RBM10, RBM17 and SON, which were suggested to regulate the alternative splicing of specific genes (Ahn et al., 2011; Bechara et al., 2013; Lu et al., 2013; De Maio

et al., 2018). A function in alternative splicing was also proposed for ZGPAT based on experiments with a generic splicing reporter (Chen et al., 2017). In addition, TFIP11 was found to participate with DHX15 in the disassembly of intron lariat spliceosomes, but a role in alternative splicing has not been addressed so far for these factors (Yoshimoto et al., 2009). Some G-patch proteins, including CHERP and RBM17, have been reproducibly detected in purified spliceosomal complexes, indicating that they are core components of the spliceosome. Other G-patch proteins, such as RBM5 and SUGP2, were suggested to be loosely associated with the spliceosome and to act only at specific stages (Bessonov et al., 2008; Wahl et al., 2009; Agafonov et al., 2011). Our results indicate that a large number of G-patch proteins function as regulators of alternative splicing, implying that the failure to detect factors such as AGGF1, GPANK1 and others in spliceosomal preparations could similarly reflect a transient association with the splicing machinery. Proteins that interact weakly with the spliceosome might be well-suited for alternative splicing regulation, which requires rapid responses to changes in cellular conditions (Agafonov et al., 2011). Interestingly, it was shown that the core spliceosome components also modulate the alternative splicing of subsets of genes and do not influence constitutive splicing, which is consistent with our findings (Saltzman et al., 2011; Papasaikas et al., 2015). Apart from the core and non-core components of the spliceosome machinery, splice site selection can also be regulated by chromatin-related factors and RNA-processing proteins, suggesting that diverse factors can participate in alternative splicing (Papasaikas et al., 2015).

One of the best studied G-patch proteins in splicing is SON, whose depletion was shown to lead predominantly to intron retention or exon skipping in specific transcripts, which is in accordance with our results (Ahn et al., 2011; Lu et al., 2013). Comparison of the SON-regulated genes detected in this study with the changes in alternative splicing reported in human embryonic stem cells after SON knockdown revealed approximately 40% overlap between the datasets, thus validating our analysis and indicating that the G-patch protein has general as well as cell type-specific substrates (Lu et al., 2013). SON was found to bind directly to some of its target pre-mRNAs and was proposed to regulate splicing by recruiting SR proteins and other factors through its N-terminal repetitive domains (Hickey et al., 2014; Lu et al., 2014). Interestingly, the G-patch domain of SON is required for the splicing function, implying that the G-patch protein might also recruit DHX15 to these target sites (Ahn et al., 2011). Consistent with a common function of SON and DHX15 in alternative splicing, our analysis identified 256 splicing events that were regulated by both proteins. Common splicing events between DHX15 and its other G-patch cofactors were also found, raising the possibility that DHX15 acts as a master regulator of alternative splicing through its network of cofactors.

In yeast, eight conserved RNA helicases participate in splicing and perform essential structural remodeling events during the assembly and catalytic activation of the spliceosome. Some of these helicases also ensure the accuracy of the process by proofreading the pre-mRNA-snRNA interactions (Cordin and Beggs, 2013). Apart from these conserved RNA helicases, additional helicases have been linked to splicing in human cells and it was suggested that they regulate alternative splicing or mediate other ATP-dependent remodeling steps required for spliceosome function (De et al., 2015; Lee et al., 2018). The only activity described for DHX15 in splicing until now was the disassembly of intron lariat spliceosomes (Yoshimoto et al., 2009). This function is conserved from yeast, where its homologue Prp43 was shown in addition to discard aberrant spliceosomes (Tanaka et al., 2007; Koodathingal et al., 2010). Interestingly, the role of Prp43 in the disassembly of late-stage and aberrant spliceosomes is tightly controlled by the G-patch protein Spp382 and by Ntr2, and in the absence of these factors the helicase can act unspecifically on properly assembled spliceosomal complexes at different stages of maturation (Fourmann et al., 2016). In human cells, this putative function of the helicase as a general disassembly factor could be regulated through interactions with its G-patch cofactors and enable DHX15 to participate at multiple steps of splicing. A recent study has shown that proofreading of splice sites by the yeast helicases Prp16 and Prp22 at the catalytic stages of splicing disengages weak substrates and allows the spliceosome to search for alternative splice sites (Semlow et al., 2016). These results provide a framework for alternative splicing regulation by DEAH/RHA helicases, suggesting that the proofreading function of yeast RNA helicases could have evolved in higher eukaryotes as a mechanism to control the choice of splice sites. Thus, DHX15 might preferentially disassemble spliceosome complexes at particular locations or stages and, in this way, lead to splicing at alternative sites. On the other hand, it is also possible that DHX15 regulates alternative splicing by resolving secondary structures in specific pre-mRNAs, as has been suggested for DDX5 (Lee et al., 2018).

The findings that DHX15 and the analyzed G-patch proteins regulate alternative splicing and that the G-patch cofactors stimulate the activity of DHX15 strongly suggest that these helicase-cofactor complexes function together in this pathway. However, in addition to splicing events that were common between DHX15 and its G-patch cofactors, unique targets for each sample were also identified. Therefore, further experiments are required to understand if the role of G-patch proteins in alternative splicing is mediated exclusively through their interacting helicase. A possible explanation for the low overlap between the splicing changes induced by DHX15 and its cofactors is functional redundancy or synergism between G-patch proteins. In line with this, the RNA-seq analysis also revealed overlapping

splicing substrates between G-patch proteins. Furthermore, a recent study found that CHERP and RBM17 have common alternative splicing targets and proposed that these G-patch proteins function together in this pathway (De Maio et al., 2018). Although our data did not confirm these observations, a high overlap between GPATCH8 and GPATCH11 was detected instead (~80%), indicating that other G-patch proteins could also act in a coordinated manner. The regulation of the same splice sites by multiple G-patch proteins suggests that the action of these cofactors might be redundant and serve as a backup mechanism. This is supported by the fact that RNA helicases probably interact in a mutually exclusive manner with their G-patch cofactors (Heininger et al., 2016). Alternatively, G-patch proteins could regulate the activity of the helicase synergistically through unknown mechanisms or could act independently of the helicase. Nevertheless, it is likely that the function of G-patch proteins is coordinated, as evidenced by the fact that in some cases they can influence each other's expression levels (Loiselle et al., 2017; Sun et al., 2017; De Maio et al., 2018).

RBM5, RBM6 and RBM10 were also found to crosslink to some of their target pre-mRNAs, similar to SON, suggesting that they have a direct role in alternative splicing regulation (Bechara et al., 2013). Our results revealed that SUGP2 binds directly to a large number of pre-mRNAs as well. Some of these targets showed changes in alternative splicing upon knockdown of the G-patch protein, supporting a direct function for SUGP2 in modulating splice site selection. Additional experiments are needed to determine the binding sites of SUGP2 on its target substrates, which would help uncover the mechanisms of alternative splicing regulation by this G-patch protein. At the same time, the other G-patch proteins identified in this study as novel regulators of alternative splicing require further investigation to confirm their function and assess if they have a direct or an indirect role in this process. It is possible that the G-patch proteins that interact directly with their RNA targets bind adjacent to the RNA helicase docking site and strengthen its interaction with the substrate in addition to stimulating its activity. This is supported by the fact that the yeast helicase Prp43 can crosslink to pre-mRNAs independent of its cofactor Spp382 (Fourmann et al., 2016). In the case of G-patch proteins that don't bind RNA on their own but still have a direct role in splice site selection, interactions with other proteins could mediate their recruitment to the target substrates as has been suggested, for example, for Spp2 (Warkocki et al., 2015).

On the other hand, the regulation of alternative splicing by G-patch proteins could take place through indirect mechanisms that are independent of their participation in the splicing reaction. In support of this, our data showed that ZGPAT does not crosslink to mRNAs, despite the numerous changes in alternative splicing observed after its knockdown, and

instead binds specific scaRNAs. These small non-coding RNAs associate with proteins into scaRNPs and generally guide the 2'-O-methylation and pseudouridylation of snRNAs by base-pairing to the target site. These two modification types alter the properties of the RNA, for example, by increasing its stability or influencing base-pairing and stacking interactions. Most 2'-O-methylations and pseudouridylations in snRNAs are found in functionally important regions that are involved in essential RNA-RNA and protein-RNA interactions (Karijolic and Yu, 2010; Bohnsack and Sloan, 2018). Consistent with this, some modifications were already shown to be required for snRNA function, although most modifications still await characterization. For example, several modifications at the 5' end of U2 snRNA were found to be essential for the formation of early spliceosomal complexes and, consequently, for efficient splicing (Donmez et al., 2004).

The mechanisms involved in the scaRNA-mediated modification of snRNAs are largely unknown, but considering that some RNA helicases were shown to modulate snoRNA-rRNA interactions, it is possible that these enzymes could have a similar role in scaRNA function. Therefore, the finding that ZGPAT crosslinks to scaRNAs, together with its role as a cofactor of DHX15, imply that this helicase-cofactor complex might be required for snRNA modification guided by scaRNAs. Analysis of the methylation status at the G11 and G25 residues in U2 snRNA revealed that knockdown of DHX15 decreased the modification level at these sites, whereas depletion of ZGPAT did not show any effects. Despite the fact that DHX15 does not interact directly with scaRNAs, it seems likely that its role in snRNA modification is mediated through scaRNAs and that ZGPAT establishes the functional connection between these components. This model is supported by the fact that knockdown of DHX15 does not affect the expression level or alternative splicing of the protein components of the scaRNP methylation machinery. Therefore, based on these results, it is possible that DHX15 and ZGPAT function together in the scaRNA-guided modification of snRNAs. Considering that the G-patch protein crosslinks to specific scaRNAs, DHX15 and ZGPAT probably influence snRNA modification only at certain sites, but this aspect needs to be further assessed.

Although it was recently shown that snRNAs are generally fully methylated at the known sites of modification across different tissues (Krogh et al., 2017), several findings have highlighted the potential of variations in the snRNA modification status to fine-tune the function of the spliceosome. In yeast, apart from the constitutive modifications, a few pseudouridine residues in U2 and U6 snRNAs were found to be induced in stress conditions or in different growth phases, and one of these novel pseudouridines decreased splicing efficiency (Wu et al., 2011; Basak and Query, 2014; Karijolic et al., 2015). In addition, substoichiometric snRNA 2'-O-methylation levels were detected at several sites during T

cell activation and the U4 snRNA C8 residue was not methylated in a specific cancer cell line as opposed to its status in normal conditions (Krogh et al., 2017). These examples of heterogeneity in snRNA modification suggest the possible existence of specialized spliceosomes, which might serve as a means to regulate alternative splicing similar to how specialized ribosomes were proposed to preferentially translate specific mRNA subsets (Krogh et al., 2017; Shi et al., 2017). Although direct evidence that variable levels of snRNA modifications influence alternative splicing is currently lacking, this hypothesis was also suggested based on the observation that decreased expression of specific scaRNAs correlates with changes in alternative splicing that contribute to cardiac disease (Patil et al., 2015; Nagasawa et al., 2018). Therefore, it is possible that DHX15 and ZGPAT modulate alternative splicing through their putative function in snRNA modification. Considering that 78 alternative splicing events were found to be regulated by both DHX15 and ZGPAT in our RNA-seq analysis, these target mRNAs could represent a starting point for confirming this model. Additional experiments will help decipher the exact function of DHX15 and ZGPAT in snRNA modification and alternative splicing regulation.

#### **4.4 The function of DHX15 and NKRF in ribosome biogenesis**

Eukaryotic ribosome biogenesis is a highly complex process involving hundreds of assembly factors that are required for the sequential maturation of rRNA precursors and their association with ribosomal proteins. In human cells, three of the four mature rRNAs are co-transcribed into a single precursor, the 47S pre-rRNA transcript, in which the sequences of the 18S, 5.8S and 28S rRNAs are separated and flanked by spacer regions (5'ETS, ITS1, ITS2 and 3'ETS). Processing of this initial transcript to release the mature rRNAs takes place through a series of endonucleolytic cleavage events at specific sites in the spacer regions that are coupled in most cases with exonucleolytic digestion. The pathway of ribosome production is generally conserved across eukaryotes, but several characteristics, such as the presence of additional cleavage sites and the larger number of assembly factors, demonstrate the increased complexity of this process in human cells compared to yeast (Henras et al., 2015; Aubert et al., 2018). The function of most ribosome assembly factors remains to be determined.

In this study, novel interactions between the RNA helicase DHX15, the G-patch protein NKRF and the 5'-3' exonuclease XRN2 were identified and the functions of these proteins in ribosome biogenesis were characterized. DHX15, NKRF and XRN2 were found to co-migrate with pre-ribosomal particles in sucrose gradients and to associate into a nucleolar subcomplex that is required for efficient pre-rRNA processing at site A' in 5'ETS. The A' cleavage event is specific for metazoans and, although it generally takes place early

in the pre-rRNA maturation pathway, it was found to not be a pre-requisite for downstream processing and can also occur at later stages or be skipped altogether (Sloan et al., 2014). This is consistent with our findings that depletion of these three factors does not affect the production of the mature ribosomal subunits despite the impaired A' cleavage. Although the role of this additional processing event in metazoans is not known, the reduced cleavage at the A' site upon knockdown of DHX15, NKRF and XRN2 leads to an accumulation of the 47S and 30SL5' precursors. Interestingly, in MCF7 cells, the levels of 30SL5' pre-rRNA are inherently higher than in other cell lines (Sloan et al., 2014). Furthermore, alternative processing pathways that generate different intermediates exist at several stages of pre-rRNA maturation and variations in the kinetics of these co-existing pathways that arise depending on the cell type or physiological conditions lead to changes in the ratios of the precursors. These different patterns of pre-rRNA processing were suggested to modulate the function of the ribosome, for example, by inducing distinct rRNA modification profiles depending on the precursors generated (Lafontaine, 2015; Aubert et al., 2018). Thus, even though the knockdown of DHX15, NKRF and XRN2 does not affect the production of the mature ribosomal subunits, it is possible that the increased levels of the 47S and 30SL5' intermediates leads to subtle differences in rRNA modification that contribute to ribosome heterogeneity. This hypothesis is supported by the fact that 2'-O-methylation and pseudouridylation, which are the most abundant types of rRNA modification, occur at early stages of ribosome biogenesis, similar to A' cleavage. Furthermore, sites with substoichiometric methylation levels as well as differences in the rRNA modification pattern between cell lines were recently discovered (Krogh et al., 2016).

The nucleolar interactions between DHX15, NKRF and XRN2 and the common A' processing defect observed upon their knockdown strongly suggests that these proteins function together at this site. Consistent with this, our results showed that NKRF acts as a cofactor of DHX15 and stimulates its RNA binding affinity, ATPase and unwinding activities, and this catalytic activity of the helicase is necessary for proper A' cleavage. On the other hand, the presence of XRN2 in this complex might serve as a quality control mechanism to enable the degradation of aberrant pre-rRNAs, similar to the function described for its mouse homologue (Wang and Pestov, 2011). The finding that the catalytic activity of DHX15 is required for efficient A' cleavage suggests that the helicase might perform a structural remodeling event that facilitates processing, for example, by enabling the access of the currently unidentified endonuclease to its target site. This function would resemble the role proposed for its yeast homologue Prp43 in promoting the cleavage of 20S pre-rRNA by the endonuclease Nob1 (Pertschy et al., 2009). Alternatively, the action of DHX15 could lead

to the release or association of other proteins, such as factors that were previously implicated in A' cleavage (Sloan et al., 2014).

The involvement of DHX15 in A' cleavage represents the first function reported for this human helicase in ribosome biogenesis. As this processing step is specific for metazoans, this activity is not performed by its yeast homologue Prp43, for which other roles in this pathway have been described instead. Prp43 was suggested to participate in the biogenesis of both ribosomal subunits by promoting the final step of 18S rRNA maturation and mediating the association or release of snoRNAs during the assembly of the large ribosomal subunit (Bohnsack et al., 2009; Pertschy et al., 2009). These distinct activities of Prp43 in ribosome biogenesis are likely regulated by the G-patch cofactors Sqs1 and Pxr1 (Pertschy et al., 2009; Robert-Paganin et al., 2017). Thus, apart from its role in ensuring efficient A' cleavage together with NKRF and XRN2, it is possible that DHX15 performs additional functions in this pathway with other G-patch cofactors, similar to Prp43. Consistent with this, our results showed that the G-patch proteins GPATCH2, GPATCH4 and PINX1 stimulate the activity of DHX15 and are localized in nucleoli, implying that they might be involved in ribosome biogenesis. Since PINX1 was found to substitute the function of its yeast homologue Pxr1 in ribosome maturation, it likely performs a similar role in human cells (Chen et al., 2014; Robert-Paganin et al., 2017).

Apart from the common function of DHX15, NKRF and XRN2 in A' cleavage, our data also revealed that knockdown of NKRF and XRN2 leads to the accumulation of several pre-rRNA fragments that are excised during processing and are normally targeted for degradation. The involvement of XRN2 in the turnover of these processing by-products has been previously described and our results further show that the similar defects induced by depletion of NKRF arise due to its role in recruiting the exonuclease to the nucleolus and to its pre-rRNA substrates (Wang and Pestov, 2011; Schillewaert et al., 2012; Sloan et al., 2013; Sloan et al., 2014). In addition, an increase in the levels of the 36S and 12S precursors was observed upon knockdown of NKRF and XRN2, which is probably a feedback effect caused by the failure to recycle the ribosome assembly factors bound to the excised pre-rRNA fragments. The common function of NKRF and XRN2 in pre-rRNA processing and turnover was also reported in a separate study, which found in addition that NKRF is upregulated during the heat shock response and restores nucleolar homeostasis by recruiting XRN2 to this subcellular location (Coccia et al., 2017).

Interestingly, it was shown that the function and localization of XRN2 are also modulated by the nucleoplasmic protein CARF, whose overexpression leads to an accumulation of XRN2 in the nucleoplasm and to similar pre-rRNA processing defects as those caused by depletion of NKRF or XRN2 (Sato et al., 2015). Similar to NKRF, CARF contains a



conserved domain that was suggested to mediate the interaction with XRN2 (XTBD; Richter et al., 2016). Although our results indicate that NKRF binds XRN2 in a different mode than CARF, taken together, these findings imply that the distribution of XRN2 between its nucleolar and nucleoplasmic functions is controlled by the interplay between NKRF and CARF. In contrast to its yeast homologue Rat1, whose exonuclease activity is stimulated by the cofactor Rai1, the regulation of XRN2 by XTBD-containing proteins does not seem to directly influence its enzymatic activity, which is probably due to the different binding mode compared to the yeast proteins (Miki et al., 2014; Sato et al., 2015; Richter et al., 2016).

Taken together, our results identify the G-patch protein NKRF as a key factor in ribosome biogenesis that mediates the assembly of a subcomplex containing DHX15 and XRN2 for facilitating A' cleavage of the pre-rRNA transcript and also functions together with XRN2 in the turnover of pre-rRNA fragments excised during processing.

#### 4.5 Conclusions and perspectives

In this study, a systematic analysis of the human G-patch protein family was performed with the aim to establish these proteins as cofactors of RNA helicases and to gain insight into the function and regulation of the G-patch cofactor-RNA helicase complexes. Our interaction studies and *in vitro* experiments showed that every human G-patch protein associates with an RNA helicase, and in most cases, they stimulate the RNA binding and/or ATPase activity of the helicase. The G-patch domain was found to be generally sufficient for the regulation. Overall, we could confirm the role as *bona fide* cofactors for 20 out of the 22 G-patch proteins, while for the remaining proteins additional experiments are required to uncover their effects on the helicase partner. Our results also revealed that only three human DEAH/RHA helicases interact with G-patch proteins: DHX16 and DHX35 have one cofactor each, while DHX15 associates with a network of 20 G-patch cofactors. It remains to be determined which characteristics enable these three RNA helicases to interact with G-patch proteins and what residues of the G-patch domain are essential for distinguishing and regulating the cognate helicase. Furthermore, the intriguing finding that DHX15 has such a large number of cofactors suggests that these interactions might be dynamically modulated and this hypothesis could be tested in different cell types or physiological conditions.

Our genome-wide analyses indicated that DHX15 and most G-patch proteins regulate the alternative splicing of subsets of genes. Based on the identification of common targets for DHX15 and its G-patch cofactors, we propose that these complexes function together in

this pathway. Similarly, the detection of overlapping splicing targets between different G-patch proteins suggests that their action might be coordinated. The function of DHX15 and of several G-patch proteins in alternative splicing was tested and confirmed, thus revealing novel regulators of this pathway. Our results also showed that SUGP2 crosslinks to pre-mRNAs, implying that this G-patch protein has a direct role in modulating splice site selection. On the other hand, ZGPAT was found to bind scaRNAs and DHX15 was shown to influence the scaRNA-guided modification of snRNAs at specific sites. Considering the role of ZGPAT as a cofactor of DHX15, this might suggest that these proteins are involved together in snRNA modification and, in this way, modulate alternative splicing. Further studies are needed to validate the function of the other G-patch proteins in alternative splicing and to understand the mechanisms through which they act.

In addition to its involvement in alternative splicing, DHX15 has a role in ribosome biogenesis together with its G-patch cofactor NKRF and the exonuclease XRN2. These proteins form a nucleolar subcomplex that is required for efficient cleavage of the pre-rRNA transcript at a specific site. The finding that the catalytic activity of DHX15 is needed for this processing step suggests that the helicase performs a structural remodeling event at this site. In addition, NKRF mediates the recruitment of XRN2 to the nucleolus and is therefore required for the function of the exonuclease in the degradation of pre-rRNA fragments excised during processing. Taken together, the data presented in this study indicate that DHX15 participates in both alternative splicing and ribosome biogenesis together with its G-patch cofactors. Future experiments will help elucidate if there is cross-regulation between these functions as has been shown for other multifunctional RNA helicases.

In conclusion, our findings validate G-patch proteins as a specific family of RNA helicase cofactors and expand the knowledge of the mode of regulation of RNA helicases by these proteins. The functional characterization of DHX15 and its G-patch cofactors revealed roles in alternative splicing and ribosome biogenesis that represent a starting point for more in-depth studies. Elucidating the function of each of these helicase-cofactor complexes would enable targeting specific activities of the helicase without affecting the others. Therefore, this study provides essential insights into the function of RNA helicases and their regulation by G-patch proteins that are significant both in normal conditions and in disease.

## References

- Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H., and Luhrmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* 31, 2667–2682.
- Ahn, E.-Y., DeKolver, R.C., Lo, M.-C., Nguyen, T.A., Matsuura, S., Boyapati, A., Pandit, S., Fu, X.-D., and Zhang, D.-E. (2011). SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol. Cell* 42, 185–198.
- Aksaas, A.K., Larsen, A.C., Rogne, M., Rosendal, K., Kvissel, A.-K., and Skalhegg, B.S. (2011). G-patch domain and KOW motifs-containing protein, GPKOW; a nuclear RNA-binding protein regulated by protein kinase A. *J. Mol. Signal.* 6, 10.
- Andersen, J.S., Lam, Y.W., Leung, A.K.L., Ong, S.-E., Lyon, C.E., Lamond, A.I., and Mann, M. (2005). Nucleolar proteome dynamics. *Nature* 433, 77–83.
- Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32, D115-9.
- Aravind, L., and Koonin, E. V (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem. Sci.* 24, 342–344.
- Arcus, V. (2002). OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* 12, 794–801.
- Aubert, M., O'Donohue, M.-F., Lebaron, S., and Gleizes, P.-E. (2018). Pre-Ribosomal RNA Processing in Human Cells: From Mechanisms to Congenital Diseases. *Biomolecules* 8.
- Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* 12, 861–869.
- Banerjee, D., McDaniel, P.M., and Rymond, B.C. (2015). Limited portability of G-patch domains in regulators of the Prp43 RNA helicase required for pre-mRNA splicing and ribosomal RNA maturation in *Saccharomyces cerevisiae*. *Genetics* 200, 135–147.
- Banroques, J., Doere, M., Dreyfus, M., Linder, P., and Tanner, N.K. (2010). Motif III in superfamily 2 “helicases” helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J. Mol. Biol.* 396, 949–966.
- Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., and Le Hir, H. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat. Struct. Mol. Biol.* 19, 983–990.

- Basak, A., and Query, C.C. (2014). A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast. *Cell Rep.* 8, 966–973.
- Bauerova-Zabranska, H., Stokrova, J., Strisovsky, K., Hunter, E., Ruml, T., and Pichova, I. (2005). The RNA binding G-patch domain in retroviral protease is important for infectivity and D-type morphogenesis of Mason-Pfizer monkey virus. *J. Biol. Chem.* 280, 42106–42112.
- Bechara, E.G., Sebestyen, E., Bernardis, I., Eyra, E., and Valcarcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell* 52, 720–733.
- 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.
- Bleichert, F., and Baserga, S.J. (2007). The long unwinding road of RNA helicases. *Mol. Cell* 27, 339–352.
- Bohnsack, M.T., and Sloan, K.E. (2018). Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biol. Chem.* 399, 1265–1276.
- Bohnsack, M.T., Kos, M., and Tollervey, D. (2008). Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase. *EMBO Rep.* 9, 1230–1236.
- Bohnsack, M.T., Martin, R., Granneman, S., Ruprecht, M., Schleiff, E., and Tollervey, D. (2009). Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol. Cell* 36, 583–592.
- Bolger, T.A., and Wente, S.R. (2011). Gle1 is a multifunctional DEAD-box protein regulator that modulates Ded1 in translation initiation. *J. Biol. Chem.* 286, 39750–39759.
- Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 426–438.
- Buchwald, G., Schussler, S., Basquin, C., Le Hir, H., and Conti, E. (2013). Crystal structure of the human eIF4AIII-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. *Proc. Natl. Acad. Sci. U. S. A.* 110, E4611–8.
- Byrd, A.K., and Raney, K.D. (2012). Superfamily 2 helicases. *Front. Biosci. (Landmark Ed.)* 17, 2070–2088.
- Capitanio, J.S., Montpetit, B., and Wozniak, R.W. (2017). Human Nup98 regulates the localization and activity of DExH/D-box helicase DHX9. *eLife*. 6:e18825.

- Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* 12, 123–133.
- Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* 63, 696–710.
- Cech, T.R. (2012). The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* 4, a006742.
- Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 157, 77–94.
- Chaker-Margot, M. (2018). Assembly of the small ribosomal subunit in yeast: mechanism and regulation. *RNA* 24, 881–891.
- Chamousset, D., Mamane, S., Boisvert, F.-M., and Trinkle-Mulcahy, L. (2010). Efficient extraction of nucleolar proteins for interactome analyses. *Proteomics* 10, 3045–3050.
- Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754.
- Chen, M.C., Tippana, R., Demeshkina, N.A., Murat, P., Balasubramanian, S., Myong, S., and Ferre-D'Amare, A.R. (2018). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* 558, 465–469.
- Chen, Y.-L., Capeyrou, R., Humbert, O., Mouffok, S., Kadri, Y. Al, Lebaron, S., Henras, A.K., and Henry, Y. (2014). The telomerase inhibitor Gno1p/PINX1 activates the helicase Prp43p during ribosome biogenesis. *Nucleic Acids Res.* 42, 7330–7345.
- Chen, Z., Gui, B., Zhang, Y., Xie, G., Li, W., Liu, S., Xu, B., Wu, C., He, L., Yang, J., et al. (2017). Identification of a 35S U4/U6.U5 tri-small nuclear ribonucleoprotein (tri-snRNP) complex intermediate in spliceosome assembly. *J. Biol. Chem.* 292, 18113–18128.
- Choudhury, P., Hackert, P., Memet, I., Sloan, K.E., and Bohnsack, M.T. The human RNA helicase DHX37 is required for release of the U3 snoRNP from pre-ribosomal particles. (*submitted*).
- Christian, H., Hofele, R. V, Urlaub, H., and Ficner, R. (2014). Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. *Nucleic Acids Res.* 42, 1162–1179.
- Ciganda, M., and Williams, N. (2011). Eukaryotic 5S rRNA biogenesis. *Wiley Interdiscip. Rev. RNA* 2, 523–533.

- Coccia, M., Rossi, A., Riccio, A., Trotta, E., and Santoro, M.G. (2017). Human NF-kappaB repressing factor acts as a stress-regulated switch for ribosomal RNA processing and nucleolar homeostasis surveillance. *Proc. Natl. Acad. Sci. U. S. A.* *114*, 1045–1050.
- Collins, R., Karlberg, T., Lehtio, L., Schutz, P., van den Berg, S., Dahlgren, L.-G., Hammarstrom, M., Weigelt, J., and Schuler, H. (2009). The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J. Biol. Chem.* *284*, 10296–10300.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* *17*, 13.
- Cordin, O., and Beggs, J.D. (2013). RNA helicases in splicing. *RNA Biol.* *10*, 83–95.
- Cordin, O., Tanner, N.K., Doere, M., Linder, P., and Banroques, J. (2004). The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J.* *23*, 2478–2487.
- Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. *Gene* *367*, 17–37.
- Cusack, S. (1999). RNA-protein complexes. *Curr. Opin. Struct. Biol.* *9*, 66–73.
- De, I., Bessonov, S., Hofele, R., dos Santos, K., Will, C.L., Urlaub, H., Luhrmann, R., and Pena, V. (2015). The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* *22*, 138–144.
- Dhote, V., Sweeney, T.R., Kim, N., Hellen, C.U.T., and Pestova, T. V (2012). Roles of individual domains in the function of DHX29, an essential factor required for translation of structured mammalian mRNAs. *Proc. Natl. Acad. Sci. U. S. A.* *109*, E3150-9.
- Diges, C.M., and Uhlenbeck, O.C. (2001). Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. *EMBO J.* *20*, 5503–5512.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.
- Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)*. *1*, 895–905.
- Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A., et al. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* *70*, 854–867.e9.

- Donmez, G., Hartmuth, K., and Luhrmann, R. (2004). Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA* 10, 1925–1933.
- Dosil, M., and Bustelo, X.R. (2004). Functional characterization of Pwp2, a WD family protein essential for the assembly of the 90 S pre-ribosomal particle. *J. Biol. Chem.* 279, 37385–37397.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498.
- Faber, Z.J., Chen, X., Gedman, A.L., Boggs, K., Cheng, J., Ma, J., Radtke, I., Chao, J.-R., Walsh, M.P., Song, G., et al. (2016). The genomic landscape of core-binding factor acute myeloid leukemias. *Nat. Genet.* 48, 1551–1556.
- Fairman-Williams, M.E., Guenther, U.-P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* 20, 313–324.
- Fairman, M.E., Maroney, P.A., Wang, W., Bowers, H.A., Gollnick, P., Nilsen, T.W., and Jankowsky, E. (2004). Protein displacement by DExH/D “RNA helicases” without duplex unwinding. *Science* 304, 730–734.
- Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.* 24, 791–799.
- Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229–234.
- Fitzgerald, M.E., Rawling, D.C., Potapova, O., Ren, X., Kohlway, A., and Pyle, A.M. (2017). Selective RNA targeting and regulated signaling by RIG-I is controlled by coordination of RNA and ATP binding. *Nucleic Acids Res.* 45, 1442–1454.
- Fourmann, J.-B., Dybkov, O., Agafonov, D.E., Tauchert, M.J., Urlaub, H., Ficner, R., Fabrizio, P., and Luhrmann, R. (2016). The target of the DEAH-box NTP triphosphatase Prp43 in *Saccharomyces cerevisiae* spliceosomes is the U2 snRNP-intron interaction. *eLife*. 5:e15564.
- Fourmann, J.-B., Tauchert, M.J., Ficner, R., Fabrizio, P., and Luhrmann, R. (2017). Regulation of Prp43-mediated disassembly of spliceosomes by its cofactors Ntr1 and Ntr2. *Nucleic Acids Res.* 45, 4068–4080.

- Frenal, K., Callebaut, I., Wecker, K., Prochnicka-Chalufour, A., Dendouga, N., Zinn-Justin, S., Delepierre, M., Tomavo, S., and Wolff, N. (2006). Structural and functional characterization of the TgDRE multidomain protein, a DNA repair enzyme from *Toxoplasma gondii*. *Biochemistry* 45, 4867–4874.
- Fu, X.-D., and Ares, M.J. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701.
- Fuller-Pace, F. V (2013). DEAD box RNA helicase functions in cancer. *RNA Biol.* 10, 121–132.
- Garcia-Montojo, M., Doucet-O'Hare, T., Henderson, L., and Nath, A. (2018). Human endogenous retrovirus-K (HML-2): a comprehensive review. *Crit. Rev. Microbiol.* 1–24.
- Gifford, R., Kabat, P., Martin, J., Lynch, C., and Tristem, M. (2005). Evolution and distribution of class II-related endogenous retroviruses. *J. Virol.* 79, 6478–6486.
- Gilman, B., Tijerina, P., and Russell, R. (2017). Distinct RNA-unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs. *Biochem. Soc. Trans.* 45, 1313–1321.
- Gorbalenya, A.E., and Koonin, E. V (1993). Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* 3, 419–429.
- Grandi, P., Rybin, V., Bassler, J., Petfalski, E., Strauss, D., Marzioch, M., Schafer, T., Kuster, B., Tschochner, H., Tollervey, D., et al. (2002). 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol. Cell* 10, 105–115.
- Granneman, S., Lin, C., Champion, E.A., Nandineni, M.R., Zorca, C., and Baserga, S.J. (2006). The nucleolar protein Esf2 interacts directly with the DExD/H box RNA helicase, Dbp8, to stimulate ATP hydrolysis. *Nucleic Acids Res.* 34, 3189–3199.
- Gripp, K.W., Hopkins, E., Johnston, J.J., Krause, C., Dobyns, W.B., and Biesecker, L.G. (2011). Long-term survival in TARP syndrome and confirmation of RBM10 as the disease-causing gene. *Am. J. Med. Genet. A* 155A, 2516–2520.
- Gross, T., Siepmann, A., Sturm, D., Windgassen, M., Scarcelli, J.J., Seedorf, M., Cole, C.N., and Krebber, H. (2007). The DEAD-box RNA helicase Dbp5 functions in translation termination. *Science* 315, 646–649.
- Gupta, G.D., Coyaud, E., Goncalves, J., Mojarad, B.A., Liu, Y., Wu, Q., Gheiratmand, L., Comartin, D., Tkach, J.M., Cheung, S.W.T., et al. (2015). A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* 163, 1484–1499.



- Gupta, N., Lorsch, J.R., and Hinnebusch, A.G. (2018). Yeast Ded1 promotes 48S translation pre-initiation complex assembly in an mRNA-specific and eIF4F-dependent manner. *eLife*. 7:e38892.
- Haag, S., Kretschmer, J., and Bohnsack, M.T. (2015). WBSR22/Merm1 is required for late nuclear pre-ribosomal RNA processing and mediates N7-methylation of G1639 in human 18S rRNA. *RNA* 21, 180–187.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M.J., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- Hanke, K., Hohn, O., and Bannert, N. (2016). HERV-K(HML-2), a seemingly silent subtenant - but still waters run deep. *APMIS* 124, 67–87.
- He, Y., Andersen, G.R., and Nielsen, K.H. (2010). Structural basis for the function of DEAH helicases. *EMBO Rep.* 11, 180–186.
- He, Y., Staley, J.P., Andersen, G.R., and Nielsen, K.H. (2017). Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. *RNA* 23, 1110–1124.
- Heininger, A.U., Hackert, P., Andreou, A.Z., Boon, K.-L., Memet, I., Prior, M., Clancy, A., Schmidt, B., Urlaub, H., Schleiff, E., et al. (2016). Protein cofactor competition regulates the action of a multifunctional RNA helicase in different pathways. *RNA Biol.* 13, 320–330.
- Henras, A.K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* 6, 225–242.
- Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341.
- Herschlag, D. (1995). RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 270, 20871–20874.
- Hickey, C.J., Kim, J.-H., and Ahn, E.-Y.E. (2014). New discoveries of old SON: a link between RNA splicing and cancer. *J. Cell. Biochem.* 115, 224–231.
- Hilbert, M., Karow, A.R., and Klostermeier, D. (2009). The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol. Chem.* 390, 1237–1250.
- Hu, L., Wang, J., Liu, Y., Zhang, Y., Zhang, L., Kong, R., Zheng, Z., Du, X., and Ke, Y. (2011). A small ribosomal subunit (SSU) processome component, the human U3 protein 14A (hUTP14A) binds p53 and promotes p53 degradation. *J. Biol. Chem.* 286, 3119–3128.

- Huen, M.S.Y., Sy, S.M.H., Leung, K.M., Ching, Y.-P., Tipoe, G.L., Man, C., Dong, S., and Chen, J. (2010). SON is a spliceosome-associated factor required for mitotic progression. *Cell Cycle* 9, 2679–2685.
- Inesta-Vaquera, F., Chaugule, V.K., Galloway, A., Chandler, L., Rojas-Fernandez, A., Weidlich, S., Pegg, M., and Cowling, V.H. (2018). DHX15 regulates CMTR1-dependent gene expression and cell proliferation. *Life Sci. Alliance* 1, e201800092.
- Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* 36, 19–29.
- Jankowsky, E., and Bowers, H. (2006). Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res.* 34, 4181–4188.
- Jankowsky, E., and Fairman, M.E. (2007). RNA helicases--one fold for many functions. *Curr. Opin. Struct. Biol.* 17, 316–324.
- Jankowsky, E., and Harris, M.E. (2015). Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* 16, 533–544.
- Jarmoskaite, I., and Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.* 83, 697–725.
- Jern, P., Sperber, G.O., and Blomberg, J. (2005). Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology* 2, 50.
- Jianfeng, D., Feng, J., Chaoneng, J., Zhongzhou, Z., Shaohua, G., Qihan, W., Liu, W., Gang, Y., Yi, X., and Mao, Y. (2003). Cloning of the correct full length cDNA of NF-kappaB-repressing factor. *Mol. Cells* 16, 397–401.
- Kaneko, H., Kitoh, H., Matsuura, T., Masuda, A., Ito, M., Mottes, M., Rauch, F., Ishiguro, N., and Ohno, K. (2011). Hyperuricemia cosegregating with osteogenesis imperfecta is associated with a mutation in GPATCH8. *Hum. Genet.* 130, 671–683.
- Karijovich, J., and Yu, Y.-T. (2010). Spliceosomal snRNA modifications and their function. *RNA Biol.* 7, 192–204.
- Karijovich, J., Yi, C., and Yu, Y.-T. (2015). Transcriptome-wide dynamics of RNA pseudouridylation. *Nat. Rev. Mol. Cell Biol.* 16, 581–585.
- Katz, Y., Wang, E.T., Airolidi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* 7, 1009–1015.

- Kiianitsa, K., Solinger, J.A., and Heyer, W.-D. (2003). NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Anal. Biochem.* 321, 266–271.
- Konikkat, S., and Woolford, J.L.J. (2017). Principles of 60S ribosomal subunit assembly emerging from recent studies in yeast. *Biochem. J.* 474, 195–214.
- Koodathingal, P., and Staley, J.P. (2013). Splicing fidelity: DEAD/H-box ATPases as molecular clocks. *RNA Biol.* 10, 1073–1079.
- Koodathingal, P., Novak, T., Piccirilli, J.A., and Staley, J.P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. *Mol. Cell* 39, 385–395.
- Kornprobst, M., Turk, M., Kellner, N., Cheng, J., Flemming, D., Kos-Braun, I., Kos, M., Thoms, M., Berninghausen, O., Beckmann, R., et al. (2016). Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell* 166, 380–393.
- Kos, M., and Tollervey, D. (2005). The Putative RNA Helicase Dbp4p Is Required for Release of the U14 snoRNA from Preribosomes in *Saccharomyces cerevisiae*. *Mol. Cell* 20, 53–64.
- Kossen, K., Karginov, F. V., and Uhlenbeck, O.C. (2002). The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA. *J. Mol. Biol.* 324, 625–636.
- Kressler, D., Hurt, E., and Bassler, J. (2010). Driving ribosome assembly. *Biochim. Biophys. Acta* 1803, 673–683.
- Kressler, D., Hurt, E., and Bassler, J. (2017). A Puzzle of Life: Crafting Ribosomal Subunits. *Trends Biochem. Sci.* 42, 640–654.
- Krizova, I., Hadravova, R., Stokrova, J., Gunterova, J., Dolezal, M., Ruml, T., Rumlova, M., and Pichova, I. (2012). The G-patch domain of Mason-Pfizer monkey virus is a part of reverse transcriptase. *J. Virol.* 86, 1988–1998.
- Krogh, N., Jansson, M.D., Hafner, S.J., Tehler, D., Birkedal, U., Christensen-Dalsgaard, M., Lund, A.H., and Nielsen, H. (2016). Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* 44, 7884–7895.
- Krogh, N., Kongsbak-Wismann, M., Geisler, C., and Nielsen, H. (2017). Substoichiometric ribose methylations in spliceosomal snRNAs. *Org. Biomol. Chem.* 15, 8872–8876.
- Kudlinzki, D., Schmitt, A., Christian, H., and Ficner, R. (2012). Structural analysis of the C-terminal domain of the spliceosomal helicase Prp22. *Biol. Chem.* 393, 1131–1140.

- Lafontaine, D.L.J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nat. Struct. Mol. Biol.* 22, 11–19.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Laplace, J.M., O'Rourke, F., Lu, X., Fein, A., Olsen, A., and Feinstein, M.B. (2000). Cloning of human Ca<sup>2+</sup> homeostasis endoplasmic reticulum protein (CHERP): regulated expression of antisense cDNA depletes CHERP, inhibits intracellular Ca<sup>2+</sup> mobilization and decreases cell proliferation. *Biochem. J.* 348 Pt 1, 189–199.
- Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A., and Nagamine, Y. (2010). Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* 38, 6219–6233.
- Lazdins, I.B., Delannoy, M., and Sollner-Webb, B. (1997). Analysis of nucleolar transcription and processing domains and pre-rRNA movements by in situ hybridization. *Chromosoma* 105, 481–495.
- Lebaron, S., Papin, C., Capeyrou, R., Chen, Y.-L., Froment, C., Monsarrat, B., Caizergues-Ferrer, M., Grigoriev, M., and Henry, Y. (2009). The ATPase and helicase activities of Prp43p are stimulated by the G-patch protein Pfa1p during yeast ribosome biogenesis. *EMBO J.* 28, 3808–3819.
- Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 84, 291–323.
- Lee, Y.J., Wang, Q., and Rio, D.C. (2018). Coordinate regulation of alternative pre-mRNA splicing events by the human RNA chaperone proteins hnRNPA1 and DDX5. *Genes Dev.* 32, 1060–1074.
- Lestrade, L., and Weber, M.J. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* 34, D158–62.
- Liang, X.-H., and Fournier, M.J. (2006). The helicase Has1p is required for snoRNA release from pre-rRNA. *Mol. Cell. Biol.* 26, 7437–7450.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Lin, M.-L., Fukukawa, C., Park, J.-H., Naito, K., Kijima, K., Shimo, A., Ajiro, M., Nishidate, T., Nakamura, Y., and Katagiri, T. (2009). Involvement of G-patch domain containing 2 overexpression in breast carcinogenesis. *Cancer Sci.* 100, 1443–1450.

- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* *12*, 505–516.
- Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 20209–20214.
- Liu, Y., Conaway, L., Rutherford Bethard, J., Al-Ayoubi, A.M., Thompson Bradley, A., Zheng, H., Weed, S.A., and Eblen, S.T. (2013). Phosphorylation of the alternative mRNA splicing factor 45 (SPF45) by Clk1 regulates its splice site utilization, cell migration and invasion. *Nucleic Acids Res.* *41*, 4949–4962.
- Loiselle, J.J., Roy, J.G., and Sutherland, L.C. (2017). RBM10 promotes transformation-associated processes in small cell lung cancer and is directly regulated by RBM5. *PLoS One* *12*, e0180258.
- Lu, X., Goke, J., Sachs, F., Jacques, P.-E., Liang, H., Feng, B., Bourque, G., Bubulya, P.A., and Ng, H.-H. (2013). SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat. Cell Biol.* *15*, 1141–1152.
- Lu, X., Ng, H.-H., and Bubulya, P.A. (2014). The role of SON in splicing, development, and disease. *Wiley Interdiscip. Rev. RNA* *5*, 637–646.
- De Maio, A., Yalamanchili, H.K., Adamski, C.J., Gennarino, V.A., Liu, Z., Qin, J., Jung, S.Y., Richman, R., Orr, H., and Zoghbi, H.Y. (2018). RBM17 Interacts with U2SURP and CHERP to Regulate Expression and Splicing of RNA-Processing Proteins. *Cell Rep.* *25*, 726–736.e7.
- Major, M.B., Roberts, B.S., Berndt, J.D., Marine, S., Anastas, J., Chung, N., Ferrer, M., Yi, X., Stoick-Cooper, C.L., von Haller, P.D., et al. (2008). New regulators of Wnt/beta-catenin signaling revealed by integrative molecular screening. *Sci. Signal.* *1*, ra12.
- Mallam, A.L., Del Campo, M., Gilman, B., Sidote, D.J., and Lambowitz, A.M. (2012). Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* *490*, 121–125.
- Martin, R. (2014). Functional Characterisation of Ribosome Biogenesis Cofactors in *Saccharomyces cerevisiae*. PhD thesis, University of Göttingen.
- Martin, R., Straub, A.U., Doebele, C., and Bohnsack, M.T. (2013). DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol.* *10*, 4–18.

- Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol. Cell* 54, 751–765.
- Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* 19, 560–567.
- Memet, I., Doebele, C., Sloan, K.E., and Bohnsack, M.T. (2017). The G-patch protein NF-kappaB-repressing factor mediates the recruitment of the exonuclease XRN2 and activation of the RNA helicase DHX15 in human ribosome biogenesis. *Nucleic Acids Res.* 45, 5359–5374.
- Mikhailova, T., Shuvalova, E., Ivanov, A., Susorov, D., Shuvalov, A., Kolosov, P.M., and Alkalaeva, E. (2017). RNA helicase DDX19 stabilizes ribosomal elongation and termination complexes. *Nucleic Acids Res.* 45, 1307–1318.
- Miki, T.S., and Grosshans, H. (2013). The multifunctional RNase XRN2. *Biochem. Soc. Trans.* 41, 825–830.
- Miki, T.S., Richter, H., Ruegger, S., and Grosshans, H. (2014). PAXT-1 promotes XRN2 activity by stabilizing it through a conserved domain. *Mol. Cell* 53, 351–360.
- Montpetit, B., Thomsen, N.D., Helmke, K.J., Seeliger, M.A., Berger, J.M., and Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* 472, 238–242.
- Morris, K. V., and Mattick, J.S. (2014). The rise of regulatory RNA. *Nat. Rev. Genet.* 15, 423–437.
- Mosallanejad, K., Sekine, Y., Ishikura-Kinoshita, S., Kumagai, K., Nagano, T., Matsuzawa, A., Takeda, K., Naguro, I., and Ichijo, H. (2014). The DEAH-box RNA helicase DHX15 activates NF-kappaB and MAPK signaling downstream of MAVS during antiviral responses. *Sci. Signal.* 7, ra40.
- Mullineux, S.-T., and Lafontaine, D.L.J. (2012). Mapping the cleavage sites on mammalian pre-rRNAs: where do we stand? *Biochimie* 94, 1521–1532.
- Nagasawa, C., Ogren, A., Kibiryeva, N., Marshall, J., O'Brien, J.E., Kenmochi, N., and Bittel, D.C. (2018). The Role of scaRNAs in Adjusting Alternative mRNA Splicing in Heart Development. *J. Cardiovasc. Dev. Dis.* 5.E26

- Neumann, B., Wu, H., Hackmann, A., and Krebber, H. (2016). Nuclear Export of Pre-Ribosomal Subunits Requires Dbp5, but Not as an RNA-Helicase as for mRNA Export. *PLoS One* 11, e0149571.
- Nie, Y., Ran, Y., Zhang, H.-Y., Huang, Z.-F., Pan, Z.-Y., Wang, S.-Y., and Wang, Y.-Y. (2017). GPATCH3 negatively regulates RLR-mediated innate antiviral responses by disrupting the assembly of VISA signalosome. *PLoS Pathog.* 13, e1006328.
- Nielsen, K.H., Chamieh, H., Andersen, C.B.F., Fredslund, F., Hamborg, K., Le Hir, H., and Andersen, G.R. (2009). Mechanism of ATP turnover inhibition in the EJC. *RNA* 15, 67–75.
- Niu, Z., Jin, W., Zhang, L., and Li, X. (2012). Tumor suppressor RBM5 directly interacts with the DExD/H-box protein DHX15 and stimulates its helicase activity. *FEBS Lett.* 586, 977–983.
- Nourbakhsh, M., and Hauser, H. (1999). Constitutive silencing of IFN-beta promoter is mediated by NRF (NF-kappaB-repressing factor), a nuclear inhibitor of NF-kappaB. *EMBO J.* 18, 6415–6425.
- Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* 4, 1487–1502.
- Ounap, K., Kasper, L., Kurg, A., and Kurg, R. (2013). The human WBSCR22 protein is involved in the biogenesis of the 40S ribosomal subunits in mammalian cells. *PLoS One* 8, e75686.
- Ozgun, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J.R., and Conti, E. (2015). The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J.* 282, 850–863.
- Pan, C., and Russell, R. (2010). Roles of DEAD-box proteins in RNA and RNP Folding. *RNA Biol.* 7, 667–676.
- Papasaikas, P., Tejedor, J.R., Vigevani, L., and Valcarcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell* 57, 7–22.
- Park, J.W., Tokheim, C., Shen, S., and Xing, Y. (2013). Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS. *Methods Mol. Biol.* 1038, 171–179.
- Patil, P., Kibiryeva, N., Uechi, T., Marshall, J., O'Brien, J.E.J., Artman, M., Kenmochi, N., and Bittel, D.C. (2015). scaRNAs regulate splicing and vertebrate heart development. *Biochim. Biophys. Acta* 1852, 1619–1629.

- Pena, C., Hurt, E., and Panse, V.G. (2017). Eukaryotic ribosome assembly, transport and quality control. *Nat. Struct. Mol. Biol.* 24, 689–699.
- Pertschy, B., Schneider, C., Gnadig, M., Schafer, T., Tollervey, D., and Hurt, E. (2009). RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1. *J. Biol. Chem.* 284, 35079–35091.
- Phipps, K.R., Charette, J.M., and Baserga, S.J. (2011). The small subunit processome in ribosome biogenesis-progress and prospects. *Wiley Interdiscip. Rev. RNA* 2, 1–21.
- Pillet, B., Mitterer, V., Kressler, D., and Pertschy, B. (2017). Hold on to your friends: Dedicated chaperones of ribosomal proteins: Dedicated chaperones mediate the safe transfer of ribosomal proteins to their site of pre-ribosome incorporation. *Bioessays* 39, 1–12.
- Preti, M., O'Donohue, M.-F., Montel-Lehry, N., Bortolin-Cavaille, M.-L., Choesmel, V., and Gleizes, P.-E. (2013). Gradual processing of the ITS1 from the nucleolus to the cytoplasm during synthesis of the human 18S rRNA. *Nucleic Acids Res.* 41, 4709–4723.
- Putnam, A.A., and Jankowsky, E. (2013). DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim. Biophys. Acta* 1829, 884–893.
- Putnam, A.A., Gao, Z., Liu, F., Jia, H., Yang, Q., and Jankowsky, E. (2015). Division of Labor in an Oligomer of the DEAD-Box RNA Helicase Ded1p. *Mol. Cell* 59, 541–552.
- Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.* 37, 317–336.
- Ranji, A., and Boris-Lawrie, K. (2010). RNA helicases: emerging roles in viral replication and the host innate response. *RNA Biol.* 7, 775–787.
- Redder, P., Hausmann, S., Khemici, V., Yasrebi, H., and Linder, P. (2015). Bacterial versatility requires DEAD-box RNA helicases. *FEMS Microbiol. Rev.* 39, 392–412.
- Richter, H., Katic, I., Gut, H., and Grosshans, H. (2016). Structural basis and function of XRN2 binding by XTB domains. *Nat. Struct. Mol. Biol.* 23, 164–171.
- Robert-Paganin, J., Rety, S., and Leulliot, N. (2015). Regulation of DEAH/RHA helicases by G-patch proteins. *Biomed Res. Int.* 2015, 931857.
- Robert-Paganin, J., Halladjian, M., Blaud, M., Lebaron, S., Delbos, L., Chardon, F., Capeyrou, R., Humbert, O., Henry, Y., Henras, A.K., et al. (2017). Functional link between DEAH/RHA helicase Prp43 activation and ATP base binding. *Nucleic Acids Res.* 45, 1539–1552.



- Robert, F., and Pelletier, J. (2013). Perturbations of RNA helicases in cancer. *Wiley Interdiscip. Rev. RNA* 4, 333–349.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
- Rodriguez-Galan, O., Garcia-Gomez, J.J., and de la Cruz, J. (2013). Yeast and human RNA helicases involved in ribosome biogenesis: current status and perspectives. *Biochim. Biophys. Acta* 1829, 775–790.
- Rohmoser, M., Holzel, M., Grimm, T., Malamoussi, A., Harasim, T., Orban, M., Pfisterer, I., Gruber-Eber, A., Kremmer, E., and Eick, D. (2007). Interdependence of Pes1, Bop1, and WDR12 controls nucleolar localization and assembly of the PeBoW complex required for maturation of the 60S ribosomal subunit. *Mol. Cell. Biol.* 27, 3682–3694.
- Rother, S., Bartels, M., Schweda, A.T., Resch, K., Pallua, N., and Nourbakhsh, M. (2016). NF-kappaB-repressing factor phosphorylation regulates transcription elongation via its interactions with 5'-->3' exoribonuclease 2 and negative elongation factor. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* 30, 174–185.
- Rudolph, M.G., and Klostermeier, D. (2015). When core competence is not enough: functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. *Biol. Chem.* 396, 849–865.
- Russell, R. (2008). RNA misfolding and the action of chaperones. *Front. Biosci.* 13, 1–20.
- Russell, R., Jarmoskaite, I., and Lambowitz, A.M. (2013). Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol.* 10, 44–55.
- Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* 25, 373–384.
- Samatanga, B., Andreou, A.Z., and Klostermeier, D. (2017). Allosteric regulation of helicase core activities of the DEAD-box helicase YxiN by RNA binding to its RNA recognition motif. *Nucleic Acids Res.* 45, 1994–2006.
- Sambrook, J.F., and Russell, D.W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory Press, New York.
- Sampath, J., Long, P.R., Shepard, R.L., Xia, X., Devanarayan, V., Sandusky, G.E., Perry, W.L. 3rd, Dantzig, A.H., Williamson, M., Rolfe, M., et al. (2003). Human SPF45, a splicing factor, has limited expression in normal tissues, is overexpressed in many tumors, and can confer a multidrug-resistant phenotype to cells. *Am. J. Pathol.* 163, 1781–1790.
- Sampson, N.D., and Hewitt, J.E. (2003). SF4 and SFRS14, two related putative splicing factors on human chromosome 19p13.11. *Gene* 305, 91–100.

- Sardana, R., Liu, X., Granneman, S., Zhu, J., Gill, M., Papoulas, O., Marcotte, E.M., Tollervey, D., Correll, C.C., and Johnson, A.W. (2015). The DEAH-box helicase Dhr1 dissociates U3 from the pre-rRNA to promote formation of the central pseudoknot. *PLoS Biol.* 13, e1002083.
- Sasaki-Osugi, K., Imoto, C., Takahara, T., Shibata, H., and Maki, M. (2013). Nuclear ALG-2 protein interacts with Ca<sup>2+</sup> homeostasis endoplasmic reticulum protein (CHERP) Ca<sup>2+</sup>-dependently and participates in regulation of alternative splicing of inositol trisphosphate receptor type 1 (IP3R1) pre-mRNA. *J. Biol. Chem.* 288, 33361–33375.
- Sato, S., Ishikawa, H., Yoshikawa, H., Izumikawa, K., Simpson, R.J., and Takahashi, N. (2015). Collaborator of alternative reading frame protein (CARF) regulates early processing of pre-ribosomal RNA by retaining XRN2 (5'-3' exoribonuclease) in the nucleoplasm. *Nucleic Acids Res.* 43, 10397–10410.
- Schillewaert, S., Wacheul, L., Lhomme, F., and Lafontaine, D.L.J. (2012). The evolutionarily conserved protein Las1 is required for pre-rRNA processing at both ends of ITS2. *Mol. Cell. Biol.* 32, 430–444.
- Schuberth-Wagner, C., Ludwig, J., Bruder, A.K., Herzner, A.-M., Zillinger, T., Goldeck, M., Schmidt, T., Schmid-Burgk, J.L., Kerber, R., Wolter, S., et al. (2015). A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2'O-Methylated Self RNA. *Immunity* 43, 41–51.
- Schuller, J.M., Falk, S., Fromm, L., Hurt, E., and Conti, E. (2018). Structure of the nuclear exosome captured on a maturing preribosome. *Science* 360, 219–222.
- Schutz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9564–9569.
- Semlow, D.R., Blanco, M.R., Walter, N.G., and Staley, J.P. (2016). Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. *Cell* 164, 985–998.
- Sen, N.D., Zhou, F., Ingolia, N.T., and Hinnebusch, A.G. (2015). Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res.* 25, 1196–1205.
- Sharma, S., Yang, J., van Nues, R., Watzinger, P., Kotter, P., Lafontaine, D.L.J., Granneman, S., and Entian, K.-D. (2017). Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genet.* 13, e1006804.
- Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* 18, 655–670.

- Shi, Z., Fujii, K., Kovary, K.M., Genuth, N.R., Rost, H.L., Teruel, M.N., and Barna, M. (2017). Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol. Cell* 67, 71–83.e7.
- Silverman, E., Edwalds-Gilbert, G., and Lin, R.-J. (2003). DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene* 312, 1–16.
- Silverman, E.J., Maeda, A., Wei, J., Smith, P., Beggs, J.D., and Lin, R.-J. (2004). Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. *Mol. Cell. Biol.* 24, 10101–10110.
- Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* 76, 23–50.
- Sloan, K.E., and Bohnsack, M.T. (2018). Unravelling the Mechanisms of RNA Helicase Regulation. *Trends Biochem. Sci.* 43, 237–250.
- Sloan, K.E., Mattijssen, S., Lebaron, S., Tollervey, D., Pruijn, G.J.M., and Watkins, N.J. (2013). Both endonucleolytic and exonucleolytic cleavage mediate ITS1 removal during human ribosomal RNA processing. *J. Cell Biol.* 200, 577–588.
- Sloan, K.E., Bohnsack, M.T., Schneider, C., and Watkins, N.J. (2014). The roles of SSU processome components and surveillance factors in the initial processing of human ribosomal RNA. *RNA* 20, 540–550.
- Sloan, K.E., Leisegang, M.S., Doebele, C., Ramirez, A.S., Simm, S., Safferthal, C., Kretschmer, J., Schorge, T., Markoutsas, S., Haag, S., et al. (2015). The association of late-acting snoRNPs with human pre-ribosomal complexes requires the RNA helicase DDX21. *Nucleic Acids Res.* 43, 553–564.
- Smietanski, M., Werner, M., Purta, E., Kaminska, K.H., Stepinski, J., Darzynkiewicz, E., Nowotny, M., and Bujnicki, J.M. (2014). Structural analysis of human 2'-O-ribose methyltransferases involved in mRNA cap structure formation. *Nat. Commun.* 5, 3004.
- Srivastava, L., Lapik, Y.R., Wang, M., and Pestov, D.G. (2010). Mammalian DEAD box protein Ddx51 acts in 3' end maturation of 28S rRNA by promoting the release of U8 snoRNA. *Mol. Cell. Biol.* 30, 2947–2956.
- Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* 3, 55–64.
- Stanek, D., Pridalova-Hnilicova, J., Novotny, I., Huranova, M., Blazikova, M., Wen, X., Sapra, A.K., and Neugebauer, K.M. (2008). Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. *Mol. Biol. Cell* 19, 2534–2543.

- Steimer, L., and Klostermeier, D. (2012). RNA helicases in infection and disease. *RNA Biol.* 9, 751–771.
- Strunk, B.S., and Karbstein, K. (2009). Powering through ribosome assembly. *RNA* 15, 2083–2104.
- Sun, Q., Zhu, X., Qi, J., An, W., Lan, P., Tan, D., Chen, R., Wang, B., Zheng, S., Zhang, C., et al. (2017a). Molecular architecture of the 90S small subunit pre-ribosome. *eLife*. 6:e22086.
- Sun, Y., Bao, Y., Han, W., Song, F., Shen, X., Zhao, J., Zuo, J., Saffen, D., Chen, W., Wang, Z., et al. (2017b). Autoregulation of RBM10 and cross-regulation of RBM10/RBM5 via alternative splicing-coupled nonsense-mediated decay. *Nucleic Acids Res.* 45, 8524–8540.
- Svec, M., Bauerova, H., Pichova, I., Konvalinka, J., and Strisovsky, K. (2004). Proteinases of betaretroviruses bind single-stranded nucleic acids through a novel interaction module, the G-patch. *FEBS Lett.* 576, 271–276.
- Tafforeau, L., Zorbas, C., Langhendries, J.-L., Mullineux, S.-T., Stamatopoulou, V., Mullier, R., Wacheul, L., and Lafontaine, D.L.J. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Mol. Cell* 51, 539–551.
- Tanaka, N., Aronova, A., and Schwer, B. (2007). Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev.* 21, 2312–2325.
- Tauchert, M.J., Fourmann, J.-B., Luhrmann, R., and Ficner, R. (2017). Structural insights into the mechanism of the DEAH-box RNA helicase Prp43. *eLife*. 6:e21510.
- Theobald, D.L., Mitton-Fry, R.M., and Wuttke, D.S. (2003). Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* 32, 115–133.
- Thoms, M., Thomson, E., Bassler, J., Gnadig, M., Griesel, S., and Hurt, E. (2015). The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* 162, 1029–1038.
- Tian, X.-L., Kadaba, R., You, S.-A., Liu, M., Timur, A.A., Yang, L., Chen, Q., Szafranski, P., Rao, S., Wu, L., et al. (2004). Identification of an angiogenic factor that when mutated causes susceptibility to Klippel-Trenaunay syndrome. *Nature* 427, 640–645.
- Tieg, B., and Krebber, H. (2013). Dbp5 - from nuclear export to translation. *Biochim. Biophys. Acta* 1829, 791–798.

- Toczydlowska-Socha, D., Zielinska, M.M., Kurkowska, M., Astha, Almeida, C.F., Stefaniak, F., Purta, E., and Bujnicki, J.M. (2018). Human RNA cap1 methyltransferase CMTr1 cooperates with RNA helicase DHX15 to modify RNAs with highly structured 5' termini. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 373.
- Tomecki, R., Sikorski, P.J., and Zakrzewska-Placzek, M. (2017). Comparison of preribosomal RNA processing pathways in yeast, plant and human cells - focus on coordinated action of endo- and exoribonucleases. *FEBS Lett.* 591, 1801–1850.
- Utans, U., and Kramer, A. (1990). Splicing factor SF4 is dispensable for the assembly of a functional splicing complex and participates in the subsequent steps of the splicing reaction. *EMBO J.* 9, 4119–4126.
- Wahl, M.C., and Luhrmann, R. (2015). SnapShot: Spliceosome Dynamics II. *Cell* 162, 456–456.e1.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718.
- Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., van Tilbeurgh, H., Henry, Y., and Leulliot, N. (2010). Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *EMBO J.* 29, 2194–2204.
- Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951.
- 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.
- Wang, Z., and Burge, C.B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14, 802–813.
- Wang, W., Li, G.-Y., Zhu, J.-Y., Huang, D.-B., Zhou, H.-C., Zhong, W., and Ji, C.-S. (2015a). Overexpression of AGGF1 is correlated with angiogenesis and poor prognosis of hepatocellular carcinoma. *Med. Oncol.* 32, 131.
- Wang, Y., Gogol-Doring, A., Hu, H., Frohler, S., Ma, Y., Jens, M., Maaskola, J., Murakawa, Y., Quedenau, C., Landthaler, M., et al. (2013). Integrative analysis revealed the molecular mechanism underlying RBM10-mediated splicing regulation. *EMBO Mol. Med.* 5, 1431–1442.

- Wang, Y., Liu, J., Huang, B.O., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M., et al. (2015b). Mechanism of alternative splicing and its regulation. *Biomed. Reports* 3, 152–158.
- Warkocki, Z., Schneider, C., Mozaffari-Jovin, S., Schmitzova, J., Hobartner, C., Fabrizio, P., and Luhrmann, R. (2015). The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes Dev.* 29, 94–107.
- Watkins, N.J., and Bohnsack, M.T. (2012). The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip. Rev. RNA* 3, 397–414.
- Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* 15, R8.
- Weick, E.-M., Puno, M.R., Januszyk, K., Zinder, J.C., DiMattia, M.A., and Lima, C.D. (2018). Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex. *Cell* 173, 1663–1677.e21.
- West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522–525.
- Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, a003707.
- Wilson, D.N., and Doudna Cate, J.H. (2012). The structure and function of the eukaryotic ribosome. *Cold Spring Harb. Perspect. Biol.* 4, a011536.
- Wu, G., Xiao, M., Yang, C., and Yu, Y.-T. (2011). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J.* 30, 79–89.
- Xing, L., Niu, M., and Kleiman, L. (2014). Role of the OB-fold of RNA helicase A in the synthesis of HIV-1 RNA. *Biochim. Biophys. Acta* 1839, 1069–1078.
- Xu, Y., Zhou, M., Wang, J., Zhao, Y., Li, S., Zhou, B., Su, Z., Xu, C., Xia, Y., Qian, H., et al. (2014). Role of microRNA-27a in down-regulation of angiogenic factor AGGF1 under hypoxia associated with high-grade bladder urothelial carcinoma. *Biochim. Biophys. Acta* 1842, 712–725.
- Yang, Q., Del Campo, M., Lambowitz, A.M., and Jankowsky, E. (2007). DEAD-box proteins unwind duplexes by local strand separation. *Mol. Cell* 28, 253–263.

- Yoshimoto, R., Kataoka, N., Okawa, K., and Ohno, M. (2009). Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res.* 37, 891–902.
- Young, C., and Karbstein, K. (2012). Analysis of cofactor effects on RNA helicases. *Methods Enzymol.* 511, 213-237.
- Young, C.L., Khoshnevis, S., and Karbstein, K. (2013). Cofactor-dependent specificity of a DEAD-box protein. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2668-76.
- Yourik, P., Aitken, C.E., Zhou, F., Gupta, N., Hinnebusch, A.G., and Lorsch, J.R. (2017). Yeast eIF4A enhances recruitment of mRNAs regardless of their structural complexity. *eLife.* 6:e31476.
- Yuan, K., Li, N., Jiang, K., Zhu, T., Huo, Y., Wang, C., Lu, J., Shaw, A., Thomas, K., Zhang, J., et al. (2009). PinX1 is a novel microtubule-binding protein essential for accurate chromosome segregation. *J. Biol. Chem.* 284, 23072–23082.
- Yusupova, G., and Yusupov, M. (2014). High-resolution structure of the eukaryotic 80S ribosome. *Annu. Rev. Biochem.* 83, 467–486.
- Zang, S., Lin, T.-Y., Chen, X., Gencheva, M., Newo, A.N.S., Yang, L., Rossi, D., Hu, J., Lin, S.-B., Huang, A., et al. (2014). GPKOW is essential for pre-mRNA splicing in vitro and suppresses splicing defect caused by dominant-negative DHX16 mutation in vivo. *Biosci. Rep.* 34, e00163.
- Zhang, Y., You, J., Wang, X., and Weber, J. (2015). The DHX33 RNA Helicase Promotes mRNA Translation Initiation. *Mol. Cell. Biol.* 35, 2918–2931.
- Zhou, X.Z., and Lu, K.P. (2001). The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor. *Cell* 107, 347–359.
- Zhu, J., Liu, X., Anjos, M., Correll, C.C., and Johnson, A.W. (2016). Utp14 Recruits and Activates the RNA Helicase Dhr1 To Undock U3 snoRNA from the Preribosome. *Mol. Cell. Biol.* 36, 965–978.

## Supplementary information

**Supplementary Table S1. Primers used for cloning**

Construct	Forward / reverse primer sequence (5'-3')	Restriction enzyme
AGGF1-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGCCTCGGAGGCGCCGTCC	HindIII
	ATATATGGATCCCTCTAAAGTCCCTTTACCCAAGG	BamHI
CHERP-His <sub>6</sub> -2xFLAG	ATATATGGTACCGCCACCATGGAGATGCCGCTGCC	Acc65I
	ATATATGCTAGCCTTACACTCGTCCCTGGCCTTC	NheI
CMTR1-His <sub>6</sub> -2xFLAG	ATTATTCTTAAGGCCACCATGAAGAGGAGAACTGACCCAGAATGC	AflIII
	ATTATTGGTACCGGCCCTGTGCATCTGGATGAAG	Acc65I
His <sub>6</sub> -2xFLAG-DHX15 <sub>1-698</sub>	ATATGCGGCCGCGCCATGTCCAAGCGGCACCGTTGGACCTAG	NotI
	ATTATTCTCGAGTCAAAAATACCCAGTAACCAAAGC	XhoI
DHX35-His <sub>6</sub> -2xFLAG	ATTATTGGATCCGCCACCATGGCTGCGCCCGTGGGACC	BamHI
	ATTATTGCTAGCCGGGTCTGGACCTTGGCCCTTTTGG	NheI
DHX35 <sub>1-597</sub> -His <sub>6</sub> -2xFLAG	ATTATTGGATCCGCCACCATGGCTGCGCCCGTGGGACC	BamHI
	ATTATTGCTAGCCTTCTGGGCACCTGAAACTTGAC	NheI
GPANK1-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGTCCCGGCCCTTGCTCATCACC	HindIII
	ATATATGGATCCGAACTCGAGTTTCATGTAAGTCCTTAG	BamHI
GPATCH1-His <sub>6</sub> -2xFLAG	ATATATGGTACCGCCACCATGGCGGCGCGGGACAGTG	Acc65I
	ATATATGCTAGCCTGCCTTCTTAGTGGAAGACTTTTC	NheI
GPATCH1 <sub>199-931</sub> -His <sub>6</sub> -2xFLAG	ATTATTGGTACCGCCACCATGAGCTCGGAAGGATCTGAGG	Acc65I
	ATATATGCTAGCCTGCCTTCTTAGTGGAAGACTTTTC	NheI
GPATCH2-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGTTCCGGGGCCGCGGGCGCCAAC	HindIII
	ATATATGCTAGCGGCGGATTTTCTGCATTGGGGGTAGTAG	NheI
GPATCH3-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGCGGTGCCCGGCGAGG	HindIII
	ATATATGGATCCGTCAGGCAATGAGGGGCTGTCTGAAGC	BamHI
GPATCH4-His <sub>6</sub> -2xFLAG	ATATATGGATCCGCCACCATGAATGTCACCCAGAGGTC	BamHI
	ATATATGCTAGCGTCTCTCTTCTTCTGTTTCTTTTGG	NheI
GPATCH8-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGCGGACCGCTTCTCC	HindIII
	ATTATTGGTACCCGTGCCATGGCTGGGGGGATG	Acc65I
GPATCH11-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGCAGAAGAAGAGGACTATATG	HindIII
	ATATATGGATCCGTCATGATCTGCAGAAATTGGTCC	BamHI
GPKOW-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGCTGACTCCAAAGAGGGTG	HindIII
	ATATATGGATCCGTCATCATCTGTGTCACTAGG	BamHI
His <sub>6</sub> -2xFLAG-NKRF	ATTATTGGATCCATGGGCTTTATGTTACCTCTC	BamHI
	ATTATTCTCGAGTCAATTTGCTTGAGGCATAACA	XhoI
His <sub>6</sub> -2xFLAG-NKRF <sub>110-705</sub> (+ nuclear localization signal)	ATATATGGATCCAAACCTTCCAAAGGTCAAAAACGCCACCTACATG TGATGGTCAAAATCCTCCTAAAAAGGACATCTACCAAGATTATACTC	BamHI
	ATTATTCTCGAGTCAATTTGCTTGAGGCATAACA	XhoI
PINX1-His <sub>6</sub> -2xFLAG	ATTATTAAGCTTATGTCTATGCTGGCTGAACG	HindIII
	ATTATTGCTAGCTTTGGAATCTTCTTCTTCTTC	NheI
RBM5-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGGTTCAGACAAAAGAGTGAG	HindIII
	ATATATGGATCCCTCCATCTCAGTGAACC	NheI

(continued on next page)



(Supplementary Table S1 continued)

His <sub>6</sub> -2xFLAG-RBM6	TAGTACGCGGCCGCATGTGGGGGATTCTCGACC	NotI
	ATTATTCTCGAGTTAATCGAGTTCTTTATATCGAGC	XhoI
RBM10-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGAGTATGAAAGACGTGG	HindIII
	ATATATGCTAGCCTGGGCCTCGTTGAAGCGGGTC	NheI
RBM17-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGTCCCTGTACGATGACC	HindIII
	ATATATGCTAGCAACTTGTCTGCCAAATCCAAG	NheI
SON-His <sub>6</sub> -2xFLAG	ATATATCCTAGGGCCACCATGGCGACCAACATCGAGC	AvrII
	ATATATCCTAGGATACCTATTCAAGAAAAACATACAATTGG	AvrII
His <sub>6</sub> -2xFLAG-SUGP1	ATTATTGGATCCATGAGTCTCAAGATGGACAACC	BamHI
	ATTATTCTCGAGTCAGTAGTAAGGCCGTCTGGGATTG	XhoI
SUGP2-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGCAGCCAGACGAATTACAC	HindIII
	ATATATGGATCCTTTGTTGGCCCGCTTGTGTCTG	BamHI
TFIP11-His <sub>6</sub> -2xFLAG	ATATATCTTAAGGCCACCATGTCAATTGTCCCACTTATACC	AflII
	ATATATCCTAGGCTTGCCCATGTCGATCAGGCTC	AvrII
ZGPAT-His <sub>6</sub> -2xFLAG	ATATATAAGCTTGCCACCATGGACGAGGAGAGCCTGGAGTC	HindIII
	ATATATGGATCCGAACCTCAGTCATCTTCTTGTGGGTG	BamHI
ZZ-AGGF1_619-665-His <sub>7</sub>	ATTATTGGTACCGATAGCAACAAAGGTCGGAAG	Acc65I
	ATTATTGTCGACTGGTTTGCTGTCCCCAAGC	Sall
ZZ-CHERP_841-891-His <sub>7</sub>	ATTATTGGTACCGAAGAGAACAAGGCCATCAG	Acc65I
	ATTATTGTCGACATCCAGAGCCACGCCAC	Sall
ZZ-CMTR1_87-133-His <sub>7</sub>	ATTATTGGTACCTATAATAGCGTCTCCAGAAGC	Acc65I
	ATTATTGTCGACCCGGAGTGTCAGACCCAAG	Sall
MBP-DHX16-His <sub>10</sub>	ATTATTGTCGACATGGCGACGCCGGCGGGTCTG	Sall
	ATTATTAGATCTCCCTAGCTCTTCTCGTGTCTTGC	BglII
ZZ-GPANK1_255-301-His <sub>7</sub>	ATTATTGGTACCATCTCCAGCCCGGGCTTC	Acc65I
	ATTATTGTCGACTGCTGATCTGTAGCCTAGTCC	Sall
ZZ-GPATCH1_152-198-His <sub>7</sub>	ATTATTGGTACCAAATTATCTGTTGGTTTCGAATTG	Acc65I
	ATTATTGTCGACTCCAGGGGGTAATGCACAG	Sall
ZZ-GPATCH2_467-513-His <sub>7</sub>	ATTATTCCATGGGGGAAAATAATTGGAAACCG	NcoI
	ATTATTGGATCCTGGTAGAGGAAATCCAAGTC	BamHI
ZZ-GPATCH3_410-458-His <sub>7</sub>	ATTATTGTCGACACCAAGGGCATTGGGCGGAAG	Sall
	ATTATTGGATCCCTCTCCATGGTACCCCAATCCAC	BamHI
ZZ-GPATCH4_11-57-His <sub>7</sub>	ATTATTGGTACCGGGATGAAGTTTGCTGAGGAGC	Acc65I
	ATTATTGGATCCGGCAGGGTCATGTCCTACC	BamHI
ZZ-GPATCH8_40-86-His <sub>7</sub>	ATTATTGGTACCTCGGATAATATTGGACACCG	Acc65I
	ATTATTGTCGACCATTTCCATGCGACCCATG	Sall
ZZ-GPATCH11_69-115-His <sub>7</sub>	ATTATTGGTACCTGTGAAAACAAAGGGTTTGCC	Acc65I
	ATTATTGTCGACTGATGCCTCATGACCAATGC	Sall
ZZ-GPKOW_164-210-His <sub>7</sub>	ATTATTCCATGGGGGTGGAGGCCTATGGGCTGG	NcoI
	ATTATTGGATCCGGTCAGGTTGGCACCCAGCC	BamHI

(continued on next page)

(Supplementary Table S1 continued)

ZZ-GPKOW-His <sub>7</sub>	TAAGCAGGTACCATGGCTGACTCCAAAG	Acc65I
	TAAGCAGGATCCGTCATCATCTGTGTAC	BamHI
ZZ-NKRF_551-596-His <sub>7</sub>	ATATATCCATGGAAGATAATATTGGAAATCAGCTG	NcoI
	ATATATGGATCCCTCTACATCCAGACCAAG	BamHI
MBP-NKRF-His <sub>10</sub>	ATATATCCATGGGCTTTATGTTACCTCTCATC	NcoI
	ATGGATCCATTTGCTTGAGGCATAACAAG	BamHI
ZZ-PINX1_26-72-His <sub>7</sub>	ATTATTCCATGGGGGATTCCAAGTTTGCCAGC	NcoI
	ATATATGGATCCATTGATGGTAGCTCCGAG	BamHI
ZZ-RBM5_743-789-His <sub>7</sub>	ATTATTGGTACCCACAGTAACATTGGCAACAAG	Acc65I
	ATTATTGTCGACGCTGCCTTTGGCTCCTAGG	Sall
ZZ-RBM6_1051-1097-His <sub>7</sub>	ATTATTCCATGGGGACTAGCAGCAAAGGAGGCTG	NcoI
	ATTATTGGATCCTCTTCTGAGGCTCCAACACTGG	BamHI
ZZ-RBM10_858-904-His <sub>7</sub>	ATTATTGGTACCACTGACAACATTGGCAGTCGGATG	Acc65I
	ATTATTGTCGACGCTGCCCCGTGCACCCAG	Sall
ZZ-RBM17_235-283-His <sub>7</sub>	ATTATTCCATGGGGGGGGGCACGGTGGCGCA	NcoI
	ATTATTGGATCCTGTGGCGTCGCCCACGATGATC	BamHI
ZZ-SON_2305-2351-His <sub>7</sub>	ATTATTGGTACCACTGGAGGAATGGGAGCC	Acc65I
	ATTATTGTCGACTTCTCCTACTGCAACAAGACC	Sall
ZZ-SUGP1_562-609-His <sub>7</sub>	ATTATTGGTACCGTGGAGAACATCGGCTACCAGATG	Acc65I
	ATTATTGTCGACCGGCCGGTCAATGCCGAAGCC	Sall
ZZ-SUGP2_1011-1057-His <sub>7</sub>	ATTATTGGTACCGATAAGAACCTGGGCTTCC	Acc65I
	ATTATTGTCGACCTGCCCGTCAGCACCCA	Sall
ZZ-TFIP11_149-195-His <sub>7</sub>	ATTATTGGTACCACAAAAGGAATTGGACAGAAGC	Acc65I
	ATTATTGTCGACGGATCCATAAGCCCCAC	Sall
ZZ-ZGPAT_313-359-His <sub>7</sub>	ATTATTGGTACCACGCGAGGTATAGGCTCCAGAC	Acc65I
	ATTATTGTCGACCTCCACACACTGGTCCAGCG	Sall

Supplementary Table S2. Primers used for site-directed mutagenesis

Construct	Forward / reverse primer sequence (5'-3')
His <sub>6</sub> -2xFLAG-DHX15-siRNA resistant	GACAGAGATCAGATTTAAAGGTGATCGTGATGAGTGCCACTCTAGATGCAGGAAAATTC
	GAATTTTCCTGCATCTAGAGTGGCACTCATCACGATCACCTTTAAATCTGATCTCTGTC
His <sub>6</sub> -2xFLAG-DHX15 <sub>E261Q</sub> / MBP-DHX15 <sub>E261Q</sub> -His <sub>10</sub>	GGTGTAATAATTCTTGATCAGGCTCATGAGAGGACACTG
	CAGTGCTCTCATGAGCCTGATCAAGAATTATTACACC
His <sub>6</sub> -2xFLAG-NKRF <sub>G1-6A</sub>	CAGCTGCTGAGAAAGATGGCTTGGACTGCTGCTTTAGCTAAATCTGGTGAGG
	CCTCACCAGATTTAGCTAAAGCAGCAGCAGTCCAAGCCATCTTTCTCAGCAGCTG

**Supplementary Table S3. Primers used for (q)RT-PCR**

Target gene	Forward / reverse primer sequence (5'-3')
ACIN1	TACTTAGGCAGCGTCTGGAACG
	CTTGGTTTCTCATCATCAGAGTCACC
AGGF1	CACAGAACGGCTGTACCAGA
	TTACTGAGTTCTTCCACCTGCG
CD46-E1	GAATGCGATAAGGGTTTTTACCTCG
	GAGACTGGAGGCTTGTAAGTAGG
CD46-E2	GCAGTAATTTGTGTTGTCCCGTAC
	GTGGTTGATTTAGTCTGGTAAGTGG
CHERP	CGCTCAGACAGGAGCAAGTGA
	ATGTCTAGCTGGGTCTCCTCC
CMTR1	TGAGCCCTGGACTATGGGAT
	CGGCCATAGTAGCAAATGTGAA
DHX15	CAGCTCCCTGTTTGGGAATAC
	TTGGGTACAGGCAACTCCTC
FAM135A	GTTGTAGCTTCATCATTGCAGACT
	TCAGTAAGTCGAGCTTCTACATCC
GPANK1	CGGAAAACCGGTCTCCTACTC
	GTGCGGTGGTTGGAATCTTG
GPATCH3	TGGATTGGGGTACCATGGAG
	GAGATGAGCCCCAAGCCATT
GPATCH8	TCTCCCGCTTCAACGAAGAC
	GCGGTGTCCAATATTATCCGATT
GPATCH11	AGCGAGCCTGTCAACAACTG
	GCCTCAACCAGTACCATGCTT
GPKOW	AGACTGGAAGGGTGGGACAT
	GGCAGATGGCATCGTAGTGA
NF1	CAGCGGAACCTCCTTCAGATG
	CACTTCCTACTGCACCGATGC
RBM5	TGACCCCAACTCGCAATACTAC
	CACGTAGGTCTCTTTTCCCA
RBM6	GTCCGCCTTACTACTGCCAA
	AATGGCGGATCAAGGTTCTGT
RBM10	AACGCCAATGACACCATCAT
	ATGGTGAGAGCTGGATGAA
RBM17	ACCTAGGAGTGGAGACCAGTG
	TTTGGCTCTTTCCTGAGTGA
RHOT1	GGATCATTACAGAGACAGACTCTCC
	GTAGCACCAAACTTGCTCGAAG
SFXN2	GAACCAGTCCTTCAATGCCTTAGTC
	TCATGGGGATATTGACACAGTTAGC

(continued on next page)

(Supplementary Table S3 continued)

SLIT2	GGACAGATCAAAAGCAAGAAATTCCG
	CAGATCCGCAAAGCAGTCTCC
SON	AGGAAAGGATTGATGCCTGGG
	CAGGCTTGGGCACCACTATT
SUGP1	TCGCTCAGAAGAAACGGGAA
	TGCATTTGTGATTCGCCAGG
SUGP2	GAGGACCCTGCTTACTGGTTT
	CTCATCCGCTGCATTTCTGC
TFIP11	ACCACCAAGGATCCAGATATAATTC
	GACATGGCCAGTCACTTAGAA
THEM4	CGTACACCTACTGAATGGATTCAAG
	GCCTGTGACATTTGTTCTTCTTC
UPF3B	GATAGACAGAATTCCAGAAAGGGAC
	TCATCTCCTTTTCTGGCTTCTTG
ZGPAT	TCCGTGTGCTTTACCTGTACC
	CAGCTCATCCAGAGAGACCAC

## List of Figures

Figure 1.1. Structural comparison of DEAD-box and DEAH/RHA RNA helicases .....	3
Figure 1.2. Overview of ribosome biogenesis in human cells .....	8
Figure 1.3. Overview of the splicing cycle .....	12
Figure 1.4. Characteristics and sequence alignment of human G-patch proteins.....	19
Figure 3.1. Confirmation of the expression of the FLAG-tagged variants of human G-patch proteins from inducible stable cell lines .....	48
Figure 3.2. Human G-patch proteins interact specifically with DHX15, DHX16 or DHX35 <i>in vivo</i> .....	51
Figure 3.3. GPATCH1 and DHX35 interact through conserved domains .....	52
Figure 3.4. Overview of the recombinantly purified G-patch domains and full-length GPKOW, DHX15 and DHX16 .....	53
Figure 3.5. The G-patch domain is generally sufficient to stimulate the ATPase activity of the interacting RNA helicase .....	54
Figure 3.6. Full-length GPKOW stimulates the ATPase activity of DHX16 .....	55
Figure 3.7. The RNA binding affinity of DHX15 is enhanced by the G-patch domains of its cofactors .....	56
Figure 3.8. The G-patch domains of GPKOW and GPATCH1 bind RNA .....	56
Figure 3.9. Human G-patch proteins are localized almost exclusively in the nucleus.....	58
Figure 3.10. The expression levels of the G-patch protein-encoding genes are efficiently reduced by siRNA treatment .....	59
Figure 3.11. Classification of alternative splicing events.....	60
Figure 3.12. Depletion of G-patch proteins and DHX15 affects both exon skipping and inclusion .....	62
Figure 3.13. Knockdown of G-patch proteins leads to changes in alternative splicing for specific genes.....	63
Figure 3.14. Specific splicing events are regulated by both DHX15 and its G-patch cofactors .....	64
Figure 3.15. Alternative splicing of the CD46 gene is regulated by multiple G-patch proteins.....	65

Figure 3.16. Knockdown of DHX15 and of most G-patch proteins does not lead to significant changes in gene expression levels .....	66
Figure 3.17. SUGP2 crosslinks to thousands of mRNAs .....	67
Figure 3.18. SUGP2 binds preferentially pre-mRNAs .....	68
Figure 3.19. Overlap between the SUGP2 bound mRNAs and its alternative splicing targets .....	68
Figure 3.20. ZGPAT crosslinks to multiple scaRNAs/snRNAs and depletion of DHX15 affects snRNA modification .....	69
Figure 3.21. DHX15, NKRF and XRN2 are associated into a complex <i>in vivo</i> .....	73
Figure 3.22. The NKRF-DHX15 interaction is mediated by the G-patch domain of NKRF and the OB-fold of the helicase .....	75
Figure 3.23. DHX15, NKRF and XRN2 interact in the nucleolus and co-migrate with pre-ribosomal complexes .....	76
Figure 3.24. NKRF crosslinks to the pre-rRNA transcript in the spacer regions and within the 28S rRNA .....	78
Figure 3.25. DHX15, NKRF and XRN2 are required for pre-rRNA processing and turnover of by-products .....	80
Figure 3.26. Knockdown of DHX15, NKRF or XRN2 does not affect pre-rRNA transcription and the production of mature rRNAs .....	81
Figure 3.27. NKRF is required for the nucleolar localization of XRN2 .....	82
Figure 3.28. The activity of DHX15 is stimulated by NKRF and is required for efficient pre-rRNA processing .....	83

## List of Tables

Table 2.1. Specific reagents used in this study .....	24
Table 2.2. Composition of standard solutions and buffers .....	25
Table 2.3. Mammalian expression plasmids used in this study .....	26
Table 2.4. Plasmids for recombinant protein expression used in this study.....	27
Table 2.5. siRNAs used in this study.....	27
Table 2.6. RNA oligonucleotides used in this study .....	28
Table 2.7. Northern blotting probes used in this study .....	28
Table 2.8. Primary antibodies used in this study .....	29
Table 2.9. Secondary antibodies used in this study .....	30
Table 2.10. PCR parameters for cloning .....	30
Table 2.11. PCR parameters for site-directed mutagenesis .....	31
Table 2.12. PCR parameters for alternative splicing analysis .....	39
Table 2.13. PCR parameters for CRAC .....	43
Table 3.1. MS analysis results of IP experiments with FLAG-tagged G-patch proteins ....	49
Table 3.2. MS analysis results of the FLAG-DHX15 IP .....	50
Table 3.3. Summary of the RNA-seq and alternative splicing analysis results .....	61
Table 3.4. Overview of the scaRNAs and snRNAs enriched in the ZGPAT RIP-seq analysis .....	70
Supplementary Table S1. Primers used for cloning .....	124
Supplementary Table S2. Primers used for site-directed mutagenesis .....	126
Supplementary Table S3. Primers used for (q)RT-PCR .....	127

## Abbreviations

A3'SS	Alternative 3' splice site
A5'SS	Alternative 5' splice site
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
bp	Base pair
BSA	Bovine serum albumin
cDNA	Complementary DNA
CE	Cassette exon
CPM	Counts per million
CRAC	Crosslinking and analysis of cDNA
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ETS	External transcribed spacer
FBS	Fetal bovine serum
GFP	Green fluorescent protein
GP	G-patch domain
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
IP	Immunoprecipitation
IPTG	Isopropyl- $\beta$ -D-thiogalactoside
ITS	Internal transcribed spacer
Kd	Dissociation constant
kDa	Kilodalton
LB	Lysogeny broth
LDS	Lithium dodecyl sulfate
lncRNA	Long non-coding RNA
LSU	Large ribosomal subunit
MBP	Maltose-binding protein
MES	2-(N-morpholino)ethanesulfonic acid
MIF4G	Middle domain of eIF4G
mRNA	Messenger RNA
MS	Mass spectrometry
MXE	Mutually exclusive exons
NAD <sup>+</sup>	$\beta$ -Nicotinamide adenine dinucleotide
NADH	$\beta$ -Nicotinamide adenine dinucleotide reduced
Ni-NTA	Nickel-nitrilotriacetic acid



---

nt	Nucleotide
NT	Non-target
NTP	Nucleoside triphosphate
OB-fold	Oligonucleotide/oligosaccharide-binding-fold
PAGE	Polyacrylamide gel electrophoresis
PAR	Photoactivatable ribonucleoside-enhanced
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
P <sub>i</sub>	Inorganic phosphate
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
Pre-	Precursor
PSI	Percentage spliced in
PVDF	Polyvinylidene fluoride
RBP	RNA-binding protein
RI	Retained intron
RIP	RNA immunoprecipitation
RNA	Ribonucleic acid
RNP	RNA-protein / ribonucleoprotein
RP	Ribosomal protein
RRM	RNA recognition motif
rRNA	Ribosomal RNA
(q)RT-PCR	(Quantitative) Reverse transcription-PCR
s.e.m	Standard error of the mean
SC	Spectral count
scaRNA	Small Cajal body-associated RNA
SDS	Sodium dodecyl sulfate
seq	Sequencing
SF	Superfamily
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SSU	Small ribosomal subunit
TBE	Tris-borate-EDTA
TBS	Tris-buffered saline
tRNA	Transfer RNA
U	Units
UTR	Untranslated region
UV	Ultraviolet
WH	Winged helix
WT	Wild-type
XTBD	XRN-Two-Binding-Domain

## Acknowledgements

I would like to thank Prof. Dr. Markus Bohnsack for the opportunity to conduct my PhD research in his lab and for his continuous support and guidance throughout my work. I am also very thankful to Prof. Dr. Marina Rodnina and Prof. Dr. Peter Rehling for being part of my thesis committee and for their valuable advice and suggestions. In addition, I would like to thank Dr. Ricarda Richter-Dennerlein, Prof. Dr. Silvio Rizzoli and Prof. Dr. Jörg Stülke for their participation in the extended examination board for my PhD defense.

I would like to acknowledge Prof. Dr. Henning Urlaub, Dr. Christof Lenz and the Proteomics Service Facility at the University Medical Center Göttingen for mass spectrometry analysis. Furthermore, I would like to thank the Transcriptome and Genome Analysis Laboratory at the University Medical Center Göttingen for next-generation sequencing. I am also thankful to Dr. Jens Kretschmer for the bioinformatics analysis of the sequencing data.

I am very grateful to be part of the IMPRS Molecular Biology program and I would like to thank the program committee as well as the program coordinators, Dr. Steffen Burkhardt and Kerstin Grüniger, for support during my studies. Furthermore, I would like to acknowledge the director of the Institute for Molecular Biology, Prof. Dr. Blanche Schwappach-Pignataro, for establishing an excellent work environment.

I am also grateful to all the colleagues in the institute and in the lab with whom I've had the opportunity to work in the past four years. Here, I would like to particularly acknowledge Gerald Aquino, Dr. Lukas Brüning, Priyanka Choudhury, Jimena Davila, Dr. Sara Haag, Philipp Hackert, Dr. Jens Kretschmer, Dr. Katherine Sloan and Dr. Ahmed Warda, who have been around for most of my time in the Bohnsack lab and have been a great support.

In the end, I want to thank my family for always being there for me.