Decoding HNRNPH-mediated splicing regulation

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

am Fachbereich Biologie der

Johannes Gutenberg-Universität in Mainz

Kerstin Ruth Hedwig Tretow

geb. am 01.05.1993 in Münster

Mainz, 2023

Dekan: Prof. Dr. Eckhard Thines

1. Berichterstatter:

2. Berichterstatterin:

Tag der mündlichen Prüfung: 23.04.2024

Zusammenfassung

Alternatives Spleißen ist einer der wichtigsten molekularen Wege für eine Zelle, um aus einem Gen mehrere protein-kodierende mRNA-Isoformen zu bilden, was zur Komplexität des Proteoms beiträgt und zelluläre Vielfalt ermöglicht. Ein Zusammenspiel von RNA-bindenden Proteinen (RBPs) steuert alternative Spleiß-Entscheidungen in menschlichen Zellen. Heterogenes nukleares Ribonukleoprotein H (HNRNPH) ist eines dieser RBPs, das als Spleißfaktor fungiert. Kürzlich zeigten Braun et al. (2018), dass HNRNPH das alternative Spleißen des Proto-Onkogens *MST1R* kooperativ reguliert. Eine solche kooperative Spleißregulation würde das zelluläre System sogar auf kleine Veränderungen der HNRNPH-Expression sehr empfindlich reagieren lassen. Inspiriert durch diese Ergebnisse hatte das hier vorgestellte Projekt zum Ziel, eine mögliche Transkriptom-weite HNRNPH-vermittelte kooperative Spleißregulation zu analysieren und den zugrundeliegenden Mechanismus zu entschlüsseln. RNA-Sequenzierungen (RNA-Seq) von Zellen mit titriertem HNRNPH-Proteinspiegel deuteten tatsächlich darauf hin, dass HNRNPH das alternative Spleißen auf Transkriptom-weiter Ebene kooperativ reguliert. Um den kooperativen Mechanismus zu entschlüsseln, zeigte ein iCLIP Experiment, dass HNRNPH an RNAs mit Guanin (G)-reichen Bereichen bindet, die sich mit vorhergesagten RNA-G-Quadruplexen (rG4) überlappen. Durch *In vitro* Reverse Transkriptase Stop Profiling wurde die rG4-Strukturbildung der HNRNPH-Bindungsstellen im Hochdurchsatz validiert. *In vitro* iCLIP belegte, dass HNRNPH eine intrinsische Präferenz für die Bindung an RNAs mit rG4-Faltungspotenzial hat, und *in vitro* Bindungsstudien zeigten, dass die Bildung von rG4-Strukturen für eine kooperative Bindung erforderlich ist. Schließlich unterstrichen *in vivo* Minigen-Studien die Beteiligung von rG4s an der kooperativen Spleißregulation durch HNRNPH in Zellen. Zusammenfassend wird in dieser Studie zum ersten Mal beschrieben, dass HNRNPH alternatives Spleißen kooperativ auf Transkriptom-weiter Ebene vermittelt und dass rG4-Strukturen bei dieser kooperativen Regulation involviert sind.

Summary

Alternative splicing is one of the most important molecular pathways for a cell to create multiple protein-coding mRNA isoforms from one gene, contributing to the proteomic complexity and enabling cellular diversity. An interplay of RNA-binding proteins (RBPs) guides alternative splicing decisions in human cells. Heterogeneous nuclear ribonucleoprotein H (HNRNPH) is one of these RBPs, which functions as a splice factor. Recently, Braun et al. (2018) showed that HNRNPH cooperatively regulates alternative splicing of the proto-oncogene *MST1R*. Such a cooperative splicing regulation would make the cellular system very sensitive to even only small changes in the HNRNPH expression. Inspired by these results, the here presented project aimed to analyse a transcriptome-wide HNRNPH-mediated cooperative splicing regulation and to decode the underlying mechanism. Initial RNA-Sequencings (RNA-Seq) from cells with titrated HNRNPH protein levels highlighted indeed that HNRNPH regulates alternative splicing cooperatively on a transcriptome-wide scale. To decode the cooperative mechanism, an Individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) experiment showed that HNRNPH binds to RNAs composed of guanine (G)-rich sequences overlapping with predicted RNA G-quadruplexes (rG4). *In vitro* Reverse Transcriptase Stop Profiling validated rG4 structure formation of HNRNPH binding sites in high-throughput. Next, *in vitro* iCLIP revealed that HNRNPH has an intrinsic preference for the binding to RNAs with rG4-folding potential and *in vitro* binding studies showed that rG4 structure formation is required for cooperative binding of HNRNPH. Finally, *in vivo* minigene studies highlighted the involvement of rG4s in cooperative splicing regulation by HNRNPH in cells. In summary, this study describes for the first time that HNRNPH mediates alternative splicing cooperatively on a transcriptome-wide scale and proposes the importance of rG4 structures in this cooperative regulation.

Preface

I hereby declare that I wrote the submitted dissertation without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages, which are appropriated verbatim or paraphrased from published and unpublished texts, as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of Johannes Gutenberg-University Mainz to insure standard scientific practice. The content of this thesis is based on a research collaboration between the laboratories of [redacted] and [redacted]. I was involved in design, execution and evaluation of all experiments and analyses. All experiments were performed by myself, with the following exception: the purification of recombinant HNRNPH1 protein was performed in cooperation with [redacted], [redacted] performed the mass spectrometry analysis and the experiment shown in Figure 37 was performed by [redacted]. [redacted] conducted all bioinformatical analyses for the RNA-sequencing splice isoform quantification, iCLIP data, *in vitro* iCLIP data and *in silico* G-quadruplex predictions. Further, [redacted] performed the Hill function modelling and the design of the *in vitro* RNA library. [redacted] performed bioinformatical analysis of the Reverse Transcriptase Stop Profiling data. [redacted] performed iCLIP- and RNA-sequencing data processing. [redacted] supervised all experimental work and [redacted] supervised all bioinformatical work.

Table of contents

1 Introduction

1.1 RNA processing of human protein-coding genes

The information for the production of functional proteins is stored as genetic information in the DNA sequence in the nucleus. Until 2019, in total 19,116 protein-coding genes have been characterised (Piovesan et al., 2019). To translate the information of the protein-coding genes into proteins, the genes are first transcribed into messenger RNA (mRNA) and then exported from the nucleus into the cytoplasm for translation. Intriguingly, the average length of a protein-coding gene is 26,018 base pairs (bp), while the average length of an mRNA is only 2,938 bp (Piovesan et al., 2019). The discrepancy in length between the genes themselves and the respective mRNAs implies that not all information within a gene is required as information for the translation of mRNAs into proteins. Hence, protein-coding genes consist of coding regions and non-coding regions. In a very simplified model, exons comprise the coding regions and introns the non-coding regions. However, one should keep in mind that the RNA world is not that simple and exons, for example in the 5'- or 3'-untranslated regions, may also comprise non-coding RNA parts. Additionally, in the aspect of alternative splicing, introns can contribute to the coding information. On average, a protein-coding human gene contains nine exons and eight introns (Piovesan et al., 2019). Exons and introns are both transcribed from the protein-coding genes, which leads to the production of precursor mRNAs (pre-mRNA). Since mainly the exons encode genetic information for protein synthesis, the introns have to be removed from the pre-mRNAs post-transcriptionally to obtain mature mRNAs. The process of intron removal is called pre-mRNA splicing. Pre-mRNA splicing is a tightly controlled and complex orchestration of proteins and regulatory RNAs.

1.1.1 The function of the spliceosome

A large protein-RNA complex, the spliceosome, facilitates pre-mRNA splicing. The spliceosome consists of the five small nuclear ribonucleoproteins (snRNP) U1, U2, U4, U5 and U6. Besides these five core snRNPs more than 100 splicing proteins orchestrate together with the spliceosome to facilitate the process of intron removal (reviewed in Gehring and Roignant, 2021). To achieve intron removal, the spliceosome assembles through the recognition of conserved RNA motifs and catalyses two transesterifications to ligate the exons of the mature mRNA covalently. The conserved sequences that define the exon-intron

structure are embedded in the introns and comprise short sequences at the exon-intron junctions, the 5’ splice site (5’ss) and 3’ splice sites (3’ss), as well as the branch point (adenosine) and the polypyrimidine tract, which are both localised close to the 3’ss (reviewed in Gehring and Roignant, 2021).

Recognition of the 5’ss is the first step of the spliceosome assembly. The U1 snRNP recognises the conserved GU motif of the 5’ss and base pairs with the nucleotide position -3 to +6 to form the E complex. Splicing factor 1 (SF1) and U2 auxiliary factor (U2AF) recognise and bind to the branch point and the polypyrimdine tract. To form the prespliceosome (A complex), SF1 is replaced by the U2 snRNP, which binds to the branch point (Figure 1). Next, the U4/U6.U5 tri-snRNP complex is recruited to form the pre-catalytic B complex and after rearrangements and release of the U1 and U4 snRNPs the catalytically active B* complex remains (Figure 1). The B* complex catalyses the first transesterification (Figure 1). Therefore, the 2’-OH group of the branch point adenosine attacks the phosphate group of the 5’ss guanine nucleotide leading to a lariat structure formation and transition to the C complex (reviewed in Shi, 2017; Figure 1). The C complex finally performs the second transesterification (Figure 1). The free 3’-OH group of the exon flanking the 5’ss attacks the phosphate group of the first nucleotide of the exon flanking the 3’ss (reviewed in Shi, 2017). Hence, the C complex releases the lariat structure and ligates the exons to form the mature mRNA (Figure 1).

Research of the last years has shown that the assembly of the spliceosome across the intronic 5’ss and 3’ss, also known as the intron definition model, only works for introns < 250 nucleotides (nt) long (reviewed in Hertel, 2008; Berget, 1995). Intriguingly, with an average length of 1747 nt the majority of human introns are much longer than the described maximum intron length for cross-intron spliceosome assembly (Piovesan et al., 2019). In contrast, in lower eukaryotes like in yeast or in *Drosophila* the majority of introns are shorter than 100 nt in size (Goguel and Rosbash, 1993; Hawkins, 1988). The appearance of long introns in higher eukaryotes like humans suggests the presence of another mechanism to recognise the splice sites. In 1990, the exon definition model was first proposed (Robberson et al., 1990; reviewed in Berget, 1995). Within the exon definition model the U1 snRNP binds to the 5’ss of an intron, while the U2 snRNP binds to the 3’ss of the respective upstream intron. This leads to a cross-exon assembly of the spliceosome, including the U4/U6.U5 tri-snRNP. Finally, the cross-exon assembly transitions into a

cross-intron complex for the final intron splicing (Berget, 1995; reviewed in Hertel, 2008; Schneider et al., 2010).

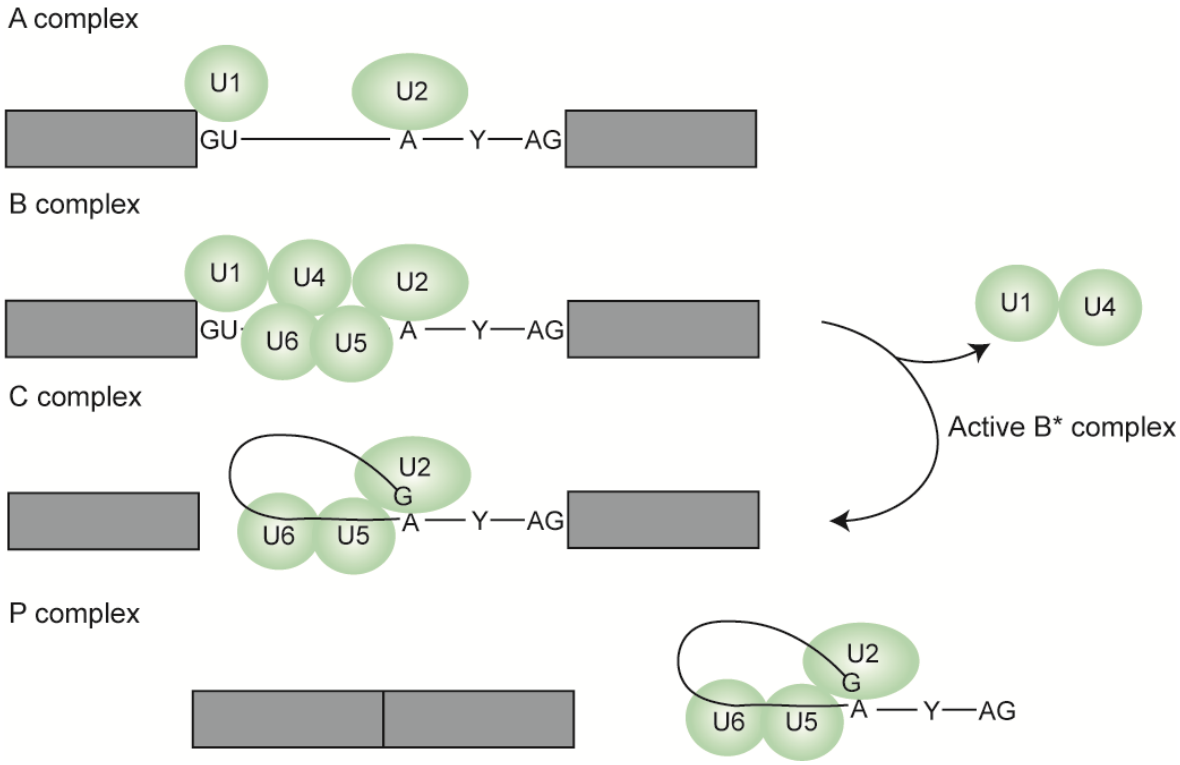

Figure 1: Constitutive pre-mRNA splicing is facilitated in a multi-step process by the spliceosome. Non-coding intronic sequences of pre-mRNAs contain conserved *cis*-regulatory sequences to define the exon-intron structure: the 5' splice site (GU), the 3' splice site (AG), the branch point (A) and the polypyrimidine tract (Y). The spliceosome, a large protein-RNA complex comprising five snRNPs, builds up in a multi-step process recognising the cis-regulatory sequences. First, to form the A complex, the snRNPs U1 and U2 bind to the 5' splice site and the branch point. Next, the B complex is formed by integration of the tri-snRNP complex U4/U6.U5. The first transesterification takes place and forms the intron lariat structure by ligating the 5' splice site to the branch point. The snRNPs U1 and U4 are leaving and the C complex remains. In the last step (P complex), the second transesterification is performed, which ligates the exons to form the mature mRNA and releases the intron lariat. Adapted from Gehring and Roignant, 2021.

1.1.2 Alternative splicing

Recognition of a splice site and the underlying exon-intron structure by the spliceosome depends not only on the splice sites and on the branch point. In humans, the region -3 to +6 of the 5'ss is not strictly conserved, which consequently leads to splice sites with varying strengths (Roca et al., 2012, reviewed in Roca et al., 2013). Hence, to recognise the exon-intron structure and to assemble, the spliceosome often relies on the support of regulatory pre-mRNA sequences, the *cis*-regulatory elements. *Trans*-acting regulatory factors, RNA-binding proteins (RBPs), bind to the *cis*-regulatory elements and thereby support or repress splice site recognition of the spliceosome (reviewed in Wahl et al., 2009).

The combination of varying splice site strengths, *cis*-regulatory elements and RBPs enables eukaryotes, mainly multicellular eukaryotes, to create complex life with a limited number of genes (reviewed in Kornblihtt et al., 2013). With an individual cross talk between splice sites, *cis*-regulatory elements and *trans*-regulating factors, a cell can generate different mRNA isoforms from one gene and thereby increases proteome complexity and achieves context-dependent protein expression. These individual splicing decisions are known as the process of alternative splicing. Extensive research has revealed that about 95% of the human genes are alternatively spliced (Wang et al., 2008; Merkin et al., 2012). The most prominent example of alternative splicing is the alternative splicing of cassette exons (CE), which results either in inclusion or skipping of the respective exons (Wang et al., 2008). Other types of alternative splicing include alternative intron retention (IR), alternative usage of 5'ss or 3'ss (A5SS or A3SS), alternative first or last exons (AFE or ALE) and the splicing of mutually exclusive exons (ME) (Figure 2).

While most types of alternative splicing are accepted as physiological mechanisms in humans, the importance of IR events was under discussion for a long time. However, research has revealed that IR events can regulate transcript expression post-transcriptionally (reviewed in Monteuuis et al., 2019). IR often associates with the regulation of cell-type and differentiation specific expression profiles (Braunschweig et al., 2014). In this context, IR mainly occurs in lowly expressed and physiologically not required transcripts (Braunschweig et al., 2014). For example, to regulate transcript expression in the context of development, premature termination codons (PTC) that are encoded in the retained introns can lead to transcript degradation via nonsense-mediated decay (NMD) (Braunschweig et al., 2014).

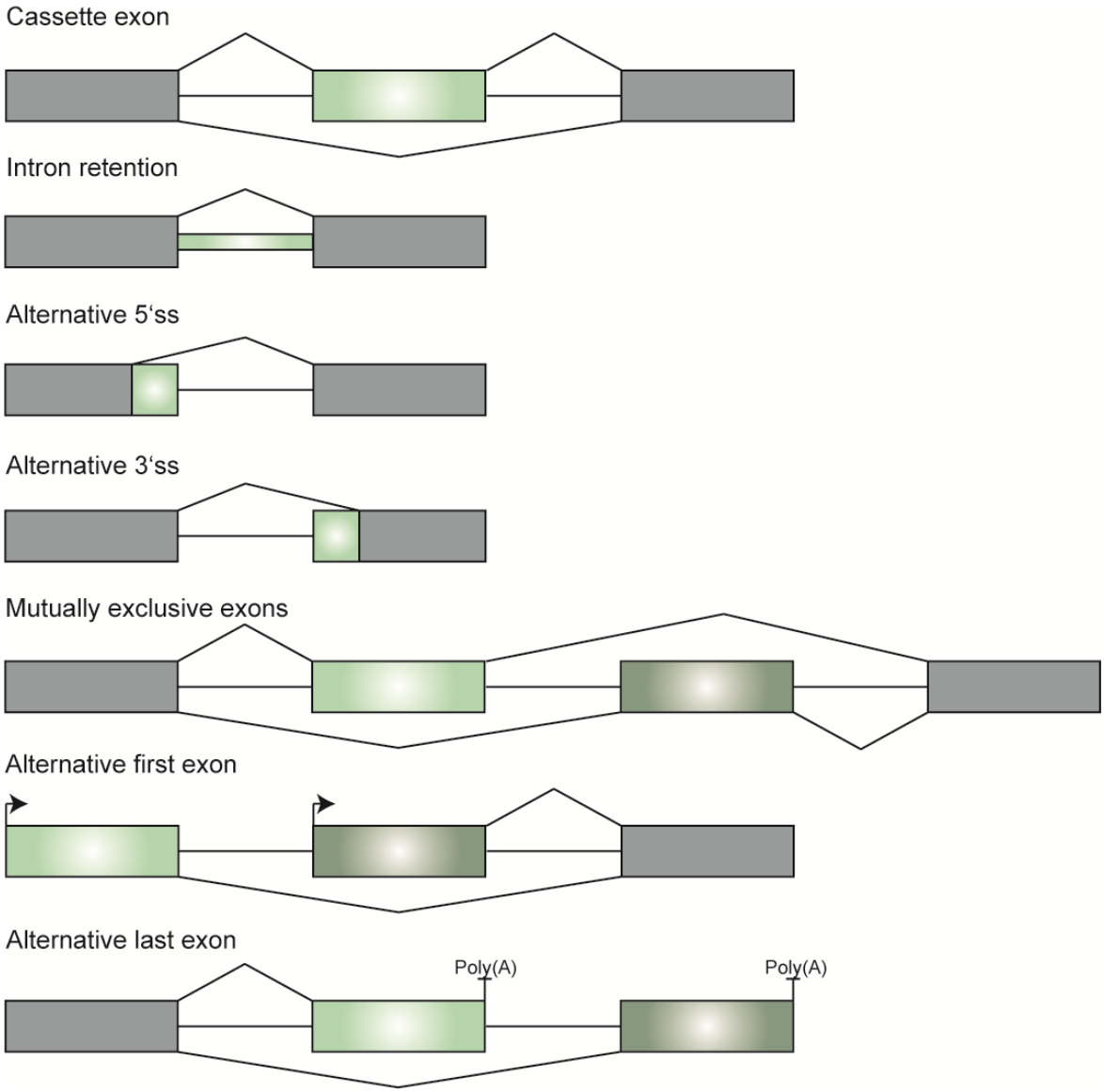

Figure 2: Types of alternative splicing.
Constitutive splicing removes non-coding introns from pre-mRNAs to generate mature mRNAs. Besides the constitutive splicing, the process of alternative splicing can produce various different mRNAs from one pre-mRNA sequence based on individual splicing decisions. Alternatively spliced sequences are shown in light or dark green. Adapted from Park et al., 2018.

1.2 *Cis*-regulatory elements

Cis-regulatory elements act as binding platforms for *trans*-acting factors and can be classified into exonic or intronic splicing enhancers (ESE or ISE) and exonic or intronic splicing silencers (ESS or ISS). A *trans*-acting factor that binds to a splicing silencer negatively effects the splice site usage by the spliceosome (reviewed in Wang and Burge, 2008). In contrast, binding of a *trans*-acting factor to a splicing enhancer increases the splice site usage by the spliceosome (reviewed in Wang and Burge, 2008). *Cis*-regulatory elements can act as binding platforms for *trans*-acting factors either via the sequence composition or via RNA structures (reviewed in Wang and Burge, 2008).

1.2.1 Long-range RNA structures as *cis*-regulatory elements

RNA structures can interact with *trans*-factor binding through different mechanisms. The RNA structure itself can be recognised, the structure can sequester binding sites of single-strand binding *trans*-acting factors or the structure can mediate interactions of *trans*-acting sites that are in large distance. In the last years, the importance of RNA structures for splicing regulation has been extensively studied. Several studies have described the importance of long-range interactions, for example via stem loops, to bring the 5'ss and the 3'ss into close proximity (Muh et al., 2002; Baraniak et al., 2003). Especially for eukaryotes that have very long introns, the mechanism of long-range interactions via RNA structures seems to be important to maintain splicing (Raker et al., 2009). Another fascinating and extensively studied example of long-range interactions is the mutually exclusive splicing of the *drosophila* pre-mRNA *Dscam1* (Xu et al., 2019). *Dscam1* contains an exon 6 cluster that consists of 48 mutually exclusive exons 6, from which always only one remains for the final mRNA sequence. Research has revealed that 66 highly conserved nt in the downstream intron of the constitutive exon 5 act as a docking site for long-range RNA interactions to mediate mutually exclusive splicing of the *drosophila Dscam1* exon 6 cluster (Graveley, 2005). Additionally to this docking site, conserved selector sequences have been identified upstream of each individual exon 6 of the exon 6 cluster (Graveley, 2005). Computational predictions revealed that the docking site could interact with each selector sequence (Graveley, 2005). However, due to the sequence similarity of the selector sequences, the docking site cannot interact with several selector sequences at a time (Graveley, 2005). Experimental validation of these docking site:selector sequence interactions is challenging. Nevertheless, long-range interactions between the docking site and selector sequences have been experimentally confirmed (May et al., 2011). The example of *Dscam1* mutually exclusive splicing highlights the potential of RNA structures to communicate complex alternative splicing decisions.

1.2.2 G-quadruplex RNA structures as *cis*-regulatory elements

A well-known local RNA structure is the RNA G-quadruplex (rG4) (Figure 3). rG4s can form from guanine (G)-rich RNA sequences that consist of several G-tracts interspersed by short loops. $G_{(3-6)}N_{(1-7)}G_{(3-6)}N_{(1-7)}G_{(3-6)}N_{(1-7)}G_{(3-6)}$ describes the consensus sequence of an rG4 that is built from four G-tracts (reviewed in Fay et al., 2017). Together, four Gs, one from each G-tract, form a planar cyclic arrangement, the G-quartet, via Hoogsten hydrogen bonds

(Figure 3). The G-quartets stack on each other and thereby form the final rG4 structure (Figure 3). Monovalent cations can intercalate in the cyclic G-quartet or between stacking G-quartets and thereby stabilise the rG4 (reviewed in Fay et al., 2017). Since the rG4 has a defined structure, the ionic radius of the monovalent cation strongly impacts to which extend the cation stabilises the rG4. Large cations like potassium (K) can strongly stabilise rG4s, since they can fit between two G-quartets (reviewed in Lane et al., 2008). In contrast, small cations like sodium (Na) can only intercalate in the centre of one G-quartet, which consequently leads to less stabilised rG4s (Guiset Miserachs et al., 2016; Wong and Wu, 2003; reviewed in Fay et al., 2017). Lithium (Li), which has an even smaller ionic radius than Na, does not stabilise rG4s at all (Guiset Miserachs et al., 2016).

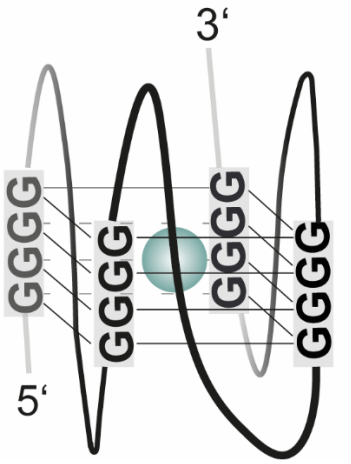

Figure 3: Structure of a parallal-stranded RNA G-quadruplex.
An RNA sequence with four G-tracts interspersed by short loops can form a G-quadruplex structure. The G-quadruplex structure consists of cyclic, planar G-quartets that stack on each other. A cyclic G-quartet is formed via Hoogsten hydrogen bonds and consists of four Gs, one G from each G-tract. To stabilise the G-quadruplex, monovalent cations like potassium intercalate in the centre of the stacking G-quartets.

The existence of rG4s *in vivo* has been extensively discussed in the last years. While the usage of computational approaches has identified a genome-wide distribution of potential rG4s in the human genome, the evidence for the existence of rG4s *in vivo* is contradictory (Huppert and Balasubramanian, 2005; Todd et al., 2005; Kikin et al., 2006). *In vitro*, the folding of rG4s has been demonstrated for individual rG4 sequences, but also on a transcriptome-wide scale. The usage of a Reverse Transcriptase (RT) Stop assay has identified thousands of rG4 sites transcriptome-wide in polyadenylated RNA *in vitro* (Kwok et al., 2016; Guo and Bartel, 2016). The RT-Stop assay relies on the stalling of the RT at sites of RNA structures like rG4s. A coupled next generation sequencing (NGS) of the reversely transcribed complementary DNAs (cDNA) can identify the positions of RT

stalling. Interestingly, the performance of an RT-Stop assay with RNA from cells that have been treated with dimethyl sulfate (DMS) could not confirm rG4 structure formation *in vivo* (Guo and Bartel, 2016). DMS can penetrate into living cells and modify unfolded RNA sequences, which inhibits a potential re-folding of rG4s *in vitro* during the RT-Stop assay. These high-throughput experiments have confirmed the existence of rG4 structures *in vitro*, but not *in vivo*. In contrast, immunofluorescence-staining using an rG4-detecting antibody revealed cytoplasmic staining of rG4s in cells (Biffi et al., 2014). The cytoplasmic staining with the rG4-detecting antibody was sensitive to ribonuclease treatment, whereas treatment with an rG4-stabiliser resulted in an increased detection of rG4s in the cytoplasm (Biffi et al., 2014). In summary, computational approaches and *in vitro* analyses have confirmed the existence of thousands of rG4s in the human transcriptome. However, whether rG4s can form *in vivo* or not is still under discussion.

Although it is still unclear, whether rG4s can form *in vivo*, rG4s have been described to be involved in all post-transcriptional processes. Of specific interest for this work is the function of rG4s in splicing regulation. In the last years, several publications have shown that rG4s can regulate alternative splicing. For example, Marcel et al. proposed the importance of an intronic rG4 for alternative splicing of the tumour suppressor gene p53 (*TP53*) (Marcel et al., 2011). With the usage of different experimental approaches like mutational constructs or rG4-stabilising experiments, the authors could show that the identified intronic rG4 in the *TP53* pre-mRNA is important for the excision of an upstream-located intron (Marcel et al., 2011). Another study has identified and described an rG4 in an intron of the cell adhesion molecule *CD44* (Huang et al., 2017). Using a combination of mutational constructs and an rG4-destabiliser, the authors reported that their described rG4 structure is required for the inclusion of an upstream-located alternative exon (AE) of *CD44* (Huang et al., 2017). The treatment of human cells with a small molecule that destabilises rG4s has highlighted the transcriptome-wide importance of rG4s for alternative splicing regulation (Zhang et al., 2019). Destabilisation of rG4s resulted in a transcriptome-wide differential splicing regulation and further computational analyses revealed an enrichment of potential rG4s within the differentially spliced events (Zhang et al., 2019).

1.2.3 Targeting *cis*-regulatory elements with antisense oligonucleotides

Studies of the recent years have shed light on the importance of splicing-dysregulating mutations (Xiong et al., 2015; reviewed in Abramowicz and Gos, 2018).

Splicing-dysregulating mutations are nucleotide variants that are located in *cis*-regulatory elements like the splice sites or splicing enhancer/silencer sequences. These mutations disrupt the splicing code by either affecting the binding of regulatory proteins or the formation of RNA structures. The difficulty in the detection and evaluation of splice-disrupting variants is due to their location: such variants can locate in deep intronic sequences, which are often not sequenced in the context of a diagnosis. Further, studies about the functions of intronic *cis*-regulatory elements are rare. Bioinformatical algorithms are the main tool to evaluate detected variants that are potentially disrupting splicing (Xiong et al., 2015; reviewed in Abramowicz and Gos, 2018).

The increased interest in splicing-disrupting mutations has started a new era in the development of small molecule drugs, the antisense oligonucleotides (ASO), to target splicing dysregulation. ASOs are short synthetic DNA or RNA sequences that are designed to bind to defined *cis*-regulatory elements within pre-mRNAs. Two different targeting approaches exist: (1) the ASO anneals to a splicing enhancer element blocking trans-factor binding, which leads to artificial exon skipping, (2) the ASO binds to a splicing silencer element and thereby induces artificial exon inclusion (reviewed in Havens and Hastings, 2016). In both scenarios, the ASOs induce a splicing switch, which is why they are often referred to as splice-switching ASOs. Many researchers have focused on the design and development of ASOs to treat diseases. The most famous examples for therapeutically used ASOs are ASOs for the therapy of Spinal Muscular Atrophy (Hoy, 2017) and Duchenne Muscular Dystrophy (Syed, 2016; Heo, 2020; Komaki et al., 2020; Lee et al., 2017) that have been successfully tested in clinical studies.

For the treatment of Spinal Muscular Atrophy, the ASO drug Nusinersen has been designed, which anneals to a splicing silencer element of the *SMN2* pre-mRNA. *SMN2* is a paralogue of the *SMN1* gene. Homozygous deletions of *SMN1*, mainly of exon 7, are the genetic cause of Spinal Muscular Atrophy (Lefebvre et al., 1995). However, in case of a non-functional *SMN1* gene, the paralogue *SMN2* cannot take over the physiological function. The *SMN2* gene carries a single nucleotide substitution variant in exon 7 that goes in hand with the expression of *SMN2* exon 7 skipping transcripts encoding for a truncated, unstable protein (Lorson et al., 1999; Monani et al., 1999; Cartegni and Krainer, 2002; Cartegni et al., 2006; Kashima and Manley, 2003; Cho and Dreyfuss, 2010). To overcome the non-functional *SMN2* transcript expression, binding of the ASO drug Nusinersen to a splicing silencer close

to *SMN2* exon 7 induces *SMN2* exon 7 inclusion, restoring properly spliced and functional *SMN2* transcripts (Hoy, 2017).

In contrast to the induction of exon inclusion via Nusinersen for the treatment of spinal musculuar atrophy, ASOs to target Duchenne Muscular Dystrophy are designed to induce exon skipping. Mutations in the *DMD* gene encoding for the protein dystrophin are the leading cause of Duchenne Muscular Dystrophy. The most prominent mutations in the *DMD* gene causing Duchenne Muscular Dystrophy are deletions, duplications or nonsense mutations that result in reading frame shifts or the occurrence of PTCs (Aartsma-Rus et al., 2006). To compensate for the reading frame shift, ASOs have been designed that induce the skipping of defined *DMD* exons restoring the original reading frame (Syed, 2016; Heo, 2020; Komaki et al., 2020; Lane et al., 2008). This results in the expression of a partially functional dystrophin protein, which improves the disease outcome.

1.3 *Trans*-acting factors: RNA-binding proteins

Alternative splicing relies on a complex interplay of *cis*-regulatory elements and RBPs, the *trans*-acting factors, which bind to these motifs. Hence, research of the last decades has focussed extensively on the discovery and decoding of functional mechanisms of RBPs in alternative splicing regulation. Besides the regulation of alternative splicing, RBPs are also involved in all other steps of mRNA processing like polyadenylation, mRNA transport, translation or decay. The most studied and best-described RBPs belong to the serine and arginine-rich (SR) proteins and the heterogeneous ribonucleoproteins (HNRNP).

In humans, the SR protein family comprises twelve members. They are characterised by one or two RNA-binding domains of the RNA recognition motif-type (RRM) and a C-terminal domain consisting of repeated arginine-serine (RS) dipeptides. The RS domain is important for protein-protein interactions but can also be involved in nuclear import. SR proteins mostly bind to ESEs and thereby guide spliceosome assembly (reviewed in Howard and Sanford, 2015).

1.3.1 Heterogeneous nuclear ribonucleoproteins

The HNRNP family consists of 16 main members that have been described to crosslink to pre-mRNAs. HNRNPs are functionally diverse, which goes hand in hand with the four different HNRNP-annotated RNA-binding domains: RRM, quasi-RRM (qRRM), RGG box

and KH domain (reviewed in Geuens et al., 2016). The RRM is the most common RNA-binding domain of HNRNPs and consists of four β-sheets and two α-helices (reviewed in Geuens et al., 2016). Additionally, the RRM contains two highly conserved RNP sequences. The RNP1 contains eight and the RNP2 six mostly aromatic and positively charged amino acids (Adam et al., 1986; Swanson et al., 1987; Dreyfuss et al., 1988). Besides the RNA-binding domains, HNRNPs further consist of auxiliary domains that are often glycine-rich (Dreyfuss et al., 2002; reviewed in Geuens et al., 2016).

1.3.2 Heterogeneous nuclear ribonucleoprotein family H/F

Within the HNRNP family, the HNRNP H/F subfamily exists. The HNRNP H/F subfamily comprises the five proteins HNRNPH1, HNRNPH2, HNRNPF, HNRNPH3 and GRSF-1. The amino acid sequence homology of the family members varies between 44 to 96% (Schaub et al., 2007). HNRNPH1 and HNRNPH2 share with 96% the highest similarity on amino acid sequence level (Schaub et al., 2007). In the following work, they are both referred to as HNRNPH unless otherwise stated. While HNRNPH1, HNRNPH2, HNRNPF and GRSF-1 all contain three RNA-binding domains and two auxiliary glycine-rich domains, HNRNPH3 encodes only two RNA-binding domains (Figure 4). The middle glycine-rich domain is called the GYR-domain (rich in the amino acids glycine, tyrosine, arginine) and the C-terminal glycine-rich domain is referred to as the GY-domain (rich in the amino acids glycine, tyrosine) (Figure 4). Both auxiliary domains communicate protein-protein interactions, while the GYR-domain additionally contains a nuclear localisation signal (NLS) that is important for shuttling of the proteins (van Dusen et al., 2010; Gueroussov et al., 2017).

The RNA-binding domains of the qRRM-type are characteristic for the HNRNP H/F family. In contrast to the classical RRMs, RNA-binding domains of the qRRM-type show no conservation of the RNP sequences RNP1 and RNP2 (Honoré et al., 1995; Dominguez and Allain, 2006). However, similar to other RRMs like the RRM of HNRNPA1, the qRRMs of the HNRNP H/F family recognise Gs of G-tract containing nucleotide sequences (Matunis et al., 1994; Burd and Dreyfuss, 1994; Caputi and Zahler, 2001). Despite this similarity, the mode of recognition of the Gs is different between classical RRMs and qRRMs. While the RRM of HNRNPA1 only recognises the first G of a G-tract, the qRRMs of HNRNPF form a molecular cage around G-tracts specifically binding to all Gs (Dominguez et al., 2010).

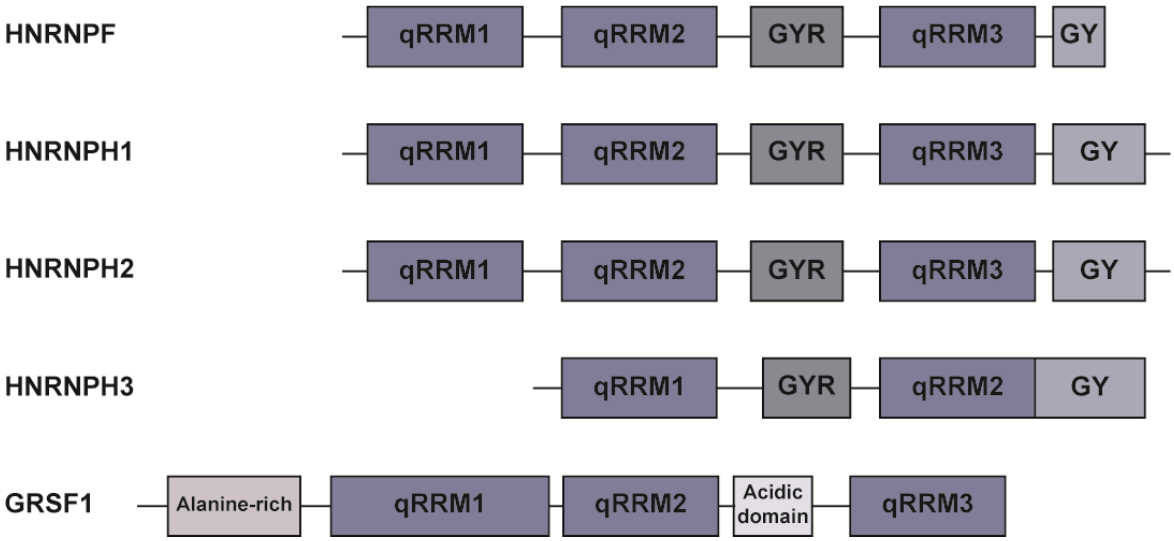

Figure 4: Protein domain structure of the HNRNP H/F protein family.
The HNRNP H/F protein family consists of quasi RNA recognition motif (qRRM) RNA-binding domains, a glycine/tyrosine/arginine-rich protein domain (GYR) and a glycine/tyrosine-rich protein domain (GY). Adapted from Brownmiller and Caplen, 2023.

1.3.3 RNA G-Quadruplex structures and HNRNPH binding

A qRRM of the HNRNP H/F family binds to an RNA G-tract and HNRNP H/F family members contain up to three of those qRRMs (Dominguez et al., 2010). With its consensus sequence $G_{(3-6)}N_{(1-7)}G_{(3-6)}N_{(1-7)}G_{(3-6)}N_{(1-7)}G_{(3-6)}$, rG4s can theoretically encode four HNRNPH binding motifs. The overlap of the HNRNPH binding motif with the consensus sequence of rG4s led to the idea that rG4 structures are involved in HNRNPH-mediated RNA processing.

With respect to the idea of rG4s being involved in HNRNPH-mediated splicing regulation, the binding of HNRNPH to rG4s in RNA regulation has been studied extensively. For example, the GGGGCC repeat expansion of the *C9ORF72* gene associated with amyotrophic lateral sclerosis (ALS) can form rG4s and these rG4s sequester HNRNPH (Conlon et al., 2016). HNRNPF was described to bind an rG4 structure in the mRNA sequence of *CD44* (Huang et al., 2017). In contrast, in combination with RNA helicases, both HNRNPH and HNRNPF prefer to bind to unfolded G-tracts (Dardenne et al., 2014; Herviou et al., 2020). Overall, the question of whether HNRNPH binds to the rG4 structure itself or prefers the unfolded G-tract is still controversial.

Independent of the ambiguity of folded or unfolded rG4 binding, many studies have proposed a correlation between rG4s and the alternative splicing regulation by HNRNPH (Dardenne et al., 2014; Conlon et al., 2016; Neckles et al., 2019).

1.3.4 HNRNPH in alternative splicing

HNRNPH is involved in different aspects of RNA biology like translation or telomere maintenance (Herviou et al., 2020; Xu et al., 2020). Independent of these diverse functions, HNRNPH's best-described biological role is its function as a *trans*-acting factor in alternative splicing. During the last decades, several studies have described HNRNPH and other members of the HNRNP H/F family as regulators of AE inclusion or skipping. To regulate alternative splicing, HNRNPH binds to G-tracts with GGG triplets as the minimum binding motif (Caputi and Zahler, 2001; Dominguez et al., 2010). Most studies have described the function of HNRNPH as a splicing repressor. Binding of HNRNPH to motifs within the AE itself determines AE skipping of transcripts such as rat *β-tropomyosin*, *TCF3*, *MADD* or *MST1R* (Chen and Helfman, 1999; Lefave et al., 2011; Yamazaki et al., 2018; Braun et al., 2018). Nevertheless, investigating for example *PLP* alternative splicing or mouse *c-src* splicing highlighted that HNRNPH can also act as a splicing enhancer, which is accompanied by binding of HNRNPH to intronic sequences (Chou et al., 1999; Wang et al., 2007).

1.3.5 HNRNPH-mediated switch-like splicing regulation

Work on *MST1R* exon 11 splicing has revealed a phenomenon of HNRNPH-mediated splicing: A switch-like splicing regulation. The study highlighted that *MST1R* exon 11 contains a cluster of HNRNPH binding sites (Braun et al., 2018). Using mutational approaches, the authors observed that with the mutation of one HNRNPH binding site the binding of HNRNPH to the surrounding binding sites was also disturbed (Braun et al., 2018). Hence, the conclusion of this study was that HNRNPH binds cooperatively to *MST1R* exon 11 to mediate a switch-like splicing regulation (Braun et al., 2018). The switch-like regulation presented itself in such a way that just a little variation in the cellular HNRNPH protein concentration was sufficient to change the splicing of *MST1R* exon 11 (Braun et al., 2018). A switch-like regulated splicing event is more sensitive to changes in the HNRNPH protein expression than a splicing event without switch-like regulation. Hence, switch-like splicing is a very interesting phenomenon, especially in the aspect of diseases with dysregulated HNRNPH protein concentrations like cancer or neurodegenerative diseases (Lefave et al., 2011; Li et al., 2016; Conlon et al., 2016).

1.4 Cooperation in cellular processes

Switch-like responses are very attractive mechanisms for cells to achieve a fast and strong signalling response. Within a switch-like response, just a little change in the signal concentration is sufficient to activate a large change of the target response. One of the best-studied switch-like regulations is the cooperative binding of ligands to allosteric proteins and the probably best-known example is the cooperative binding of oxygen to haemoglobin. Haemoglobin is an allosteric protein consisting of four subunits each carrying one binding site for oxygen. Already in 1904, Christian Bohr observed that oxygen-haemoglobin binding (ligand-protein complex), plotted against the oxygen partial pressure (ligand), follows a sigmoidal binding curve rather than the, so far, classically known hyperbolic binding curve (Bohr et al., 1904). Considering a fixed ligand concentration range, a classical hyperbolic binding curve cannot reach an as strong ligand-protein complex formation as a sigmoidal binding curve (Figure 5). Inspired by the observation of a sigmoidal haemoglobin-oxygen binding, many researchers have worked on this phenomenon and thereby showed that binding of one oxygen molecule increases the affinity of other oxygen molecules to bind to haemoglobin (Adair et al., 1925). When the binding of one ligand increases the affinity for the binding of other ligands, this phenomenon reflects positive cooperation. However, binding of one ligand can also lower the affinity for the binding of other ligands, which then refers to negative cooperativity. For example, referring again to the observations of Christian Bohr in 1904, he described that binding of carbon dioxide to haemoglobin leads to a reduced affinity for the binding of oxygen molecules (Bohr et al., 1904).

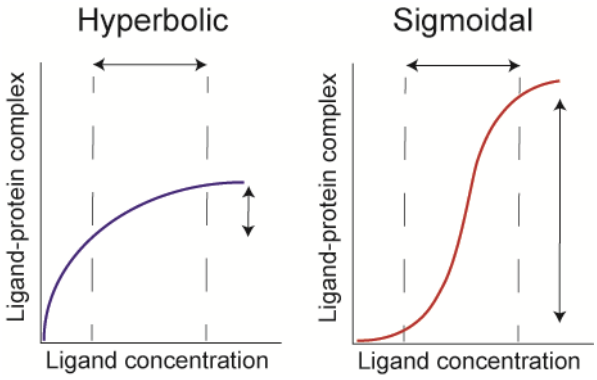

Figure 5: Schematic illustration of hyperbolic (blue) and sigmoidal (red) ligand-protein binding curves.

1.4.1 The Hill equation allows to measure cooperativity

Sigmoidal dose-response curves, which are characteristic for switch-like regulative processes, are often described by the Hill equation (Somvanshi and Venkatesh, 2013).

$$Y = \left(\frac{I^{nH}}{K_{0.5}^{nH} + I^{nH}}\right)$$

Applying the Hill equation for cooperative ligand-protein binding, *Y* is the concentration of the ligand-protein complex and *I* the free ligand concentration. Besides these two variables, the Hill equation additionally contains two parameters: The Hill coefficient (nH) and the half-saturation constant ($K_{0.5}$). These two parameters can be extracted via linearization of the Hill equation (Somvanshi and Venkatesh, 2013).

$$\ln\left(\frac{Y}{1-Y}\right) = nH \ln I - nH \ln K_{0.5}$$

Having extracted the Hill coefficient from the Hill equation, the Hill coefficient can be used as a measure for cooperativity. The Hill coefficient reflects the steepness of the respective dose-response curve. A hyperbolic, non-cooperative dose-response curve has a Hill coefficient nH = 1 (Somvanshi and Venkatesh, 2013). Positive cooperativity, described by sigmoidal dose-response curves, has a Hill coefficient nH >|1|. In contrast, negative cooperativity is described by Hill coefficients nH <|1| (Somvanshi and Venkatesh, 2013).

1.4.2 Cooperation in RNA metabolism

Cooperation has been described in different aspects of RNA biology like RBP binding to microRNAs, nuclear export of viral RNAs or the formation of high-order complexes to control alternative splicing (Daugherty et al., 2010; Xue et al., 2013; HafezQorani et al., 2016; Gueroussov et al., 2017). Independent of the mode of action, cooperativity of RNA-protein complexes can be either based on direct or indirect cooperation.

1.4.3 Direct cooperation

Direct cooperation, as the name already implies, comprises direct interactions of proteins. To communicate direct cooperation, two or more proteins interact with each other via protein-protein interaction domains (Figure 6). These direct protein interactions lead to

stabilised RNA-protein complexes. Several studies have described direct cooperation of RBPs to regulate RNA metabolism. Probably one of the best-studied examples is the direct cooperation of RBPs mediated by glycine-rich protein domain of HNRNPs to regulate splicing (Cartegni et al., 1996; Gueroussov et al., 2017). Direct cooperation of RBPs with RNAs is further involved in the nuclear export of HIV RNAs or in the avoidance of aggregate formation of the RBP TDP-43 for which aggregate formation has been described in neurodegenerative diseases (Daugherty et al., 2010; Rengifo-Gonzalez et al., 2021).

1.4.4 Indirect cooperation

Indirect cooperation is not mediated via any direct protein-protein interactions. In contrast, RNA structures are the mediators of indirect cooperation. An RNA structure that overlaps with RBP binding sites consequently hides the RBP binding sites when the RNA structure is folded (Figure 6). Accordingly, when the RNA is unfolded the binding of one protein can prevent re-folding of the RNA, which then indirectly increases the accessibility for other proteins to bind to the RNA (Figure 6; Lin and Bundschuh, 2015; Becker et al., 2019).

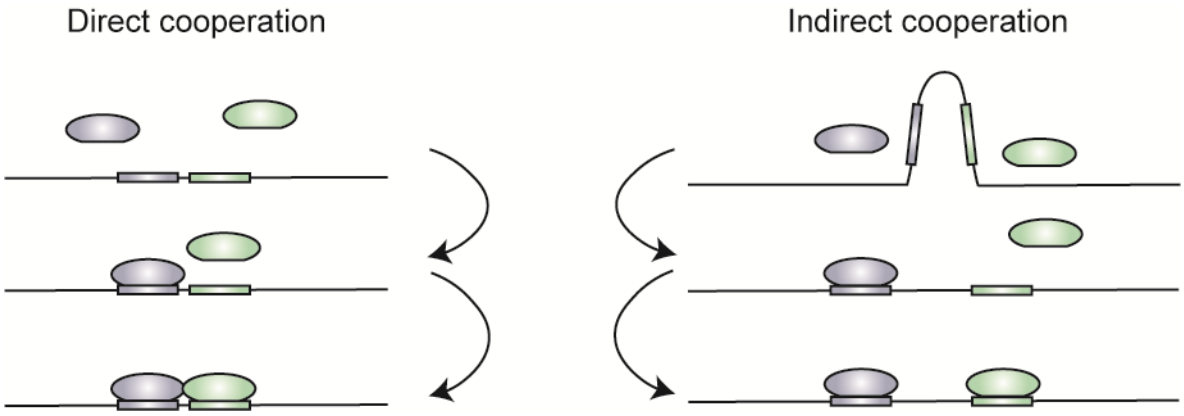

Figure 6: Schematic illustration of direct and indirect cooperation on RNAs.
In case of direct cooperation, the binding of one protein to RNA directly increases the affinity of another protein to bind via protein-protein interactions (left graph). In contrast, RNA structures hide protein binding sites in case of indirect cooperation (right graph). If the RNA is unfolded, the binding of one protein can interfere with RNA-refolding, which then indirectly increases the RNA accessibility for other proteins. Adapted from Becker et al., 2019.

1.5 Aim of the work

The aim of this work is to identify features that control HNRNPH-mediated cooperative splicing regulation, which has been suggested in the literature (Braun et al., 2018). With respect to this aim, it is of special interest to understand if direct or indirect cooperativity controls the HNRNPH-mediated cooperative splicing regulation. In addition, the question of

whether and, if so, how rG4 structures can participate in cooperative splicing will be addressed.

First, an *in vivo* titration of HNRNPH protein levels will be performed, followed by an RNA-Sequencing (RNA-Seq) analysis to characterise HNRNPH-regulated alternative splicing events. The titration of HNRNPH protein levels in combination with the RNA-Seq will allow assessing a potential cooperative splicing regulation by HNRNPH on a transcriptome-wide scale. High-throughput experiments like Individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP), *in vitro* iCLIP and Reverse Transcriptase Stop Profiling will be used to achieve a better understanding of HNRNPH binding sites, especially with respect to rG4 structures. *In vitro* binding studies will further deliver insights into the potential role of rG4 structures for HNRNPH-mediated cooperative splicing regulation. Overall, the study will investigate variables that regulate cooperative splicing regulation by HNRNPH with a specific focus on the function of rG4 structures. Besides understanding the cooperative splicing mechanism, the study will assess the potential of ASOs to control HNRNPH expression in pathophysiological conditions like breast cancer. In sum, this PhD thesis is expected to unravel and explain an HNRNPH-mediated cooperative splicing regulation on a transcriptome-wide scale. Further, the study aims to suggest a new therapeutic potential for cancer patients by controlling the HNRNPH expression.

2 Materials and Methods

2.1 Materials

2.1.1 Plasmids

Table 1: Plasmids

Plasmid ID	Name	Backbone	Description
pKT1	pcDNA3.1 (+)_EV	pcDNA3.1 (+)	Empty pcDNA3.1 (+) vector
pKT2	HNRNPH1_OE	pcDNA3.1 (+)	Expression vector for *HNRNPH1* (Braun et al., 2018)
pKT4	HNRNPH1_OE_mut-siRNA24	pcDNA3.1 (+)	Expression vector for *HNRNPH1* with mutated siRNA 24 binding site

pKT13	MBP-HNRNPH1-His	pET	*HNRNPH1* with N-terminal MBP-tag and C-terminal His-tag
pKT16	HNRNPH1_del_Cterm_GY-domain_mut-siRNA24	pcDNA3.1 (+)	*HNRNPH1* with deleted C-terminal GY-domain
pKT20	AKR1A1 minigene	pcDNA3.1 (+)	*AKR1A1* minigene
pKT34	AKR1A1 minigene_mutated 3'SS_2	pcDNA3.1 (+)	*AKR1A1* minigene with G549T mutation
pKT35	AKR1A1 minigene_mutated 3'SS_2_mutBS_UI	pcDNA3.1 (+)	*AKR1A1* minigene with mutated HNRNPH binding site in the upstream intron
pKT42	AKR1A1 minigene_mutated 3'SS_2_mutBS_DI	pcDNA3.1 (+)	*AKR1A1* minigene with mutated HNRNPH binding site in the downstream intron
pKT43	AKR1A1 minigene_mutated 3'SS_2_mutBS_UI-DI	pcDNA3.1 (+)	*AKR1A1* minigene with mutated HNRNPH binding site in upstream and downstream intron
pKT52	AKR1A1 minigene_mutated 3'SS_2_noG4-BS-1_DI	pcDNA3.1 (+)	*AKR1A1* minigene with new HNRNPH binding site without G-quadruplex
pKT53	AKR1A1 minigene_mutated 3'SS_2_noG4-BS-2_DI	pcDNA3.1 (+)	*AKR1A1* minigene with new HNRNPH binding site without G-quadruplex
pKT54	AKR1A1 minigene_mutated 3'SS_2_G4-BS-1_DI	pcDNA3.1 (+)	*AKR1A1* minigene with new HNRNPH binding site with G-quadruplex
pKT11	HNRNPH1_EGFP	pEGFP-C1	HNRNPH1 with N-terminal EGFP-tag (gift from Cyril Dominguez)
pL37	pMX-Dest53-IP-GFP	pDEST	Emtpy vector containing GFP

2.1.2 Oligonucleotides

2.1.2.1 PCR primer sequences

Table 2: PCR primer sequences

No.	Name	Sequence (5'-3')	Purpose
oKT79	PIP4K2C_rev	CCACATCTCTCTTCAGCTTCTCC	RT-PCR for PIP4K2C exon 6
oKT78	PIP4K2C_for	TTGTGAAGTGCCATGGCAAC	RT-PCR for PIP4K2C exon 6
oKT59	TEV_His_HNRNPH1_r	AAGCTTAGTGATGGTGATGGTGATGTGCAATGTTTGATTGAAAATCACTGG	Gibson cloning MBP-HNRNPH1-His
oKT58	pMAL_TEV_His_r	CTCCACCTTCCGTGCCCAACATCATGCTTTGAAAATATAAATTTTCCTCG	Gibson cloning MBP-HNRNPH1-His

oKT57	pMAL_TEV_His_f	CATCACCATCACCATCACTAAG	Gibson cloning MBP-HNRNPH1-His
oKT55	SMT3_MAL_HNRNPH1_f	ATGATGTTGGGCACGGAAGG	Gibson cloning MBP-HNRNPH1-His
oKT46	MYL6_rev	CTGACGGCAAACATCATCC	RT-PCR for MYL6 exon 6
oKT45	MYL6_for	GGCATGAGGACAGCAATG	RT-PCR for MYL6 exon 6
oKT39	endo_Ron_validation_r	CTAGCTGCTTCCTCCGCCACCAGTA	RT-PCR MST1R exon 11
oKT38	endo_Ron_validation_f	CCTGAATATGTGGTCCGAGACCCCCAG	RT-PCR MST1R exon 11
oKT35	AKR1A1_rev_ex8_RT-PCR	AGATCACTTTCCGCTGGACC	RT-PCR AKR1A1 exon 7
oKT34	AKR1A1_for_ex6_RT-PCR	TCCACCCACTACAAGGAGAC	RT-PCR AKR1A1 exon 7
oKT336	YBX3_DI_rev	CCACCCGTCAACTCAGGGCTCTCTC	Exchange of the HNRNPH binding site in the downstream intron of AKR1A1 minigene with HNRNPH binding site from YBX3 (Non-G4#2)
oKT335	YBX3_DI_for	TTGGTTAGAGATTTCTTATTTCAGTGTCTG	Exchange of the HNRNPH binding site in the downstream intron of AKR1A1 minigene with HNRNPH binding site from YBX3 (Non-G4#2)
oKT334	SYTL2_DI_rev	CAAATCACCCCAGTCAACTCAGGGCTCTCTC	Exchange of the HNRNPH binding site in the downstream intron of AKR1A1 minigene with HNRNPH binding site from SYTL2 (Non-G4#1)
oKT333	SYTL2_DI_for	GGCAGTCTCCGGGTTCAGTGTCTGAGTGAGG	Exchange of the HNRNPH binding site in the downstream intron of AKR1A1 minigene with HNRNPH binding site from SYTL2 (Non-G4#1)
oKT331	SNHG21_rev	GGGACAGATTTACCTCAAGG	RT-PCR for SNHG21 exon 3
oKT330	SNHG21_for	TTGCAGTATAGCCACTGAAG	RT-PCR for SNHG21 exon 3
oKT209	AKR1A1_mini_ver3_3SS_rev	CCCATCACCATGAAGTGG	Single point mutation G549T
oKT208	AKR1A1_mini_ver3_3SS_for	TTATTCTTTGTCTCAGGTGGAATG	Single point mutation G549T

oKT195	AKR1A1_mini_UImut_rev	ACTCCTCCACTCTTCTTCAGCCTGCTTTTCTCACCATTTCTCATCGGCAATTTCATTCTCTCC	Mutation of the HNRNPH binding site in the upstream intron
oKT194	AKR1A1_mini_UImut_for	CAGCAATAGAAAGTGGTCCAGACATGCATGTCTGAGATGAGCCAAGCAAGCTGGGTGTC	Mutation of the HNRNPH binding site in the upstream intron
oKT193	AKR1A1_mini_DImut_rev	CCTCTCAACTCAGGGCTCTCTCTCTAAGACTGCTTCATACCTGAGCAAGATCTGAG	Mutation of the HNRNPH binding site in the downstream intron
oKT192	AKR1A1_mini_DImut_for	CAAGAGTTAAGAGATTTCTTATTTCAGTGTCTGAGTGAGGCTGAGGATCTTGC	Mutation of the HNRNPH binding site in the downstream intron
oKT187	AKR1A1_minigene_rev	AGATGGCTGGCAACTAGAAGG	Reverse primer for RT-PCR of pcDNA3.1 (+) based minigenes
oKT176	AKR1A1_mini_XbaI_rev	ACGACTTCTAGACCAAGAAAATGTCCCAAAGAGC	AKR1A1 minigene cloning
oKT175	AKR1A1_mini_HindIII_for	TAAGCAAAGCTTGTAGAAGCCTGGCATTTGTGTC	AKR1A1 minigene cloning
oKT129	Actin_rev_qPCR	TGGAGTTGAAGGTAGTTCGTG	qPCR analysis Actin
oKT128	Actin_fwd_qPCR	TCCTCCCTGGAGAAGAGCTAC	qPCR analysis Actin
oKT 93	H1_GYdel_new_rev	TGCTGTAGAATTCAAGAAGAGTTCTACATATCTGTGTTGCAT	Deletion of C-terminal GY-domain of HNRNPH1
oKT 92	H1_GYdel_new_for	TAGTCTAGAGGGCCCTATTCTATAGTGTCACC	Deletion of C-terminal GY-domain of HNRNPH1
	RToligo	GGATCCTGAACCGCT	Primer for reverse transcription for iCLIP, invitro iCLIP, RT-Stop profiling
	P5Solexa_s	ACACGACGCTCTTCCGATCT	Pre-amplification PCR iCLIP, invitro iCLIP
	P3Solexa_s	CTGAACCGCTCTTCCGATCT	Pre-amplification PCR iCLIP, invitro iCLIP
	P5Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	Final library PCR iCLIP, invitro iCLIP
	P3Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT	Final library PCR iCLIP, invitro iCLIP
oS153	hnRNPH1_qPCR_f	GTACACATGCGGGGATTACC	qPCR analysis HNRNPH1
oS154	hnRNPH1_qPCR_r	CGAACTCGACATCTGCTTCA	qPCR analysis HNRNPH1

oKT74	Oligo_MYL6_WT_rev	GAGCCGTACCCG	Primer for PCR amplifcation of ssDNA oligo MYL6_WT
oKT40	T7_promoter_for	TAATACGACTCACTATAG	Primer for PCR amplification of ssDNA with 5' T7-promotor
oKT41	L3-Linker_Rev	GCTGAACCGCTCTT	Primer for PCR amplification of ssDNA with 3' L3-Linker
oKT271	MYL6_mut_rev	GGAGTCGTACATGTAAC	Primer for PCR amplifcation of ssDNA oligo MYL6 mutant

2.1.2.2 Oligonucleotides for *in vitro* transcription

Table 3: Synthetic ssDNA oligonucleotides

No.	Name	Sequence (5'-3')	Purpose
oKT122	ARPC2.von.Hacht	TAATACGACTCACTATAGTAAATAGATTTGCGTTACTGTCTAAGGGGGCTGGGCGGGGACCGGGTTTTTCTTTTATTTTCTTTTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAACCCCCTGCCCCCATCCCTGCAGCTTGGCCAGGATGGTGCTGCGCGGAAACCGACAGATCGGAAGAGCGGTTCAGC	ssDNA oligo encoding ARPC2.von.Hacht G-quadruplex sequence with 5' T7-promotor and 3' L3-Linker
oKT225	CD46	TAATACGACTCACTATAGTAAAATCCTTCAAATTTCTGGTGATGTTCACTTATTTTTAGTAATGAAAGGAGGGAGGACATTTGTATAGCCAAGATCGGAAGAGCGGTTCAGC	ssDNA oligo encoding HNRNPH binding site from CD46 with 5' T7-promotor and 3' L3-Linker
oKT72	MYL6 WT	TAATACGACTCACTATAGTTTGTGAGGCATATCCTGTCGGGGTGACGGGCCCATGGGGCGGGTACGGCTCC	ssDNA oligo encoding HNRNPH binding site from MYL6 with 5' T7-promotor
oKT272	MYL6 mutant	TAATACGACTCACTATAGTTTGTAAGACATATCCTGTCTATCTGACGTACCCATGTTACATGTACGACTCC	ssDNA oligo encoding mutated HNRNPH binding site from MYL6 with 5' T7-promotor

2.1.2.3 RNA oligonucleotides

Table 4: Synthetic RNA oligonucleotides

Name	Sequence
MYL6	UGUCGGGGUGACGGGCCCAUGGGGCGGGUACGGCUCC
ARPC2.von.Hacht	UCUAAGGGGGCUGGGCGGGGACCGGGUUU
CD46	UAGUAAUGAAAGGAGGGAGGACAUU
FAM-ARPC2.von.Hacht	FAM-AGGGGGCUGGGCGGGGACCGGGU
FAM-TAMRA-ARPC2.von.Hacht	FAM-AGGGGGCUGGGCGGGGACCGGGU-TAMRA

2.1.3 siRNAs

Table 5: siRNA sequences

No.	Target	Sequence
sR18	no target	UGGUUUACAUGUCGACUAA[dT][dT]
sR24	*HNRNPH1*	GGAGCUGGCUUUGAGAGGA[dT][dT]

2.1.4 ASOs

Table 6: Antisense RNA oligonucleotides

Number	Sequence	Purpose
ASO2	CCUUUUCAGCUAUUUCCUGUGA	ASO to inhibit inclusion of HNRNPH1 exon (Chr5:179,619,407-179,619,269), Skipping isoform undergoes NMD
Ctrl ASO2	CAAUCCGCAAUCCGCAAUCC	FAM-labelled, scrambled RNA, kind gift from Maria Ines Alvelos (Center for Diabetes Research, Université Libre de Bruxelles)

2.1.5 Antibodies

Table 7: List of primary and secondary antibodies

Antigen	Dilution	Origin	Product number	Supplier
HNRNPH	1:10,000	rabbit	AB10374	Abcam
Vinculin	1:10,000	mouse	V9624	Sigma-Aldrich
Anti-rabbit IgG	1:5,000	goat	7074S	Cell Signalling
Anti-mouse IgG	1:5,000	horse	7076S	Cell Signalling

2.1.6 Buffers

Table 8: Modified RIPA buffer (mRIPA)

Component	Final concentration
Tris HCl, pH 7.5	50 mM
NaCl	150 mM
EDTA	1 mM
NP-40	1% (v/v)
Sodium deoxycholate	0.1% (w/v)
+ cOmplete™ EDTA-free Protease Inhibitor Cocktail	1 tablet/ 10 mL
+ Turbo DNase	40 µU/µL

Table 9: Lysis buffer

Component	Final concentration
Tris HCl, pH 7.5	50 mM
NaCl	500 mM
Glycerol	5% (v/v)
Imidazole	10 mM
Triton X-100	0.1% (v/v)
+ cOmplete™ EDTA-free Protease Inhibitor Cocktail	1 tablet/ 10 mL
+ Sm Nuclease	6 cU/µL
+ DTT	1 mM
+ $MgCl_2$	1 mM

Table 10: IMAC-A

Component	Final concentration
Tris HCl, pH 7.5	50 mM
NaCl	500 mM
Glycerol	5% (v/v)
Imidazole	10 mM

Table 11: IMAC-B

Component	Final concentration
Tris HCl, pH 7.5	50 mM
NaCl	500 mM
Glycerol	5% (v/v)
Imidazole	500 mM

Table 12: Gel filtration buffer

Component	Final concentration
Na-HEPES, pH 7.4	30 mM
NaCl	150 mM
Glycerol	5% (v/v)
+ DTT	1 mM

Table 13: 10x binding buffer

Component	Final concentration
HEPES, pH 7.2	100 mM
$MgCl_2$	30 mM
Glycerol	30% (v/v)
DTT	10 mM

Table 14: Ponceau Red staining

Component	Final concentration
Ponceau S	0.1% (w/v)
Acetic acid	5% (v/v)

Table 15: Reverse transcription buffer

Component	Final concentration
Tris HCl, pH 7.5	20 mM
KCl or NaCl	150 mM
$MgCl_2$	3 mM

Table 16: 0.01% PBST

Component	Final concentration
ddH2O	8 L
5x PBS	2 L
Tween 20	1 mL

Table 17: SILAC DMEM Light

Component	Final concentration
DMEM without Arginine and Lysine	445 mL
Dialyzed Fetal Bovine Serum	50 mL
Penicillin/Streptomycin (100x)	5 mL
L-Arginine-0 (84 mg/mL)	250 µL
L-Lysine-0 (146 mg/mL)	250 µL

Table 18: SILAC DMEM Medium

Component	Final concentration
DMEM without Arginine and Lysine	445 mL
Dialyzed Fetal Bovine Serum	50 mL
Penicillin/Streptomycin (100x)	5 mL
L-Arginine-6 (84 mg/mL)	250 µL
L-Lysine-4 (146 mg/mL)	250 µL

Table 19: SILAC DMEM Heavy

Component	Final concentration
DMEM without Arginine and Lysine	445 mL
Dialyzed Fetal Bovine Serum	50 mL
Penicillin/Streptomycin (100x)	5 mL
L-Arginine-10 (84 mg/mL)	250 µL
L-Lysine-8 (146 mg/mL)	250 µL

Table 20: Interactome lysis buffer

Component	Final concentration
Tris HCl, pH 7.5	50 mM
NaCl	150 mM
EDTA	1 mM
NP-40	1% (v/v)
Sodium deoxycholate	0.1% (w/v)
β-Glycerophosphate	5 mM
+ Sodium fluoride	5 mM
+ Na-Orthovandate	1 mM
+ N-Ethylmaleimide	10 mM
+ cOmplete™ EDTA-free Protease Inhibitor Cocktail	1 tablet/ 10 mL

2.2 Methods

2.2.1 Cell culture

MCF7 cells were purchased from ATCC and cultured in Roswell Park Memorial Institute (RPMI) 1640 medium (Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies), 1% L-Glutamine (Life Technologies) and 10 µg/mL Insulin (Sigma-Aldrich). The cells were splitted twice per week. To split the cells, they were washed two times with DPBS (Life Technologies) and then detached with 1.5 mL 1% Trypsin (Life Technologies) for three min at 37°C. After detaching, the cells were collected adding

4.5 mL complete RPMI medium and centrifuged for 3 min at 500xg. Then, the supernatant was carefully removed and the cell pellet was resuspended in 1 mL complete RPMI medium. From this cell suspension, the cells were splitted in a 1:4 ratio in a new cell culture dish.

HEK293T cells were grown in Dulbecco's modified Eagle medium (DMEM; Life Technologies) that was supplemented with 10% fetal bovine serum and 1% L-Glutamine. HEK293T cells were splitted twice per week following the same procedure as described above, except for the usage of complete DMEM medium. HEK293T cells were splitted in a 1:10 ratio in a new cell culture dish.

All cells were kept at 37°C with 5% CO_2 and regularly tested for mycoplasma infection.

2.2.2 Transfection of plasmids

For transfection of plasmids, cells were seeded one day prior transfection with $5*10^5$ cells (MCF7 cells) or $3*10^5$ cells (HEK293T cells) per well in a 6-well plate. On the next day, plasmid transfection was performed using Lipofectamine 2000 (Invitrogen). First, 5 µL Lipofectamine 2000 was mixed in 95 µL OPTI-MEM (Invitrogen). In parallel, 2 µg of the respective plasmid (Table 1) was set up in 100 µL OPTI-MEM. The plasmid mixture was added to the Lipofectamine 2000 approach and the final 200 µL transfection mix was incubated for 20 min at room temperature. Finally, the transfection mix was added to cells dropwise. For HNRNPH1 protein expression, cells were transfected with a pcDNA3.1 (+) *HNRNPH1* overexpression construct, pKT2, which was taken from a previous publication (Braun et al., 2018; Table 1). The cells were harvested 48 h after plasmid transfection. In the case of minigene analyses, cells were harvested 24 h after plasmid transfection. As control, empty pcDNA3.1 (+) vector, pKT1, was transfected (Table 1).

2.2.3 Transfection of siRNAs

RNA interference with single small interfering RNAs (siRNA) was used to achieve protein knockdown (KD). For the transfection of siRNAs, cells were seeded one day prior transfection with $4*10^5$ cells (MCF7 cells) per well in a 6-well plate. 24 h later, the siRNAs were transfected using Lipofectamine RNAiMax (Invitrogen). 5 µL Lipofectamine RNAiMax was combined with 95 µL OPTI-MEM (Invitrogen). At the same time, the siRNA (Table 5) was diluted in 100 µL OPTI-MEM to end up with a final concentration of

20 nM siRNA per 6-well. The siRNA mixture was combined with the Lipofectamine RNAiMax mix, followed by an incubation at room temperature for 20 min. Finally, the 200 µL final transfection mix was added dropwise to the cells that were kept in 1.8 mL growth medium. Cells were harvested 48 h after siRNA transfection.

In case of a combined KD and plasmid transfection, $4*10^5$ cells per well were seeded in a 6-well plate and transfected 24 h later with siRNAs. Another 24 h later, the cells were transfected with 2 µg plasmid using Lipofectamine 2000 as described above (see chapter 2.2.2). The cells were harvested 72 h after seeding.

2.2.4 Transfection of antisense oligonucleotides

ASOs were ordered from Eurogentec and dissolved in nuclease-free water to a final concentration of 100 µM. The ASOs consisted of 2'-O-Methyl RNA bases with phosphorothioate linkage. For the transfection of ASOs, $8*10^5$ MCF7 cells per well were seeded in 6-well plates. On the next day, the cells were transfected with ASOs using Lipofectamine 2000 (Invitrogen). 2 µL ASO (100 µM, Eurogentec) was mixed with 98 µL OPTI-MEM (Invitrogen). In parallel, 5 µL Lipofectamine 2000 (Invitrogen) was set up in 95 µL OPTI-MEM (Invitrogen). The ASO reaction was combined to the Lipofectamine 2000 approach and the complete transfection reaction was incubated at room temperature for 20 min. Finally, the transfection reaction was added to the cells dropwise, resulting in a final ASO concentration of 100 nM.

To test the effect of the ASOs on MCF7 cell proliferation, the cells were transfected with ASOs using Lipofectamine 3000 (Invitrogen). For the proliferation assay, Lipofectamine 3000 (Invitrogen) was used instead of Lipofectamine 2000 (Invitrogen), since a very high transfection efficiency was aimed. 1 µL ASO (10 µM, Eurogentec) was mixed with 9 µL OPTI-MEM (Invitrogen). In parallel, 0.15 µL Lipofectamine 3000 (Invitrogen) was set up in 9.85 µL OPTI-MEM (Invitrogen). The ASO reaction was combined to the Lipofectamine 3000 approach and the complete transfection reaction was incubated at room temperature for 10 min. Finally, the transfection reaction was added to the cells dropwise, resulting in a final ASO concentration of 100 nM.

2.2.5 Proliferation assay

To measure cell proliferation, the Cell Proliferation Kit I (Roche) was used according to the manufacturer's instructions. In brief, $2*10^4$ MCF7 cells per well were seeded in 96-well plates in 100 µL RPMI complete growth medium. On the next day, the cells were transfected with siRNAs or ASOs (see chapters 2.2.3; 2.2.4). 72 h after transfection, the cells were treated with 10 µL MTT Labelling Reagent (Roche) for a period of 4 h. Metabolically active cells cleave 3-(4, 5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide) (MTT) to the water-insoluble formazan salt. Addition of 100 µL Solubilization Solution (Roche) solubilises the formazan salt and after incubation overnight, the solubilised formazan was quantified. Therefore, absorbance at 565 nm (bandwidth: 9 nm) was measured in a Tecan Infinite M200 Pro (Tecan) with a reference wavelength of 660 nm (bandwidth: 9 nm).

2.2.6 Cycloheximide treatment

Cycloheximide (CHX) was used to inhibit NMD of mRNAs. Therefore, MCF7 cells were treated with 30 µg/mL CHX (30 mg/mL, Sigma-Aldrich) for 4 h and then harvested (see chapter 2.2.8). For control, cells were treated with DMSO. To validate inhibition of NMD, an RT-PCR analysis was used. The RT-PCR monitored an alternative splicing event of *HNRNPH1* from which the skipping isoform is an NMD-target (see chapter 2.2.13).

2.2.7 Emetine dihydrochloride Treatment

To investigate the effect of rG4s on alternative splicing in cells, MCF7 cells were treated with emetine dihydrochloride (Merck Chemicals). Emetine diyhdrochloride was dissolved in DMSO. On the first day, $1*10^6$ MCF7 cells per well were seeded in 6-well plates. 24 h later, cells were treated with increasing concentrations of emetine dihydrochloride. Another 24 h later, the cells were harvested and used for RT-PCR analysis.

2.2.8 Cell harvest

Cells were harvested on ice. First, the cells were washed two times with 1 mL ice-cold DPBS. After washing, the cells were harvested in 1 mL ice-cold DPBS using a cell scraper. The harvested cells were centrifuged for 5 min at 0.8xg and 4°C. The supernatant was removed and the cell pellets were stored at -80°C for further usage.

2.2.9 Gibson cloning

The *HNRNPH1* expression construct for recombinant protein production was generated using Gibson cloning. To generate the MBP-HNRNPH1-His_6 expression construct (pKT13), the *HNRNPH1* open reading frame (ORF) was cloned into a pET vector backbone carrying a C-terminal His_6-tag and an N-terminal MBP-tag, a kind gift from the IMB Protein Production Core Facility (Table 1). All primers for Gibson cloning reactions were designed with the NEBuilder Assembly tool (New England Biolabs). The *HNRNPH1* ORF was PCR-amplified from pKT2 using the forward primer oKT55 and the reverse primer oKT59 with Phusion DNA polymerase (New England Biolabs), an annealing temperature of 65°C and an elongation time of 50 sec. The pET vector backbone was prepared for Gibson assembly performing a PCR reaction with the forward primer oKT57 and the reverse primer oKT58 using the Phusion DNA polymerase (New England Biolabs) with an annealing temperature of 59°C and an elongation time of 3 min. Both PCR products were gel-purified from a 1% agarose gel using the QIAquick Gel Extraction Kit (QIAGEN). Gibson reactions were set up mixing in total 100 ng DNA (vector:insert ratio 1:3) with 10 µL of Gibson assembly 2x master mix from the IMB Protein Production Core Facility. The Gibson reactions were incubated at 50°C for 1 h and afterwards transformed in Dh5α *E. coli*.

2.2.10 Cloning

To clone the *AKR1A1* minigene construct (pKT20), the genomic region of interest (Chromosome [Chr.]1 45,567,931 to 45,569,050) was PCR-amplified from human genomic DNA using the Phusion DNA polymerase (New England Biolabs). The primers were extended with either a *Hind*III (oKT175) or an *Xba*I (oKT176) restriction enzyme site. The PCR reaction was performed with Phusion DNA polymerase following the manufacturer's instructions with an annealing temperature of 63°C and 30 PCR cycles. The ~1,100 bp long PCR product was gel-purified from a 1% agarose gel with the QIAquick Gel Extraction Kit (QIAGEN) following the manufacturer's instructions. 1 µg of the PCR product as well as 1 µg of the donor vector pcDNA 3.1 (+) were digested with 20 units (U) *Hind*III-HF and 20 U *Xba*I restriction endonucleases (20,000 U/mL, New England Biolabs) in a total reaction volume of 50 µL with CutSmart Buffer (New England Biolabs). The restriction digestion reaction was incubated at 37°C for 1 h and then the digested products were again gel-purified from a 1% agarose gel. The ligation reaction was performed in 20 µL using 1 µL T4 DNA Ligase (400,000 U/mL, New England Biolabs), a 1:3 vector:insert ratio and

2 µL T4 DNA Ligase Buffer (10x, New England Biolabs). The ligation reaction was performed at 16°C overnight followed by a heat inactivation step at 65°C for 10 min. Finally, 5 µL of the ligation reaction was transformed in 50 µL chemically competent Dh5α *E. coli*.

2.2.11 *AKR1A1* minigene mutagenesis cloning

To achieve a minigene splicing pattern that was comparable to the endogenous *AKR1A1* splicing, the 3'ss of the AE from the *AKR1A1* minigene (pKT34) was strengthened by introduction of one point mutation (G549T) using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's recommendations (forward primer: oKT208, reverse primer: oKT209).

For the splicing analysis of the *AKR1A1* minigene, two HNRNPH binding sites (Chr.1 45,568,349 to 45,568,431 and Chr.1 45,568,702 to 45,568,761), in the upstream and downstream intron, were mutated. Therefore, for each triple G-stretch (GGG) the middle G was substituted and for each quadruple G-stretch (GGGG) the two middle Gs were substituted. The mutations were performed using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's recommendations (Upstream Intron [pKT35]: forward primer: oKT194, reverse primer: oKT195; Downstream intron [pKT42]: forward primer: oKT193, reverse primer: oKT192; Upstream and downstream intron [pKT43]). Then, using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs), new HNRNPH binding sites were cloned into the downstream intron of pKT43. These new HNRNPH1 binding sites either form or do not form an rG4 (Non-rG4 #1 [pKT52]: forward primer: oKT333, reverse primer: oKT334; Non-rG4 #2 [pKT53]: forward primer: oKT335, reverse primer: oKT336; rG4 [pKT54]: forward primer: oKT357, reverse primer: oKT358). Aim of these mutations was, to replace the endogenous HNRNPH binding site of the downstream intron.

2.2.12 *HNRNPH1* expression construct mutagenesis cloning

The binding site of the HNRNPH1-targeting siRNA (sR24) was mutated in the *HNRNPH1* expression construct pKT2. The *HNRNPH1* expression construct carrying silent mutations of the sR24-binding site (pKT4) was generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) following the manufacturer's recommendations with the forward primer oKT16 and the reverse primer oKT17.

To evaluate the effect of the GY-domain for HNRNPH-mediated splicing regulation, the GY-domain was deleted from pKT4 (HNRNPH1 w/o GY: pKT16). The deletion was performed using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) following the manufacturer's recommendations with the forward primer oKT92 and the reverse primer oKT93.

2.2.13 Semiquantitative RT-PCR for splice isoform quantification

Semiquantitative RT-PCR was conducted to quantify splicing patterns of endogenous mRNA or minigene splicing events. Total RNA was isolated from cells with the RNeasy Plus Mini Kit (Qiagen) following the manufacturer's instructions. Then, the reverse transcription reaction was performed in a total volume of 20 µL to generate cDNA. First, 500 ng RNA was incubated with 1 µL Oligo$(dT)_{18}$ primer (100 µM, Thermo Fisher Scientific) in nuclease free water for 5 min at 70°C. Then, 1 µL dNTPs (10 mM, New England Biolabs), 4 µL RT reaction buffer (5x, Thermo Fisher Scientific) and 1 µL RevertAid Reverse Transcriptase (200 U/µL, Thermo Fisher Scientific) was added to the reaction. The final reactions were incubated for 5 min at 25°C, 60 min at 42°C, 10 min at 45°C and 5 min at 70°C.

To quantify splicing patterns, 1 µL of cDNA was used as input for the semiquantitative PCR reaction in a total reaction volume of 20 µL. First, 1 µL cDNA was mixed with 0.5 µL forward primer (10 µM), 0.5 µL reverse primer (10 µM), 0.5 µL dNTPs (10 mM, New England Biolabs), 4 µL OneTaq GC Reaction Buffer (5x, New England Biolabs) and 0.125 µL OneTaq DNA Polymerase (5,000 U/mL, New England Biolabs). PCR conditions for *AKR1A1* splicing event quantification were as following: 94°C for 30 s, 30 cycles (endogenous) or 25 cycles (minigene) of (94°C for 20 sec, 55°C for 30 sec, 68°C for 30 sec) and final extension at 68°C for 5 min. To amplify the *AKR1A1* splicing isoforms the forward primer oKT34 was either combined with the endogenously binding reverse primer oKT35 or with the reverse primer oKT187 binding the pcDNA3.1 (+) backbone upstream of the polyadenylation site. To monitor endogenous *MST1R* splicing, the forward primer oKT38 and the reverse primer oKT39 were used in a 30-cycle PCR reaction with an annealing temperature of 61°C and OneTaq DNA polymerase (New England Biolabs) following the above described PCR conditions.

To validate splicing events from the RNA-Seq analysis of the gradual downregulation and gradual overexpression of HNRNPH in MCF7 cells, RT-PCR was performed for *MYL6*, *PIP4K2C* and *SNHG21* (gradual downregulation from Braun et al., 2018). Therefore, 1 µL of cDNA was used as input for a 30-cycle PCR reaction with OneTaq DNA polymerase (New England Biolabs) (PCR conditions described above using adapted annealing temperatures). *MYL6* was amplified using the forward primer oKT45 and the reverse primer oKT46 with an annealing temperature of 53°C. For amplification of *PIP4K2C*, oKT78 (forward primer) and oKT79 (reverse primer) with an annealing temperature of 56°C were used. *SNHG21* was amplified with oKT330 (forward primer) and oKT331 (reverse primer) and annealing at 51°C.

HNRNPH1 encodes an mRNA isoform that is an NMD-target. Expression of this NMD isoform depends on the skipping of an AE (Chr5:179,619,407-179,619,269) of *HNRNPH1*. To control for NMD, this *HNRNPH1* exon-skipping event can be monitored via semiquantitative RT-PCR. Therefore, 1 µL of cDNA was used as input for a 27-cycle PCR reaction with OneTaq DNA polymerase (New England Biolabs) using the forward primer oKT273 and the reverse primer oKT274. Annealing was performed at 53.5°C.

The TapeStation 2200/4400 capillary gel electrophoresis instrument (Agilent) was used for isoform quantification of the PCR products on D1000 ScreenTapes (Agilent).

2.2.14 Quantitative real-time PCR (qPCR)

For quantitative determination of mRNA levels, 500 ng total RNA was reversely transcribed into cDNA using the RevertAid Reverse Transcriptase (Thermo Fisher Scientific) and Oligo(dT)$_{18}$ primer (Thermo Fisher Scientific) as described in chapter 2.2.13. In accordance to the manufacturer's instruction, qPCR reactions were performed using 10 µL Luminaris HiGreen qPCR Master Mix, low ROX (2x, Thermo Fisher Scientific) with 0.3 µM forward primer, 0.3 µM reverse primer and 2 µL of 1:10 diluted cDNA as template in a total reaction volume of 20 µL. All qPCR reactions were run on a ViiA 7 Real-Time PCR System (Applied Biosystems) in technical triplicates. To quantify *HNRNPH1* mRNA levels the forward primer oS153 and the reverse primer oS154 were used. For normalisation, mRNA levels of the housekeeping gene *actin beta* were determined (forward primer: oKT128, reverse primer: oKT129). The quantification of mRNA

expression levels was performed with the delta-delta Ct method (Livak and Schmittgen, 2001).

2.2.15 *In vitro* oligonucleotide library (ivRNA library)

The sequences included in the oligonucleotide library were designed as following: For CE events, windows for the 3' end of the upstream exon, the 5' end and the 3' end of the upstream and downstream introns, the 5' end and the 3' end of the AE, and the 5' end of the downstream exon were defined. Within these defined windows, regions with HNRNPH binding sites from the HNRNPH iCLIP analysis and regions with predicted rG4s were detected. For intronic regions, the windows were 400 nt long and for exonic regions, the windows spanned 100 nt, while the centre of the exons or introns was never crossed to avoid overlapping of two flanking windows. HNRNPH binding sites with overlapping predicted rG4s were merged into one single region. For each region (HNRNPH binding site, predicted rG4, HNRNPH binding site with predicted rG4), one oligonucleotide construct was designed having the respective region at construct position 51. In the end, each oligonucleotide construct was 200 nt long with the T7 promotor sequence at the 5' end, an oligonucleotide individual barcode sequence at the 3' end and the sequence of the iCLIP2 L3-linker at the 3' end (Buchbender et al., 2020). In total, 1666 regions from HNRNPH-regulated CE splicing events and 1334 regions from not regulated CE splicing events were designed into oligonucleotide constructs. For the final ivRNA library, each oligonucleotide was included in quadruplicates each of it having an individual barcode at the 3' end. The ivRNA library was ordered as a single stranded (ss) DNA oligonucleotide library from TWIST Bioscience as an oligonucleotide pool, Tier 3L ssDNA with 12,000 individual oligonucleotides consisting of 200 nt each.

2.2.16 PCR amplification of ssDNA oligonucleotides

To synthesise RNA *in vitro*, the sequences of interest were ordered as ssDNA oligos from Integrated DNA Technology (IDT) or as an ssDNA oligo pool from TWIST Bioscience. All sequences encoded the T7 promotor sequence at their 5'-ends. The ssDNA was PCR-amplified with Phusion DNA-polymerase to have dsDNA for *in vitro* transcription (PCR conditions: 98°C for 30 sec, 25 cycles [individual ssDNA oligonucleotides] or 10 cycles [ssDNA oligonucleotide library] of [98°C for 30 sec, 50°C for 30 sec, 72°C for 20 sec] and final extension at 72°C for 10 min). The forward primer used for the PCR amplification was always oKT40. The respective reverse primers were oKT41 for

ARPC2.von.Hacht or *CD46* or the ssDNA oligo library, oKT74 for *MYL6* WT and oKT271 for *MYL6* mutant (Table 3). The PCR reactions were set up with 25 ng ssDNA oligo or 10 ng ssDNA oligonucleotide library, 1 U/50 µL Phusion DNA polymerase (New England Biolabs), 1x Phusion GC buffer (New England Biolabs), 200 µM dNTPs (New England Biolabs), 0.5 µM forward primer and 0.5 µM reverse primer in total 50 µL (ssDNA oligonucleotide) or 200 µL (ssDNA oligonucleotide library) reaction volume. The PCR products were purified using a MyONE Silane bead clean-up (see chapter 2.2.17).

2.2.17 MyONE Silane bead clean-up

After PCR amplification, the dsDNA was purified using MyONE Silane beads (Invitrogen). Briefly, 20 µL MyONE Silane beads per sample were magnetically separated and washed with 500 µL RLT buffer (Qiagen). After washing, the beads were again magnetically separated and the supernatant was removed. The beads were resuspended in 250 µL RLT buffer per sample. The beads were added to the PCR reaction and mixed. Then, one volume of 100% ethanol (50 µL PCR product + 250 µL beads: 300 µL 100% ethanol) was added, carefully mixed and incubated for 5 min at room temperature. After incubation, the reaction was again mixed and incubated for 5 min. To remove the supernatant, the beads were magnetically separated and afterwards washed with 1 mL of 80% ethanol. After transferring the samples to a new tube, the beads were washed twice in 1 mL 80% ethanol. Finally, the beads were magnetically separated and the supernatant was removed completely. The beads were air-dried for 5 min at room temperature. Then, the beads were resuspended in 12 µL nuclease-free H_2O and incubated for 5 min at room temperature. The beads were magnetically separated and the eluted dsDNA was transferred to a new tube for *in vitro* transcription. Eluted DNA was quality-controlled using the TapeStation 2200 capillary gel electrophoresis instrument (Agilent) with D1000 ScreenTapes (Agilent).

2.2.18 *In vitro* transcription

For *in vitro* transcription, the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) was used following the manufacturer's recommendation for short transcripts (< 0.3 kb) in a total volume of 40 µL with 500-1000 ng of template dsDNA. The dsDNA was mixed with 3 µL Reaction Buffer (10x, New England Biolabs), 3 µL ATP (100 mM, New England Biolabs), 3 µL CTP (100 mM, New England Biolabs), 3 µL UTP (100 mM, New England Biolabs), 3 µL GTP (100 mM, New England Biolabs) and 3 µL T7 RNA Polymerase Mix (New England Biolabs). To generate RNA with

7-deaza-GTP, 3 µL GTP was replaced with 2.7 µL 7-deaza-GTP (100 mM, TriLink Biotechnologies) and 0.3 µL GTP (100 mM, New England Biolabs).

All *in vitro* transcription reactions were incubated in a thermocycler at 37°C for 16 h. To remove template DNA, the *in vitro* transcription reactions were treated with 2 µL Turbo DNase (2 U/µL, Invitrogen) in 100 µL reaction volumes for 15 min at 37°C. Subsequently, the RNA was purified using the RNeasy Plus Mini Kit (Qiagen) following the manufacturer's instructions for the purification of total RNA containing miRNA. The RNA was eluted in 20 µL nuclease-free water and the length was validated using the TapeStation 2200 system with RNA ScreenTapes (Agilent). RNAs were stored at -80°C.

2.2.19 Western blot analysis

For protein level quantification, cells from 6-well cell culture plates were harvested and lysed in 80 µL ice-cold mRIPA buffer (Table 8). The protein lysates were kept at 37°C for 5 min to degrade the DNA. Then, the cell lysates were cleared by centrifugation at 16,000xg for 15 min at 4°C. The cleared supernatant was collected and the protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) following the manufacturer's instructions. After quantification, 30 µg total protein was set up in 15 µL mRIPA buffer and 5 µL NuPAGE™ LDS Sample Buffer (4x, Invitrogen). The protein lysates were incubated at 70°C for 10 min. 20 µL protein lysate was run on NuPAGE™ 4 to 12%, Bis-Tris, Mini Protein Gels (Invitrogen) at 180 V for 1 h using 1x NuPAGE™ MOPS SDS Running Buffer (20x, Invitrogen). After size separation, the proteins were transferred to a nitrocellulose membrane (0.45 µm pore size, GE Healthcare) for western blot (WB) analysis using the XCell II™ Blot Module (Invitrogen) in 1x NuPAGE™ Transfer Buffer (20x, Invitrogen) with 10% methanol (v/v). The transfer was performed at 35 V for 1 h. Then, the nitrocellulose membrane was washed in PBST (Table 16) for 3 min. The protein transfer was quality-controlled with a Ponceau Red staining of the membrane for 2 min (Table 14). PBST was used to destain the membrane after Ponceau Red staining. Before WB analysis with primary antibodies, the membrane was blocked with 5% milk in PBST for 30 min at room temperature. After washing the membrane three times in PBST, incubation with primary antibodies was performed overnight at 4°C (Table 7). For the detection of HNRNPH, rabbit polyclonal anti-HNRNPH (AB10374, Abcam) was used in a 1:10,000 dilution in PBST supplemented with 1% BSA. The protein Vinculin was used as

the loading control. To detect Vinculin, mouse monoclonal anti-Vinculin (V9624, Sigma-Aldrich) was used in a 1:10,000 dilution in PBST supplemented with 1% BSA. After incubation with the primary antibodies, the membrane was washed three times in PBST, followed by incubation with the secondary antibodies (Table 7). Secondary antibodies were diluted 1:5,000 in PBST supplemented with 5% milk and incubated with the membrane for 1 h at room temperature. Next, the membrane was washed again three times in PBST and finally the proteins were detected using SuperSignal™ West Pico Chemiluminscent Substrate (Thermo Fisher Scientific).

2.2.20 HNRNPH iCLIP in MCF7 cells

iCLIP was used to gain information about the RNA-binding profile of HNRNPH in MCF7 cells. The iCLIP experiment was performed according to the previously published iCLIP2 protocol (Buchbender et al., 2020). Briefly, the iCLIP library was made from MCF7 cells with approximately $1.5*10^6$ cells per immunoprecipitation in seven technical replicates. The cells were irradiated with 150 mJ/cm^2 at 254 nm in a UV-C crosslinker (CL-1000 Ultraviolet Crosslinker, UVP) to covalently crosslink protein that was bound to RNA. Before immunoprecipitation of the HNRNPH-RNA complexes, the RNA was digested with 10 µL of a 1:250 diluted RNase I (100 U/µL, Ambion). To pull down HNRNPH-RNA complexes from the cell lysates, 5 µg of polyclonal rabbit anti-HNRNPH antibody (AB10374) was combined with Protein G Dynabeads (Invitrogen) (Table 7). The final iCLIP samples were amplified with in total 16 PCR cycles. After PCR amplification, the samples were pooled with equimolarity. The final iCLIP library was sequenced using NGS on an Illumina NextSeq 500 platform with high output and 92 cycles (single-end).

2.2.21 Reverse Transcriptase Stop profiling (RT-Stop profiling)

To map the positions of rG4s *in vitro* in high-throughput, the RT Stop assay protocol was adapted (Guo and Bartel, 2016). For high-throughput rG4 mapping, RNA of the ivRNA library (see chapter 2.2.18) was used as a template for the RT-Stop profiling. First, 1 µg RNA of the ivRNA library was mixed with 0.5 pmol RToligo (0.5 pmol/µL, Table 2) and 0.5 mM dNTPs (10 mM, New England Biolabs) in reverse transcription buffer in a total reaction volume of 19.5 µL (Table 15). The reaction was incubated at 80°C for 2 min, followed by an incubation at 25°C for 2 min. Then, 100 U of SuperScript™ III Reverse Transcriptase (200 U/µL, Invitrogen) was added and the reaction was incubated for 10 min at 42°C. To hydrolyse the template RNA, 1.65 µL of 1 M NaOH was added, followed by an

incubation at 98 °C for 20 min. Addition of 20 µL 1M HEPES-NaOH pH 7.3 stopped the hydrolysis reaction. From here on, the library preparation of the iCLIP2 protocol was used continuing with second adapter ligation to the 3' end of the cDNA (Buchbender et al., 2020). In brief, after MyONE Silane bead clean up the second adaptor was ligated to the 3' end of the cDNA. Then, the cDNA was purified via MyONE Silane beads and amplified by PCR using the forward primer P5Solexa and the reverse primer P3Solexa (Table 2, chapter 2.2.17). The cDNA pre-amplification of the iCLIP2 protocol was omitted for the library preparation after RT-Stop profiling. After PCR amplification, all samples were pooled in equimolarity. The final library was size-selected two times with ProNex Chemistry (Promega GmbH) with a sample-to-ProNex (v/v) ratio of 1:2.4. Finally, the library was paired-end sequenced on an Illumina NextSeq 500 platform with mid output and 150 cycles (read 1: 141 nt, read 2: 25 nt).

2.2.22 Recombinant His_6-HNRNPH1-MBP protein

Recombinant His_6-MBP-tagged, full-length HNRNPH1 was expressed in *E. coli* BL21 (DE3) pLysS using a pET vector (pKT13). Cells were grown at 37°C and 160 rpm until an OD_{600} of 0.6. Subsequently, recombinant protein expression was induced by addition of IPTG (1 mM final concentration, Thermo Fisher Scientific). The expression culture was incubated at 37°C and 160 rpm for 3 h. Cells were harvested by centrifugation at 4000xg for 15 min at 4°C and resuspended in lysis buffer (Table 9). Then, cells were lysed by sonication, the lysate was cleared by centrifugation (45,000xg, 30 min, 4°C) and then passed over a HisTrap 5 mL column (GE Healthcare) to retain the recombinant protein. After washing the column with lysis buffer, the bound recombinant protein was eluted using a gradient from 10 to 500 mM imidazole (IMAC-A and IMAC-B buffer, Table 10, Table 11) over 10 column volumes. The HNRNPH1-containing elution fractions were pooled and run on a HiLoad 16/60 pg Superdex 200 column (GE Healthcare) in gel filtration buffer (Table 12). The fractions containing the recombinant protein were pooled and snap-frozen in liquid nitrogen for further storage at -80°C.

2.2.23 *In vitro* binding assay

2.2.23.1 5'-dephosphorylation of RNA

Before radioactive labelling, *in vitro* transcribed RNA was dephosphorylated at the 5'-end. Therefore, the RNA was incubated with 0.5 µL RNasin Plus (40 U/µL, Promega) and

5 µL Antarctic Phosphatase (5000 U/mL, New England Biolabs) in 20 µL for 30 min at 37°C. Subsequently, the RNA was purified using the RNA Clean & Concentrator-5 Kit (Zymo Research) following the manufacturer's instructions and aiming for a final RNA concentration of 15 µM.

2.2.23.2 Radioactive RNA labelling

To analyse the binding conditions of HNRNPH1 to RNA, the RNA was first radioactively labelled. Therefore, 5 µM of 5'-dephosphorylated RNA was incubated with 1 µL ^{32}P-γ-ATP (Perkin Elmer), 1 µL T4 Polynucleotide Kinase (10,000 U/mL, New England Biolabs) and 0.5 µL RNasin Plus (40 U/µL, Promega) in a total volume of 20 µL for 20 min at 37°C and 1100 rpm. Then, the radiolabelled RNA was purified using Sephadex G-25 quick spin columns (Roche) following the manufacturer's instructions.

2.2.23.3 RNA re-folding

To re-fold the RNA, 2 µL radioactively labelled RNA (1 µM) was incubated with 150 mM KCl or 150 mM NaCl and 20 mM Tris-HCl pH 7.5 in a reaction volume of 13 µL for 5 min at 80°C, followed by 15 min at 25°C.

2.2.23.4 *In vitro* binding of recombinant HNRNPH1 protein to RNA

After re-folding the RNA, 2 µL 10x binding buffer (table 13) and 5 µL recombinant HNRNPH1 protein (4x stock) were added to the 13 µL re-folded RNA reaction, reaching a final RNA concentration of 100 nM. For the *in vitro* titration, increasing concentrations of recombinant HNRNPH1 protein were combined with 100 nM RNA. The protein-RNA reaction was incubated at 37°C for 10 min and 1,100 rpm. Then, the reaction was irradiated with 250 mJ/cm^2 in a UV-C crosslinker (CL-1000 Ultraviolet Crosslinker, UVP) at 254 nm. The samples were run on NuPAGE 3-8% Tris-Acetate gels (Invitrogen) for 60 min at 150 V to separate protein-RNA complexes from unbound RNA. Next, the protein-RNA complexes were transferred to a nitrocellulose membrane (0.45 µm pore size, GE Healthcare) for 60 min at 30 V using the XCell II™ Blot Module (Invitrogen) in 1x NuPAGE™ Transfer Buffer (20x, Invitrogen) with 10% methanol (v/v). Finally, the membrane was exposed to a BAS Storage Phosphor Screen (GE Healthcare) for 2 h and then scanned on a Typhoon scanner (GE Healthcare). The relative intensities of the protein-RNA complexes were quantified with the ImageJ software. To test the *in vitro* titrations for cooperativity, a three parametric log-logistic function (Hill equation with lower limit equal to 0) was fitted to the

scatterplots plotting the relative intensities of the protein-RNA complexes against recombinant HNRNPH1 protein concentrations. The three parametric log-logistic functions were fitted with the R package “drc” using the “LL.3” package function (Ritz et al., 2015).

2.2.23.5 *In vitro* iCLIP of recombinant HNRNPH1 protein to the oligonucleotide library

The *in vitro* binding of recombinant HNRNPH1 protein to RNA was analysed in high-throughput using an adaption of the iCLIP2 and *in vitro* iCLIP protocols (Buchbender et al., 2020; Sutandy et al., 2018). Therefore, recombinant HNRNPH1 protein was UV-crosslinked to the ivRNA library (see chapter 2.2.23.4) and iCLIP library preparation was performed. In brief, 5 µM ivRNA library was radioactively labelled (see chapter 2.2.23.2). Next, 100 nM radioactively labelled ivRNA library was re-folded in 150 mM KCl (see chapter 2.2.23.3). Then, the RNA was incubated with 500 nM recombinant HNRNPH1 protein to allow protein-RNA-binding, followed by UV-crosslinking (see chapter 2.2.23.4). To normalise the final *in vitro* binding signal of HNRNPH1, a spike-in RNA pool (10 transcripts in equimolarity) was included. Therefore, the spike-in RNA pool was treated in the same way as the ivRNA library for RNA re-folding and UV-crosslinked with 500 nM recombinant HNRNPH1 protein. After UV-crosslinking, 1 µL of the spike-in RNA pool (2 nM) was added to each sample of the UV-crosslinked ivRNA library. Then, the samples were run on NuPAGE 3-8% Tris-Acetate gels (Invitrogen) to size-separate the protein-RNA complexes from the unbound RNA and afterwards the protein-RNA complexes were transferred to a nitrocellulose membrane (see chapter 2.2.23.4). Next, the library preparation for NGS sequencing was performed according to the previously published iCLIP2 protocol, continuing with RNA isolation from the nitrocellulose membrane (Buchbender et al., 2020). The experiment was done in four technical replicates, for the ivRNA library *in vitro* transcribed with GTP and for the ivRNA library *in vitro* transcribed with 7-deaza-GTP. The final library was paired-end sequenced on an Illumina NextSeq 500 platform with mid output and 150 cycles (read 1: 141 nt, read 2: 25 nt).

2.2.24 Circular dichroism (CD) spectroscopy

All CD spectra were measured with a Jasco J-1500 CD spectrometer (Jasco). The RNA oligonucleotides were ordered from IDT and dissolved in nuclease-free water to a final concentration of 100 µM (Table 4). Prior to the spectroscopic analysis, 10 µM RNA in 20 mM Tris-HCl pH 7.5 in a total volume of 200 µL was heated to 95°C for 3 min in the

presence of either 150 mM KCl, 150 mM NaCl or without any monovalent cation. Then, the reactions were gradually cooled to room temperature for 1 h. CD spectra of these samples were scanned at 25°C at a scan rate of 200 nm/min, with a 1-sec response time, 1 nm bandwidth and in continuous scan mode in 1 mm cuvettes. The CD spectra presented are averages of six individual spectra corrected for buffer background.

2.2.25 Fluorescence Resonance Energy Transfer (FRET)

All fluorescence measurements were performed with a Tecan Infinite M200 Pro (Tecan). The FRET-RNA oligonucleotides were ordered from Integrated DNA Technologies and dissolved in nuclease-free water (Table 4). To re-fold the RNA, 0.44 µM RNA in 20 mM Tris-HCl pH 7.5 was heated at 95°C for 3 min in a volume of 60 µL in the presence of either 150 mM KCl or 150 mM LiCl and then gradually cooled to room temperature for 1 h. Next, 0.4 µM re-folded RNA was incubated with 5 µL recombinant HNRNPH1 protein (10x stock) in a total reaction volume of 50 µL for 10 min at room temperature. As a control, 0.4 µM re-folded RNA was incubated with 5 µL gel filtration buffer (Table 12) in a total reaction volume of 50 µL for 10 min at room temperature. The RNA-protein mixture was transferred into a white 96-well plate (Thermo Fisher Scientific) and fluorescence was measured from top. Excitation (bandwidth: 9 nm) was performed with 25 flashes at 485 nm and emission (bandwidth: 20 nm) was detected at 520 nm. Each well was measured four times using an optimal gain. All measurements were performed in five independent replicates.

2.2.26 Interactome of HNRNPH1

2.2.26.1 Preparation of SILAC HEK293T cells

For analysis of the HNRNPH1 interactome, HEK293T cells were labelled with the method of stable isotope labelling using amino acids in cell culture (SILAC). Therefore, HEK293T cells were grown in SILAC DMEM Light (Table 17), SILAC DMEM Medium (Table 18) or SILAC DMEM Heavy (Table 19) growth medium for four weeks and sub-cultured twice per week in the respective growth medium. In the beginning, $1*10^7$ HEK293T cells per dish were plated in 15 cm cell culture dishes.

2.2.26.2 Transfection of plasmid in SILAC HEK293T cells

To transfect plasmids in SILAC HEK293T cells, for each sample 20 µg plasmid was set up in 1 mL OPTI-MEM (Invitrogen). In parallel, 40 µL Lipofectamine 2000 (Invitrogen) per sample was mixed in 960 µL OPTI-MEM (Invitrogen). The plasmid approach was added to the Lipofectamine 2000 approach and the transfection reaction was incubated at room temperature for 20 min. Finally, the transfection reaction was added to the cells dropwise. SILAC DMEM Light (light condition)-labelled HEK293T cells were transfected with the empty GFP-vector pL37 as the control condition (Table 1). SILAC DMEM Medium (medium condition)-labelled and SILAC DMEM Heavy (heavy condition)-labelled HEK293T cells were transfected with pKT11, which contains the open reading frame of *HNRNPH1* fused with an N-terminal EGFP (Table 1). The plasmid pKT11 was a kind gift from Cyril Dominguez (Department of Molecular Cell Biology, University of Leicester). The cells were harvested 48 h after transfection to prepare the samples for mass-spectrometric analysis.

2.2.26.3 GFP pulldown for mass-spectrometry analysis

To harvest, cells were washed twice with ice-cold PBS and then lysed in 500 µL Interactome lysis buffer (Table 20). The cell lysates were cleared by centrifugation at 16,000xg for 15 min at 4°C. The supernatant was transferred to new tubes and the protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) following the manufacturer's instructions. Before GFP pulldown, 25 µg total protein per sample was kept as the input control with 5 µL NuPAGE™ LDS Sample Buffer (4x, Invitrogen) in a total reaction volume of 20 µL. For the pulldown of GFP, 20 µL Magnetic agarose GFP binder beads (GFP-beads, Protein Production Core Facility, Institute of Molecular Biology [IMB], Mainz) per sample were washed three times with 900 µL Interactome lysis buffer (Table 20). After washing, the GFP-beads were dissolved in 100 µL Interactome lysis buffer per sample (Table 20). To pull down GFP, 4 mg total protein lysate was prepared in a final reaction volume of 900 µL with Interactome lysis buffer and then combined with 100 µL GFP-beads. The pulldown reaction was incubated at 4°C for 1 h and continuous rotation. After pulldown, the samples were washed twice with 900 µL Interactome lysis buffer (Table 20). Then, the heavy condition was treated with RNase for a partial RNase digestion (see chapter 2.2.26.4), while the light and medium conditions were kept on ice. After RNase digestion, 20% (volume) per sample of each

condition was taken to test efficient GFP-pulldown using WB. Therefore, the GFP-beads were eluted with 20 µL 2x NuPAGE™ LDS Sample Buffer (4x, Invitrogen) and the eluat was used for WB analysis (see chapter 2.2.19). For the remaining sample volumes, the light, medium and heavy conditions were combined per replicate and washed one more time with 900 µL Interactome lysis buffer (Table 20). Finally, the samples were eluted from the GFP-beads with 45 µL 2x NuPAGE™ LDS Sample Buffer (4x, Invitrogen) and 5 µL dithiothreitol (10 mM, Sigma-Aldrich). The eluates were heated to 70°C for 10 min. After cool-down, 0.5 µL 2-chloroacetamide (550 mM, Sigma-Aldrich) was added and the samples were incubated in the dark for 30 min at room temperature. Finally, [redacted] performed the mass spectrometry analysis [redacted].

2.2.26.4 Partial RNA digestion after HNRNPH1-EGFP pulldown

The pulldown samples of the heavy condition were subjected to a partial RNA digestion. Therefore, after pulldown and two washes with Interactome lysis buffer, the GFP-beads were resuspended in 100 µL Interactome lysis buffer (Table 20). For partial RNA digestion, the heavy condition was treated with 7.14 µL RNase A (7 U/µL, Qiagen) and 2 µL RNase T1 (1000 U/µL, Thermo Fisher Scientific) for 30 min at 4°C. Then, the heavy condition was washed three times with Interactome lysis buffer, followed by continuing with the combination of the three conditions light, medium and heavy (see chapter 2.2.26.3).

2.2.27 Downstream analyses of NGS data

[redacted] performed pre-processing of all NGS data. All downstream analyses of the NGS data was done by [redacted] (iCLIP, *in vitro* iCLIP, rG4 prediction) and [redacted] (RT-Stop profiling) [redacted]. To predict rG4s *in silico*, the pqsfinder R package was used with the following setting parameters (Hon et al., 2017):

I) Maximum predicted quadruplex sequence length: 30 bases
II) Minimal loop length: 1base
III) Maximal loop length: 7 bases
IV) Maximum number of defects: 0
V) Minimal PQS Score: 25.

3 Results

3.1 Gradual *HNRNPH1* overexpression in MCF7 cells

A recent study has suggested that HNRNPH can mediate splicing regulation of *MST1R* in a cooperative manner (Braun et al., 2018). Following up on this study the question arose whether cooperative splicing regulation is a global feature of HNRNPH. Further, the here presented PhD thesis aims to identify variables that are involved in the process of HNRNPH-mediated cooperative splicing regulation.

First, it was of interest to get a transcriptome-wide overview of HNRNPH-regulated splicing events. Therefore, the gradual *HNRNPH1* downregulation (HNRNPH KD) from Braun et al. was expanded with a gradual overexpression of *HNRNPH1* (HNRNPH OE) (Braun et al., 2018). A gradual titration can give insights into the dynamics of splicing responses. Additionally, the gradual titration offers a better simulation of physiologically relevant HNRNPH expression level dynamics than just comparing a strong HNRNPH KD with a strong HNRNPH OE. The gradual HNRNPH KD was performed with transfection of increasing concentrations of a siRNA targeting *HNRNPH1* in MCF7 cells (Braun et al., 2018). For the gradual HNRNPH OE increasing amounts of the pcDNA3.1(+)-HNRNPH1 plasmid (pKT2) were transfected into MCF7 cells (Figure 7A). To validate the gradual HNRNPH protein increase, total protein was isolated and the HNRNPH protein expression level was quantified with WB. With increasing amounts of pKT2, the protein expression level of HNRNPH increased (Figure 7B). The gradual overexpression of *HNRNPH1* resulted in HNRNPH protein levels up to 170% of wild type (WT) HNRNPH protein levels (Figure 7B).

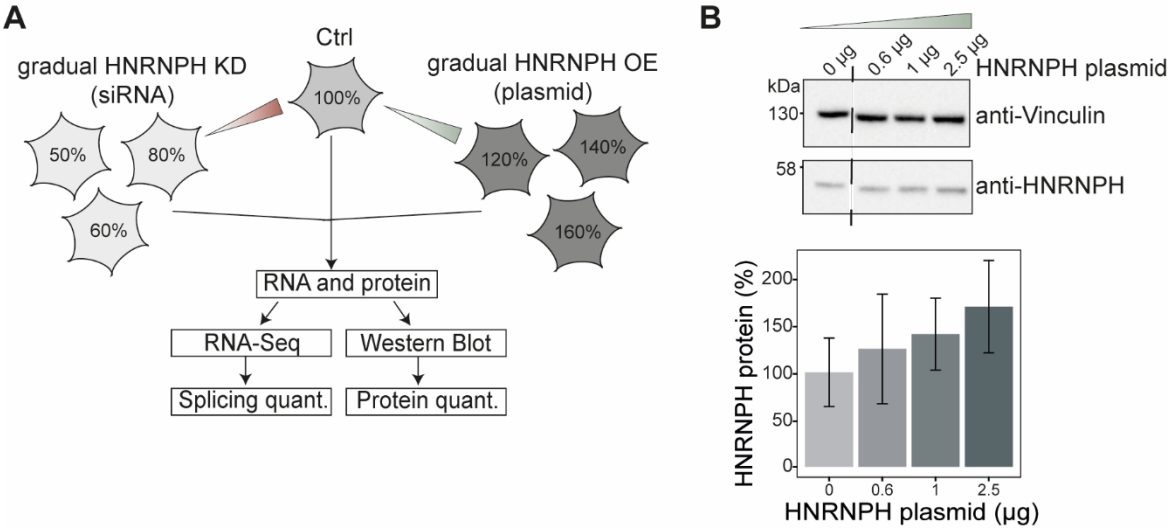

Figure 7: HNRNPH protein level titration in MCF7 cells.
(A) Graphical illustration of the experimental procedure. Data for the gradual *HNRNPH1* knockdown (KD) were taken from Braun et al. (Braun et al., 2018). Gradual *HNRNPH1* overexpression (OE) was performed with transfection of increasing amounts of pcDNA3.1(+)-HNRNPH1 plasmid (pKT2, see Table 1). (B) Western blot confirms the gradual overexpression of HNRNPH in MCF7 cells. For quantification, HNRNPH protein levels were normalised to the expression of the housekeeping protein Vinculin. Bar plot shows the mean percentage HNRNPH protein expression +/- standard deviation (n=3).

3.1.1 Validation of HNRNPH overexpression using *MST1R* splicing

It was shown that HNRNPH regulates alternative splicing of exon 11 of *MST1R* (Lefave et al., 2011; Braun et al., 2018). To validate the functional impact of the gradual HNRNPH OE for alternative splicing regulation, the splicing regulation of *MST1R* exon 11 was quantified using RT-PCR. With increasing HNRNPH OE the percent spliced in (PSI) of *MST1R* exon 11 in MCF7 cells decreased gradually from 68% to 32% exon inclusion frequency (Figure 8). This indicates that HNRNPH controls the skipping of *MST1R* exon 11 (Figure 8).

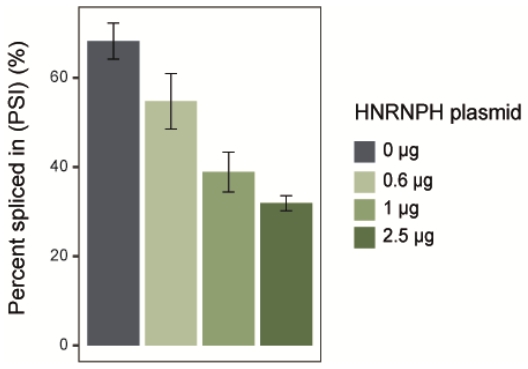

Figure 8: *MST1R* exon 11 inclusion decreases with HNRNPH overexpression.
Bar plot shows the mean percent spliced in (PSI) of *MST1R* exon 11 +/- standard deviation in response to gradual pcDNA3.1(+)-HNRNPH1 plasmid overexpression (n=3).

3.1.2 Analysis of transcriptome-wide splicing regulation of HNRNPH

To analyse the effect of the HNRNPH protein titration for alternative splicing regulation, RNA-Seq was performed for the eight gradual HNRNPH KD samples from Braun et al. and for the here described four gradual HNRNPH OE samples in MCF7 cells (Braun et al., 2018). Using two different tools, MAJIQ and Voila, the RNA-Seq data was analysed and local splicing variations (LSVs) were quantified (Vaquero-Garcia et al., 2016). The titration of HNRNPH protein levels in MCF7 cells affected hundreds of alternative splicing events (Figure 9A). The biggest effect on alternative splicing regulation was detected in the strongest HNRNPH KD condition (Figure 9A). Overall, HNRNPH KD had a stronger effect on changes in the alternative splicing regulation than HNRNPH OE (Figure 9A). The HNRNPH titration series affected in total 1,828 alternative splicing events, belonging to all types of splicing regulation, while the majority referred to CE events (Figure 9A). Transcriptome-wide, the HNRNPH protein titration reached up to +/-40% changes of the PSI (ΔPSI) of splicing events (Figure 9B). For the majority of regulated CE events, HNRNPH repressed the inclusion of the respective CE (Figure 9B). However, the HNRNPH titration also identified several CE events, for which HNRNPH enhanced the inclusion of the respective CE (Figure 9B). Overall, the HNRNPH titration revealed that HNRNPH can act as a splicing repressor and a splicing enhancer in MCF7 cells and thereby controls alternative splicing of hundreds of splicing events.

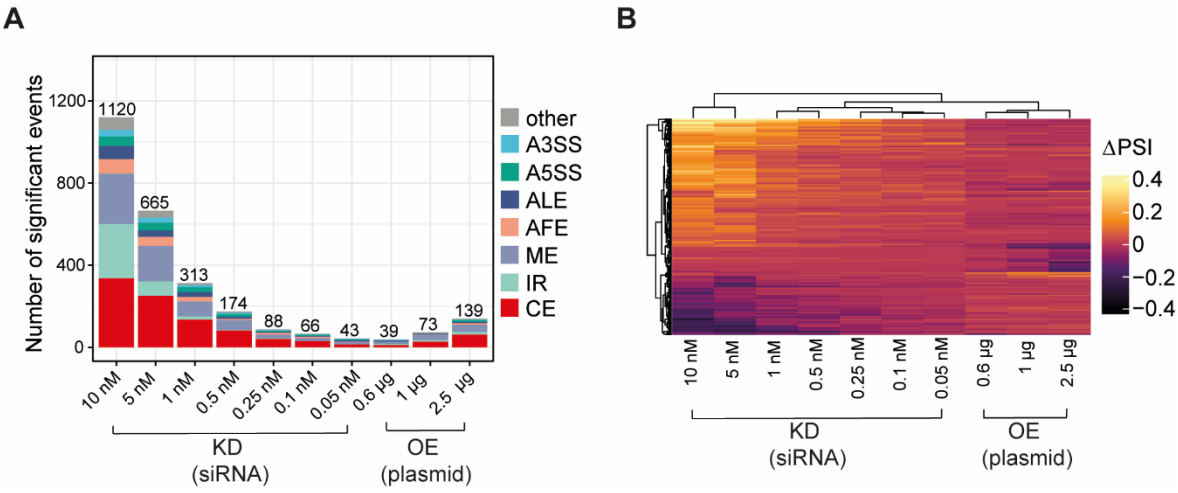

Figure 9: HNRNPH titration affects hundreds of alternative splicing events in MCF7 cells.
(A) Stacked bar plot shows the number of significantly changed splicing events quantified from the RNA-Seq of the HNRNPH titration in MCF7 cells. (B) Heatmap showing the delta PSI (ΔPSI) range of all significantly changed cassette exon (CE) events from the HNRNPH titration. The computational analysis was done [redacted].

3.1.3 Biological function of HNRNPH-regulated splicing events

The titration of the HNRNPH protein level in MCF7 cells revealed a large set of splicing events that are significantly regulated by HNRNPH. To identify the biological functions of genes that are alternatively spliced by HNRNPH, a gene ontology (GO) analysis using the gprofiler2 R package was performed (Kolberg et al., 2020). The gene names of all CE events that were significantly regulated by HNRNPH were used as input for the GO analysis. The GO analysis revealed that genes of HNRNPH-regulated CE events significantly associate with the RNA metabolism, like mRNA processing, RNA splicing and alternative mRNA splicing via the spliceosome (Table 21).

Table 21: Gene ontology analysis of HNRNPH-regulated alternative splicing events

GO	term_name	term_size	intersection_size		p_value
GO:0051252	regulation of RNA metabolic process	3753	288		2.3e−09
GO:0043484	regulation of RNA splicing	157	33		2.3e−09
GO:0048024	regulation of mRNA splicing, via spliceosome	105	23		7.0e−07
GO:0050684	regulation of mRNA processing	141	27		7.5e−07
GO:0006412	translation	707	74		2.3e−06
GO:0032774	RNA biosynthetic process	3600	259		6.9e−06
GO:0051254	positive regulation of RNA metabolic process	1769	145		7.5e−06
GO:1903311	regulation of mRNA metabolic process	326	42		9.4e−06
GO:0000380	alternative mRNA splicing, via spliceosome	79	16		1.7e−04
GO:0050686	negative regulation of mRNA processing	30	9		7.2e−04

3.2 Switch-like splicing regulation

As highlighted in before, HNRNPH controls the splicing of hundreds of splicing events (Figure 9). Following up on the cooperative splicing regulation of *MST1R* suggested by Braun et al., the question was addressed whether HNRNPH cooperatively mediates alternative splicing regulation on a transcriptome-wide scale (Braun et al., 2018).

3.2.1 Hill function modelling of transcriptome-wide splicing targets

It was the aim to investigate whether HNRNPH controls switch-like splicing of other splicing events besides *MST1R* exon 11. Therefore, Hill function modelling was performed for the dose-response curves of all CEs that have been identified within the RNA-Seq analysis of the HNRNPH titration in MCF7 cells. The nH value, which was extracted from the Hill function, provided information about the steepness of the splicing

regulation and thereby served as an indicator for cooperativity. The distribution of the retrieved nH values highlighted that most of the analysed CEs were cooperatively regulated by HNRNPH (Figure 10A). All CEs with nH values > |2| were defined as cooperatively regulated splicing events. HNRNPH cooperatively spliced 301 CEs, while HNRNPH acted as a splicing repressor for 223 of these splicing events (Figure 10B). Only a few splicing events were not cooperatively (non-cooperative) regulated by HNRNPH (Figure 10B). Figure 10 shows Hill equation fittings for the alternative splicing events of *PIP4K2C*, *MYL6* and *SNHG21* using ΔPSI values quantified from the RNA-Seq analysis (Figure 10C,D). The alternative splicing events of *PIP4K2C* and *MYL6* are both described by nH values larger than |2|, therefore belonging to the category of cooperatively regulated splicing events (Figure 10C). In contrast, the Hill equation fitting for *SNHG21* resulted in an nH value smaller than |2| (Figure 10D). Hence, *SNHG21* exon 3 is not cooperatively regulated by HNRNPH. Overall, the transcriptome-wide RNA-Seq analysis of the HNRNPH titration series identified that HNRNPH globally regulates alternative splicing in a cooperative manner.

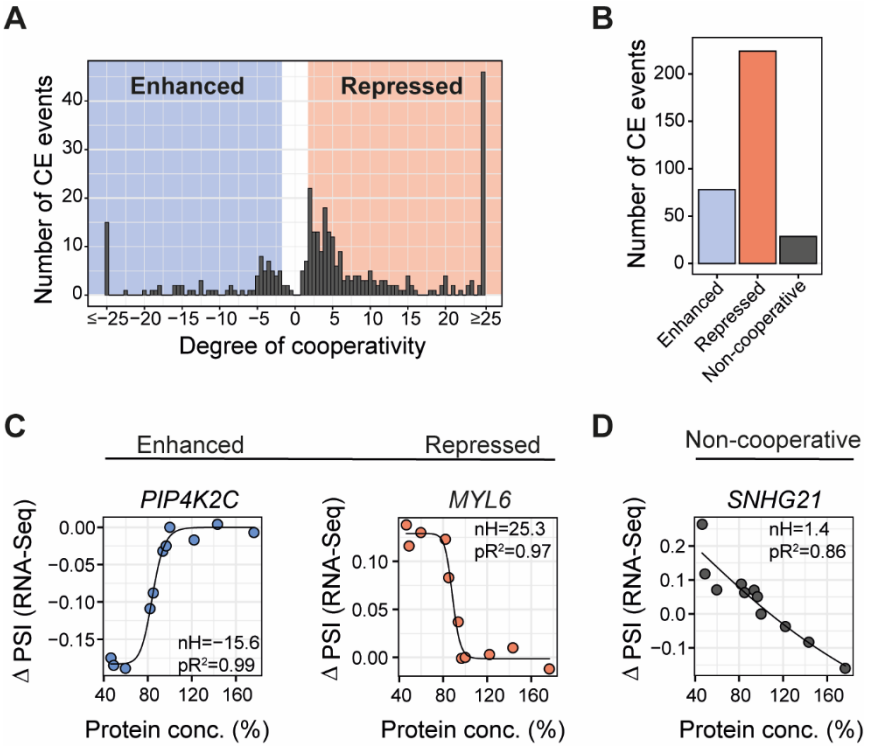

Figure 10: Hill function modelling shows transcriptome-wide HNRNPH-mediated cooperative splicing.
(A) Distribution of the Hill coefficients extracted from the Hill function modelling for the CE that were alternatively spliced in response to the HNRNPH titration in MCF7 cells. Negative Hill coefficients refer to HNRNPH as a splicing enhancer (Enhanced). Positive Hill coefficients refer to HNRNPH as a splicing repressor (Repressed). (B) Total number of CE events with cooperative enhanced, cooperative repressed and non-cooperative splicing regulation by HNRNPH. (C) Hill functions, modelling the dose-response curves for the RNA-Seq quantified ΔPSIs of *PIP4K2C* exon 6 (left curve) and *MYL6* exon 6 (right curve) in response to HNRNPH protein titration. (D) Hill function, modelling the dose-response curve for the RNA-Seq quantified ΔPSI of *SNHG21* exon 3 in response to HNRNPH protein titration. The computational analysis was done by [redacted].

3.2.2 RT-PCR analysis validates switch-like splicing regulation

RT-PCR experiments were performed to validate the cooperative splicing regulation that has been identified from the RNA-Seq analysis of the HNRNPH titration series. The RT-PCR experiments were performed using the RNA samples from the HNRNPH titration series in MCF7 cells. Indeed, RT-PCR confirmed the switch-like splicing regulation of individual splicing events in response to the HNRNPH protein titration in MCF7 cells. For *PIP4K2C* exon 6 and *MYL6* exon 6, a small reduction of the HNRNPH protein level was sufficient to reach the full HNRNPH-mediated splicing regulation (Figure 11A). In contrast, splicing of *SNHG21* exon 3 did not show a switch-like splicing regulation in response to the HNRNPH titration (Figure 11B). To evaluate the steepness of the switch, Hill functions were modelled for the individual dose-response curves. The dose-response curves for *PIP4K2C* and *MYL6* splicing both gave nH values that were larger than |2| (Figure 11A). Hence, both splicing events are cooperatively regulated by HNRNPH. With an nH value close to 1,

SNHG21 did not show a cooperatively-mediated splicing regulation by HNRNPH (Figure 11B). In summary, the RT-PCR experiments successfully validated the cooperative splicing regulation by HNRNPH, which has been identified from the RNA-Seq data (Figure 10).

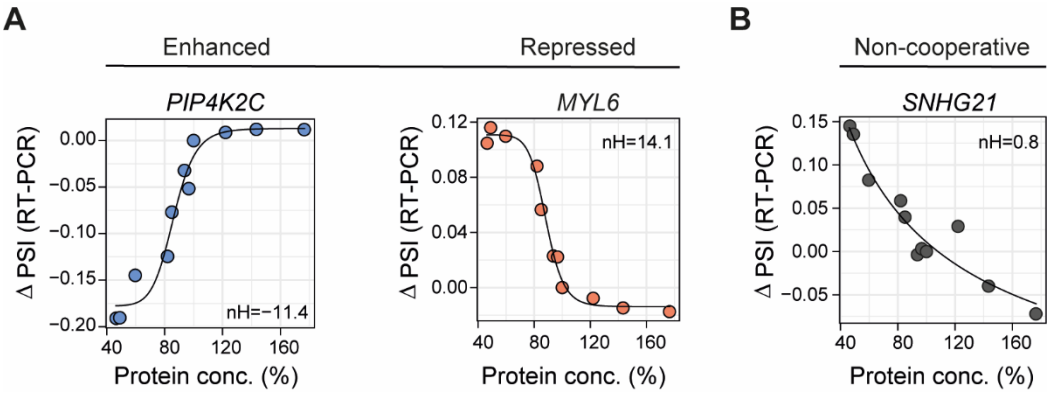

Figure 11: RT-PCR shows switch-like splicing regulation for HNRNPH-mediated splicing of *PIP4K2C* and *MYL6*.
(A) Hill function modelling of the dose-response curve for the delta PSI (ΔPSI) of *PIP4K2C* exon 6 (left plot) and *MYL6* exon 6 (right plot) in response to HNRNPH protein titration. (B) Hill function modelling of the dose-response curve for the ΔPSI of *SNHG21* exon 3 in response to HNRNPH protein titration. The Hill function modelling was done by [redacted].

3.3 Direct cooperativity

The phenomenon that HNRNPH regulates splicing cooperatively in a transcriptome-wide manner raised the question of which molecular mechanisms mediate this cooperation. Interestingly, HNRNPs have the intrinsic feature to form multivalent complexes (Gueroussov et al., 2017). The GY-domains of HNRNPs facilitate the formation of these higher-order complexes via direct protein-protein interactions (Gueroussov et al., 2017). Direct protein-protein interactions have been described to mediate cooperative binding of HNRNPs to RNA (Cartegni et al., 1996; Cieniková et al., 2015). To investigate the role of direct cooperative binding in HNRNPH-mediated cooperative splicing regulation, several approaches were combined:

I) HNRNPH1-pulldown followed by mass spectrometry to identify interactors of HNRNPH1.
II) *In vivo* splicing analysis of a GY-domain-deleted *HNRNPH1* expression construct.

3.3.1 Interactome of HNRNPH1

To characterise interactors of HNRNPH1, an *EGFP-HNRNPH1* expression construct was expressed in HEK293T cells and immunoprecipitated (Figure 12A). The cells were labelled with SILAC, which allowed the identification of interactors using mass spectrometry.

Comparison of EGFP-HNRNPH1 interactors with empty vector (EV) interactors identified proteins that were specifically enriched for EGFP-HNRNPH1. Especially, other HNRNPs like HNRNPC, HNRNPU and HNRNPA0 were enriched for EGFP-HNRNPH1 (Figure 12B). Moreover, EGFP-HNRNPH1 interacted with the highly similar paralogue HNRNPH2 (96% similarity on amino acid level) (Figure 12B). Besides the interaction with other HNRNPs, the EGFP-HNRNPH1 interactome revealed an interaction of HNRNPH1 with the RBPs SMU1 and IK (RED) that are involved in activation of the spliceosome (Keiper et al., 2019) (Figure 12B). The complete set of interactors significantly enriched for EGFP-HNRNPH1 compared to EV is shown in Table 22. Next to the association with RBPs, the EGFP-HNRNPH1 interactome revealed interactions with mitochondrial transporters like members of the SLC25 family or the inner mitochondrial membrane transporter TIMM23 (Figure 12B). In sum, the interactome of HNRNPH1 highlighted an interaction of HNRNPH1 with several other RBPs.

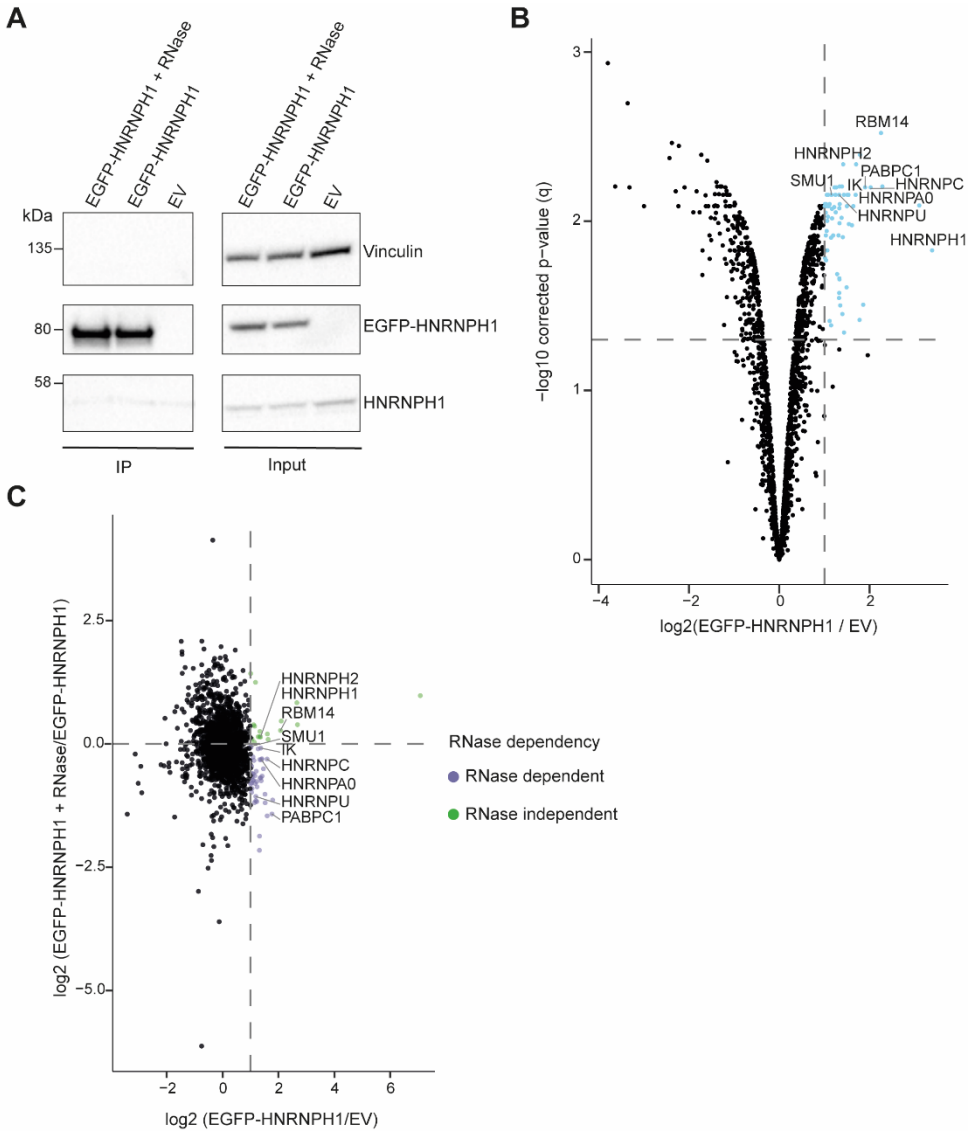

Figure 12: Interactome of EGFP-HNRNPH1 reveals interactions with other RBPs.
(A) WB showing the overexpression of EGFP-HNRNPH1 (pKT11, see Table 1) in the input and the selective pulldown of EGFP-HNRNPH1 after immunoprecipitation (IP) using GFP agarose beads. (B) Protein interactome of EGFP-HNRNPH1 identified using mass-spectrometry analysis. The log2-transformed ratio of EGFP-HNRNPH1/empty vector (EV) after z-score normalisation is plotted against the -log10-transformed corrected p-value. All proteins enriched for EGFP-HNRNPH1 (≥ log2(1)) with a corrected p-value < -log10(0.05) are highlighted in blue. (C) EGFP-HNRNPH1 interactions can be independent of RNA. The log2-transformed ratio of EGFP-HNRNPH1/EV is plotted against the log2-transformed ratio of EGFP-HNRNPH1 + RNase/EGFP-HNRNPH1. All EGFP-HNRNPH1 interactors that are enriched for EGFP-HNRNPH1 + RNase are classified as RNase independent interactors (coloured in green).

It was of interest to identify HNRNPH1 protein-protein interactions that are independent of RNA. Therefore, the interactome of EGFP-HNRNPH1 + RNase treatment was characterised and compared to the interactome of EGFP-HNRNPH1. From the RBP interactors that have been identified from EGFP-HNRNPH1, the interactions with HNRNPH2 and RBM14 were insensitive to the RNase treatment (Figure 12C). Interactions with other HNRNP members and with SMU1, IK and PABPC1 were lost in response to RNase treatment (Figure 12C).

Hence, only interactions with the paralogue HNRNPH2 and the RBP RBM14 are RNA-independent protein-protein interactions. The detected RNA-independent interaction of HNRNPH1 with the highly similar HNRNPH2 could hint towards direct cooperativity mediated by RNA-independent protein-protein interactions.

Table 22: Complete set of interactors significantly enriched for EGFP-HNRNPH1 (corrected p-value < 0.05)

Gene	log2(HNRNPH1/EV)	p-value	corrected p-value (q)
HSDL2	1.003291419	0.00034771	0.00814853
PRKAR1A	1.003509683	0.00032486	0.00814853
ARL8A	1.011427665	0.00272168	0.0164813
TIAL1	1.011851539	0.00027444	0.00792274
SLC25A1	1.029422274	0.00048577	0.00890724
GLMN	1.031140096	0.00115653	0.01213313
SFXN1	1.035708017	0.00300133	0.0170191
GCAT	1.036438894	0.00030272	0.00806999
DPM1	1.036621113	0.00026713	0.00792274
RPS29	1.037049717	0.00038537	0.00814853
SLC25A3	1.048929584	0.00104783	0.01142796
GPX8	1.054112386	0.00018113	0.00698652
RAB8A	1.056385061	0.00021246	0.00698652
EBP	1.061612722	0.00071964	0.0099565
SSR1	1.064403489	0.00205905	0.01487854
DLD	1.075009616	0.00020393	0.00698652
NAPA	1.078708987	0.00057552	0.00941363
SSR4	1.080918283	0.00021492	0.00698652
SLC25A4	1.092917232	0.00031604	0.00814853
NDUFA10	1.114526571	0.01265668	0.03886136
AGK	1.114937412	0.00028052	0.00792274
SACM1L	1.146876245	0.00132171	0.01252753
TTC38	1.147527415	0.01413058	0.04125623
SMU1	1.15718515	0.00021722	0.00698652
SLC25A10	1.167750977	0.00043537	0.00845731
FASTKD5	1.169155391	0.00065234	0.00962327
RAB1B;RAB1A;RAB1C	1.193511096	0.0011511	0.01213313
DARS2	1.215703033	0.00041246	0.00825771
PHB	1.224952281	0.00012127	0.00631483
TIMM23;TIMM23B	1.244355425	0.00019025	0.00698652
PHB2	1.250807761	0.00047138	0.00886016
HSD17B4	1.261924305	0.00012231	0.00631483
RAB7A	1.281362413	0.00453312	0.02054118
SLC25A24	1.312494811	0.00113707	0.01212461
ACTA1;ACTC1	1.323385268	0.00783886	0.02842856
HNRNPU	1.32584162	0.00016645	0.00698652
ZADH2	1.32705513	0.00690541	0.02653248
GLS	1.329806951	0.00042739	0.00838545

SLC25A21	1.330680022	0.00501385	0.02176368
ACTG1	1.333193538	0.0051666	0.02205204
ATAD3A	1.334006679	0.00080077	0.01028091
PTPRF	1.337899033	8.40E-05	0.00623823
SLC25A13	1.338447813	0.00028267	0.00792274
ABCC4	1.349739199	0.00913577	0.03144129
SLC25A6	1.380527049	0.0006131	0.0095469
MTX2	1.387897887	8.00E-05	0.00623823
RAN	1.398776224	0.01081259	0.03535717
PLK1	1.406915513	0.00125811	0.01242679
HNRNPH2	1.414795627	5.17E-05	0.00461124
HADH	1.431284517	0.00019264	0.00698652
CRAT	1.433359381	0.01654662	0.04546845
RAB13	1.490896666	0.00618147	0.02460637
MRPS21	1.49735253	0.00037983	0.00814853
IK	1.507810084	0.00017673	0.00698652
SLC25A5	1.531772823	0.00021453	0.00698652
SCO2	1.541521986	0.00082152	0.01039879
OAT	1.609616373	0.00090223	0.010567
SLC25A12	1.640600622	0.00038075	0.00814853
HNRNPA0	1.690985796	0.00021319	0.00698652
SEC16A	1.702236618	4.69E-05	0.00461124
PPOX	1.77768079	0.01238123	0.03828929
CMAS	1.789060679	3.30E-05	0.00405242
COX5B	1.859081324	0.00899968	0.03116839
PABPC1;PABPC4	1.901325697	0.0001029	0.00630921
HNRNPC	2.02320627	0.00010626	0.00631483
RBM14	2.252797682	1.69E-05	0.00301192
ASPH	2.283995463	7.91E-05	0.00623823
PLOD1	3.100839808	0.00029603	0.00806999
HNRNPH1	3.384983732	0.00202489	0.01487854

3.3.2 RT-PCR analysis of *HNRNPH1* mutant expression constructs

The interactome of EGFP-HNRNPH1 identified interactions with other HNRNPs like HNRNPH2, HNRNPC or HNRNPU. This led to the idea that HNRNPH1's C-terminal GY-domain could play a role in cooperative splicing regulation. To investigate this question, an *HNRNPH1* expression construct with deletion of the C-terminal GY-domain (*HNRNPH1* w/o GY) was generated. The GYR-domain was not mutated, since this domain has been described to be involved in nucleocytoplasmic shuttling of HNRNPH (van Dusen et al., 2010).

The *HNRNPH1* w/o GY construct was expressed in HEK293T cells that have been transfected with a siRNA targeting endogenous *HNRNPH1*. To guarantee that the siRNA

only targeted endogenously expressed *HNRNPH1*, the siRNA target site was mutated in the *HNRNPH1* expression constructs. The ectopic expression of *HNRNPH1* in an endogenous *HNRNPH1* KD allowed analysing the rescue capacity of the *HNRNPH1* deletion construct in comparison to a WT *HNRNPH1* expression construct. The effect of ectopic *HNRNPH1* expression for splicing decisions was exemplary investigated by analysing *MST1R* exon 11 inclusion levels.

With increasing concentrations of *HNRNPH1* expression constructs, the HNRNPH1 protein level increased, although the cells were treated with the siRNA for *HNRNPH1* KD (Figure 13A). This confirmed that the *HNRNPH1* expression constructs were indeed insensitive to the siRNA treatment. With higher HNRNPH1 protein levels, the inclusion frequency of *MST1R* exon 11 decreased (Figure 13B). The HNRNPH1 protein with deleted C-terminal GY-domain showed rescue of endogenous *HNRNPH1* KD similar to the rescue capacity of WT HNRNPH1 protein constructs (Figure 13B).

To investigate an effect of the GY-domain on the switch-like splicing regulation by HNRNPH1, the HNRNPH1 protein level was log10-transformed. This allowed getting an impression about the steepness of the splicing responses (Figure 13B). Without the C-terminal GY-domain, the steepness of the splicing response mediated by HNRNPH1 did not change severely compared to the WT HNRNPH1 (Figure 13B). In summary, deletion of the C-terminal GY-domain does not affect HNRNPH's splice repressor function for *MST1R* exon 11 splicing.

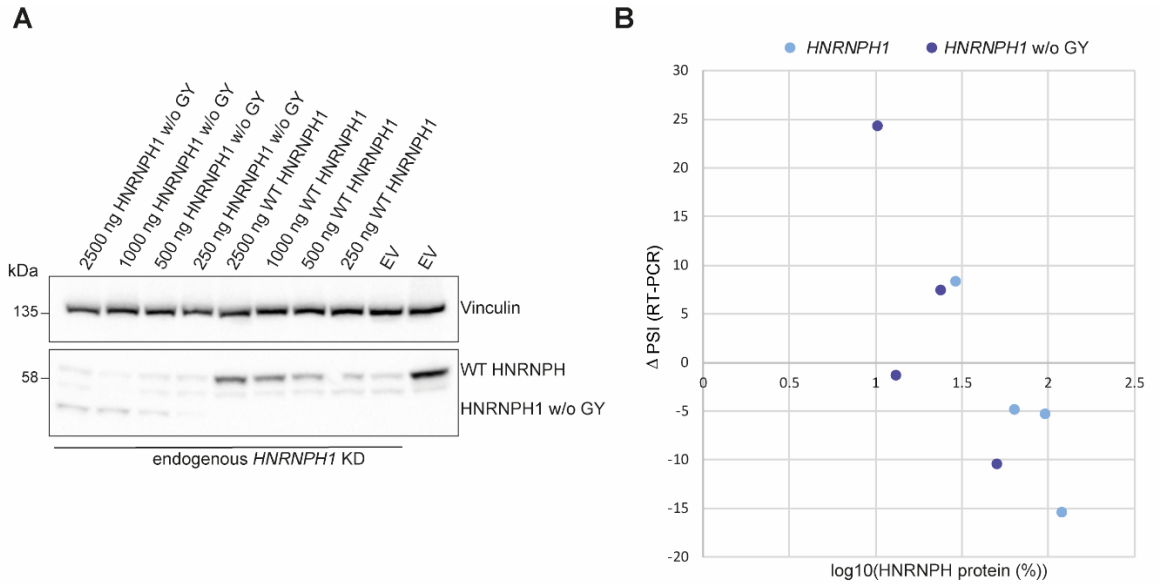

Figure 13: With deletion of the C-terminal GY-domain HNRNPH1 can rescue endogenous HNRNPH1 KD.
(A) WB showing the protein expression of WT HNRNPH (pKT4, see Table 1) and HNRNPH1 protein with deleted C-terminal GY-domain (HNRNPH1 w/o GY; pKT16, see Table 1). Endogenous HNRNPH1 knockdown (KD) is validated using transfection of an empty vector (EV). (B) Expression of WT *HNRNPH1* or *HNRNPH1* without the GY-domain both affect *MST1R* exon 11 inclusion frequency. The log10-transformed HNRNPH protein level is plotted against the change of the PSI (ΔPSI) of *MST1R* exon 11 in comparison to WT cells (no *HNRNPH1* KD, no *HNRNPH1* expression construct).

3.5 HNRNPH iCLIP

HNRNPH controls hundreds of splicing events in MCF7 cells cooperatively (Figure 10). A mass spectrometry analysis revealed several interactions of HNRNPH1 with other HNRNPs (Figure 12). However, deletion of the GY-domain, which has been described as an important protein-protein interaction domain for other HNRNPs, did not severely change the splice effector function of HNRNPH1 (Figure 13). Hence, to gain detailed insights into the characteristics of HNRNPH binding to RNA, an HNRNPH iCLIP experiment was performed in MCF7 cells using the recently published iCLIP2 protocol (Buchbender et al., 2020). The iCLIP experiment was performed in MCF7 cells with seven technical replicates and sufficiently generated protein-RNA complexes that immunoprecipitated with the HNRNPH antibody (Figure 14). After RNA preparation from these protein-RNA complexes, the replicates were pooled into one final iCLIP library. Sequencing of the iCLIP library resulted in 26-43 million reads per sample.

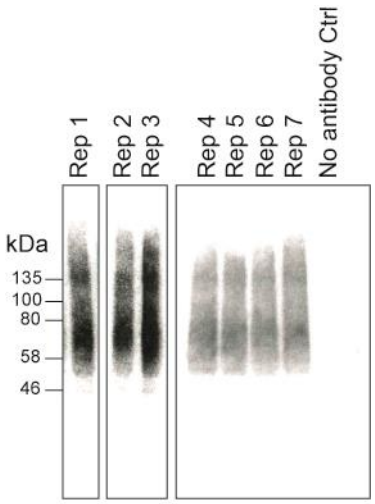

Figure 14: Autoradiogram of HNRNPH iCLIP in MCF7 cells.
The autoradiogram shows the protein-RNA complexes of a HNRNPH iCLIP experiment in MCF7 cells for seven technical replicates (Rep). For the "No antibody Ctrl" condition, the iCLIP experiment was performed without addition of the HNRNPH antibody.

3.5.1 HNRNPH binding sites across the genome

Having sequenced the HNRNPH iCLIP library, PureCLIP identified in total 468,452 binding sites across the genome (Krakau et al., 2017). The majority of the HNRNPH binding sites, 85.7%, located to protein-coding genes (Figure 15A). Within the protein-coding genes, 88.6% of the binding sites mapped to intronic regions (Figure 15B). In contrast, only 2% of all HNRNPH binding sites within protein-coding genes were located in the coding sequence (CDS) (Figure 15B). Overall, the HNRNPH iCLIP experiment revealed that HNRNPH preferentially binds the introns of protein-coding genes.

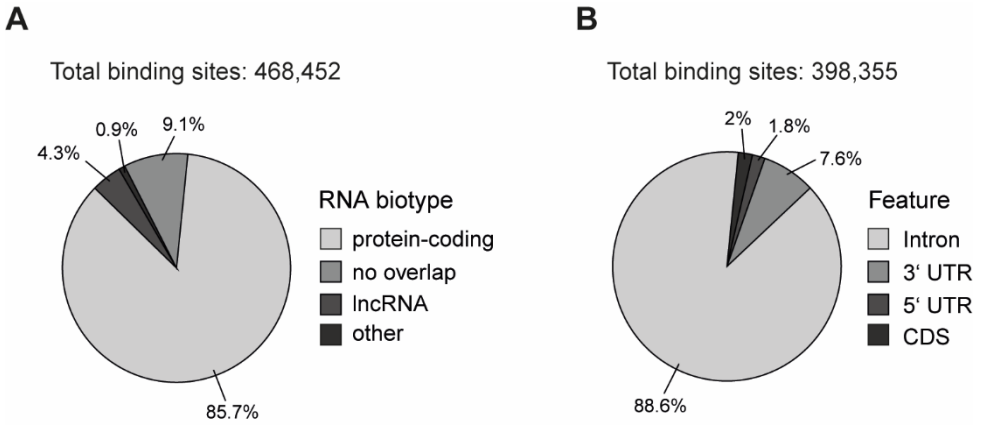

Figure 15: HNRNPH mainly binds to the introns of protein-coding genes.
(A) Pie chart shows the distribution of HNRNPH binding sites across the whole genome. (B) Pie chart illustrates the binding of HNRNPH across the gene body of a protein-coding gene. The computational analysis was done by [redacted].

3.5.2 RNAmaps highlight effect of HNRNPH positioning for splicing regulation

The majority of HNRNPH binding sites localised to protein-coding genes and the RNA-Seq analysis revealed hundreds of HNRNPH-sensitive splicing events (Figure 9; Figure 15). Hence, the question arose whether the positioning of HNRNPH within the gene-body can function as a discriminator for its splicing activity. To investigate this, RNA maps were generated for CE events. RNA maps visualise the binding landscape of an RBP (i.e. by using the iCLIP binding signal) around a defined region of the mRNA architecture (i.e. a CE event). Hence, RNA maps can illustrate the RBP binding landscape for mRNAs and thereby give an impression of preferred RBP binding positions across a defined region of the gene-body. To construct the RNA maps for this PhD thesis, CE events for which HNRNPH acts as a splicing repressor were separated from events for which HNRNPH acts as a splicing enhancer. To visualise the HNRNPH binding landscape, the HNRNPH iCLIP signal from CE events with HNRNPH-mediated splicing regulation was compared to the iCLIP signal of a control group that showed no splicing change in response to HNRNPH ($\Delta PSI \leq 0.025$).

When HNRNPH acts as a splicing repressor, HNRNPH bound significantly stronger to the AE and its 5'ss and 3'ss than in comparison to the control group (corrected p-value ≤ 0.01) (Figure 16A). The HNRNPH binding in the flanking introns did not differ between CE events repressed by HNRNPH and CE events without HNRNPH-mediated splicing regulation (Figure 16A). In contrast, for exons that are enhanced by HNRNPH, HNRNPH bound significantly stronger to the introns compared to the control group (corrected p-value ≤ 0.01) (Figure 16B). In more detail, HNRNPH especially preferred binding to the first 300 nt of the upstream intron as well as to the downstream intron (first 300 nt and last 300 nt) (Figure 16B). Additionally, when HNRNPH functions as a splicing enhancer the binding strength to the AE itself did not severely differ in comparison to the non-regulated CE events (Figure 16B). Hence, positioning of HNRNPH binding can discriminate the splicing regulation by HNRNPH. HNRNPH mainly binds to the region and close surrounding of the AE itself when it functions as a splicing repressor. In contrast, for splicing enhancement, HNRNPH preferentially binds to the introns and especially to the flanking downstream intron of the respective CE event.

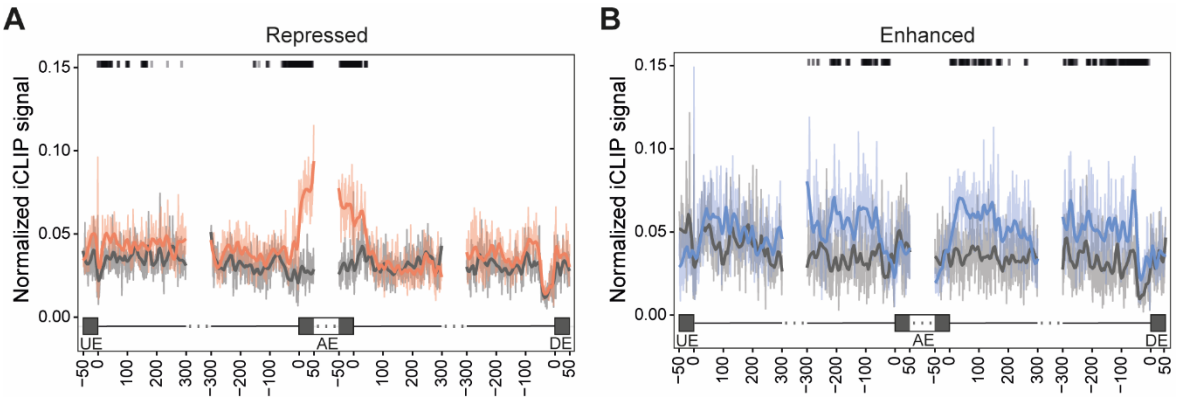

Figure 16: RNAmaps identify a position-dependency of HNRNPH's splice factor function.
(A,B) RNAmap for cassette exon (CE) events that are repressed (A) or enhanced (B) by HNRNPH. iCLIP signal is plotted for the last 50 nt of the upstream exon (UE), the first and last 50 nt of alternative exon (AE) and the first 50 nt of the downstream exon (DE). Further, starting from each splice site, the iCLIP signal for the first 300 nt of the introns is plotted. Normalised iCLIP signal for repressed CE events (orange) or enhanced CE events (blue) is compared to the normalised iCLIP signal of non-regulated CE events (grey). CE events are classified as non-regulated when they responded to the HNRNPH titration with a $\Delta PSI \leq 0.025$. The sets of non-regulated CEs and regulated CEs are matched to have an equal mean WT PSI. A Wilcoxon signed rank test is performed for the averaged iCLIP signal of 10 nt bins to compare the iCLIP signal of regulated and non-regulated CE events. Darkgrey ticks on top of the profiles mark bins with a significant difference (corrected p-value ≤ 0.01) between the two groups. The computational analysis was done by [redacted].

3.5.3 HNRNPH motif enrichment analysis

The positioning of HNRNPH in the exon-intron structure can determine the splice factor function of HNRNPH (Figure 16). However, the question whether HNRNPH-RNA interactions include features that regulate cooperative splicing remained open. To address this, the RNA-binding motif for HNRNPH was extracted from the iCLIP data. Other studies have published that HNRNPH binds to G-rich motifs (Swanson and Dreyfuss, 1988; Caputi and Zahler, 2001; Schaub et al., 2007; Dominguez et al., 2010). The iCLIP data generated for this study also showed that HNRNPH preferentially binds to a G within a G-stretch of up to five Gs (Figure 17A).

To get an impression about the most frequently used RNA-binding motifs, the frequency of tetramer motifs was compared between the 20% strongest and 20% weakest HNRNPH binding sites. The strongest HNRNPH binding sites showed an enriched usage of G-rich tetramer motifs (Figure 17B). These tetramer motifs mostly consisted of three Gs in a row (Figure 17B). The tetramer motif that had the highest frequency in the 20% strongest HNRNPH binding sites was the GGGG motif (Figure 17B). In comparison, the 20% weakest HNRNPH binding sites did not show high frequencies for G-rich motifs (Figure 17B). Hence, strong HNRNPH binding sites preferentially contain G-rich tetramers.

Next, the question was addressed, whether the usage of HNRNPH binding motifs differs between cooperatively regulated and non-regulated splicing events. Therefore, the frequency of tetramer motifs was compared between cooperatively regulated events and non-regulated events. Non-regulated events showed slightly lower frequencies of HNRNPH binding sites with a GG or GGG-containing tetramer motif compared to the HNRNPH binding sites of the cooperatively regulated events (Figure 17C). Furthermore, the non-regulated events had a clear enrichment for the UUUU tetramer, which most likely referred to a UV crosslinking bias for U-rich RNA sequences (Figure 17C). The tetramer usage between cooperatively regulated and non-regulated splicing events did not show a striking difference. Nevertheless, the motif enrichment analysis reveals that HNRNPH binds to G-rich sequences and further that strong HNRNPH binding sites have an enrichment for tetramers with G-tracts consisting of three or four Gs in a row. Moreover, the iCLIP data reveal an enrichment for Gs up to 25 nt upstream and downstream of the respective HNRNPH binding site (Figure 17A).

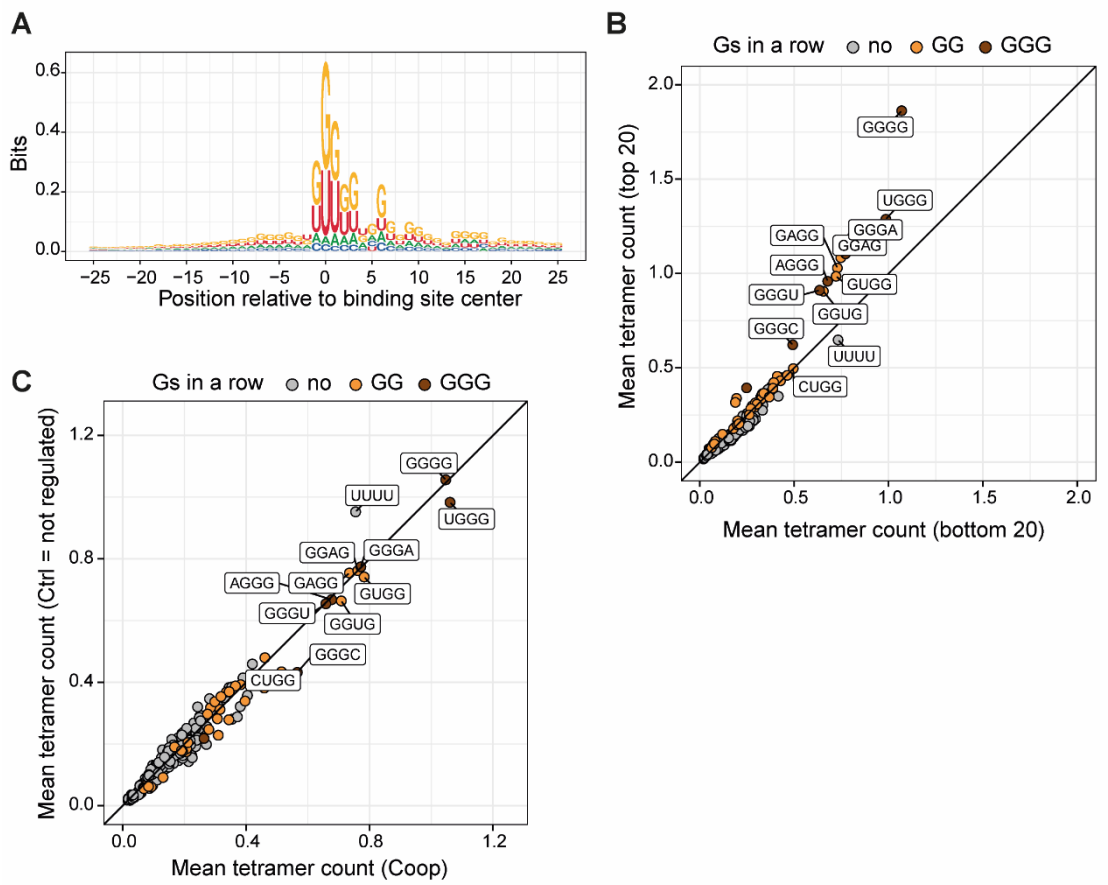

Figure 17: HNRNPH binds to G-rich RNAs.
(A) Motif enrichment analysis showing the RNA-binding motif of HNRNPH 25 nt upstream and downstream to the respective HNRNPH binding site center. (B) Scatter plot compares the mean tetramer frequency between the 20% strongest (top 20) and 20% weakest (bottom 20) HNRNPH binding sites. Tetramers having two or three Gs in a row are coloured. (C) Scatter plot compares the mean tetramer frequency for HNRNPH between not regulated (Ctrl) and HNRNPH-cooperatively regulated CE events (Coop). Tetramers having two or three Gs in a row are coloured. The computational analysis was done by [redacted].

3.5.4 Impact of predicted rG4s for HNRNPH binding

The HNRNPH motif enrichment analysis confirmed HNRNPH's preference to bind to G-tract RNAs (Figure 17). Interestingly, the canonical motif for an rG4 consists of the $G_{(3-6)}N_{(1-7)}G_{(3-6)}N_{(1-7)}G_{(3-6)}N_{(1-7)}G_{(3-6)}$ consensus sequence, which includes the G-tract tetramer motif enriched in the 20% strongest HNRNPH binding sites (Dumas et al., 2021). Hence, the role of rG4s in HNRNPH binding was investigated. To predict rG4s *in silico*, the pqsfinder R package was used (Hon et al., 2017).

Having the information about predicted rG4s, the question was addressed whether the co-occurrence of predicted rG4s affected the binding strength of HNRNPH. Indeed, HNRNPH bound stronger to binding sites overlapping with a predicted rG4 (Figure 18). Even more intriguing, HNRNPH binding sites with a G-tract motif (GGG) and an overlapping predicted rG4 showed the highest PureClip scores (Figure 18). In summary, the co-occurrence of predicted rG4s with the HNRNPH binding motif characterises the strongest HNRNPH binding sites.

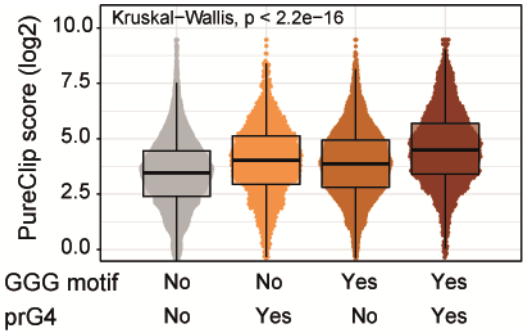

Figure 18: HNRNPH binds stronger to binding sites that overlap with predicted rG4s.
Shown are log2-transformed PureClip scores of HNRNPH binding sites. The binding sites were assigned to four groups based on the presence of the GGG motif and on the presence of predicted rG4s (prGs). A Kruskal-Wallis test was performed to compare the PureClip scores between the four different groups. The computational analysis was done by [redacted].

3.5.5 Predicted rG4s correlate with cooperativity

The iCLIP data analysis revealed a correlation between strong HNRNPH binding sites and the co-occurrence of predicted rG4s (Figure 18). Furthermore, RNAmaps showed that HNRNPH acts as a splicing repressor when it binds to the AE itself, while binding to the flanking introns correlates with a splicing enhancer function of HNRNPH (Figure 16). Next, it was investigated whether the localisation of predicted rG4s also affects the HNRNPH-mediated splicing regulation. Similar to the RNAmaps approach, genomic regions of HNRNPH-regulated CE events were compared to non-regulated CE events (Ctrl).

Then, within the exon-intron structure of the CE events, the fraction of events having a predicted rG4 was quantified. Interestingly, 35% of the cooperatively repressed events had at least one predicted rG4 within the AE itself, while for the cooperatively enhanced and the Ctrl events less than 12% showed a predicted rG4 in the AE (Figure 19A). On the other hand, 82% of the cooperatively enhanced events contained at least one predicted rG4 in the intronic regions flanking the AE (Figure 19A). In contrast, from the cooperatively repressed and the Ctrl events only around 50% had a predicted rG4 within the intronic region (Figure 19A). In summary, these results highlight that not only the position of HNRNPH binding determines the splicing regulation, but also the occurrence of rG4s correlates with HNRNPH-mediated splicing regulation.

In view of these results, the question remained whether the presence of rG4s plays a role in the strength of the cooperative splicing regulation by HNRNPH. Therefore, the cooperatively regulated CE events were clustered into five groups based on the quantile of their absolute nH values (0-20%, 20-40%, 40-60%, 60-80%, 80-100%). Within these quantiles, the fraction of predicted rG4s was compared. Based on the results from Figure 19A, for the cooperatively repressed CE events the predicted rG4s were counted in the AE, while for cooperatively enhanced CE events the predicted rG4s were quantified in the intronic regions. Interestingly, for the cooperatively repressed CE events the fraction of predicted rG4s increased with higher nH values (Figure 19B). In contrast, for the cooperatively enhanced events the fraction of predicted rG4s did not increase with higher nH values (Figure 19B). Introns are often much longer than exons and therefore can potentially contain many predicted rG4s. However, most likely not all these predicted rG4s are involved in the splicing regulation by HNRNPH. The rG4s that are important for HNRNPH splicing regulation cannot be separated from non-regulative rG4s. Therefore, non-regulative intronic rG4s probably have masked a correlation between the fraction of introns with predicted rG4s and the strength of cooperativity for the cooperatively enhanced events. However, in summary, AE that are cooperatively regulated by HNRNPH show an enrichment for predicted rG4s in the respective regulative regions (cooperatively repressed events: AE; cooperatively enhanced events: Introns). Moreover, for cooperatively repressed exons, the fraction of predicted rG4s correlates with the strength of the HNRNPH-mediated cooperative splicing regulation.

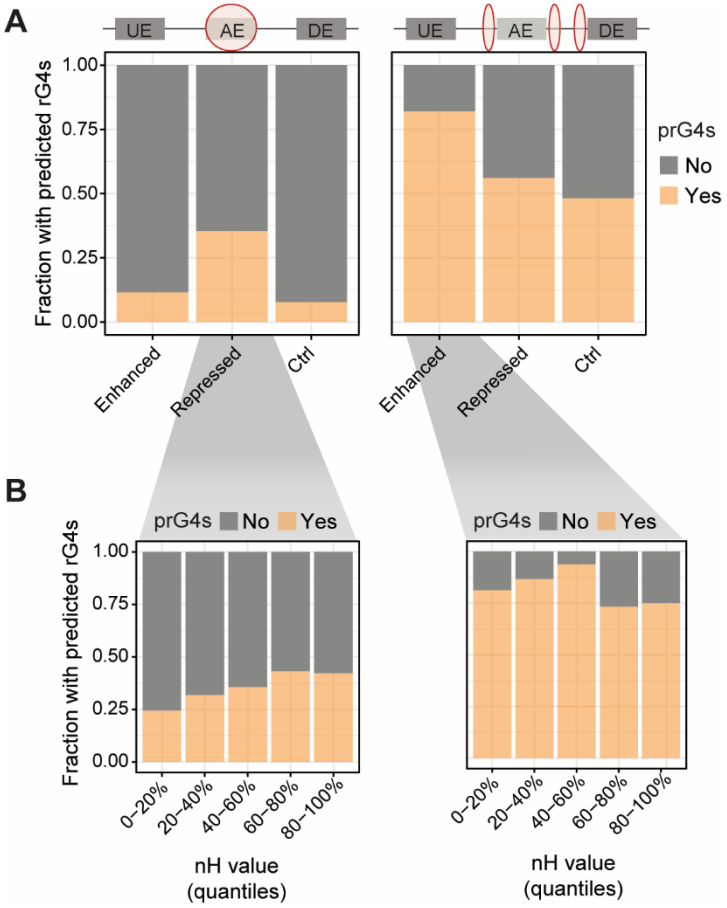

Figure 19: Predicted rG4s overlap with the position-dependent splicing regulation of HNRNPH and correlate with cooperative splicing regulation.
(A) Fraction of cassette exon (CE) events with predicted rG4s (prG4s) from cooperatively enhanced (Enhanced), cooperatively repressed (Repressed) or non-regulated (Ctrl) events. The fractions with prG4s were quantified for regions important for HNRNPH repressive splice factor function (left plot) and for regions important for HNRNPH enhancing splice factor function (right plot). (B) Left plot: Fraction of alternative exons from repressed CE events with prG4s. The CE events were separated into five groups based on the quantile of their absolute Hill coefficient (nH) values. Right plot: Fraction of intronic regions from enhanced CE events with prG4s. The events were separated into five groups based on the quantile of their absolute nH values. The computational analysis was done by [redacted].

Overall, the HNRNPH iCLIP analysis revealed that the majority of HNRNPH binding sites are located in protein-coding genes. The binding site position of HNRNPH determines whether HNRNPH acts as a splicing enhancer or a splicing repressor. Further, the analysis confirmed that HNRNPH preferentially binds to G-tract RNAs and that binding is stronger when the HNRNPH binding sites overlap with predicted rG4s. Moreover, predicted rG4s co-occur with the position-dependent splicing regulation of HNRNPH and correlate with the strength of HNRNPH's cooperative splicing control.

3.6 Emetine dihydrochloride treatment

The iCLIP analysis uncovered a potential interaction between HNRNPH binding and rG4s. However, these analyses were based on predicted rG4s without any experimental evidence for an involvement of rG4s in HNRNPH-mediated splicing regulation. To get a first impression of an interaction between rG4s and HNRNPH-mediated splicing regulation, MCF7 cells were treated with increasing concentrations of the small molecule emetine dihydrochloride. Emetine dihydrochloride has been identified as a drug that stabilises RNA rG4s (Zhang et al, 2019). To validate the treatment, *AKR1A1* splicing was analysed, since alternative splicing of *AKR1A1* exon 7 was described to depend on an rG4 (Dardenne et al., 2014; Zhang et al., 2019). Indeed, treatment with the rG4 stabiliser emetine dihydrochloride reduced the inclusion frequency of the exon 7 of *AKR1A1* (Zhang et al, 2019). (Figure 20A). Having validated the rG4-stabilising effect of the drug emetine dihydrochloride, its influence on the HNRNPH-mediated splicing event *MST1R* exon 11 was investigated. As shown earlier in this work and referring to literature, HNRNPH acts as a splicing repressor promoting *MST1R* exon 11 skipping (Lefave et al., 2011; Braun et al., 2018). With increasing concentrations of the small molecule emetine dihydrochloride, the skipping frequency of *MST1R* exon 11 increased (Figure 20B). Besides this CE splicing event, *MST1R* exon 11 splicing comprises IR isoforms (Braun et al., 2018). Emetine dihydrochloride treatment decreased the expression of these IR splice variants (Figure 20B). In summary, the treatment of MCF7 cells with the small molecule emetine dihydrochloride experimentally validated that rG4-stabilisation can affect the splicing decision of the HNRNPH-mediated splicing event *MST1R* (Figure 20B).

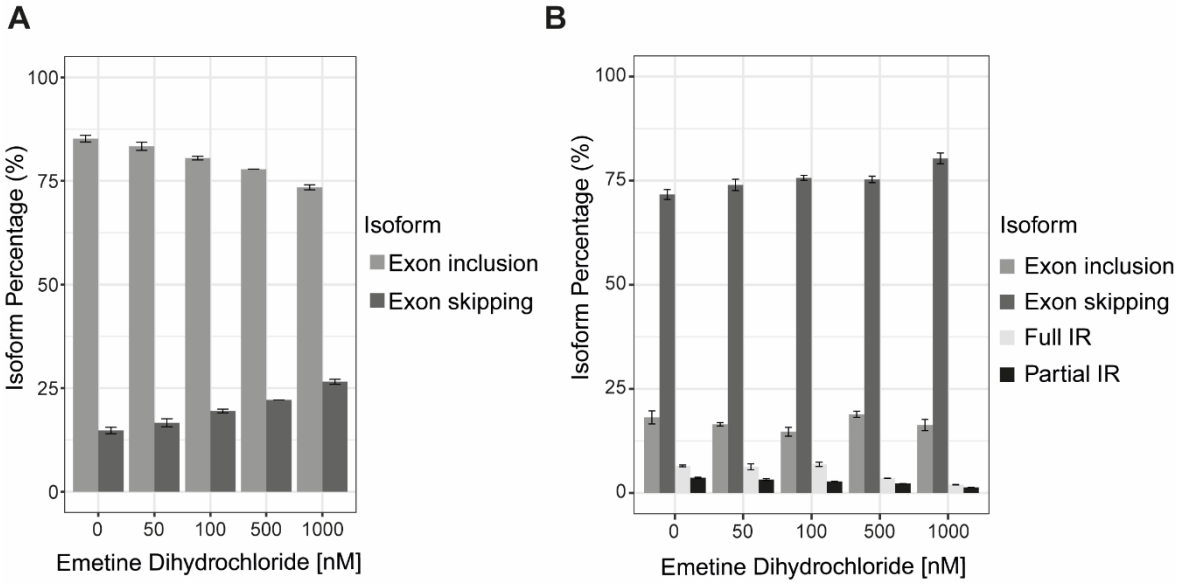

Figure 20: The rG4-stabiliser emetine dihydrochloride affects alternative splicing decisions.
MCF7 cells were treated with increasing concentrations of the small molecule emetine dihydrochloride in two biological replicates (n=2). (A,B) Bar diagram shows the isoform expression of *AKR1A1* exon 7 splice isoforms (A) or *MST1R* exon 11 splice isoforms (B) in response to emetine dihydrochloride treatment. Bars show the mean of two biological replicates and error bars indicate the standard deviation.

3.7 HNRNPH ivRNA library

The analysis of the HNRNPH iCLIP data revealed that the strongest HNRNPH binding sites overlap with predicted rG4s (Figure 18). Hence, it was aimed to experimentally study rG4 formation of HNRNPH binding sites. Instead of studying only individually selected HNRNPH binding sites, rG4 formation of HNRNPH binding sites was studied in a high-throughput manner. Therefore, an *in vitro* oligonucleotide library (ivRNA library) was designed (see chapter 2.2.15). In total, 3,000 genomic regions were included in the ivRNA library and each region was represented by four individually barcoded constructs, finally ending up with an ivRNA library containing 12,000 constructs. The library consisted of 1,666 regions from HNRNPH-regulated CE splicing events that are cooperatively regulated in the *in vivo* titration of HNRNPH in MCF7 cells. Additionally, the ivRNA library contained 1,334 regions from CE events that are not regulated by HNRNPH. The regions included in the ivRNA library were chosen from windows spanning from all splice sites of the respective CE event (Figure 21). Within these windows it was screened for regions with HNRNPH binding sites (finally included regions: 1409), predicted rG4s (finally included regions: 145) or HNRNPH binding sites overlapping with predicted rG4s (finally included regions: 1340) (Figure 21). The chosen regions were put into individual oligonucleotides at position 51 including 146 nt of the respective flanking genomic sequence (Figure 21). The

ivRNA library was used for high-throughput studies to experimentally validate rG4 formation and to study intrinsic characteristics of HNRNPH binding to RNAs.

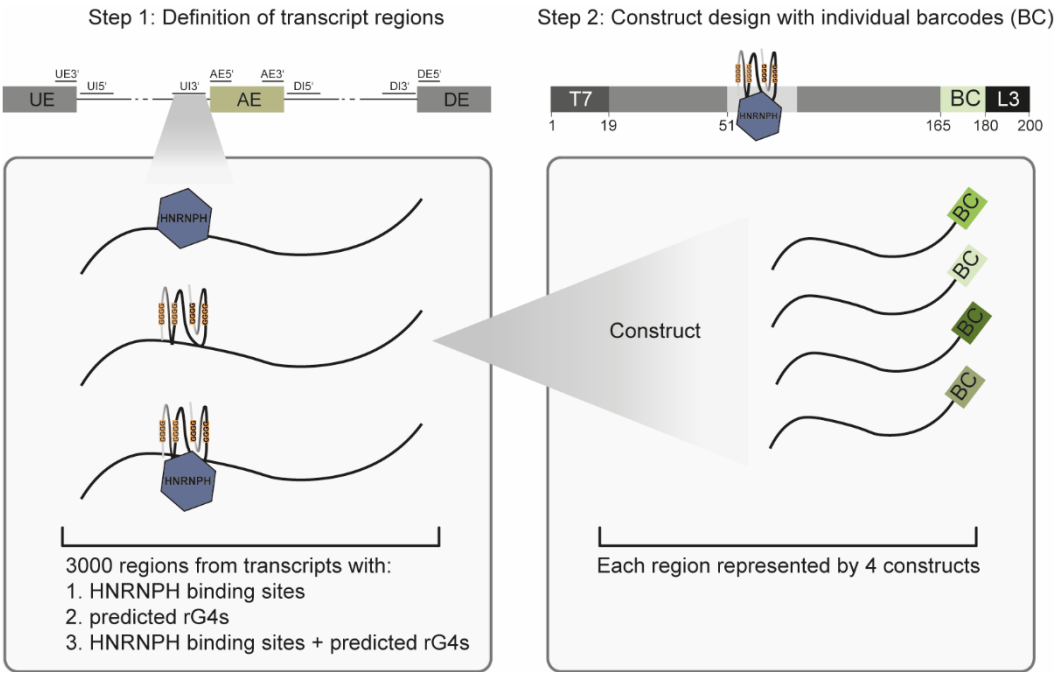

Figure 21: Design of the oligonucleotide library (ivRNA library) to analyse HNRNPH splicing regulation.
Cassette exon events with significant HNRNPH regulation were chosen for the oligonucleotide library design. To choose the sequences for the oligonucleotide library, windows were defined for the 3' end of the upstream exon (UE3'), the 5' end (UI5'; DI5') and the 3' end (UI3', DI3') of the upstream and downstream introns, the 5' end (AE5') and the 3' end (AE3') of the alternative exon, and the 5' end (DE5') of the downstream exon. Within these windows, regions with HNRNPH binding sites, predicted rG4s or a combination of both criteria were chosen for the final library design. Each identified region was designed into an individual oligonucleotide construct putting the respective region at construct position 51. The final constructs were 200 nucleotides long with the T7 promotor a position 1-19, an oligonucleotide individual barcode (BC) at position 165-180 and the iCLIP2-L3 linker sequence at position 180-200. Each region was represented by four constructs with individual BCs.

3.8 RT-Stop profiling

Predicted rG4s accumulate in regions that are important for HNRNPH-mediated splicing regulation (Figure 19). To experimentally validate the occurrence of rG4s in HNRNPH binding sites, the approach of the RT-Stop assay was adapted to map rG4 formation in high-throughput *in vitro* (Guo and Bartel, 2016; see chapter 2.2.21). The adapted protocol (RT-Stop profiling) allowed it to map rG4 formation of 12,000 synthetic RNA oligonucleotides in parallel using the *in vitro* transcribed ivRNA library as input (see chapter 2.2.15). The RT-Stop profiling experiment was performed in two set ups:

I) ivRNA library *in vitro* transcribed with GTP.
II) ivRNA library *in vitro* transcribed with 7-deaza-GTP.

Within both set ups the RT-Stop profiling experiments were performed in an rG4-stabilising condition (KCl) and in an rG4-destabilising condition (NaCl). The rG4-stabilising condition was expected to result in RT-truncation (RT-Stop) at rG4-forming sites because the RT cannot pass the rG4 structure (Figure 22A, Guiset Miserachs et al., 2016). In both experimental set-ups the RT-Stop profiling was performed in three technical replicates for all conditions. In the end, all samples were pooled into one final RT-Stop profiling library for NGS.

Inspecting the cDNA length profile after RT of the ivRNA library, the cDNAs from the RT under rG4-stabilising conditions were overall shorter in length than the cDNAs from the RT under rG4-disfavouring conditions, indicating the occurrence of RT-Stops (Figure 22B). To validate that the shorter cDNAs in rG4-stabilising conditions really arose from rG4s, the RT was also performed with an ivRNA library for which GTP has been exchanged with 7-deaza-GTP. 7-deaza-GTP is missing the N7 atom in the guanine base, hence Hoogsten base pairing cannot appear anymore and consequently rG4s are completely abolished. Without the possibility of rG4 formation, the cDNA length profile of the ivRNA library did not show any difference between rG4-stabilising and rG4-destabilising conditions, demonstrating the validity of the approach (Figure 22B). In summary, RT of the ivRNA library showed that RT-Stops occur under rG4-stabilising conditions.

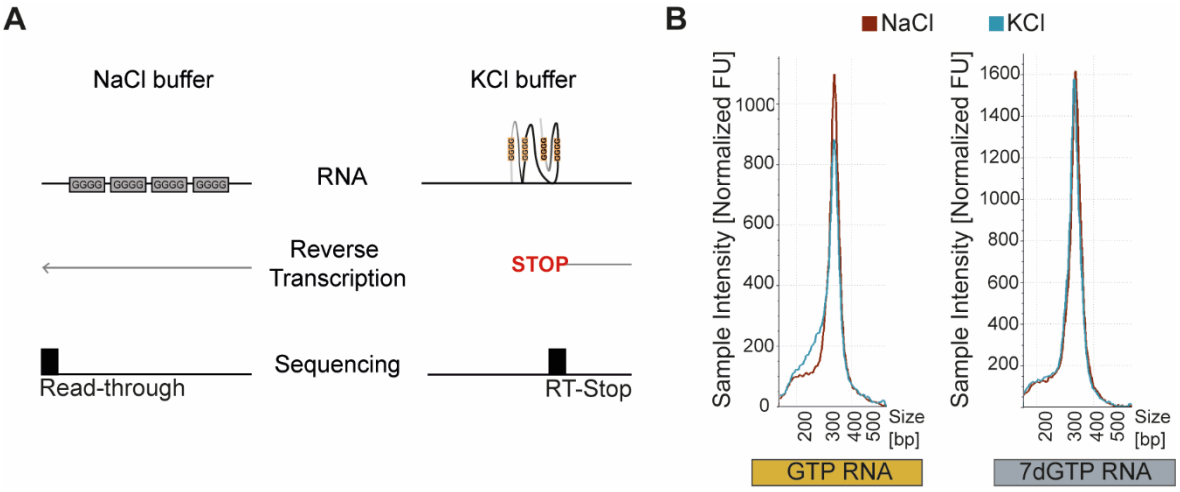

Figure 22: Reverse Transcriptase Stop Profiling of the ivRNA library reveals RT truncations in the presence of KCl.
(A) Reverse transcription of the ivRNA library was performed in NaCl and KCl buffer conditions. Then, library preparation was performed for the cDNA products and next generation sequencing was used to map the position of the last nucleotide of each cDNA product to the original ivRNA oligonucleotide sequence. (B) Electropherogram of ivRNA library cDNA products after reverse transcription with KCl (red) or NaCl (blue) buffer using GTP ivRNA library or 7-deaza-GTP (7dGTP) ivRNA library as input material for the reverse transcription.

3.8.1 RT-Stops align with G-tracts

Globally, the cDNA length profile after RT of the ivRNA library in rG4-stabilising conditions showed the appearance of RT-Stops (Figure 22B). Next, the RT-Stop profiling library was sequenced, resulting in total 138,868,479 reads. The reads were mapped to the 12,000 ivRNA library constructs.

Exemplary visual inspection of RT-Stops after RT-Stop profiling showed that RT-Stops mainly occurred in rG4-stabilising conditions with KCl and disappeared with the usage of NaCl for the RT (Figure 23). The truncated RT-Stops aligned very well with G-tracts that were interspersed by short loops (Figure 23). RT-Stop profiling of the ivRNA library with 7-deaza-GTP did not result in RT-Stops that mapped to G-tracts, indicating that rG4 structures do not form with 7-deaza-GTP (Figure 23). Additionally, using the 7-deaza-GTP ivRNA library, the RT-Stop profiles of KCl and NaCl conditions looked overall very similar (Figure 23). Overall, the visual data analysis confirmed that the RT-Stop profiling generates RT-Stops, which preferentially form in rG4-favouring experimental conditions and align with G-tract sequences.

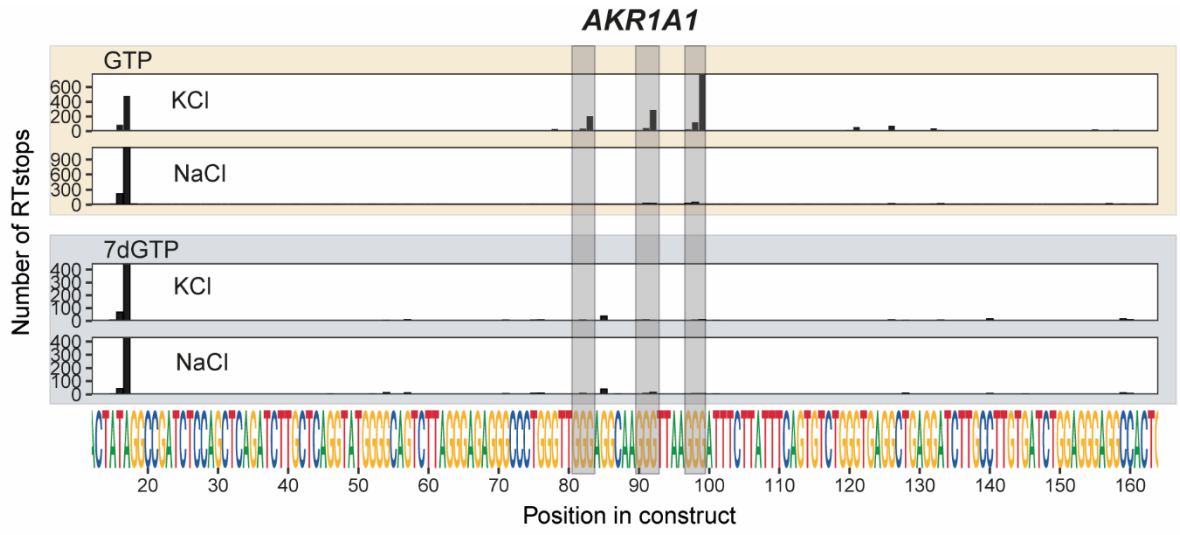

Figure 23: Genome browser screenshot shows alignment of RT-Stops with G-tracts.
The upper panel (light yellow) shows the number of RT-Stops for *AKR1A1* after RT-Stop profiling of the GTP ivRNA library. RT-Stops that align with G-rich sequences are highlighted in grey. The lower panel (light grey) shows the RT-Stops for *AKR1A1* after RT-Stop profiling using 7-deaza-GTP (7dGTP) ivRNA library.

3.8.2 RT-Stop profiling identifies rG4 structures in the ivRNA library

The visual data analysis highlighted the formation of G-tract-aligned RT-Stops that form in the rG4-favouring KCl condition and disappear in the NaCl condition (Figure 23). The disappearance of RT-Stops at G-tracts in the NaCl condition allows determining positions

of rG4s within the ivRNA library. For the final analysis, only ivRNA library constructs that did not show more than a 2-fold global expression change between the experimental RT-Stop profiling conditions were used, ending up with in total 7573 analysed constructs (Figure 24A). To determine rG4s within the ivRNA library, all RT-Stops (RT-Stop peaks) were extracted from the sequenced RT-Stop profiling library; for the KCl and for the NaCl conditions. Now, log2-fold changes were calculated for RT-Stop peaks comparing NaCl with KCl, which showed an enrichment for RT-Stop peaks that disappeared in the NaCl condition (Figure 24B). More than 25% of all detected RT-Stop peaks significantly went down in the NaCl condition, classifying these peaks as rG4 peaks (Figure 24B). The remaining RT-Stop peaks either significantly went up with NaCl (< 12.5%) or did not show a significant change between KCl and NaCl (Figure 24B). In total, from the 7573 analysed constructs, 3576 constructs contained at least one significant rG4 peak, classifying these constructs as rG4 constructs. The ivRNA library consisted of 12,000 constructs based on 3000 defined genomic regions, while each genomic region was represented by four constructs (n=4) (see chapter 3.7). The information about the rG4 formation on construct level was merged for the four construct replicates to classify the 3000 genomic regions into rG4 and non-rG4 regions. Only regions, for which at least two out of the four construct replicates have been classified as rG4 constructs, were finally determined as rG4 regions (Figure 24C). In summary, RT-Stop profiling of the ivRNA library identified 1,109 non-rG4 regions and 1,060 rG4 regions, while for 70% of the analysed HNRNPH binding sites overlapping with predicted rG4s the rG4 formation could be experimentally validated (Figure 24C). The classification of 1,060 rG4 regions within the ivRNA library highlights that HNRNPH binding sites indeed contain sequences that can form rG4s.

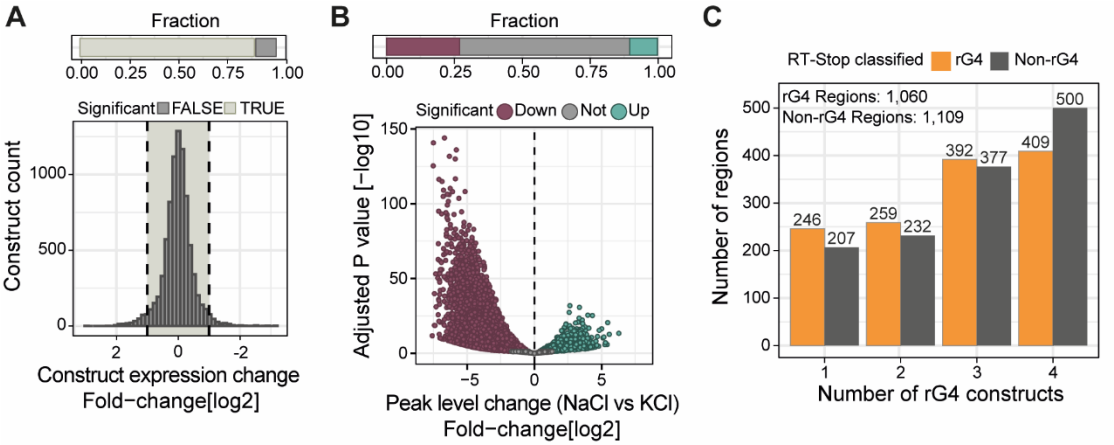

Figure 24: RT-Stop profiling analysis of the ivRNA library experimentally maps rG4s.
(A) The histogram shows the log2-transformed expression fold-changes of all ivRNA library constructs calculated with DESeq2 from the NGS sequencing data of the ivRNA RT-Stop profiling. For further analyses, all constructs with more than a 2-fold global expression changes were removed. (B) Volcano plot showing the DESeq2-calculated significance levels of all log2-transformed RT-Stop peak level fold-changes between experimental conditions (NaCl condition vs. KCl condition). RT-Stop peaks that are significantly reduced in the NaCl condition get the label "Down" and are classified as rG4 constructs. In contrast, RT-Stop peaks that are significantly higher in the NaCl condition are labelled with "Up" and are classified as non-rG4 constructs. The Significance threshold was set to a p-value < 0.1. (C) All ivRNA library constructs are represented as quadruplicates in the ivRNA library and each construct quadruplicate reflects one genomic input region. The bar plot describes the construct agreement of all construct quadruplicates regarding their classification as an rG4 or a non-rG4 construct. Genomic input regions, for which at least two out of the four construct quadruplicates agree in their rG4 classification, are referred as rG4 region or non-rG4 region. The computational analysis was done by [redacted].

3.9 *In vitro* iCLIP

RT-Stop profiling of the ivRNA library experimentally confirmed the formation of rG4 structures around HNRNPH binding sites (Figure 24). Having experimentally validated the accumulation of rG4s around HNRNPH binding sites, it was of interest to understand whether HNRNPH binds differentially to rG4 and non-rG4 RNAs. To address this question, an adapted *in vitro* iCLIP experiment (ivCLIP) was performed, analysing the binding of recombinant HNRNPH1 protein to the ivRNA library (Sutandy et al., 2018, see chapter 2.2.23.5). In short, after radioactive labelling, the ivRNA library was incubated with recombinant HNRNPH1 protein and next, the protein-RNA complexes were covalently crosslinked using UV-light (Figure 25). After size-separation, the protein-RNA complexes were isolated, followed by ivCLIP library preparation for high-throughput sequencing as it has been described by Sutandy and colleagues (Sutandy et al., 2018; Figure 25). The ivCLIP experiment was performed for the ivRNA library *in vitro* transcribed with GTP and for the ivRNA library *in vitro* transcribed with 7-deaza-GTP. Both experimental conditions were conducted with four technical replicates and for sequencing all samples were pooled into one final ivCLIP library. Sequencing of the ivCLIP library resulted in 16-25 million reads per sample.

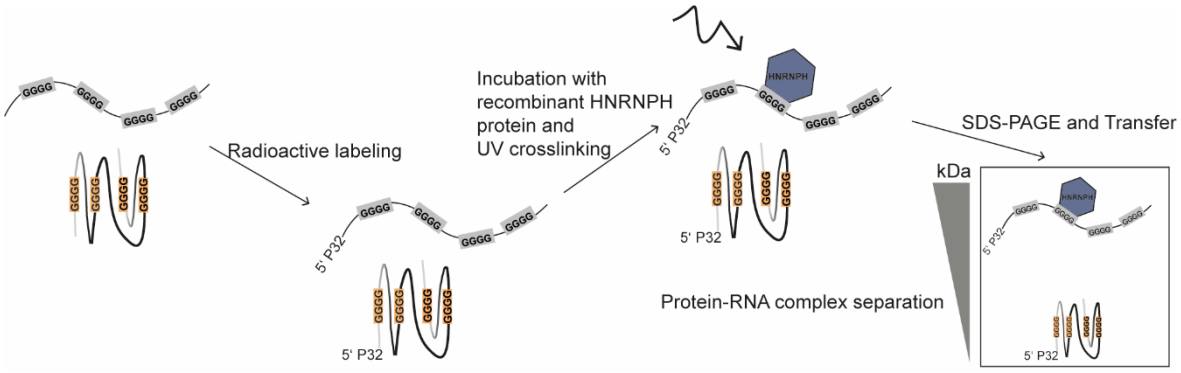

Figure 25: Schematic illustration of the HNRNPH1 *in vitro* binding assay.
First, the RNA of interest is radioactively labelled. Next, the RNA is incubated with recombinant HNRNPH1 protein and formed protein-RNA complexes are covalently crosslinked using UV light. The protein-RNA complexes are size-separated from unbound RNAs using an SDS-PAGE. Finally, the size-separated protein-RNA complexes are transferred to a nitrocellulose membrane and the radioactively labelled RNA is visualised using a Storage Phosphor Screen (GE Healthcare).

To elucidate whether HNRNPH1 binding differs between rG4 and non-rG4 RNAs, the HNRNPH1 binding signal was quantified from the ivCLIP experiment, grouping the ivRNA library into rG4 and non-rG4 RNAs based on the RT-Stop profiling results. Recombinant HNRNPH1 bound stronger to rG4 RNAs than to non-rG4 RNAs (Figure 26A). However, in rG4-disfavouring condition (7-deaza-GTP ivRNA library), the recombinant HNRNPH1 binding did not differ that strongly anymore between rG4 and non-rG4 RNAs (Figure 26B). Direct comparison between rG4-favouring (GTP ivRNA library) and rG4-disfavouring experimental conditions (7-deaza-GTP ivRNA library) showed that in the rG4-favouring conditions recombinant HNRNPH1 protein bound enriched to RNAs with strong rG4s (Figure 26C). Moreover, the enriched HNRNPH1 binding to rG4 RNAs only appeared for RNAs with an experimentally validated rG4 structure and not for RNAs that had just a predicted rG4 (Figure 26D). Overall, integration of the RT-Stop profiling analysis with the HNRNPH1 ivCLIP highlights that HNRNPH1 has an intrinsic binding preference for rG4 RNAs.

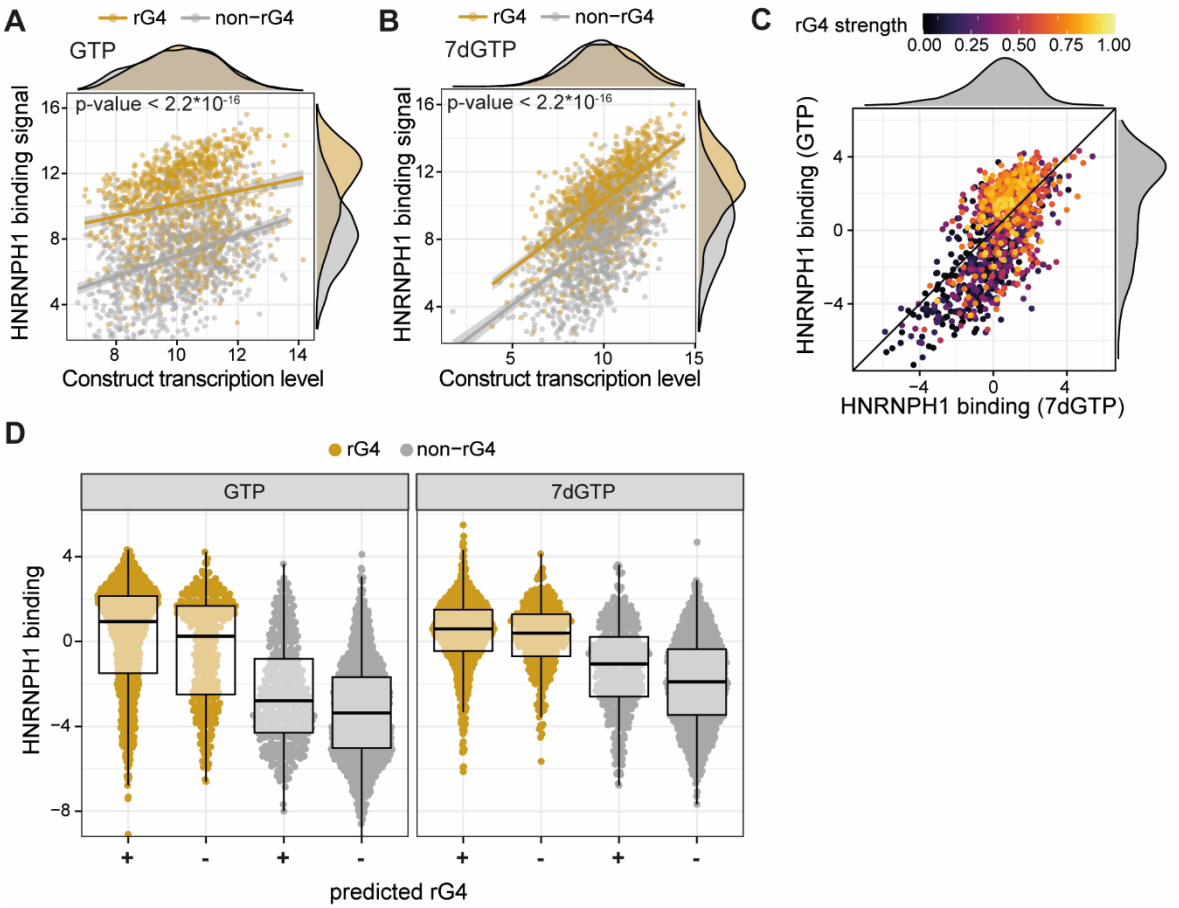

Figure 26: HNRNPH1 *in vitro* iCLIP (ivCLIP) identifies enriched binding of HNRNPH1 to RNAs with rG4s.
(A, B) The scatterplots show the recombinant HNRNPH1 binding signal, quantified from the ivCLIP experiment, to RT-Stop profiling classified rG4 and non-rG4 constructs in relation to the construct transcription levels. The construct transcription levels were taken from the ivRNA library RT-Stop profiling. The recombinant HNRNPH1 binding is shown for GTP RNA (A) and for 7-deaza-GTP (7dGTP) RNA conditions (B). All values are log2-transformed. (C) The recombinant HNNRPH1 binding signal was normalised to the construct expression levels. The scatterplot plots the normalised HNRNPH1 binding of the GTP condition against the 7dGTP condition. The data points (constructs) were coloured by their rG4 strength, which has been quantified from the RT-Stop profiling. (D) The plots show the normalised HNRNPH1 binding, for GTP and 7dGTP conditions, to rG4 and non-rG4 constructs, which are further splitted into groups with *in silico* predicted rG4s and groups without *in silico* predicted rG4s. The computational analysis was done by [redacted].

3.10 *In vitro* binding studies to extract effects of rG4s for HNRNPH binding

The high-throughput RT-Stop profiling experimentally validated that HNRNPH binding sites are likely to form rG4s *in vitro* (Figure 24). Moreover, ivCLIP experiments using recombinant HNRNPH1 protein revealed that HNRNPH1 has a preference to bind to rG4 RNAs (Figure 26). For the cooperative binding of RBPs to RNAs, it has been shown that RNA structure could act as the mediator of cooperativity (Lin and Bundschuh, 2013; Lin and Bundschuh, 2015). Since HNRNPH controls alternative splicing cooperatively on a transcriptome-wide scale, it was investigated, whether rG4 structures are involved in the

cooperative splicing regulation of HNRNPH. To address this question, several *in vitro* binding studies using recombinant HNRNPH1 protein were performed:

I) *In vitro* titration of recombinant HNRNPH1 protein with rG4 and non-rG4 RNAs to extract Hill fittings.
II) *In vitro* binding of recombinant HNRNPH1 protein in different buffer conditions to understand binding preferences of HNRNPH.
III) *In vitro* FRET analyses to further validate binding preferences of HNRNPH.

3.10.1 *In vitro* titration of recombinant HNRNPH1 protein

In order to understand whether rG4 formation of RNAs can mediate cooperative binding of HNRNPH, the *in vitro* binding of recombinant HNRNPH1 protein to RNAs with and without rG4s was investigated using increasing concentrations of the protein (*in vitro* titration). To determine the effect of rG4s for the binding of HNRNPH, the binding of recombinant HNRNPH1 protein to RNA was always compared between RNAs *in vitro* transcribed with GTP and RNAs *in vitro* transcribed with 7-deaza-GTP. In total, the *in vitro* titration of HNRNPH1 was performed for three different RNAs:

I) rG4 RNA: *ARPC2*
II) rG4 RNA: *MYL6*
III) non-rG4 RNA: *CD46.*

The well-known rG4 structure *ARPC2*, for which RNA pulldown-assays have shown HNRNPH binding to the RNA, served as an rG4 positive control (Hacht et al., 2014). The sequences of *MYL6* and *CD46* were chosen based on the RT-Stop profiling results. While the RT-Stop profiling revealed an rG4 structure for *MYL6*, *CD46* did not show any rG4 structures in the RT-Stop profiling (Figure 27A,B).

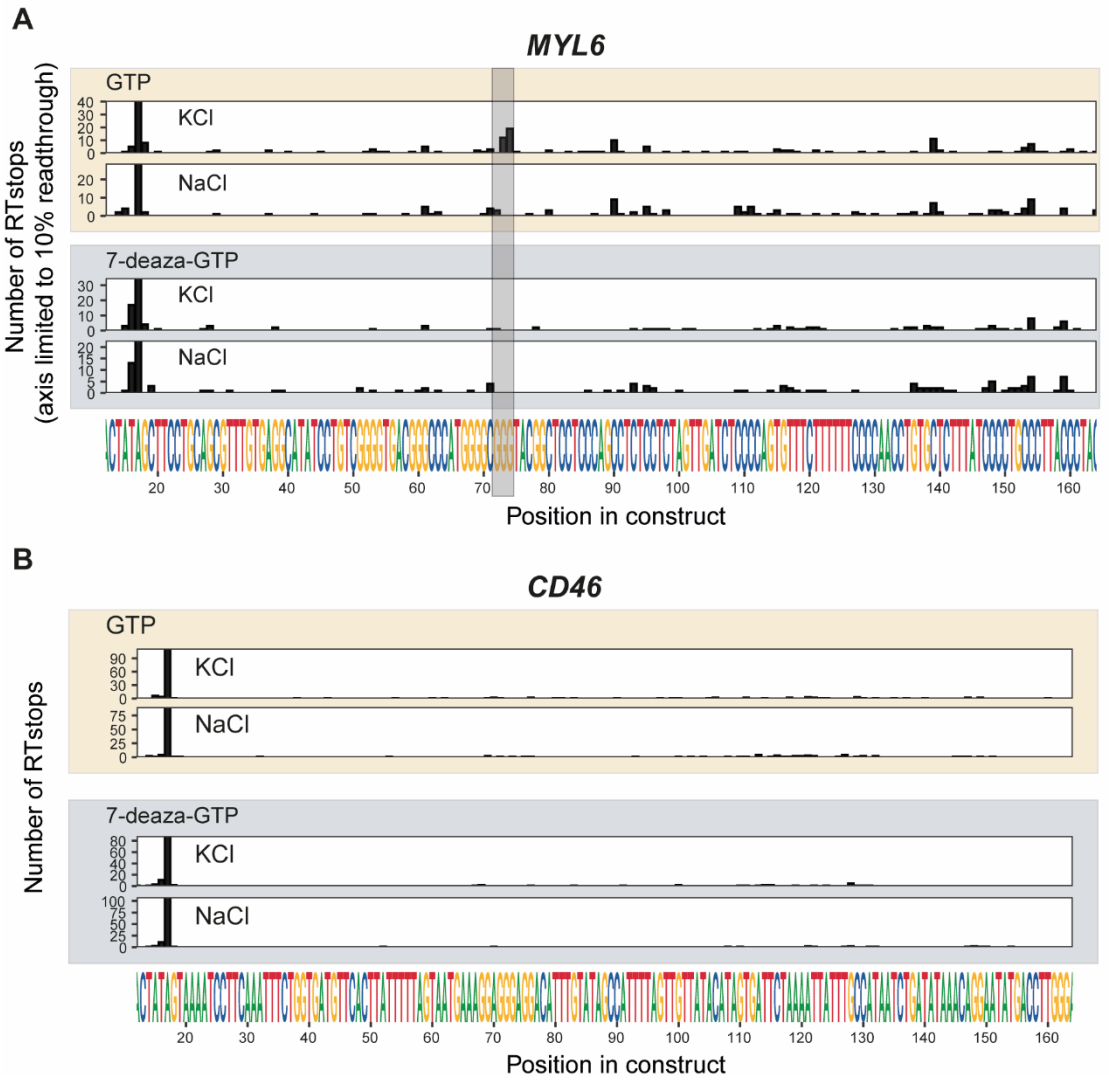

Figure 27: Genome browser screenshots showing RT-Stops for the constructs *MYL6* (A) and *CD46* (B).
The upper panel (light yellow) shows the RT-Stops after RT-Stop profiling of the GTP ivRNA library. RT-Stops that align with G-rich sequences are highlighted in grey. The lower panel (light grey) shows the RT-Stops after RT-Stop profiling using 7-deaza-GTP ivRNA library.

Further, rG4 formation of *ARPC2* and *MYL6* was validated with CD spectroscopy. *ARPC2* and *MYL6* both showed the characteristic CD profile for parallel-stranded rG4s with a positive peak at ~264 nm and a negative peak at ~240 nm and the peaks were most prominent in the presence of KCl (Figure 28A,B) (Vorlíčková et al., 2012; Randazzo et al., 2013). In presence of KCl, the peak intensity at 264 nm was the highest for *ARPC2* and *MYL6*, confirming that KCl strongly stabilises rG4 structures (Figure 28A,B) (Guiset Miserachs et al., 2016). In summary, CD spectroscopy confirmed that the RNAs *ARPC2* and *MYL6* form parallel-stranded rG4s *in vitro* that are most stable in the presence of KCl. *CD46* RNA also showed a CD spectrum with a positive and negative peak and both peaks showed a dependency on KCl (Figure 28C). However, the positive and negative peaks of the *CD46*

RNA CD spectrum were slightly shifted compared to a CD spectrum typical for a parallel-stranded rG4 (Figure 28C). Moreover, the overall signal of the measured CD spectra for *CD46* was lower than the respective signals from *ARPC2* or *MYL6* (Figure 28C). In summary, the CD spectrum of *CD46* indicates that *CD46* RNA is not forming a stable, parallel-stranded rG4.

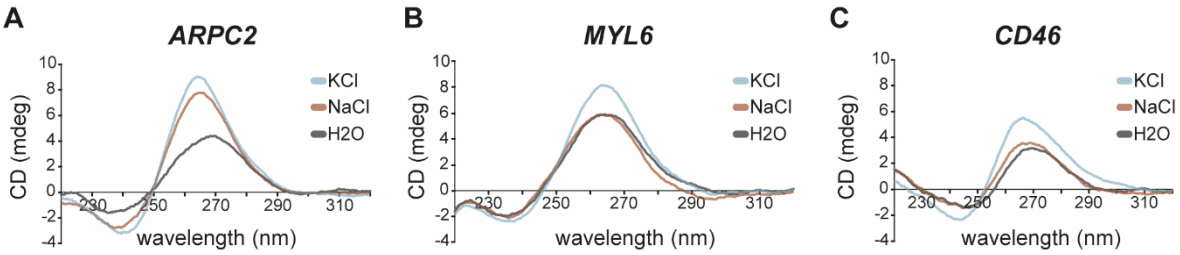

Figure 28: Circular dichroism spectra for the RNA oligonucleotides *ARPC2* (A), *MYL6* (B) and *CD46* (C).
The circular dichroism (CD) spectra for the RNA oligonucleotides *ARPC2.von.Hacht (ARPC2)* and *MYL6* show characteristics of parallel-stranded rG4 structures with a positive peak at ~264 nm and a negative peak at ~240 nm (see Table 4). CD spectra of the RNA oligonucleotides were measured with monovalent cations (150 mM KCl or 150 mM NaCl) or without the addition of salt (H2O). In the presence of KCl, the rG4-characteristic peaks are most prominent.

Having tested rG4 formation for *ARPC2* and *MYL6*, the *in vitro* binding analysis of titrated recombinant HNRNPH1 protein to RNA was performed as shown in Figure 25. The titration of recombinant HNRNPH1 protein to RNA and the subsequent quantification of the *in vitro* formed protein-RNA complexes allowed extracting dose-response curves using the Hill function. For the rG4-forming RNAs *ARPC2* and *MYL6*, the *in vitro* titrations revealed that protein-RNA complex formation appeared in a switch-like manner (Figure 29A,C). However, such a switch-like protein-RNA-binding was only detectable for GTP RNA and not for 7-deaza-GTP RNA (Figure 29B,D). To extract the degree of cooperativity, Hill fittings were performed for the *in vitro* titrations. The Hill fittings showed that HNRNPH1 bound with positive cooperativity to the rG4 RNAs *ARPC2* and *MYL6* (*ARPC2*: mean nH value =1.86; *MYL6*: mean nH value =1.8) (Figure 29A,C). Contrary, without rG4s (7-deaza-GTP conditions) the degree of cooperativity for HNRNPH1 binding to *ARPC2* or *MYL6* RNA was significantly lower with Hill coefficients closer to one (Figure 29B,D). Overall, these data indicate that rG4 formation is involved in the cooperative binding of HNRNPH1 to RNAs.

To validate this hypothesis further, the *in vitro* titration of recombinant HNRNPH1 was also performed for the non-rG4 RNA *CD46*. HNRNPH1 did not bind in a switch-like manner to *CD46*, neither to GTP nor to 7-deaza-GTP RNA (Figure 29E,F). In addition, the Hill fittings, having Hill coefficients close to one, confirmed that HNRNPH1 did not bind cooperatively to *CD46* RNA (GTP mean nH value = 0.88; 7-deaza-GTP RNA mean nH value =1.16) (Figure 29E,F). In summary, the *in vitro* titration analyses highlight that HNRNPH1 can bind cooperatively to RNAs, but the cooperativity is driven by the RNA's ability to form an rG4 (Figure 26G). However, it is still unclear whether HNRNPH in the end binds to folded or to unfolded rG4 sequences.

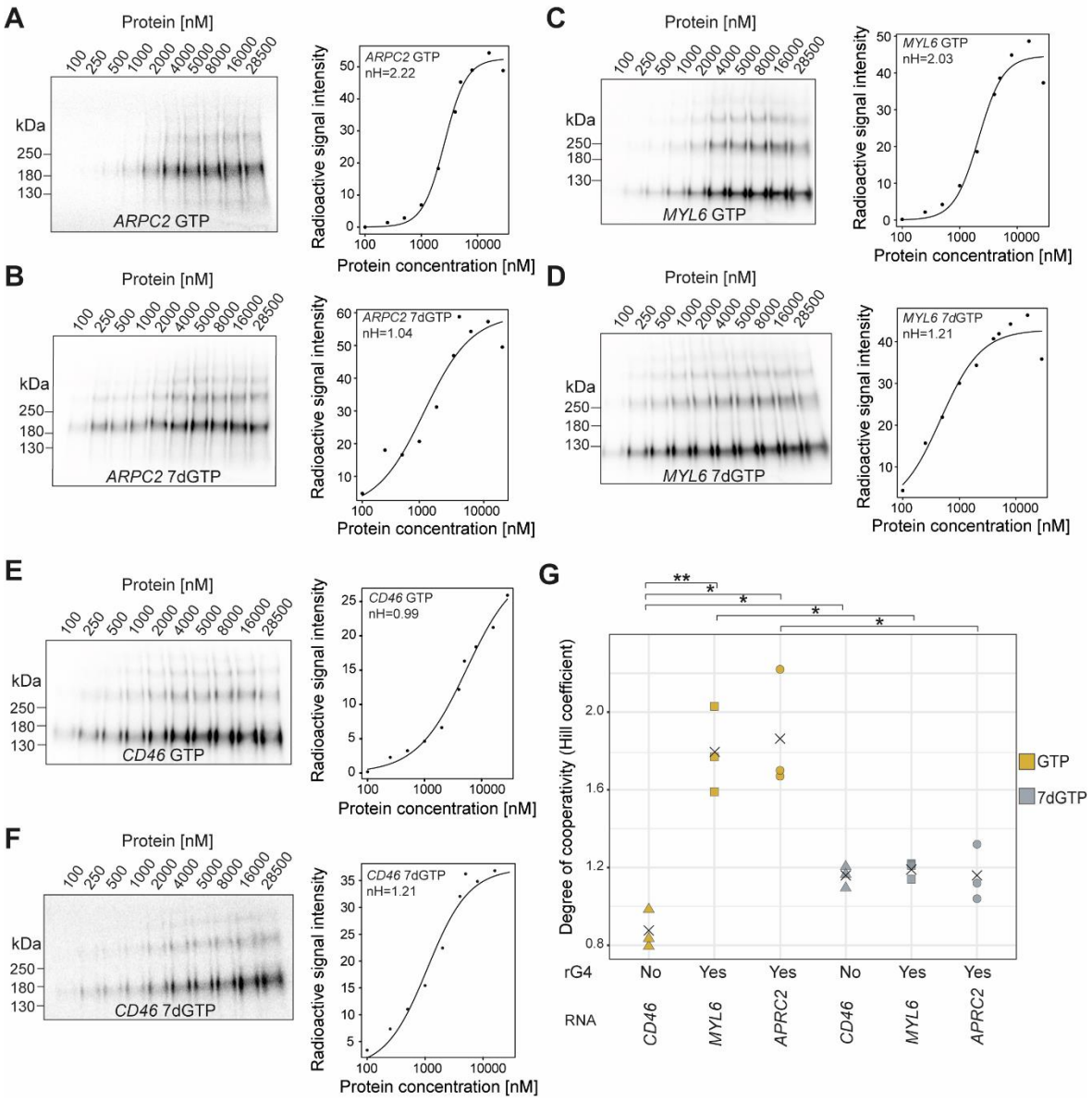

Figure 29: Recombinant HNRNPH1 protein binds cooperatively to RNAs that can form rG4s.
(A, B) Representative autoradiograms of the *in vitro* titration series of recombinant HNRNPH1 protein with *ARPC2* RNA (o KT122, see Table 3) that was *in vitro* transcribed with GTP (A) or 7-deaza-GTP (7dGTP) (B). Scatterplots show the signal intensity of the radioactively labelled RNA from the protein-RNA complexes, quantified from the autoradiograms, plotted against recombinant HNRNPH1 protein concentration. Hill equations were fitted to the scatterplots to test for cooperativity using the Hill coefficient (nH). (C, D) Representative autoradiograms show the binding of recombinant HNRNPH1 in increasing concentrations to the radioactively labelled RNA *MYL6* (o KT72, see Table 3) that was *in vitro* transcribed with GTP (C) or 7dGTP (D). Scatterplots show the signal intensity of the radioactively labelled RNA from the protein-RNA complexes, quantified from the autoradiograms, plotted against recombinant HNRNPH1 protein concentration. Hill equations were fitted to the scatterplots to test for cooperativity using the Hill coefficient (nH). (E, F) Representative autoradiograms show the binding of recombinant HNRNPH1 in increasing concentrations to the radioactively labelled RNA *CD46* (o KT 225, see Table 3) that was *in vitro* transcribed with GTP (E) or 7dGTP (F). Scatterplots show the signal intensity of the radioactively labelled RNA from the protein-RNA complexes, quantified from the autoradiograms, plotted against recombinant HNRNPH1 protein concentration. Hill equations were fitted to the scatterplots to test for cooperativity using the Hill coefficient (nH). (G) Plot shows the degree of cooperativity (Hill coefficient) for recombinant HNRNPH1 protein titration to rG4 and non-rG4 RNAs. The experiments were performed in three individual replicates using RNA *in vitro* transcribed with GTP (yellow) and RNA *in vitro* transcribed with 7dGTP (grey). The plot shows the individual values of the three replicates and the mean of the replicates (indicated with a cross). To test for significance, a Welch Two-Sample t-test was performed. Significance levels: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

3.10.2 HNRNPH1 binds stronger to unfolded rG4s *in vitro*

The *in vitro* titration of recombinant HNRNPH1 protein revealed that HNRNPH1 binds cooperatively to RNAs that can form rG4s. Further, the ivCLIP experiment has shown that recombinant HNRNPH1 binds enriched to rG4 RNAs. Next, it was investigated whether HNRNPH1 in the end binds to the folded rG4s or to the respective unfolded RNA sequence. To address this question, recombinant HNRNPH1 protein binding to RNAs was compared between rG4-favouring conditions (KCl) and rG4-disfavouring conditions (NaCl). The *in vitro* binding experiment showed that even without any Gs in the RNA sequence, HNRNPH1 binds weaker to RNAs in the KCl binding buffer (Mean fraction HNRNPH1 binding KCl/NaCl=0.79) (Figure 30A,B). This indicates that recombinant HNRNPH1 protein has an intrinsic preference for the NaCl binding buffer compared to KCl binding buffer, which is independent of rG4s. For the RNA *CD46* that was not forming a strong parallel-stranded rG4 in previous validation experiments (RT-Stop profiling and CD spectroscopy), HNRNPH1 bound even weaker to the RNA in KCl (Figure 30A,B). With a KCl/NaCl fraction of 0.57, HNRNPH1 binding to *CD46* RNA was significantly lower in the KCl binding buffer compared to the G-independent HNRNPH1 binding preferences (p-value < 0.05, Figure 30B). Inspecting HNRNPH1 binding to the strong rG4-forming RNA *ARPC2* revealed that here the binding of HNRNPH1 to the RNA in KCl binding buffer was even lower than for *CD46* (*ARPC2*: Mean fraction HNRNPH1 binding KCl/NaCl=0.43, p-value < 0.01) (Figure 30B). Overall, the data show that the recombinant HNRNPH1 protein generally prefers NaCl for RNA-binding. However, in the presence of potential rG4-forming RNAs, the preference of HNRNPH1 for the NaCl binding buffer significantly increases. NaCl disfavours stable rG4-formation; hence, the results here indicate that HNRNPH1 binds stronger to unfolded rG4s.

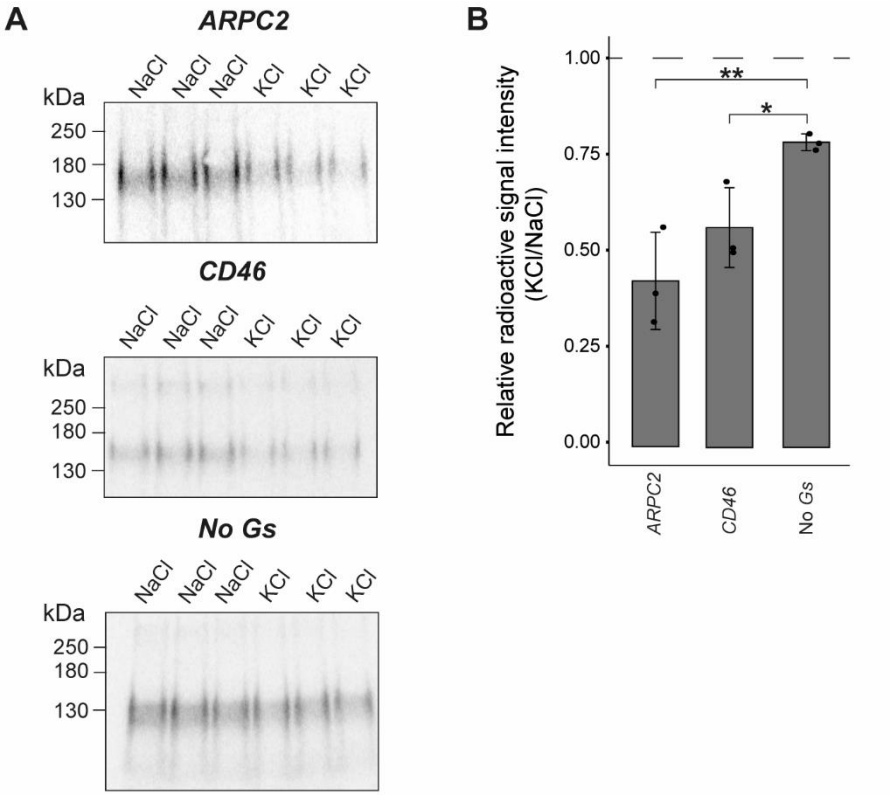

Figure 30: Recombinant HNRNPH1 prefers NaCl for RNA-binding and this preference increases for RNAs with potential rG4s.
(A) Autoradiograms showing the binding of recombinant HNRNPH1 protein to the RNA oligonucleotides *ARPC2* (o KT122, see Table 3) and *CD46* (o KT225, see Table 3) and to an RNA oligonucleotide without guanine nucleotides (No Gs; o KT272, see Table 3). The RNA oligonucleotides were radioactively labelled. The binding of recombinant HNRNPH1 protein to RNA was performed in 150 mM NaCl or in 150 mM KCl binding buffer. (B) Recombinant HNRNPH1 protein binds stronger to RNAs without rG4s. Barplot shows the signal intensity of the radioactively labelled RNA from the protein-RNA complexes in binding buffer containing KCl relative to the signal intensity of the protein-RNA complexes in binding buffer containing NaCl. The experiments were performed in three replicates. A test for significance between the results for the different RNAs was performed using the Welch Two-Sample t-test. Significance levels: * p-value < 0.05, ** p-value < 0.01.

3.10.3 HNRNPH1 can unfold rG4s *in vitro*

The analysis of the RNA-binding preferences for recombinant HNRNPH1 protein showed that HNRNPH1 binds stronger to a potential rG4 RNA in rG4-disfavouring conditions (Figure 30). However, a direct readout between HNRNPH1 binding and the unfolded state of rG4s was missing. Hence, an rG4 RNA was designed containing the donor fluorophore FAM at the 5'-end and the acceptor fluorophore TAMRA at the 3'-end (FRET-RNA) (Figure 31A). In case of a folded rG4, excited FAM transfers the emitted light to the acceptor TAMRA via FRET, which then quenches the FAM emission signal (Cian et al., 2007; Malina et al., 2020; Figure 31A). Therefore, measuring the FAM emission can indicate the state of rG4 folding. To guarantee unfolded rG4 RNA as a positive control, the FRET-RNA emission was measured in LiCl buffer, since LiCl has the lowest stabilising effect on rG4 structure formation (Millevoi et al., 2012; Guiset Miserachs et al., 2016). Relative to the

FAM emission of folded FRET-RNA (KCl), the FAM emission for the positive control (LiCl) was significantly higher (Mean relative FAM emission LiCl=1.2, p-value < 0.05) (Figure 31B). Next, increasing concentrations of recombinant HNRNPH1 protein were combined with folded FRET-RNA (KCl). In comparison to folded FRET-RNA without protein, the FAM emission increased after addition of the HNRNPH1 protein (Figure 31B). With addition of HNRNPH1, the FAM emission of the folded FRET-RNA significantly increased by 1.2 (lowest HNRNPH1 concentration, p-value < 0.01) to 1.5 times (highest HNRNPH1 concentration, p-value < 0.001) (Figure 31B). In total, using a direct readout for rG4 structure formation the results further confirm that recombinant HNRNPH1 binds to unfolded rG4 RNAs. Additionally, the data suggest that recombinant HNRNPH1 protein can destabilise rG4 RNAs and prevents folding of the rG4 by binding to the unfolded RNA.

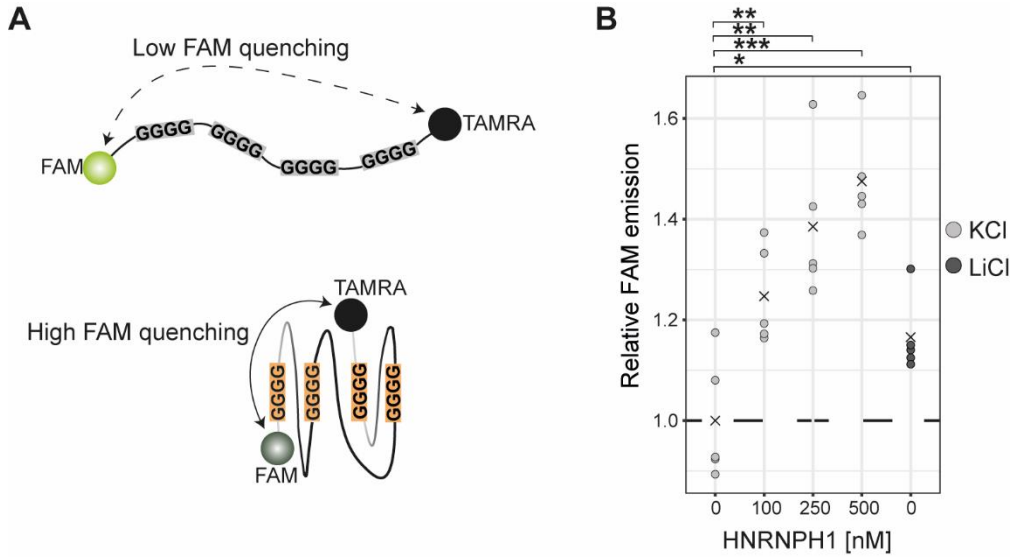

Figure 31: G-quadruplex structures induce TAMRA-dependent FAM-quenching.
(A) Schematic illustration showing how an rG4 affects the emission of the fluorophore FAM for an RNA that is labelled at the 5'-end with FAM and at the 3'-end with TAMRA. With an unfolded RNA the distance between FAM and TAMRA results in low TAMRA-dependent FAM quenching. rG4 formation brings FAM and TAMRA in close proximity and TAMRA can act as quencher of FAM. Adapted from: Vo et al., 2022. (B) Plot shows relative FAM emission of an rG4-forming FAM-TAMRA labelled RNA oligonucleotide (FAM-TAMRA-ARPC2.von.Hacht, see Table 4) with and without addition of recombinant HNRNPH1 protein in 150 mM LiCl (dark grey) or 150 mM KCl (light grey) buffer. FAM excitation was performed at 485 nm and FAM emission was measured at 520 nm. All experiments were performed in five individual replicates, which are shown in the plot together with the respective mean (indicated with a cross). A test for significance was performed using the Welch Two-Sample t-test. Significance levels: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

3.11 *AKR1A1* minigene to study rG4s for switch-like splicing *in vivo*

As shown with the *in vitro* experiments, HNRNPH binds cooperatively to RNAs that intrinsically encode rG4s (Figure 29). Although rG4 structures mediate cooperative HNRNPH binding, HNRNPH binds stronger to unfolded RNAs and additionally even seems

to destabilise rG4s (Figure 31). Building on these results, the question came up whether rG4s are not only the mediators of cooperative HNRNPH binding *in vitro*, but also mediate the cooperative splicing regulation *in vivo*. To address this question, a splicing minigene construct was designed for the CE splicing event *AKR1A1* exon 6. The *AKR1A1* exon 6 CE splicing event is a well-known G4-dependent splicing event, which has been already connected to HNRNPH (Dardenne et al., 2014; Zhang et al., 2019). Moreover, the *in vivo* titration of HNRNPH protein levels in MCF7 cells showed that HNRNPH regulates *AKR1A1* exon 6 splicing cooperatively (Figure 32).

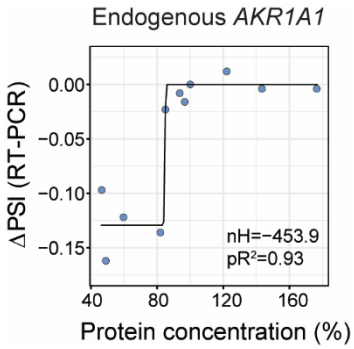

Figure 32: HNRNPH cooperatively regulates alternative splicing of endogenous *AKR1A1* exon 6. The plot shows the Hill function modelling of the dose-response curve for the RNA-Seq quantified ΔPSI of *AKR1A1* exon 6 in response to HNRNPH protein titration in MCF7 cells. The Hill function modelling was done by [redacted].

3.11.1 Construction of the *AKR1A1* minigene

To construct the *AKR1A1* minigene, the respective genomic region from the *AKR1A1* gene locus (Chr.1: 45,567,932-45,569,049) was PCR-amplified from human genomic DNA. The genomic fragment comprised the alternatively spliced exon 6, the flanking upstream and downstream constitutive exons 5 and 7 as well as the full introns in between the three exons (Figure 33). The 1120 bp long PCR product and the destination vector pcDNA3.1 (+) were both digested with sticky end restriction enzymes, followed by ligation of the PCR product into the destination vector (Figure 33). The resulting *AKR1A1* minigene was analysed for its splicing regulation. Endogenous *AKR1A1* exon 6 splicing showed PSI values around 67% (Figure 33). Contrary, for the *AKR1A1* minigene, exon 6 had inclusion levels around 7% (Figure 33). To better simulate the endogenous *AKR1A1* splicing pattern, a point mutation (G549T) was introduced into the AE 5'ss of the *AKR1A1* minigene (Figure 33). Introduction of this point mutation shifted the *AKR1A1* minigene splicing pattern from PSI values around 7% to PSI values around 79%, while splicing regulation by HNRNPH was not affected (Figure 33). In summary, the *AKR1A1* minigene with 5'ss point mutation (from here on

referred to as "*AKR1A1* minigene") shows a splicing pattern that is comparable to endogenous *AKR1A1* exon 6 splicing.

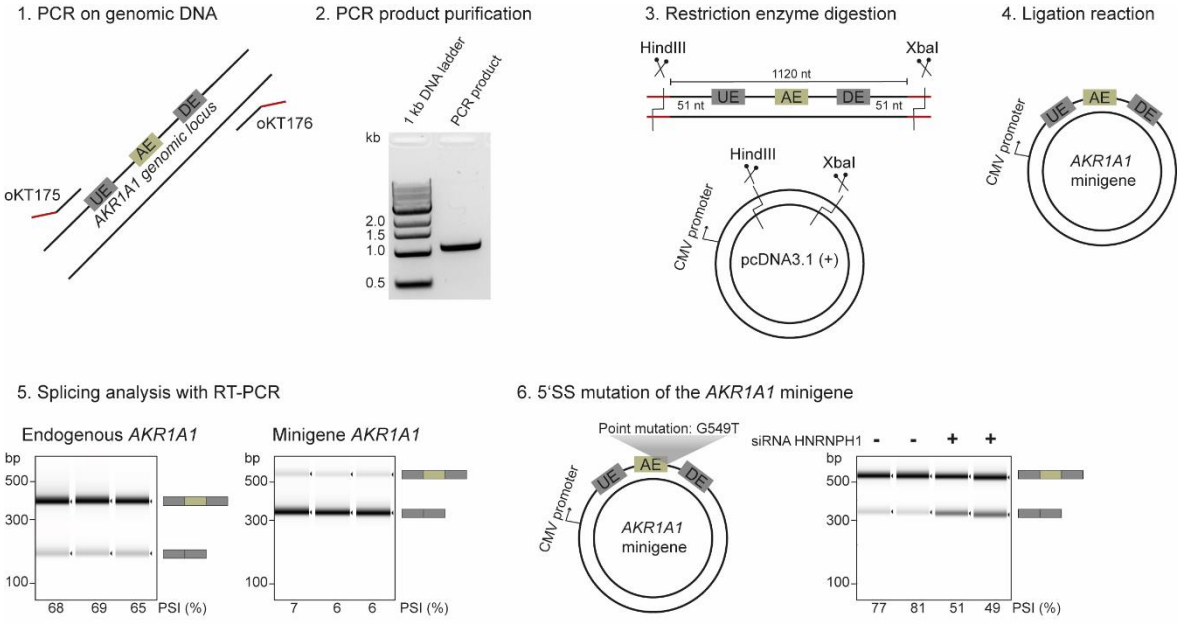

Figure 33: Schematic illustration of *AKR1A1* minigene construct generation.
The genomic region around *AKR1A1* exon 6 (alternative exon, AE) comprising the flanking upstream (UE) and downstream (DE) exons, was PCR-amplified and, after agarose gel purification, cloned into pcDNA3.1 (+) using the restriction enzymes HindIII and XbaI. The splicing pattern of the resulting *AKR1A1* minigene (pKT20, see Table 1) differed from endogenous *AKR1A1* exon 6 splicing. Introducing a point mutation (G549T) into the minigene at the 5'SS of the AE (pKT34, see Table 1), resulted in an *AKR1A1* minigene-splicing pattern comparable to the endogenous splicing.

3.11.2 Switch-like splicing regulation of the *ARK1A1* minigene

The endogenous *AKR1A1* exon 6 splicing is cooperatively regulated by HNRNPH (Figure 32). It was analysed whether the *AKR1A1* minigene is also cooperatively regulated. Therefore, after transfection of the *AKR1A1* minigene, the *HNRNPH1* mRNA level was gradually decreased using a siRNA-mediated KD. The resulting dose-response curve showed a sigmoidal curve characteristic for cooperativity (Figure 34). Hill function modelling retrieved a Hill coefficient (nH value = -13.6) that described cooperative splicing regulation of the *AKR1A1* minigene in response to HNRNPH1 KD (Figure 34). Overall, comparable to endogenous *AKR1A1* exon 6 splicing, the *AKR1A1* minigene construct is cooperatively spliced by HNRNPH.

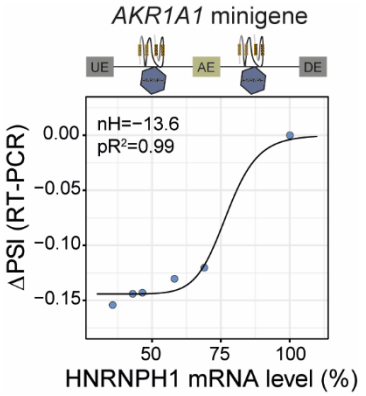

Figure 34: HNRNPH1 cooperatively regulates alternative splicing of the *AKR1A1* minigene.
The plot shows the Hill function modelling of the dose-response curve for the RT-PCR quantified ΔPSI of the *AKR1A1* minigene (pKT34, see Table 1) in response to titrated siRNA-mediated HNRNPH1 KD in MCF7 cells. The Hill function modelling was done by [redacted].

3.11.3 Mutations of HNRNPH binding sites in the *AKR1A1* minigene

The HNRNPH iCLIP showed that HNRNPH binds to the upstream and downstream introns flanking the AE 6 of *AKR1A1* (Figure 42, Appendix). Further, an *in silico* rG4 prediction revealed that the HNRNPH binding site in the downstream intron contains predicted rG4s (Figure 42, Appendix). RT-Stop profiling of the ivRNA library experimentally confirmed rG4 formation of the HNRNPH binding sites in the upstream and downstream introns (Figure 23; Figure 42, Appendix). To retrieve the HNRNPH-dependent regulative regions for *AKR1A1* exon 6 splicing, the HNRNPH binding sites in the *AKR1A1* minigene were mutated (Figure 35A). Simultaneous mutation of the HNRNPH binding sites in the upstream and downstream intron (UI/DI) completely removed any HNRNPH-dependent splicing regulation (Figure 35B,C). Single binding site mutations showed that mutating the upstream intron HNRNPH binding site (UI) still left HNRNPH-mediated splicing regulation (Figure 35B,C). In contrast, single mutation of the downstream intron HNRNPH binding site (DI) massively reduced inclusion levels of the *AKR1A1* exon 6, which was accompanied by a parallel reduction of the HNRNPH-dependent splicing regulation (Figure 35B,C). In summary, the here described data indicate that the HNRNPH binding site in the DI of the *AKR1A1* minigene is most important for *AKR1A1* exon 6 inclusion. In 2014, Dardenne et al. have already suggested a regulative role for HNRNPH binding sites with potential rG4s at the 5'ss of the *AKR1A1* exon 6 CE event (Dardenne et al., 2014). However, this study in 2014 did not use mutational constructs or further binding analyses to investigate the importance of rG4 structures for cooperative splicing regulation by HNRNPH.

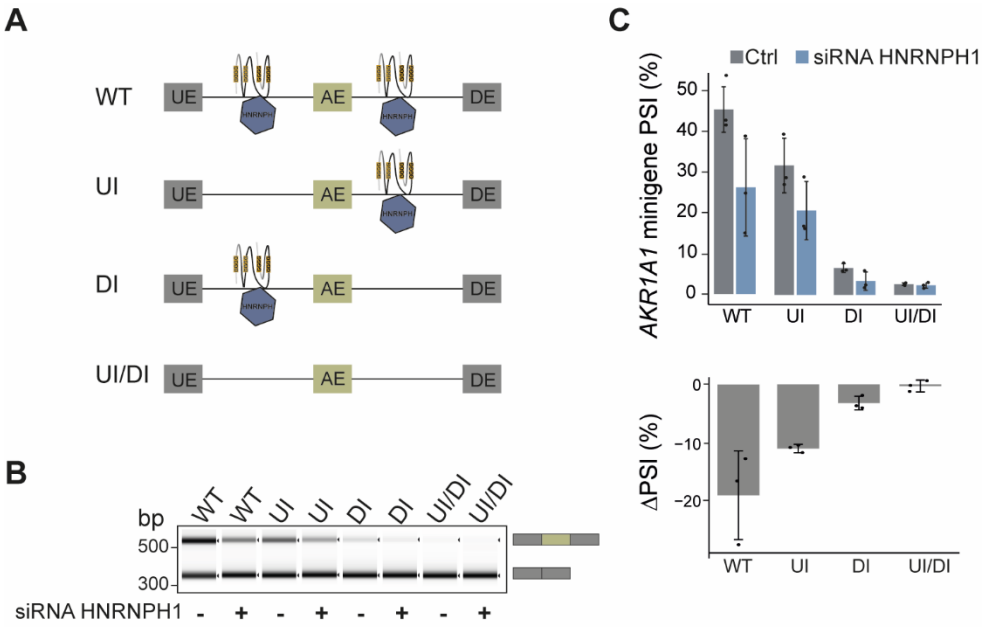

Figure 35: RT-PCR analyses of *AKR1A1* minigene mutations reveal an important splicing enhancer element in the downstream intron.
(A) *AKR1A1* minigene variants with mutation of the rG4-forming HNRNPH binding sites in the upstream intron (UI; pKT35, see Table 1), in the downstream (DI; pKT42, see Table 1) intron or in both introns simultaneously (UI/DI; pKT43, see Table 1). (B) Exemplary TapeStation 2200 electrophoresis profile of RT-PCR analyses of the *AKR1A1* minigene variants with (+) and without (-) HNRNPH1 KD. (C) Quantification of the *AKR1A1* exon 6 percent spliced in (PSI) of the different *AKR1A1* minigene variants (upper plot). The lower plot shows the change of the PSI (ΔPSI) of the *AKR1A1* minigene variants in response to a siRNA-mediated HNRNPH1 KD. All experiments were performed in three individual replicates. The bar plots indicate mean and standard deviation of the three replicates and further show the individual data points.

3.11.4 Insertion of new HNRNPH binding sites in the *AKR1A1* minigene

The question remained whether the HNRNPH binding site in the DI of the *AKR1A1* minigene needs to overlap with an rG4 to regulate *AKR1A1* minigene splicing. Hence, independently from each other, two new HNRNPH binding sites were introduced in the DI. Both binding sites did not show rG4 formation (non-rG4 *AKR1A1* minigene constructs) in the RT-Stop profiling (Figure 36A; Figure 43, Appendix). RT-PCR analysis showed that neither of the two non-rG4 HNRNPH binding sites could reconstitute HNRNPH-mediated splicing regulation for the *AKR1A1* minigene (Figure 36B,C). However, insertion of an rG4-forming HNRNPH binding site (rG4 *AKR1A1* minigene construct) in the DI restored the splicing regulation of HNRNPH for the *AKR1A1* minigene (Figure 36B,C; Figure 43, Appendix). In summary, the *AKR1A1* minigene analysis revealed that a strong HNRNPH-mediated splicing regulation of *AKR1A1* exon 6 can only be achieved with an intronic rG4-forming HNRNPH binding site located downstream of the alternatively spliced exon 6.

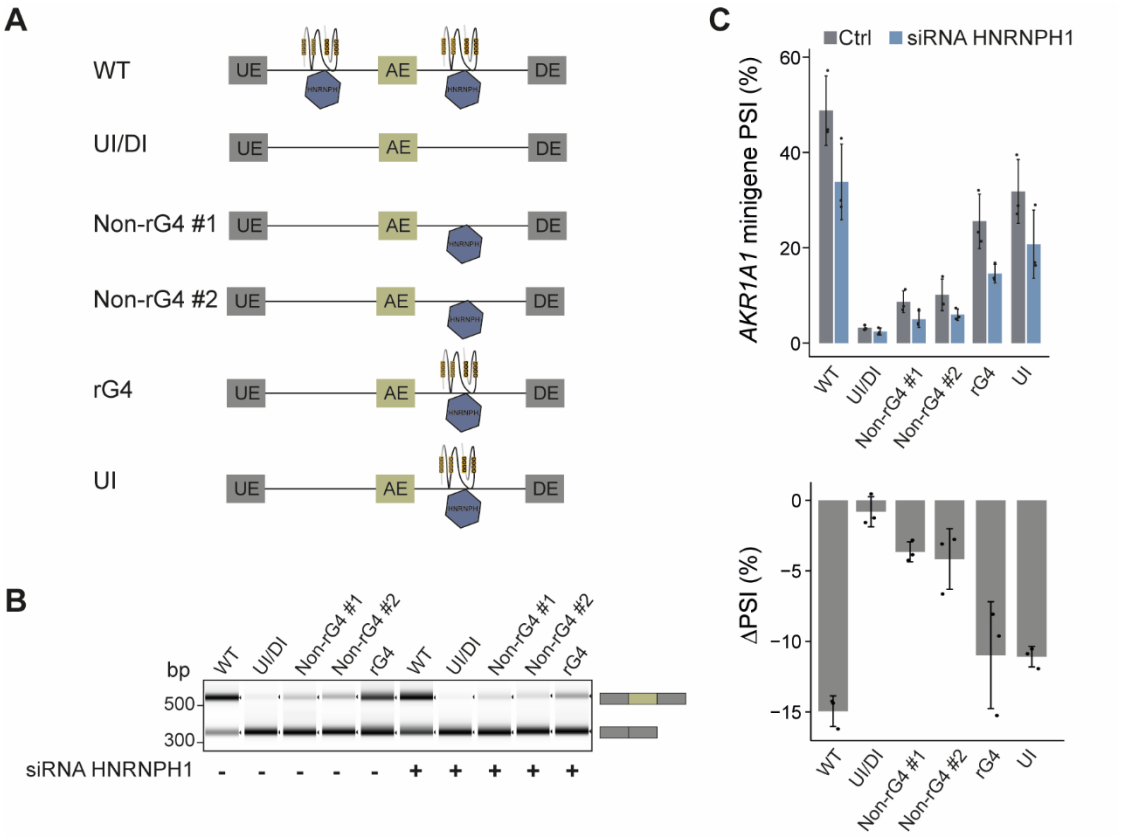

Figure 36: Insertion of an artificial rG4-forming HNRNPH binding site reconstitutes HNRNPH1-mediated *AKR1A1* minigene splicing.
(A) *AKR1A1* minigene variants with insertion of an rG4-forming (rG4; pKT54, see Table 1) and non-rG4-forming (Non-rG4 #1 [pKT52, see Table 1], Non-rG4 #2 [pKT53, see Table 1]) HNRNPH binding sites in the downstream intron using the UI/DI *AKR1A1* minigene construct as vector backbone. (B) Exemplary TapeStation 2200 electrophoresis profile of RT-PCR analyses of the *AKR1A1* minigene variants with (+) and without (-) HNRNPH1 KD. (C) Quantification of the *AKR1A1* exon 6 percent spliced in (PSI) of the different *AKR1A1* minigene variants (upper plot). The lower plot shows the change of the PSI (ΔPSI) of the *AKR1A1* minigene variants in response to a siRNA-mediated HNRNPH1 KD. All experiments were performed in three individual replicates. The bar plots indicate mean and standard deviation of the three replicates and further show the individual data points.

3.11.5 Switch-like splicing of the *AKR1A1* minigene with an artificial rG4 site

Since analysis of the rG4 *AKR1A1* minigene construct showed that insertion of the artificial rG4-forming HNRNPH binding site could reconstitute the HNRNPH-mediated splicing of *AKR1A1* exon 6, the question was addressed, whether this *AKR1A1* minigene mutation is cooperatively spliced. To do so, an HNRNPH titration was performed in MCF7 cells followed by quantification of the splicing changes of the rG4 *AKR1A1* minigene construct. Again, Hill function modelling was used to model the resulting sigmoidal dose-response curve (Figure 37). The extracted Hill coefficient (nH value = -40.4) indicated a cooperative splicing regulation of the respective minigene (Figure 37). So, in summary, the rG4 *AKR1A1* minigene construct shows, comparable to endogenous *AKR1A1* and WT *AKR1A1* minigene splicing, a cooperatively-mediated splicing regulation by HNRNPH.

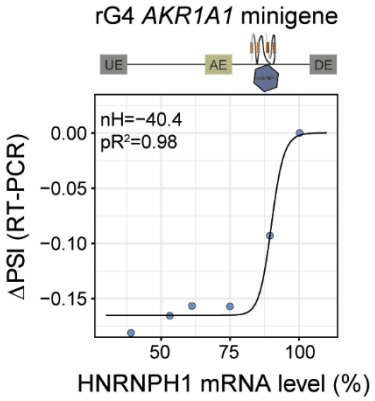

Figure 37: HNRNPH1 cooperatively regulates alternative splicing of the rG4 *AKR1A1* minigene. The plot shows the Hill function modelling of the dose-response curve for the RT-PCR quantified ΔPSI of the rG4 *AKR1A1* minigene (pKT54, see Table 1) in response to titrated siRNA-mediated HNRNPH1 KD in MCF7 cells. The experiment was performed by [redacted]. The Hill function modelling was done by [redacted].

3.12 HNRNPH expression and MCF7 cell proliferation

The titration of HNRNPH revealed that HNRNPH is an important regulator of alternative splicing events in MCF7 cells (Figure 9). The MCF7 cell line is a commonly used breast cancer cell line for breast cancer research. High HNRNPH expression levels have been associated with different cancers like gliomas or hepatocellular carcinomas (Lefave et al., 2011; Li et al., 2016; Herviou et al., 2020; Le Bras et al., 2022). To get an impression about the impact of HNRNPH expression levels for breast cancer, the role of *HNRNPH1* expression levels for MCF7 cell proliferation was investigated. Therefore, the *HNRNPH1* expression level was reduced using a siRNA-mediated KD and the proliferation capacity of the cells was monitored using an MTT treatment. Living cells metabolize MTT to the dye formazan, which can be quantified photometrically. 72 h after siRNA treatment, MCF7 cells with *HNRNPH1* KD showed a significantly lower relative formazan absorbance than MCF7 cells treated with a non-targeting siRNA (Figure 38, p-value < 0.001). In sum, the reduced relative formazan absorbance in response to an *HNRNPH1* KD indicates that MCF7 cells proliferate slower when the *HNRNPH1* expression level is decreased.

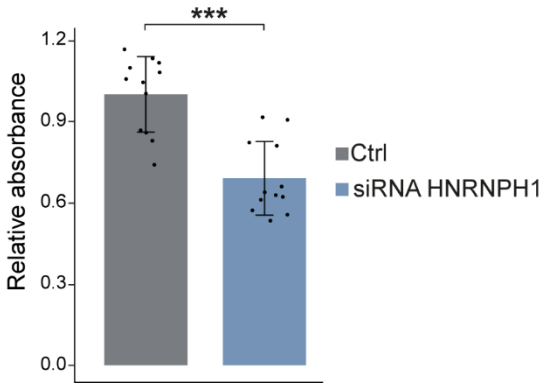

Figure 38: KD of *HNRNPH1* reduces the proliferation rate of the MCF7 breast cancer cell line.
An MTT-based proliferation assay was used to monitor the MCF7 cell proliferation rate 72 h after siRNA-mediated KD of *HNRNPH1*. The experiment was performed in three individual replicates consisting of four technical replicates each. The bar plot indicates the mean relative MTT absorbance and standard deviation of the three individual replicates and further shows all individual data points. A test for significance was performed using the Welch Two-Sample t-test. Significance level: *** p-value < 0.001.

3.12.1 Isoform of *HNRNPH1* is target of NMD

The reduced proliferative capacity of MCF7 cells after *HNRNPH1* KD raised the idea to use the reduction of *HNRNPH1* expression levels for therapeutic breast cancer treatment. *HNRNPH1* comprises an alternative splicing event that encodes an isoform that is a target of NMD (Pararajalingam et al., 2020). Skipping of *HNRNPH1* exon 4 disrupts the reading frame, which leads to a PTC in exon 5. Due to this PTC, the *HNRNPH1* exon 4 skipping mRNA isoform is an NMD-target (Pararajalingam et al., 2020). First, NMD of the *HNRNPH1* exon 4 skipping isoform was validated in MCF7 cells. Therefore, MCF7 cells were treated with CHX. CHX is a translational inhibitor, which thereby indirectly inhibits the translation-dependent NMD of mRNAs. Indeed, RT-PCR analysis of *HNRNPH1* exon 4 splicing showed that in untreated MCF7 cells the inclusion levels of exon 4 are close to 100% (Figure 39A,B). In contrast, after CHX treatment the *HNRNPH1* exon 4 inclusion levels decreased by approximately 30%, indicating accumulation of the *HNRNPH1* exon 4 skipping isoform (Figure 39C). Overall, the CHX treatment successfully validated that the *HNRNPH1* exon 4 skipping isoform is an NMD-target, which is expressed in MCF7 cells.

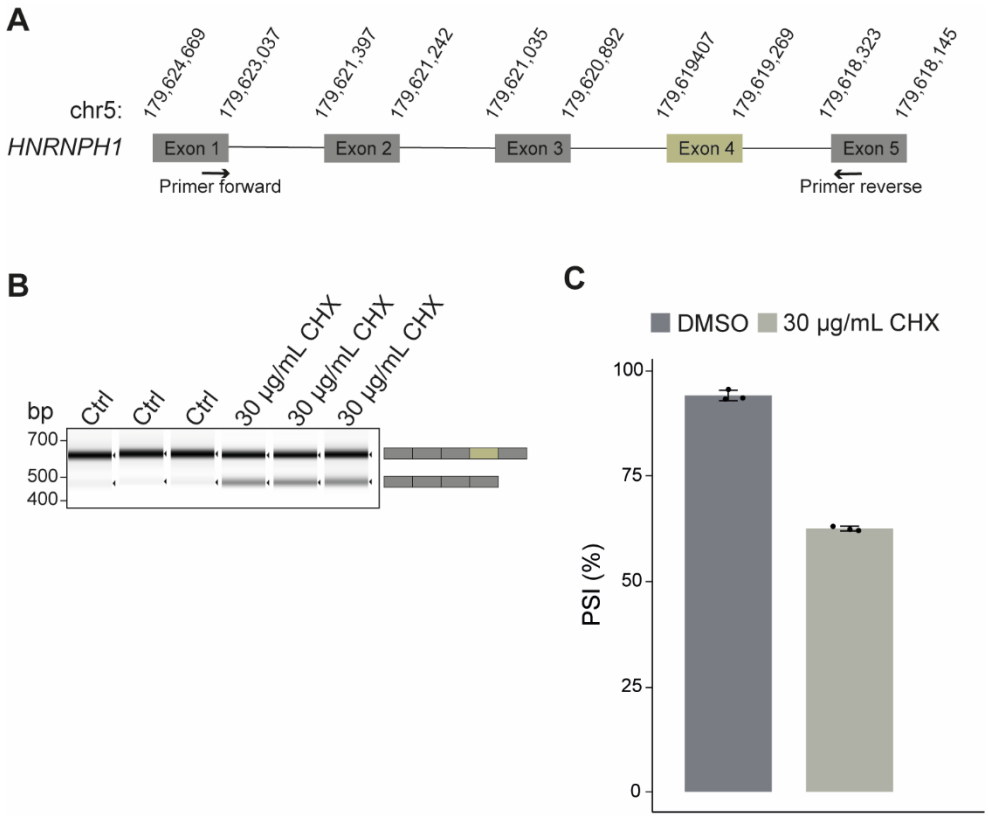

Figure 39: Inhibition of NMD validates the expression of an *HNRNPH1* exon 4 skipping isoform in MCF7 cells.
(A) Schematic illustration of the primer localisation sites to quantify *HNRNPH1* exon 4 alternative splicing with RT-PCR. (B) TapeStation 2200 electrophoresis profile of an RT-PCR analysing *HNRNPH1* exon 4 splicing in response to the inhibition of nonsense-mediated mRNA decay (NMD) after 4 h of cycloheximide (CHX) treatment. (C) Quantification of *HNRNPH1* exon 4 percent spliced in (PSI) levels in MCF7 cells in response to CHX-treatment. All experiments were performed in three individual replicates. The bar plots indicate the mean PSI and standard deviation of the three replicates and further show the individual data points.

3.12.2 ASO transfection induces *HNRNPH1* NMD isoform expression

To decrease *HNRNPH1* expression levels in MCF7 cells, the idea arose to force *HNRNPH1* exon 4 skipping. With forced expression of the *HNRNPH1* exon 4 skipping isoform, which is degraded via NMD, the overall cellular *HNRNPH1* expression levels should decrease. Splice-switching ASOs are an attractive mechanism to control splicing exogenously. To force *HNRNPH1* exon 4 skipping, an ASO was designed, which base pairs with a splicing enhancer motif within *HNRNPH1* exon 4 (Aartsma-Rus, 2012; Figure 40A). MCF7 cells were transfected with the *HNRNPH1*-targeting ASO or a scrambled ASO (Ctrl). After transfection of the ASOs, the *HNRNPH1* expression levels were quantified using qPCR. Transfection of the *HNRNPH1*-targeting ASO reduced *HNRNPH1* expression levels by 42% (Figure 40B). The ASO-dependent reduction of the *HNRNPH1* expression levels was reversible under CHX treatment (Figure 40B). In summary, the *HNRNPH1*-targeting ASO efficiently reduces *HNRNPH1* expression levels. The loss of the *HNRNPH1* expression level

reduction after CHX treatment confirms that the ASO-dependent decreased *HNRNPH1* levels arose from forced NMD isoform expression. Hence, the *HNRNPH1*-targeting ASO can be used to reduce the cellular *HNRNPH1* expression via modulating the splicing of *HNRNPH1*.

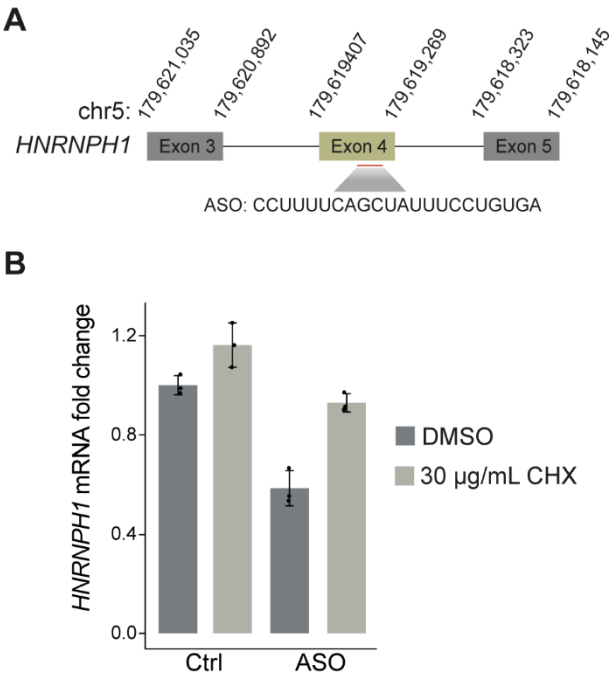

Figure 40: A splice-switching antisense oligonucleotide (ASO) can reduce the *HNRNPH1* expression in MCF7 cells.
(A) Schematic illustration, indicating the ASO (ASO2, see Table 6) annealing site within *HNRNPH1*. (B) Quantification of *HNRNPH1* mRNA expression fold changes in MCF7 cells 48 h after ASO-transfection. Nonsense-mediated decay was inhibited using 4 h of cycloheximide (CHX) treatment. All experiments were performed in three individual replicates. The bar plot indicates the mean *HNRNPH1* mRNA fold changes and standard deviation of the three replicates and further shows the individual data points.

3.12.3 *HNRNPH1* NMD isoform expression reduces MCF7 cell proliferation

SiRNA-mediated KD of *HNRNPH1* was sufficient to significantly decrease the proliferative capacity of MCF7 cells. To find a therapy-suitable way to reduce *HNRNPH1* expression in breast cancer, an ASO was designed that reduces *HNRNPH1* levels via forced expression of an NMD-sensitive *HNRNPH1* isoform. After designing and validating the ASO, the question remained whether ASO-dependent reduced *HNRNPH1* expression is sufficient to affect MCF7 cell proliferation. Using an MTT-dependent proliferation assay, it was observed that cells treated with the *HNRNPH1*-targeting ASO had a significantly lower relative formazan absorbance than cells treated with a scrambled ASO (Figure 41, p-value < 0.01). Hence, cells transfected with the *HNRNPH1*-targeting ASO metabolized lower amounts of MTT than MCF7 cells transfected with a scrambled ASO (Ctrl). Overall, these results indicate that MCF7 cells proliferate slower after transfection of the *HNRNPH1*-targeting ASO, which

probably correlates with lower *HNRNPH1* expression levels in response to the ASO transfection.

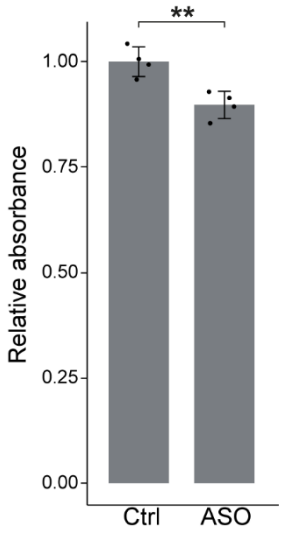

Figure 41: The *HNRNPH1*-targeting ASO reduces the proliferation rate of MCF7 cells.
An MTT-based proliferation assay was used to monitor the MCF7 cell proliferation rate 72 h after treatment of the cells with an ASO that forces the expression of an *HNRNPH1* NMD isoform. The experiment was performed in four replicates. The bar plot indicates mean relative MTT absorbance and standard deviation of the replicates and further shows all individual data points. A test for significance was performed using the Welch Two-Sample t-test. Significance level: ** p-value < 0.01.

4 Discussion

In this work, transcriptome-wide RNA-Seq and iCLIP experiments were combined with high-throughput *in vitro* techniques to decipher features of HNRNPH binding to RNAs. By integrating *in vivo* analyses with *in vitro* high-throughput studies, this PhD work extends the existing knowledge on HNRNPH-mediated splicing regulation from the understanding of individual splicing events towards a transcriptome-wide decoding of the HNRNPH splicing code. Importantly, this study highlights that HNRNPH mediates splicing regulation cooperatively on a transcriptome-wide scale and that the cooperative effect relies on the formation of the much debated and often questioned rG4 structure.

4.1 Cooperative splicing is a new transcriptome-wide feature of HNRNPH

Alternative splicing regulation is an intensively studied aspect of RNA processing biology. In respect to cooperative splicing regulation, several studies described splice isoform switches in physiological embryogenesis or in pathophysiological conditions like cancer.

During embryogenesis, pluripotent stem cells need to perform a complex phenotypical change to differentiate into all flavours of existing cell types that finally create a living organism. To facilitate such complex and rapid phenotypical transitions, splice isoform switches, which affect functionality of the resulting protein, could result in a way quicker and more precise phenotypical change than a gradual isoform expression change. Indeed, to maintain pluripotency of embryonic stem cells, the expression of an embryonic stem cell specific isoform of the transcription factor *FOXP1* is important (Gabut et al., 2011). Within the differentiation process of pluripotent embryonic stem cells, a splice isoform switch for *FOXP1* occurs, changing towards the expression of a *FOXP1* isoform that does not further maintain pluripotency (Gabut et al., 2011). Besides this specific example, several other studies described splice isoform switches that are involved in the progression from embryonic pluripotent stem cells towards differentiation (Pritsker et al., 2005; Yeo et al., 2007; Atlasi et al., 2008; Rao et al., 2010; Jin et al., 2020).

Interestingly, Yamazaki and colleagues described a connection between HNRNPH-mediated splicing and the maintenance of pluripotent embryonic stem cells (Yamazaki et al., 2018). They connected high HNRNPH expression levels in human embryonic stem cells with the regulation of a *TCF3* transcription factor splicing switch,

which is involved in the differentiation process of embryonic stem cells (Yamazaki et al., 2018). Besides this correlation between the maintenance of pluripotency and HNRNPH-mediated splicing, other studies showed the involvement of HNRNPH in the control of splicing switches in the context of cancers (Lefave et al., 2011; Li et al., 2016).

Using a transcriptome-wide RNA-Seq approach, this study could reveal that switch-like splicing regulation is not a rare feature of HNRNPH-mediated splicing (Figure 10). *In vivo* titration of HNRNPH protein levels in the breast cancer cell line MCF7 identified a large set of genes alternatively spliced by HNRNPH (Figure 9). A GO analysis of all CE events that were alternatively spliced in response to the HNRNPH titration revealed an enrichment for splicing-related genes (Table 21). These results support the data from Uren and colleagues, who showed that HNRNPH1 binding sites, identified from iCLIP experiments, are enriched at genes involved in mRNA processing (Uren et al., 2016). Hence, the presented data support that one important biological function of HNRNPH is the alternative splicing regulation of splicing-related genes. Interestingly, most of the genes that are alternatively spliced by HNRNPH in MCF7 cells show a switch-like splicing change (Figure 10). To assess switch-like splicing changes qualitatively, Hill functions were modelled for every single splicing event identified within the HNRNPH titration series. The nH value, which can be extracted from the Hill function, provides information about the steepness of the splicing switch and thereby can serve as qualitative indicator for cooperative, switch-like processes. The Hill function modelling showed that the vast majority of HNRNPH-regulated splicing events underlies a cooperative regulative process (Figure 10).

Taken together, the tight control of splice isoform switches is crucial for physiological conditions like embryonic development, but also for pathophysiological situations like cancer. Several studies related HNRNPH to the control of individual splice isoform switches (Dardenne et al., 2014, Conlon et al., 2016; Braun et al., 2018; Neckles et al., 2019). However, in the here presented study it was described for the first time that switch-like splicing regulation is a transcriptome-wide feature of HNRNPH. In summary, HNRNPH globally controls alternative splicing in a cooperative manner.

4.2 Indirect cooperativity can explain switch-like splicing by HNRNPH

Thinking of cooperativity in biological processes, one of the probably best-known switch-like regulations is the cooperative binding of oxygen to haemoglobin, characterised

by its typical sigmoidal binding curve. However, also in the field of RNA biology, cooperative binding is no stranger. As introduced in chapter 1.4.2 of this study, cooperative binding of proteins to RNA can be either direct or indirect. To achieve direct cooperativity, proteins directly interact with each other and thereby form higher-order complexes on RNAs. In contrast, indirect cooperativity is mediated via RNA structures that, if the structure is folded, could hide potential RBP-binding sites. Since the presented work showed that HNRNPH regulates alternative splicing cooperatively, this study further addressed the question of the underlying cooperative mechanism. To do so, *in vivo* experiments, like an HNRNPH interactome analysis and minigene studies, were combined with *in vitro* binding assays.

4.2.1 HNRNPH's GY-rich protein domain does not mediate direct cooperativity

In the sense of direct cooperativity, several authors described a cooperative property for GY-rich protein domains of HNRNPs (Cartegni et al., 1996; Gueroussov et al., 2017). HNRNPs with GY-rich protein domains can directly interact with itself or with other HNRNPs that carry GY-rich protein domains. Gueroussov and colleagues showed that HNRNPs could form higher-order protein complexes via their GY-rich protein domains (Gueroussov et al., 2017). Further, they highlighted the importance of these high-order protein assemblies for the regulation of alternative splicing (Gueroussov et al., 2017).

HNRNPH carries two GY-rich protein domains (the GYR- and the GY-domain) that are interspersed by HNRNPH's qRRM3 (Figure 4). Van Dusen and colleagues characterised a nucleocytoplasmic shuttling function for the GYR-domain of HNRNPH (van Dusen et al., 2010). In contrast, the C-terminal GY-domain of HNRNPH was described to take part in direct protein-protein interactions with GY-domains of other RBPs like HNRNPA1 or HNRNPD and thereby seems to modulate splicing decisions (Fisette et al., 2010; Gueroussov et al., 2017).

To investigate the role of direct protein interactions for HNRNPH's cooperative splicing regulation, an HNRNPH1 interactome was determined and further, the impact of a GY-domain deletion for HNRNPH1-mediated splicing regulation was analysed. The interactome of HNRNPH1 identified interactions with the highly related paralogue HNRNPH2 and with several other HNRNPs like HNRNPC, HNRNPU or HNRNPA0 (Figure 12). Further, an interaction with other RBPs like RBM14, SMU1 and IK was

detected (Figure 12). All the identified interacting RBPs are associated with the process of splicing, which supports HNRNPH's function in alternative splicing regulation (Xiao et al., 2012; Ye et al., 2015; Meininger et al., 2016; Sapir et al., 2022; Sharma et al., 2008; Thibault et al., 2021; Xing et al., 2022; Zarnack et al., 2013; Zhou et al., 2017; Li et al., 2020; Agafonov et al., 2011). With the loss of RNA integrity due to an RNA digestion, only interactions of HNRNPH1 with its paralogue HNRNPH2 and with RBM14 remained stable, which implies that most of the here identified RBP-HNRNPH interactions depend on the presence of RNA. The RNA-independent interaction of HNRNPH1 with its paralogue HNRNPH2 underlines the idea that protein-protein interactions of HNRNPH could mediate HNRNPH's cooperative splicing regulation via direct cooperativity. However, deletion of HNRNPH1's GY-domain did not severely affect the switch-like splicing characteristic of HNRNPH-mediated *MST1R* exon 11 splicing (Figure 13).

In conclusion, the presented work described several interactions with other RBPs, especially other HNRNPs, which underlines HNRNPH's function as a splicing regulator and further implies the cooperation of HNRNPH with other RBPs to regulate alternative splicing. Nevertheless, depletion of HNRNPH1's C-terminal protein-protein interaction domain (the GY-domain) did not result in a change of the switch-like, cooperative splicing regulation of *MST1R* exon 11, which was shown to be mediated by HNRNPH (Braun et al., 2018). Hence, the here presented results do not support the hypothesis that direct cooperativity of HNRNPHs among each other communicates the cooperative splicing regulation by HNRNPH.

4.2.2 rG4 structures can mediate indirect cooperativity of HNRNPH

RNA structures can communicate cooperative binding of RBPs indirectly. Lin and Bundschuh proposed this idea of indirect cooperativity between RBPs mediated via RNA structures (Lin and Bundschuh, 2013; Lin and Bundschuh, 2015). They proposed that a cooperative interaction of single-stranded binding RBPs between distant RNA-binding sites could arise from RNA structures. In case of single-stranded binding RBPs, a formed RNA structure hides the respective RBP binding sites. In the context of an unfolded RNA, single-stranded binding RBPs can bind to the RNA and thereby interfere with RNA re-folding. Preventing the RNA re-folding then indirectly increases the RNA's accessibility for other single-stranded binding RBPs, which explains the idea of an indirect cooperation between RBPs (Lin and Bundschuh, 2013; Lin and Bundschuh, 2015). Becker and

colleagues considered indirect cooperativity of RBPs not in the aspect of long-distance cooperation, but with respect to clustered binding sites of the RBP PUM2 (Becker et al., 2019). They could show that RNA structures indirectly mediate cooperative binding of PUM2 to a respective RNA carrying PUM2 binding site clusters (Becker et al., 2019).

HNRNPH binds to G-tract RNA sequences, which was described in several studies and is further supported by the results of the HNRNPH iCLIP experiment performed as part of the presented PhD thesis (Honoré et al., 1995; Chen et al., 1999; Caputi and Zahler, 2001; Uren et al., 2016). Interestingly, a motif enrichment analysis of the here experimentally generated HNRNPH iCLIP data revealed the appearance of G-tract clusters around the HNRNPH binding sites (Figure 17). Comparable observations of G-tract clusters flanking HNRNPH binding sites were published in an earlier study analysing RNA-binding characteristics of HNRNPH (Uren et al., 2016). When talking about G-tract clusters, it tempts one to think about a possible connection between HNRNPH and the rG4 RNA structure. As introduced in chapter 1.2.2, rG4 RNA structures can build from RNA sequences with G-tracts interspersed by short loops. Hence, the clustering of G-tracts around HNRNPH binding sites implies that indirect cooperativity via rG4 RNA structures could be involved in HNRNPH-mediated cooperative splicing regulation. Indeed, several studies already addressed a potential link between rG4 structures and HNRNPH-regulated alternative splicing (Dardenne et al., 2014; Conlon et al., 2016; Neckles et al., 2019).

4.2.2.1 rG4s mediate cooperative binding of HNRNPH *in vitro*

The presented PhD thesis is the first work that investigated a potential role of rG4-structures in mediating indirect cooperativity in alternative splicing. To address this question, different *in vitro* techniques were combined. RT-Stop profiling of the ivRNA library, consisting of 3000 genomic regions with HNRNPH binding sites, revealed that 1060 of the finally analysed 2169 genomic regions carry an intrinsic capacity to form an rG4 structure (Figure 24). Moreover, ivCLIP experiments using recombinant HNRNPH1 protein showed that, if rG4 structures can form, HNRNPH1 binds enriched to these rG4 RNAs (Figure 26). All these results underline what is also discussed in the literature, namely that RNA-binding sites of HNRNPH or the related RBP HNRNPF tend to co-appear with rG4s (Dominguez et al., 2010; Decorsière et al., 2011; Dardenne et al., 2014).

Next, *in vitro* titration series of recombinant HNRNPH1 protein showed for the first time that HNRNPH binds cooperatively to RNAs (Figure 29). Importantly, the cooperative binding of HNRNPH depends on the RNA's intrinsic capacity to form an rG4 (Figure 29). Although cooperative HNRNPH binding relies on the RNA's ability to form an rG4, further *in vitro* binding analyses revealed that in the end HNRNPH binds stronger to RNAs in experimental conditions that destabilise rG4s (Figure 30). So far, the present data imply that HNRNPH binds cooperatively to rG4-forming RNAs, even though stronger binding correlates with experimental conditions preferring unfolded rG4s. The question remained whether HNRNPH relies on an unfolded structure for binding. FRET experiments, performed to address the aforementioned question, unravelled that HNRNPH seems to encode an intrinsic potential to destabilise rG4 structures *in vitro*, thereby keeping the RNA in the favoured single-stranded state (Figure 31). This finding is in line with the very recently published study by Vo and colleagues, who showed that the qRRM1-qRRM2 domains of HNRNPH can destabilise rG4 structures in *EWSR1* transcripts (Vo et al., 2022).

In summary, the HNRNPH *in vitro* binding studies of this PhD work highlight:

I) that rG4s accumulate around HNRNPH binding sites,
II) that HNRNPH binds enriched to rG4 RNAs,
III) that cooperative binding of HNRNPH depends on the RNA's ability to form rG4s,
IV) that HNRNPH binds stronger to unfolded rG4s and
V) that HNRNPH can destabilise rG4 structures.

4.2.2.2 Does HNRNPH bind to rG4s or to single-stranded G-tract RNAs?

Although an interplay between HNRNPH and rG4s is well accepted in the literature, it is still under discussion, whether folded or unfolded rG4s are bound. Indeed, HNRNPF, which also binds to G-tract RNAs, was proposed to bind to unfolded rG4s, thereby keeping the G-tract RNAs in a single-stranded state (Dominguez et al., 2010; Samatanga et al., 2013). In addition, it was suggested that HNRNPH requires unfolded rG4s for its binding (Dardenne et al., 2014). To generate unfolded rG4s for HNRNPH or HNRNPF binding, the authors of the aforementioned study postulate a combinatorial interplay between the rG4-unwinding RNA helicases DDX5 and DDX17 and the splicing factors HNRNPH/F (Dardenne et al., 2014).

In contrast, other studies proposed the binding of HNRNPH or HNRNPF to folded rG4s (Huang et al., 2017; Neckles et al., 2019). One big limitation of the studies, showing that HNRNPH binds to folded rG4s, is their usage of mutated RNA sequences to prevent rG4 formation. Mutations that are introduced to prevent rG4 formation often include deletion or substitution of guanines, which not only inhibits the folding of rG4s. The mutation of guanines also affects the G-tract binding motif of HNRNPH, which potentially weakens HNRNPH binding to these sequences. Hence, results from experiments that use mutations of rG4 RNA sequences might be inconclusive and cannot finally prove that HNRNPH binds to folded rG4s. To overcome this issue, it became popular in the field of rG4 studies to use 7-deaza-GTP as a GTP analogue to prevent rG4 folding. Using 7-deaza-GTP, Herviou and colleagues showed that HNRNPH preferentially binds to unfolded rG4s and suggested that the RNA helicase DHX36 is involved in rG4 unwinding (Herviou et al., 2020). Also, for the experiments in the presented PhD thesis, 7-deaza-GTP was used to prevent rG4 folding. Accordingly, the presented results support the hypothesis that HNRNPH preferentially binds to unfolded rG4s, although the results also indicate that a folded rG4 structure plays a role for cooperative binding of HNRNPH and that HNRNPH can destabilise folded rG4s.

In sum, since all *in vitro* experiments in the underlying PhD thesis were performed without any additional proteins except for recombinant HNRNPH protein, the data, in combination with recent literature, lead to the following hypothesis: To recognise and bind cooperatively to RNAs, HNRNPH benefits from rG4 structures, which it can destabilise to bind to single-stranded G-tract RNAs in the end.

4.2.2.3 Do rG4s exist *in vivo*?

While the formation of rG4 structures *in vitro* is a well-accepted phenomenon, the appearance of rG4s *in vivo* is highly debated. *In vitro*, RT-Stop assays confirmed the formation of rG4s in polyadenylated RNAs from cells across the whole transcriptome (Kwok et al., 2016; Guo and Bartel, 2016). Contrary, performance of an *in vitro* RT-Stop assay after treating living cells with DMS, which labels unfolded RNAs, disproved the existence of folded rG4s *in vivo* (Guo and Bartel, 2016). These studies opened the discussion whether folded rG4s do exist *in vivo* or not. Although the RT-Stop assay in combination with DMS treatment refuted the formation of rG4s *in vivo*, this conclusion should be viewed critically. *In vivo*, rG4 structures might form, then being potentially restored to an unfolded state by the cellular machinery (Guo and Bartel, 2016). Shifting the equilibrium from folded to

unfolded rG4s *in vivo*, would explain why an RT-Stop assay after DMS treatment could not confirm a transcriptome-wide formation of rG4s. Nevertheless, in such a scenario, the short-term formed rG4s could still serve as structural landmarks for the binding of RBPs, even though the binding in the end might result in single-stranded G-tracts. The description of RBPs and helicases that specifically recognise and bind to rG4 structures underlines the hypothesis that folded rG4s do exist *in vivo* (Dardenne et al., 2014; Herviou et al., 2020).

As an additional hint for the existence of folded rG4s *in vivo*, many studies attributed regulatory functions in all kinds of post-transcriptional processes to rG4s. For example, it was shown that rG4 sequences can be present at polyadenylation signalling sites and interact there with HNRNPH/F to regulate polyadenylation of mRNAs like *TP53* (Bagga et al., 1995; Dalziel et al., 2007; Decorsière et al., 2011). Subramanian and colleagues presented an involvement of rG4s in the cellular transport and localisation of mRNAs in cortical neurites (Subramanian et al., 2011). Interestingly, another study, investigating an RNA-transporting granule associated to the transport molecule kinesin in neurons, identified the presence of RNA helicases like DDX1 and DDX3 as well as HNRNPs like HNRNPC and HNRNPA0 (Kanai et al., 2004). In the here presented study, the HNRNPH interactome revealed an interaction of HNRNPH with HNRNPC and HNRNPA0 (Figure 12). Hence, it could be an interesting future topic to study a potential function of HNRNPH in the neurite transport of rG4-containing mRNAs (Lee et al., 2018). In addition, in the context of mRNA translation, many studies suggested a regulatory role of rG4s in mRNA translation regulation. Here, the regulatory function of rG4s is bidirectional, they can either inhibit or stimulate translation (Bonnal et al., 2003; Kumari et al., 2007; Halder et al., 2009; Shahid et al., 2010; Morris et al., 2010; Bugaut and Balasubramanian, 2012; Herviou et al., 2020). Further, as already introduced in the chapter 1.3.3, studies described the importance of rG4s for alternative splicing regulation, often connecting these splicing events to HNRNPH (Gomez et al., 2004; Didiot et al., 2008; Marcel et al., 2011; Huang et al., 2017; Neckles et al., 2019; Zhang et al., 2019; Georgakopoulos-Soares et al., 2022).

4.2.2.4 Cooperative splicing regulation of *AKR1A1* requires an rG4 *in vivo*

All the aforementioned studies underline that *in vivo* rG4s participate in the regulation of many mRNA processing steps, supporting the hypothesis that rG4s do exist *in vivo* as a regulatory RNA structure. As shown in the presented work, HNRNPH binding sites encode the intrinsic potential to form rG4s *in vitro* (Figure 24). Moreover, *in vitro* this intrinsic

potential to form rG4s is important for cooperative binding of HNRNPH to the RNAs (Figure 29). To address the importance of rG4s for HNRNPH-mediated cooperative splicing regulation *in vivo*, minigene experiments were performed. For this purpose, a minigene of the *AKR1A1* exon 6 splicing event was constructed. This splicing event was chosen, because I) literature associated *AKR1A1* splicing with an rG4, II) the HNRNPH titration from the presented study identified a cooperative splicing regulation of *AKR1A1* and III) the HNRNPH iCLIP and the RT-Stop profiling revealed rG4-forming HNRNPH binding sites in the CE-flanking introns (Dardenne et al., 2014; Zhang et al., 2019; Figure 42, Appendix).

Mutation of the respective rG4-forming HNRNPH binding sites confirmed that *AKR1A1* exon 6 inclusion depends on these *cis*-regulatory elements (Figure 35). Interestingly, removal of the UI HNRNPH binding site just had a slight effect on the *AKR1A1* exon 6 inclusion level, while deletion of the DI rG4-forming HNRNPH binding site nearly completely abolished the exon inclusion (Figure 35). Comparison of these two mutational constructs implies that the *cis*-regulatory element in the DI comprises an important regulatory role for *AKR1A1* exon 6 inclusion. In line with this, Dardenne and colleagues also described the importance of potential rG4-rich sequences in the DI for the exon inclusion level of *AKR1A1* exon 6 (Dardenne et al., 2014). To test if HNRNPH-mediated cooperative splicing regulation relies on an rG4-forming HNRNPH binding site downstream of *AKR1A1* exon 6, the *cis*-regulatory effect of non-rG4 and rG4 HNRNPH binding sites was compared. Indeed, only the insertion of an artificial rG4-forming HNRNPH binding site reconstitutes the HNRNPH-mediated splicing change of the endogenously encoded rG4-forming binding site (Figure 36). Further, comparable to the WT *AKR1A1* exon 6 construct, insertion of the artificial rG4-forming HNRNPH binding site results in a switch-like, cooperative splicing regulation mediated by HNRNPH (Figure 37). Overall, the *AKR1A1* minigene analysis highlights that the intrinsic ability to form an rG4 not only contributes to the HNRNPH-mediated cooperative splicing regulation *in vitro* but also *in vivo*.

In summary, the combination of *in vitro* and *in vivo* experiments presented here suggests the importance of an RNA's intrinsic rG4-folding propensity for HNRNPH-mediated cooperative splicing regulation. The results highlight that the rG4 structure serves as a regulatory landmark, which HNRNPH can destabilise to keep RNAs in a single-stranded state. The data may help shed more light on whether HNRNPH binds to folded or unfolded

rG4. They underline that, regarding HNRNPH, the answer to this question is probably not just black or white, but rather a complex interplay of HNRNPH with folded rG4 structures, destabilisation of these structures, and indirect cooperative binding to single-stranded RNAs.

4.4 Clinical relevance of *HNRNPH* expression level regulation

With the understanding of the complexity and importance of alternative splicing for the diversity of cellular phenotypes, the interest for the role of RBPs in pathophysiological conditions like cancers arose. In respect to this, many researchers investigated the expression and function of RBPs in cancer and thereby revealed that RBPs show an aberrant expression in cancers, which often affects patient survival (reviewed in Pereira et al., 2017). Functionally, dysregulated expression of RBPs in cancer interferes with all hallmarks of cancer (reviewed in Kang et al., 2020).

Recapitulating the observation of this study that HNRNPH mediates splicing cooperatively, increased a particular interest in the role of HNRNPH expression levels in cancers. The cooperative, switch-like splicing regulation by HNRNPH makes the whole system very sensitive to even only small changes in the HNRNPH expression. Interestingly, Wang and colleagues showed that the dysregulation of RBP expression levels across different cancer types was unexpectedly low in comparison to the aberrant expression of genes, which are commonly affected in cancers (Wang et al., 2015). Nevertheless, especially for HNRNPH, using cooperativity to regulate alternative splicing, even those small expression changes could largely affect the cellular phenotype. In line with this, several studies described increased HNRNPH expression levels in cancers like gliomas or hepatocellular carcinomas, showing that HNRNPH expression correlates with patient survival (Lefave et al., 2011; Li et al., 2016; Herviou et al., 2020, Le Bras et al., 2022).

In breast cancer cells lines, which were used for the *in vivo* experiments presented in this study, not much is known about the role of HNRNPH expression levels for cancer cell survival. Tyson-Capper and colleagues showed that a combined KD of HNRNPH1/F and HNRNPK was sufficient to reduce cellular viability of breast cancer cell lines, linking this effect to a switch in the alternative splicing of the cancer-related gene *MCL-1* (Tyson-Capper and Gautrey, 2018). Another study proposed that HNRNPH1 is involved in the regulation of the *HER2* splicing hotspot in breast cancer cells (Gautrey et al., 2015). To expand the current knowledge about the role of HNRNPH expression levels for breast cancer cell lines,

an HNRNPH KD was performed in MCF7 cells. In line with published data, HNRNPH KD significantly reduces the proliferative capacity of MCF7 cells (Figure 38; Tyson-Capper and Gautrey, 2018). Overall, the here presented data and the evidence from the literature indicate that a reduction of HNRNPH expression levels is sufficient to reduce the viability and the proliferative potential of breast cancer cells.

The therapeutic targeting of RBPs to treat cancer patients is a hot topic in the field of cancer research. To control RBP expression levels specifically, researchers investigated the potential of ASOs as nucleic acid drugs (Watts and Corey, 2012; Havens and Hastings, 2016, see chapter 1.2.3). ASOs are short, synthetic nucleic acids that are designed to base pair with a targeted pre-mRNA region to interfere with RNA structure formation, the splicing machinery, or other protein-RNA interactions. As introduced in chapter 1.2.3 of this thesis, the induction of splicing switches is a promising therapeutic feature of ASOs. To induce a splicing switch, ASOs bind to splicing silencer or splicing enhancer elements, thereby either increasing or decreasing a targeted splicing event (Havens and Hastings, 2016).

Pararajalingam and colleagues characterised an *HNRNPH1* splice isoform that lacks exon 4 and is a target of NMD (Pararajalingam et al., 2020). NMD inhibition confirmed the expression of this *HNRNPH*1 NMD splice isoform in MCF7 cells (Figure 39). Considering that downregulated *HNRNPH* expression levels reduce the proliferation rate of MCF7 cells, the idea of switching the *HNRNPH1* mRNA isoform expression towards the *HNRNPH1* NMD splice isoform via ASO treatment emerged. Indeed, treatment of MCF7 cells with an ASO designed to base pair with a splicing enhancer element within *HNRNPH1* exon 4 was sufficient to reduce the overall *HNRNPH1* expression level (Figure 40). Rescue of the *HNRNPH1* expression in response to NMD inhibition implies that the ASO treatment lowered the *HNRNPH1* expression via a splicing switch favouring the *HNRNPH1* NMD splice isoform (Figure 40).

In conclusion, the here presented data show how a splice-switching ASO can be used to decrease cellular *HNRNPH1* expression levels. It was shown for the first time that treatment with an *HNRNPH1*-targeting splice-switching ASO is sufficient to decrease the proliferative capacity of the breast cancer cell line MCF7 (Figure 41). All together, these data open new doors for the investigation of the therapeutic potential of splice-switching ASOs to target dysregulated *HNRNPH* expression in cancers.

5 Outlook

The HNRNPH interactome highlighted the interaction of HNRNPH with other RBPs. However, the presented PhD thesis did not further address the question, whether direct cooperation takes part in HNRNPH-mediated cooperative splicing. To elucidate this question in more detail, it would be interesting to perform *in vitro* experiments with mutational HNRNPH proteins. Electrophoretic mobility shift assays or RNaseH protection assays, as Gueroussov and colleagues have performed for HNRNPD, could deliver new insights into the formation of high-order protein complexes of HNRNPH proteins with each other (Gueroussov et al., 2017). Further, it would be interesting to perform HNRNPH titration series in cells using genetically modified cells lines that endogenously express mutant versions of HNRNPH. These HNRNPH titration series could reveal whether the removal of specific functional domains of HNRNPH, like i.e., the C-terminal GY-domain, could disrupt the here presented transcriptome-wide HNRNPH-mediated cooperative splicing regulation.

In order to extend the analysis for the role of rG4s in HNRNPH-mediated cooperative splicing regulation *in vivo*, the splicing analysis of additional minigenes, containing HNRNPH-regulated splicing events, would allow a more detailed conclusion about the importance of rG4s for the cooperative splicing regulation *in vivo*. To address this question, it could be interesting to use a minigene high-throughput screening approach comprising a library with HNRNPH-mediated splicing events (as in Braun et al., 2018). Comparable to the mutagenic screening approach performed in the presented work for the *AKR1A1* minigene, a minigene library could include minigene variations with rG4 and non-rG4 HNRNPH binding sites. Combining such a minigene library with an HNRNPH titration *in vivo*, could deliver detailed information about features that are important for HNRNPH-mediated cooperative splicing regulation.

A transcriptome-wide cooperative splicing regulation, as shown here for HNRNPH, suggests an increased sensitivity of cells for an imbalanced HNRNPH expression. Therefore, this study proposed the usage of a splice-switching ASO to control HNRNPH1 expression in MCF7 cells. Future studies should try to characterise how an ASO-treatment affects HNRNPH-mediated splicing regulation and cellular functionality. Since increased expression levels of HNRNPH correlate with an overall reduced patient survival in

glioblastoma and hepatocellular carcinoma patients, it could be of special interest to transfer the splice-switching ASO treatment of HNRNPH to these cancer types (Lefave et al., 2011; Li et al., 2016; Herviou et al., 2020; Le Bras et al., 2022). The usage of splice-switching ASOs could serve as an elegant way to correct increased HNRNPH expression levels in these diseases.

6 References

Aartsma-Rus, A. Overview on AON design. *Methods in molecular biology (Clifton, N.J.)* **867,** 117–129; 10.1007/978-1-61779-767-5_8 (2012).

Aartsma-Rus, A., van Deutekom, J. C. T., Fokkema, I. F., van Ommen, G.-J. B. & Dunnen, J. T. den. Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle & nerve* **34,** 135–144; 10.1002/mus.20586 (2006).

Abramowicz, A. & Gos, M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of applied genetics* **59,** 253–268; 10.1007/s13353-018-0444-7 (2018)

Adair, G. S., Bock, A. V. & Field, H. THE HEMOGLOBIN SYSTEM. *The Journal of biological chemistry* **63,** 529–545; 10.1016/S0021-9258(18)85018-9 (1925).

Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K. & Dreyfuss, G. mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Molecular and cellular biology* **6,** 2932–2943; 10.1128/mcb.6.8.2932-2943.1986 (1986).

Agafonov, D. E. *et al.* Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Molecular and cellular biology* **31,** 2667–2682; 10.1128/MCB.05266-11 (2011).

Atlasi, Y., Mowla, S. J., Ziaee, S. A. M., Gokhale, P. J. & Andrews, P. W. OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem cells (Dayton, Ohio)* **26,** 3068–3074; 10.1634/stemcells.2008-0530 (2008).

Bagga, P. S., Ford, L. P., Chen, F. & Wilusz, J. The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. *Nucleic acids research* **23,** 1625–1631; 10.1093/nar/23.9.1625 (1995).

Baraniak, A. P., Lasda, E. L., Wagner, E. J. & Garcia-Blanco, M. A. A stem structure in fibroblast growth factor receptor 2 transcripts mediates cell-type-specific splicing by approximating intronic control elements. *Molecular and cellular biology* **23,** 9327–9337; 10.1128/MCB.23.24.9327-9337.2003 (2003).

Becker, W. R., Jarmoskaite, I., Vaidyanathan, P. P., Greenleaf, W. J. & Herschlag, D. Demonstration of protein cooperativity mediated by RNA structure using the human protein PUM2. *RNA (New York, N.Y.)* **25,** 702–712; 10.1261/rna.068585.118 (2019).

Berget, S. M. Exon recognition in vertebrate splicing. *The Journal of biological chemistry* **270,** 2411–2414; 10.1074/jbc.270.6.2411 (1995).

Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature chemistry* **6,** 75–80; 10.1038/nchem.1805 (2014).

Bohr, C., Hasselbalch, K. & Krogh, A. Ueber einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt1. *Skandinavisches Archiv Für Physiologie* **16,** 402–412; 10.1111/j.1748-1716.1904.tb01382.x (1904).

Bonnal, S. *et al.* A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *The Journal of biological chemistry* **278,** 39330–39336; 10.1074/jbc.M305580200 (2003).

Braun, S. *et al.* Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis. *Nature communications* **9,** 3315; 10.1038/s41467-018-05748-7 (2018).

Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome research* **24,** 1774–1786; 10.1101/gr.177790.114 (2014).

Brownmiller, T. & Caplen, N. J. The HNRNPF/H RNA binding proteins and disease. *Wiley interdisciplinary reviews. RNA* **14,** e1788; 10.1002/wrna.1788 (2023).

Buchbender, A. *et al.* Improved library preparation with the new iCLIP2 protocol. *Methods (San Diego, Calif.)* **178,** 33–48; 10.1016/j.ymeth.2019.10.003 (2020).

Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic acids research* **40,** 4727–4741; 10.1093/nar/gks068 (2012).

Burd, C. G. & Dreyfuss, G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *The EMBO journal* **13,** 1197–1204; 10.1002/j.1460-2075.1994.tb06369.x (1994).

Caputi, M. & Zahler, A. M. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *The Journal of biological chemistry* **276,** 43850–43859; 10.1074/jbc.M102861200 (2001).

Cartegni, L. *et al.* hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *Journal of molecular biology* **259,** 337–348; 10.1006/jmbi.1996.0324 (1996).

Cartegni, L., Hastings, M. L., Calarco, J. A., Stanchina, E. de & Krainer, A. R. Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *American journal of human genetics* **78,** 63–77; 10.1086/498853 (2006).

Cartegni, L. & Krainer, A. R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature genetics* **30,** 377–384; 10.1038/ng854 (2002).

Chen, C. D. & Helfman, D. M. Donor site competition is involved in the regulation of alternative splicing of the rat beta-tropomyosin pre-mRNA. *RNA (New York, N.Y.)* **5,** 290–301; 10.1017/s1355838299980743 (1999).

Chen, C. D., Kobayashi, R. & Helfman, D. M. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes & development* **13,** 593–606; 10.1101/gad.13.5.593 (1999).

Cho, S. & Dreyfuss, G. A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity. *Genes & development* **24,** 438–442; 10.1101/gad.1884910 (2010).

Chou, M. Y., Rooke, N., Turck, C. W. & Black, D. L. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Molecular and cellular biology* **19,** 69–77; 10.1128/MCB.19.1.69 (1999).

Cian, A. de *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods (San Diego, Calif.)* **42,** 183–195; 10.1016/j.ymeth.2006.10.004 (2007).

Cieniková, Z., Jayne, S., Damberger, F. F., Allain, F. H.-T. & Maris, C. Evidence for cooperative tandem binding of hnRNP C RRMs in mRNA processing. *RNA (New York, N.Y.)* **21,** 1931–1942; 10.1261/rna.052373.115 (2015).

Conlon, E. G. *et al.* The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *eLife* **5**; 10.7554/eLife.17820 (2016).

Dalziel, M., Nunes, N. M. & Furger, A. Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. *Molecular and cellular biology* **27,** 1568–1580; 10.1128/MCB.01821-06 (2007).

Dardenne, E. *et al.* RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell reports* **7,** 1900–1913; 10.1016/j.celrep.2014.05.010 (2014).

Daugherty, M. D., Liu, B. & Frankel, A. D. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nature structural & molecular biology* **17,** 1337–1342; 10.1038/nsmb.1902 (2010).

Decorsière, A., Cayrel, A., Vagner, S. & Millevoi, S. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes & development* **25,** 220–225; 10.1101/gad.607011 (2011).

Didiot, M.-C. *et al.* The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic acids research* **36,** 4902–4912; 10.1093/nar/gkn472 (2008).

Dominguez, C. & Allain, F. H.-T. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic acids research* **34,** 3634–3645; 10.1093/nar/gkl488 (2006).

Dominguez, C., Fisette, J.-F., Chabot, B. & Allain, F. H.-T. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nature structural & molecular biology* **17,** 853–861; 10.1038/nsmb.1814 (2010).

Dreyfuss, G., Swanson, M. S. & Piñol-Roma, S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends in biochemical sciences* **13,** 86–91; 10.1016/0968-0004(88)90046-1 (1988).

Dreyfuss, G., Kim, V. N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nature reviews. Molecular cell biology* **3,** 195–205; 10.1038/nrm760 (2002).

Dumas, L., Herviou, P., Dassi, E., Cammas, A. & Millevoi, S. G-Quadruplexes in RNA Biology: Recent Advances and Future Directions. *Trends in biochemical sciences* **46,** 270–283; 10.1016/j.tibs.2020.11.001 (2021).

Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *Journal of molecular biology* **429,** 2127–2147; 10.1016/j.jmb.2017.05.017 (2017).

Fisette, J.-F., Toutant, J., Dugré-Brisson, S., Desgroseillers, L. & Chabot, B. hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *RNA (New York, N.Y.)* **16,** 228–238; 10.1261/rna.1890310 (2010).

Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147,** 132–146; 10.1016/j.cell.2011.08.023 (2011).

Gautrey, H. *et al.* SRSF3 and hnRNP H1 regulate a splicing hotspot of HER2 in breast cancer cells. *RNA biology* **12,** 1139–1151; 10.1080/15476286.2015.1076610 (2015).

Gehring, N. H. & Roignant, J.-Y. Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends in genetics : TIG* **37,** 355–372; 10.1016/j.tig.2020.10.008 (2021).

Georgakopoulos-Soares, I. *et al.* Alternative splicing modulation by G-quadruplexes. *Nature communications* **13,** 2404; 10.1038/s41467-022-30071-7 (2022).

Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Human genetics* **135,** 851–867; 10.1007/s00439-016-1683-5 (2016).

Goguel, V. & Rosbash, M. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* **72,** 893–901; 10.1016/0092-8674(93)90578-e (1993).

Gomez, D. *et al.* Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic acids research* **32,** 371–379; 10.1093/nar/gkh181 (2004).

Graveley, B. R. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123,** 65–73; 10.1016/j.cell.2005.07.028 (2005).

Gueroussov, S. *et al.* Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* **170,** 324-339.e23; 10.1016/j.cell.2017.06.037 (2017).

Guiset Miserachs, H., Donghi, D., Börner, R., Johannsen, S. & Sigel, R. K. O. Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry* **21,** 975–986; 10.1007/s00775-016-1393-4 (2016).

Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science (New York, N.Y.)* **353**; 10.1126/science.aaf5371 (2016).

Hacht, A. von *et al.* Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic acids research* **42,** 6630–6644; 10.1093/nar/gku290 (2014).

HafezQorani, S. *et al.* Modeling the combined effect of RNA-binding proteins and microRNAs in post-transcriptional regulation. *Nucleic acids research* **44,** e83; 10.1093/nar/gkw048 (2016).

Halder, K., Wieland, M. & Hartig, J. S. Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic acids research* **37,** 6811–6817; 10.1093/nar/gkp696 (2009).

Havens, M. A. & Hastings, M. L. Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic acids research* **44,** 6549–6563; 10.1093/nar/gkw533 (2016).

Hawkins, J. D. A survey on intron and exon lengths. *Nucleic acids research* **16,** 9893–9908; 10.1093/nar/16.21.9893 (1988).

Heo, Y.-A. Golodirsen: First Approval. *Drugs* **80,** 329–333; 10.1007/s40265-020-01267-2 (2020).

Hertel, K. J. Combinatorial control of exon recognition. *The Journal of biological chemistry* **283,** 1211–1215; 10.1074/jbc.R700035200 (2008).

Herviou, P. *et al.* hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nature communications* **11,** 2661; 10.1038/s41467-020-16168-x (2020).

Hon, J., Martínek, T., Zendulka, J. & Lexa, M. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics (Oxford, England)* **33,** 3373–3379; 10.1093/bioinformatics/btx413 (2017).

Honoré, B. *et al.* Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. *The Journal of biological chemistry* **270,** 28780–28789; 10.1074/jbc.270.48.28780 (1995).

Howard, J. M. & Sanford, J. R. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley interdisciplinary reviews. RNA* **6,** 93–110; 10.1002/wrna.1260 (2015).

Hoy, S. M. Nusinersen: First Global Approval. *Drugs* **77,** 473–479; 10.1007/s40265-017-0711-7 (2017).

Huang, H., Zhang, J., Harvey, S. E., Hu, X. & Cheng, C. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes & development* **31,** 2296–2309; 10.1101/gad.305862.117 (2017).

Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic acids research* **33,** 2908–2916; 10.1093/nar/gki609 (2005).

Jin, L. *et al.* STRAP regulates alternative splicing fidelity during lineage commitment of mouse embryonic stem cells. *Nature communications* **11,** 5941; 10.1038/s41467-020-19698-6 (2020).

Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* **43,** 513–525; 10.1016/j.neuron.2004.07.022 (2004).

Kang, D., Lee, Y. & Lee, J.-S. RNA-Binding Proteins in Cancer: Functional and Therapeutic Perspectives. *Cancers* **12**; 10.3390/cancers12092699 (2020).

Kashima, T. & Manley, J. L. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nature genetics* **34,** 460–463; 10.1038/ng1207 (2003).

Keiper, S. *et al.* Smu1 and RED are required for activation of spliceosomal B complexes assembled on short introns. *Nature communications* **10,** 3639; 10.1038/s41467-019-11293-8 (2019).

Kikin, O., D'Antonio, L. & Bagga, P. S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic acids research* **34,** W676-82; 10.1093/nar/gkl253 (2006).

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* **9**; 10.12688/f1000research.24956.2 (2020).

Komaki, H. *et al.* Viltolarsen in Japanese Duchenne muscular dystrophy patients: A phase 1/2 study. *Annals of clinical and translational neurology* **7,** 2393–2408; 10.1002/acn3.51235 (2020).

Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature reviews. Molecular cell biology* **14,** 153–165; 10.1038/nrm3525 (2013).

Krakau, S., Richard, H. & Marsico, A. PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome biology* **18,** 240; 10.1186/s13059-017-1364-2 (2017).

Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature chemical biology* **3,** 218–221; 10.1038/nchembio864 (2007).

Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nature methods* **13,** 841–844; 10.1038/nmeth.3965 (2016).

Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic acids research* **36,** 5482–5515; 10.1093/nar/gkn517 (2008).

Le Bras, M. *et al.* Translational Regulation by hnRNP H/F Is Essential for the Proliferation and Survival of Glioblastoma. *Cancers* **14**; 10.3390/cancers14051283 (2022).

Lee, S. J. *et al.* hnRNPs Interacting with mRNA Localization Motifs Define Axonal RNA Regulons. *Molecular & cellular proteomics : MCP* **17,** 2091–2106; 10.1074/mcp.RA118.000603 (2018).

Lee, T. *et al.* 2'-O-Methyl RNA/Ethylene-Bridged Nucleic Acid Chimera Antisense Oligonucleotides to Induce Dystrophin Exon 45 Skipping. *Genes* **8**; 10.3390/genes8020067 (2017).

Lefave, C. V. *et al.* Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *The EMBO journal* **30,** 4084–4097; 10.1038/emboj.2011.259 (2011).

Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80,** 155–165; 10.1016/0092-8674(95)90460-3 (1995).

Li, J. *et al.* Rbm14 maintains the integrity of genomic DNA during early mouse embryogenesis via mediating alternative splicing. *Cell proliferation* **53,** e12724; 10.1111/cpr.12724 (2020).

Li, X. *et al.* A splicing switch from ketohexokinase-C to ketohexokinase-A drives hepatocellular carcinoma formation. *Nature cell biology* **18,** 561–571; 10.1038/ncb3338 (2016).

Lin, Y.-H. & Bundschuh, R. Interplay between single-stranded binding proteins on RNA secondary structure. *Physical review. E, Statistical, nonlinear, and soft matter physics* **88,** 52707; 10.1103/PhysRevE.88.052707 (2013).

Lin, Y.-H. & Bundschuh, R. RNA structure generates natural cooperativity between single-stranded RNA binding proteins targeting 5' and 3'UTRs. *Nucleic acids research* **43,** 1160–1169; 10.1093/nar/gku1320 (2015).

Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* **25,** 402–408; 10.1006/meth.2001.1262 (2001).

Lorson, C. L., Hahnen, E., Androphy, E. J. & Wirth, B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 6307–6311; 10.1073/pnas.96.11.6307 (1999).

Malina, J., Scott, P. & Brabec, V. Stabilization of human telomeric RNA G-quadruplex by the water-compatible optically pure and biologically-active metallohelices. *Scientific reports* **10,** 14543; 10.1038/s41598-020-71429-5 (2020).

Marcel, V. *et al.* G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **32,** 271–278; 10.1093/carcin/bgq253 (2011).

Matunis, M. J., Xing, J. & Dreyfuss, G. The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. *Nucleic acids research* **22,** 1059–1067; 10.1093/nar/22.6.1059 (1994).

May, G. E., Olson, S., McManus, C. J. & Graveley, B. R. Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA (New York, N.Y.)* **17,** 222–229; 10.1261/rna.2521311 (2011).

Meininger, I. *et al.* Alternative splicing of MALT1 controls signalling and activation of CD4(+) T cells. *Nature communications* **7,** 11292; 10.1038/ncomms11292 (2016).

Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science (New York, N.Y.)* **338,** 1593–1599; 10.1126/science.1228186 (2012).

Millevoi, S., Moine, H. & Vagner, S. G-quadruplexes in RNA biology. *Wiley interdisciplinary reviews. RNA* **3,** 495–507; 10.1002/wrna.1113 (2012).

Monani, U. R. *et al.* A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Human molecular genetics* **8,** 1177–1183; 10.1093/hmg/8.7.1177 (1999).

Monteuuis, G., Wong, J. J. L., Bailey, C. G., Schmitz, U. & Rasko, J. E. J. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic acids research* **47,** 11497–11513; 10.1093/nar/gkz1068 (2019).

Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D. & Basu, S. An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *Journal of the American Chemical Society* **132,** 17831–17839; 10.1021/ja106287x (2010).

Muh, S. J., Hovhannisyan, R. H. & Carstens, R. P. A Non-sequence-specific double-stranded RNA structural element regulates splicing of two mutually exclusive exons of fibroblast growth factor receptor 2 (FGFR2). *The Journal of biological chemistry* **277,** 50143–50154; 10.1074/jbc.M207409200 (2002).

Neckles, C. *et al.* HNRNPH1-dependent splicing of a fusion oncogene reveals a targetable RNA G-quadruplex interaction. *RNA (New York, N.Y.)* **25,** 1731–1750; 10.1261/rna.072454.119 (2019).

Pararajalingam, P. *et al.* Coding and noncoding drivers of mantle cell lymphoma identified through exome and genome sequencing. *Blood* **136,** 572–584; 10.1182/blood.2019002385 (2020).

Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. The Expanding Landscape of Alternative Splicing Variation in Human Populations. *American journal of human genetics* **102,** 11–26; 10.1016/j.ajhg.2017.11.002 (2018).

Pereira, B., Billaud, M. & Almeida, R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in cancer* **3,** 506–528; 10.1016/j.trecan.2017.05.003 (2017).

Piovesan, A. *et al.* Human protein-coding genes and gene feature statistics in 2019. *BMC research notes* **12,** 315; 10.1186/s13104-019-4343-8 (2019).

Pritsker, M., Doniger, T. T., Kramer, L. C., Westcot, S. E. & Lemischka, I. R. Diversification of stem cell molecular repertoire by alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 14290–14295; 10.1073/pnas.0502132102 (2005).

Raker, V. A., Mironov, A. A., Gelfand, M. S. & Pervouchine, D. D. Modulation of alternative splicing by long-range RNA structures in Drosophila. *Nucleic acids research* **37,** 4533–4544; 10.1093/nar/gkp407 (2009).

Randazzo, A., Spada, G. P. & Da Silva, M. W. Circular dichroism of quadruplex structures. *Topics in current chemistry* **330,** 67–86; 10.1007/128_2012_331 (2013).

Rao, S. *et al.* Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. *Molecular and cellular biology* **30,** 5364–5380; 10.1128/MCB.00419-10 (2010).

Rengifo-Gonzalez, J. C. *et al.* The cooperative binding of TDP-43 to GU-rich RNA repeats antagonizes TDP-43 aggregation. *eLife* **10**; 10.7554/eLife.67605 (2021).

Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PloS one* **10,** e0146021; 10.1371/journal.pone.0146021 (2015).

Robberson, B. L., Cote, G. J. & Berget, S. M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and cellular biology* **10,** 84–94; 10.1128/mcb.10.1.84-94.1990 (1990).

Roca, X. *et al.* Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes & development* **26,** 1098–1109; 10.1101/gad.190173.112 (2012).

Roca, X., Krainer, A. R. & Eperon, I. C. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes & development* **27,** 129–144; 10.1101/gad.209759.112 (2013).

Samatanga, B., Dominguez, C., Jelesarov, I. & Allain, F. H.-T. The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic acids research* **41,** 2505–2516; 10.1093/nar/gks1289 (2013).

Sapir, T. *et al.* Heterogeneous nuclear ribonucleoprotein U (HNRNPU) safeguards the developing mouse cortex. *Nature communications* **13,** 4209; 10.1038/s41467-022-31752-z (2022).

Schaub, M. C., Lopez, S. R. & Caputi, M. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *The Journal of biological chemistry* **282,** 13617–13626; 10.1074/jbc.M700774200 (2007).

Schneider, M. *et al.* Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Molecular cell* **38,** 223–235; 10.1016/j.molcel.2010.02.027 (2010).

Shahid, R., Bugaut, A. & Balasubramanian, S. The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **49,** 8300–8306; 10.1021/bi100957h (2010).

Sharma, S., Kohlstaedt, L. A., Damianov, A., Rio, D. C. & Black, D. L. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined

spliceosome. *Nature structural & molecular biology* **15,** 183–191; 10.1038/nsmb.1375 (2008).

Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nature reviews. Molecular cell biology* **18,** 655–670; 10.1038/nrm.2017.86 (2017).

Somvanshi, P. R. & Venkatesh, K. V. Hill Equation. In *Encyclopedia of Systems Biology*, edited by W. Dubitzky, O. Wolkenhauer, K.-H. Cho & H. Yokota (Springer New York, New York, NY, 2013), pp. 892–895 (2013).

Subramanian, M. *et al.* G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO reports* **12,** 697–704; 10.1038/embor.2011.76 (2011).

Sutandy, F. X. R. *et al.* In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome research* **28,** 699–713; 10.1101/gr.229757.117 (2018).

Swanson, M. S. & Dreyfuss, G. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Molecular and cellular biology* **8,** 2237–2241; 10.1128/mcb.8.5.2237-2241.1988 (1988).

Swanson, M. S., Nakagawa, T. Y., LeVan, K. & Dreyfuss, G. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Molecular and cellular biology* **7,** 1731–1739; 10.1128/mcb.7.5.1731-1739.1987 (1987).

Syed, Y. Y. Eteplirsen: First Global Approval. *Drugs* **76,** 1699–1704; 10.1007/s40265-016-0657-1 (2016).

Thibault, P. A. *et al.* hnRNP A/B Proteins: An Encyclopedic Assessment of Their Roles in Homeostasis and Disease. *Biology* **10**; 10.3390/biology10080712 (2021).

Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research* **33,** 2901–2907; 10.1093/nar/gki553 (2005).

Tyson-Capper, A. & Gautrey, H. Regulation of Mcl-1 alternative splicing by hnRNP F, H1 and K in breast cancer cells. *RNA biology* **15,** 1448–1457; 10.1080/15476286.2018.1551692 (2018).

Uren, P. J. *et al.* High-throughput analyses of hnRNP H1 dissects its multi-functional aspect. *RNA biology* **13,** 400–411; 10.1080/15476286.2015.1138030 (2016).

van Dusen, C. M., Yee, L., McNally, L. M. & McNally, M. T. A glycine-rich domain of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an interaction with transportin 1. *Molecular and cellular biology* **30,** 2552–2562; 10.1128/MCB.00230-09 (2010).

Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife* **5,** e11752; 10.7554/eLife.11752 (2016).

Vo, T. *et al.* HNRNPH1 destabilizes the G-quadruplex structures formed by G-rich RNA sequences that regulate the alternative splicing of an oncogenic fusion transcript. *Nucleic acids research* **50,** 6474–6496; 10.1093/nar/gkac409 (2022).

Vorlíčková, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods (San Diego, Calif.)* **57,** 64–75; 10.1016/j.ymeth.2012.03.011 (2012).

Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–718; 10.1016/j.cell.2009.02.009 (2009).

Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476; 10.1038/nature07509 (2008).

Wang, E., Dimova, N. & Cambi, F. PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes. *Nucleic acids research* **35,** 4164–4178; 10.1093/nar/gkm387 (2007).

Wang, J., Liu, Q. & Shyr, Y. Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC genomics* **16 Suppl 7,** S5; 10.1186/1471-2164-16-S7-S5 (2015).

Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA (New York, N.Y.)* **14,** 802–813; 10.1261/rna.876308 (2008).

Watts, J. K. & Corey, D. R. Silencing disease genes in the laboratory and the clinic. *The Journal of pathology* **226,** 365–379; 10.1002/path.2993 (2012).

Wong, A. & Wu, G. Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5'-monophosphate: a solid-state NMR study. *Journal of the American Chemical Society* **125,** 13895–13905; 10.1021/ja0302174 (2003).

Xiao, R. *et al.* Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Molecular cell* **45,** 656–668; 10.1016/j.molcel.2012.01.009 (2012).

Xing, S. *et al.* Identification of HnRNPC as a novel Tau exon 10 splicing factor using RNA antisense purification mass spectrometry. *RNA biology* **19,** 104–116; 10.1080/15476286.2021.2015175 (2022).

Xiong, H. Y. *et al.* RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science (New York, N.Y.)* **347,** 1254806; 10.1126/science.1254806 (2015).

Xu, B., Shi, Y., Wu, Y., Meng, Y. & Jin, Y. Role of RNA secondary structures in regulating Dscam alternative splicing. *Biochimica et biophysica acta. Gene regulatory mechanisms* **1862,** 194381; 10.1016/j.bbagrm.2019.04.008 (2019).

Xu, C. *et al.* HnRNP F/H associate with hTERC and telomerase holoenzyme to modulate telomerase function and promote cell proliferation. *Cell death and differentiation* **27,** 1998–2013; 10.1038/s41418-019-0483-6 (2020).

Xue, Y. *et al.* Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* **152,** 82–96; 10.1016/j.cell.2012.11.045 (2013).

Yamazaki, T. *et al.* TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes & development* **32,** 1161–1174; 10.1101/gad.316984.118 (2018).

Ye, J. *et al.* hnRNP U protein is required for normal pre-mRNA splicing and postnatal heart development and function. *Proceedings of the National Academy of Sciences of the United States of America* **112,** E3020-9; 10.1073/pnas.1508461112 (2015).

Yeo, G. W. *et al.* Alternative splicing events identified in human embryonic stem cells and neural progenitors. *PLoS computational biology* **3,** 1951–1967; 10.1371/journal.pcbi.0030196 (2007).

Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152,** 453–466; 10.1016/j.cell.2012.12.023 (2013).

Zhang, J., Harvey, S. E. & Cheng, C. A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. *Nucleic acids research* **47,** 3667–3679; 10.1093/nar/gkz036 (2019).

Zhou, L.-T. *et al.* A novel role of fragile X mental retardation protein in pre-mRNA alternative splicing through RNA-binding protein 14. *Neuroscience* **349,** 64–75; 10.1016/j.neuroscience.2017.02.044 (2017).

7 Appendix

7.1 *AKR1A1* iCLIP and RT-Stop profiling browser shots

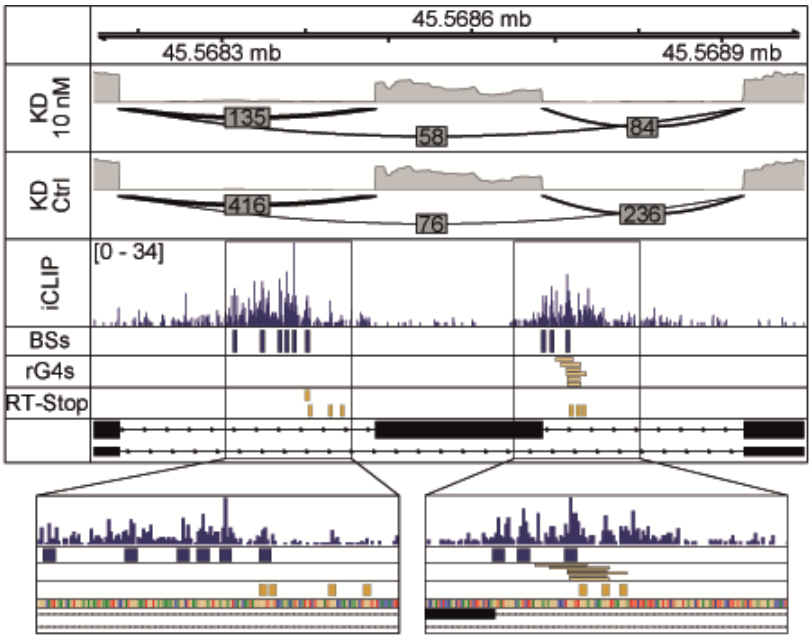

Figure 42: Genome browser screenshots for *AKR1A1* showing read counts from the HNRNPH *in vivo* CLIP in MCF7 cells and from the RT-Stop profiling experiment using the ivRNA library.
The plot shows the splice junction reads for alternative splicing of *AKR1A1* exon 6 from RNA-Seq data of HNRNPH knockdown (KD 10 nM) and Ctrl (KD Ctrl) experiments in MCF7 cells. Further, the plot illustrates the HNRNPH iCLIP reads and the Pure-CLIP called binding sites (BSs) for the respective genomic region. Additionally, *in silico* rG4 predictions (pqrsfinder) are shown for the two intronic HNRNPH BS clusters. Lastly, positions of RT-Stop reads from RT-Stop profiling experiments using the ivRNA library are shown. Illustration was done by [redacted].

7.2 Genome browser screenshots of rG4 and non-rG4 constructs

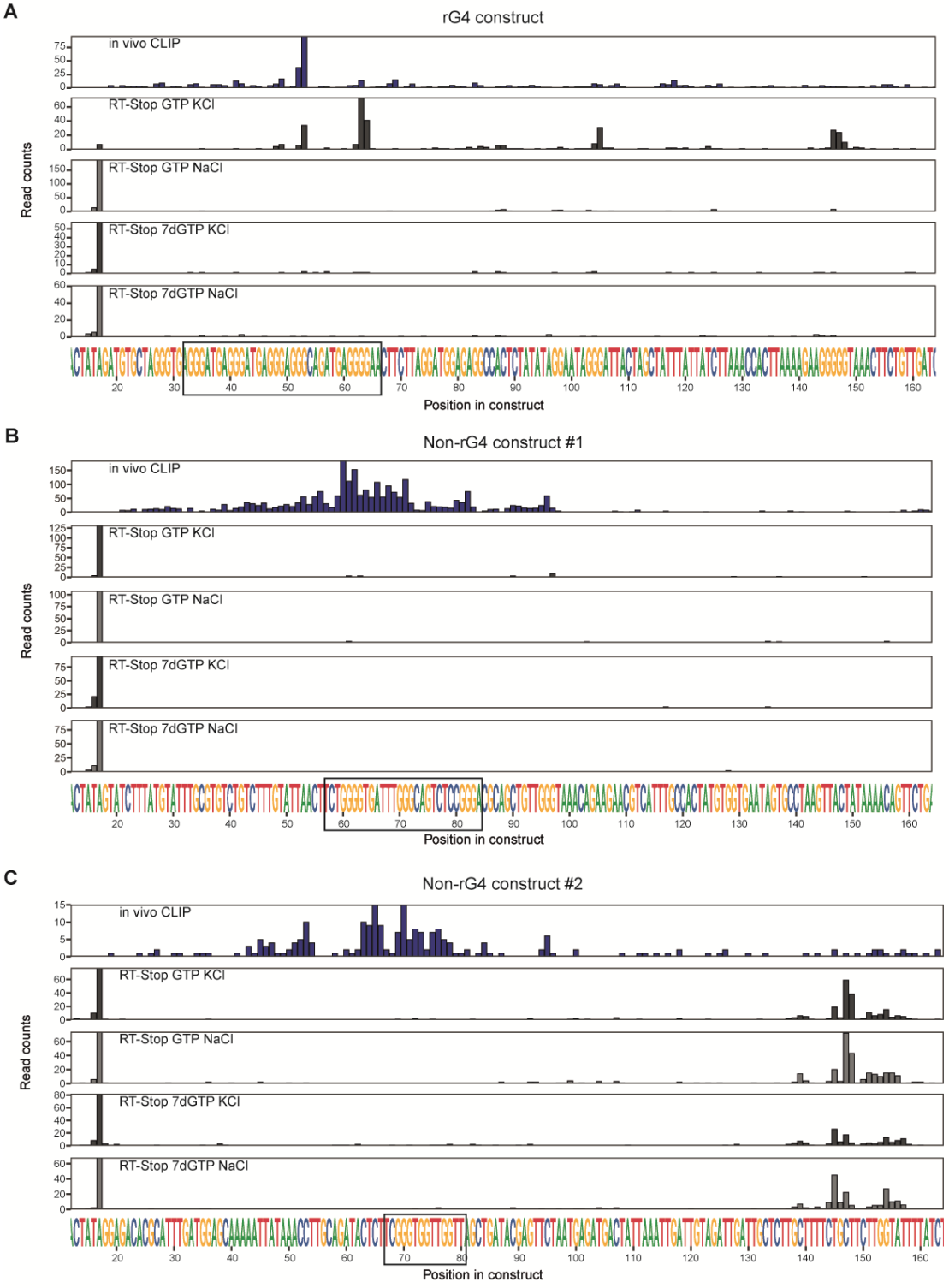

Figure 43: Genome browser screenshots showing RT-Stop read counts and the *in vivo* HNRNPH iCLIP signal for exemplary rG4 and non-rG4 constructs from the ivRNA library.
The black boxes highlight the sequences that were used to clone artificial non-rG4 and rG4 *AKR1A1* minigene constructs (Figure 36).

7.3 Abbreviations

Table 23: List of Abbreviations

Abbreviation	Explanation
ΔPSI	Change of the PSI
3'ss	3' Splice site
5'ss	5' Splice site
7dGTP	7-deaza-GTP
A3SS	Alternative 3' splice site usage
A5SS	Alternative 5' splice site usage
AE	Alternative exon
AE3'	3' End of the alternative exon
AE5'	5' End of the alternative exon
AFE	Alternative first exon
ALE	Alternative last exon
ALS	Amyotrophic lateral sclerosis
ASO	Antisense oligonucleotide
BC	Barcode
BMLS	Buchmann Institute for Molecular Life Sciences & Faculty Biological Sciences
bp	Base pairs
BS	HNRNPH binding sites
CD	Circular dichroism
CD44	Cell adhesion molecule
cDNA	Complementary DNA
CDS	Coding sequence
CE	Cassette exon
Chr.	Chromosome
CHX	Cycloheximide
Coop	Cooperatively regulated
Ctrl	Control
DE	Downstream exon
DE5'	5' end of the downstream exon
DI	Downstream intron
DI3'	3' End of the downstream intron

DI5’	5’ End of the downstream intron
DMEM	Dulbecco’s modified Eagle medium
DMS	Dimethyl sulfate
dPSI	Delta percent spliced in
enh	Enhanced
Enhanced	HNRNPH acts as a splicing enhancer
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
EV	Empty vector
FRET	Fluorescence Resonance Energy Transfer
G	Guanine
GGG	Triple G-stretch
GGGG	Quadruple G-stretch
GO	Gene ontology
GY	Glycine-tyrosine
GY	Glycine/tyrosine rich protein domain
GYR	Glycine/tyrosine/arginine rich protein domain
HNRNP	Heterogeneous ribonucleoproteins
HNRNPH	Geterogeneous nuclear ribonucleoprotein H
HNRNPH KD	HNRNPH1 downregulation
HNRNPH OE	HNRNPH1 overexpression
HNRNPH1 w/o GY	HNRNPH1 protein with deleted GY-domain
iCLIP	Individual-nucleotide resolution UV crosslinking and immunoprecipitation
IDT	Integrated DNA Technology
IMB	Institute of Molecular Biology
IP	Immunoprecipitation
IR	Intron retention
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
ivCLIP	*In vitro* iCLIP
ivRNA library	*In vitro* oligonucleotide library
K	Potassium
$K_{0.5}$	Half-saturation constant
KD	Knockdown

Li	Lithium
LSVs	Local splicing variations
ME	Mutually exclusive exons
mRIPA	Modified RIPA buffer
mRNA	Messenger RNA
MST1R	Macrophage stimulating 1 receptor
MTT	3-(4, 5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide)
n	Replicate
Na	Sodium
NGS	Next-Generation Sequencing
nH	Hill coefficient
NLS	Nuclear localisation signal
NMD	Nonsense-mediated decay
Non-cooperative	Not cooperatively regulated by HNRNPH
nt	Nucleotides
OE	Overexpression
ORF	Open reading frame
pre-mRNA	Precursor mRNA
prG4	Predicted rG4
PSI	Percent spliced in
PTC	Premature termination codons
qPCR	Quantitative real-time PCR
qRRM	Quasi-RRM
RBP	RNA-binding protein
rep	Repressed
Rep	Replicate
Repressed	HNRNPH acts as a splicing repressor
rG4	RNA G-quadruplex
RNA-Seq	RNA-Sequencing
RPMI	Roswell Park Memorial Institute
RRM	RNA Recognition Motif-type
RS	Arginine-Serine
RT	Reverse Transcriptase
RT-Stop	RT-Truncation
RT-Stop profiling	Reverse Transcriptase Stop profiling

SF1	Splicing Factor 1
SILAC	Stable isotope labelling using amino acids in cell culture
siRNA	Single small interfering RNA
snRNP	Small nuclear ribonucleoprotein
ss	Single stranded
TP53	Tumour suppressor gene p53
U	Units
U2AF	U2 auxiliary factor
UE	Upstream exon
UE3’	3’ End of the upstream exon
UI	Upstream intron
UI3’	3’ End of the upstream intron
UI5’	5’ End of the upstream intron
WB	Western blot
WT	Wild type

7.4 List of Figures

7.5 List of Tables