Human circular RNAs:

design, development and functional analysis

of protein sponges

Dissertation

vorgelegt von

Stephen Sukumar Nuthalapati

(Master of Science in Applied Microbiology)

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

Institut für Biochemie, Fachbereich 08 Biologie und Chemie

der Justus-Liebig-Universität Gießen

Gießen, Dezember 2021

Die vorliegende Arbeit wurde am Institut für Biochemie des Fachbereichs 08 der Justus-Liebig-Universität Gießen in der Zeit von September 2018 bis Oktober 2021 unter der Leitung von Prof. Dr. Albrecht Bindereif angefertigt.

Dekan: Prof. Dr. Thomas Wilke
Institut für Tierökologie und spezielle Ökologie
Fachbereich für Biologie und Chemie (FB08)
Justus-Liebig-Universität Gießen

1. Gutachter: Prof. Dr. Albrecht Bindereif
Institut für Biochemie
Fachbereich für Biologie und Chemie (FB08)
Justus-Liebig-Universität Gießen

2. Gutachter: Prof. Dr. Reinhard Dammann
Institut für Genetik
Fachbereich für Biologie und Chemie (FB08)
Justus-Liebig-Universität Gießen

Contents

Zusammenfassung

Zirkuläre RNAs (circRNAs) sind eine verhältnismäßig neue Klasse von stabilen nichtkodierenden RNAs, die in vielen Fällen von proteinkodierenden Genen exprimiert werden. Sie entstehen durch einen sogenannten „*Backsplicing*"-Mechanismus, wobei ein Exon oder mehrere benachbarte Exons in zirkulärer Konfiguration aus einer Prä-mRNA herausgeschnitten werden. Die Funktion der meisten circRNAs ist nicht bekannt und wahrscheinlich divers. Mehrere mögliche Funktionen wurden zwar für diese RNA-Klasse vorgeschlagen, wie zum Beispiel „*Sponging"* von miRNAs oder RNA-Bindeproteinen (RBPs), oder auch Proteintranslation, aber nur wenige davon wurden sorgfältig validiert. Diese Arbeit befasst sich mit dem *Sponging* von RBPs als einer neuen Funktion von circRNAs.

Der primäre Fokus dieser Arbeit lag darin, artifiziell entworfene circRNAs als molekulare Werkzeuge for RBP-Sequestrierung einzusetzen, mit dem Ziel, wichtige zelluläre Prozesse wie Spleißen zu modulieren. Zuerst untersuchten wir die kombinatorische Erkennung von RNA-Motiven durch RBPs auf der Grundlage von SELEX (Systematische Evolution von Liganden durch EXponentielle Anreicherung), in Kombination mit Hochdurchsatz-Sequenzierung und Motivanalyse. Dadurch wurden hochspezifische und maßgeschneiderte RBP-Bindungssequenzen identifiziert. Diese wurden in circRNAs integriert, wo sie als RBP-*Sponge*s wirkten. Um die Machbarkeit zu demonstrieren, setzten wir diese Strategie bei zwei Regulatoren des alternativen Spleißens, hnRNP L (heterogenes nukleäres RiboNukleoProtein L) und RBM24. HnRNP L ist ein globaler Regulator von alternativem Spleißen, der bevorzugt an CA-repetitive und CA-reiche RNA-Sequenzen bindet, RBM24 ist ein muskelspezifischer Regulator von alternativem Spleißen, der an GU-reiche Sequenzen von Ziel-RNAs bindet. Wir beobachteten, dass zirkuläre RNAs einen klaren Vorteil gegenüber den entsprechenden linearen RNAs besitzen, dies aufgrund ihrer höheren Stabilität und Resistenz gegenüber zellulären RNasen.

Zweitens wurde die Bindungsaffinität von SELEX-abgeleiteten, hochspezifischen Konsensus-Sequenzen und der RNA-Protein-Wechselwirkung mittels elektrophoretischer Mobilitäts- und *in vitro Pulldown*-Assays validiert. Überraschenderweise zeigten unsere Ergebnisse, dass die RNA-Bindungspräferenz von hnRNP L nicht nur von Sequenzspezifität abhängt, sondern auch von der relativen Positionierung (*Spacing*) der RNA-Motive. Die vier Domänen des hnRNP L-Proteins binden die entsprechenden RNA-Elemente mit präferierten Abständen .

Zusätzlich wurden auch direkte *in vivo* hnRNP L-RNA Interaktionen durch RNA-Immunpräzipitation (RIP) aus HeLa-Zelllysaten erfasst. Um *in vivo*-Sponging von hnRNP L in vivo nachzuweisen, sowie Effekte auf Spleißen, konzentrierten wir uns schließlich auf circRNA-Überexpression. Ein tRNA-basiertes Tornado-Ribozym-System wurde hier für die circRNA-Expression eingesetzt, wie auch *in vitro*-generierte circRNAs. Alternatives Spleißen von hnRNP L-Zielgenen wurde mittels RT-PCR analysiert. Interessanterweise beobachteten

wir hnRNP L-Verlagerung von der überwiegenden nukleären Lokalisation, hervorgerufen durch hnRNP L-*Sponging*, zu einer stärker cytoplasmatischen Verteilung in HeLa-Zellen.

Diese Arbeit demonstrierte ein beträchtliches Potential von Designer-circRNAs, die Aktivitäten von RNA-Bindeproteinen und alternative Spleißprozesse zu modulieren. Solche Designer-circRNAs können als eine vielversprechende Alternative zu pharmakologischer Hemmung von Proteine angewandt werden, ähnlich wie spleißregulierende Oligonukleotide, die auf individuelle Spleißereignisse abzielen. Dies bildet die Grundlage für das Design von optimalen circRNA-Protein-Schwämmen (*Sponges*), die für jedes RBP mit klinischer Relevanz angewandt werden können.

Summary

Circular RNAs (circRNAs) are a relatively new class of stable, non-coding RNAs often derived from protein-coding genes. They are generated by the so-called backsplicing mechanism, whereby one single or several adjacent exons are excised in circular configuration from a pre-mRNA. The functional roles of the vast majority of circRNAs are largely unknown and likely diverse. While several putative functions have been suggested for this class of RNAs, such as miRNA sponging, templates for translation and RBP (RNA binding protein) sponging. Only a few of these proposed functions have been thoroughly validated. This work explores RBP sponging as a novel function of circRNAs.

The primary focus of this work was on employing artificially designed circRNAs as molecular tools for RBP sequestration, with a view to influence key cellular processes such as splicing. *Firstly*, we investigated the combinatorial recognition of RNA motifs by RBPs using SELEX (Systematic Evolution of Ligands by EXponential enrichment) coupled with high-throughput sequencing and motif analysis, to identify highly specific and tailored RBP-binding sequences. These were implemented into a circular RNA, which functioned as an RBP sponge. As a proof of principle, we utilised this strategy for two RBPs regulating alternative splicing - hnRNP L (heterogeneous nuclear RiboNucleoProtein L) and RBM24. HnRNP L is a global regulator of alternative splicing, binding preferentially to CA-rich RNA sequences, while RBM24 is a muscle-specific alternative splice regulator that binds to GU-rich stretches on target mRNAs. We observed that circular RNAs have a clear advantage over their conventional linear counterparts due to their higher stability and resistance to degradation by cellular RNases.

Secondly, the binding affinity of the SELEX-derived, highly specific consensus sequences and the protein-RNA interaction was validated by electrophoretic mobility shift assay and *in vitro* pulldown assays. Surprisingly, our findings revealed that the binding preference of hnRNP L with its cognate RNA not only depends on sequence specificity, but also on the spacing of the motifs within the RNA. The four domains of hnRNP L appear to prefer specific nucleotide spacing for binding.

Finally, to demonstrate *in vivo* sponging of hnRNP L and its effects on splicing we focused on circRNA overexpression and applied a tRNA-based, ribozyme-driven Tornado circRNA expression system, as well as *in vitro*-generated circRNAs. Alternative splicing of hnRNP L target genes was analysed by RT-PCR. Interestingly, as a result of sponging, we observed hnRNP L translocation from the predominant nuclear localisation to a more cytoplasmic distribution in HeLa cells.

This work demonstrates the considerable potential of designer circRNAs for modulating activities of RBPs or for alterations of particular alternative splicing events. As a promising alternative to pharmacological inhibition of proteins, they can be applied in a way similar to antisense splice-switching oligonucleotides targeting individual splicing events. This forms the basis for the design of optimal circRNA-protein sponges, which can be applied to any RBP of clinical importance.

1. Introduction

The human genome is designed such that it contains only about 25,000 genes that code for at least 100,000 proteins, which constitute the proteome of the human body. One of the reasons for such large diversity in gene expression lies in the fact that, in contrast to bacterial gene organization, higher eukaryotes are made up of so-called 'split genes' (Berget *et al*., 1977; Chow *et al*., 1977). The protein-coding regions of a gene, termed 'exons' (expressed regions) are interrupted by long non-coding sequences, called 'introns' (intervening regions). The entirety of this array of coding and non-coding elements is transcribed as a pre-mRNA (pre- messenger RNA), by RNA polymerases. This discovery of the discontinuous architecture of eukaryotic genes is one of the most unanticipated findings in molecular biology (Lee & Rio, 2015).

Active genomic regions of the DNA where protein-coding genes are stored, are transcribed by RNAP II (RNA polymerase II) into precursor RNA (pre-mRNA), while ribosomal RNA (rRNA) and transfer RNA (tRNA) are transcribed by RNAP I and III, respectively (RNA polymerases I and III, reviewed by White, 2008). Transcription of RNA is followed by processing of RNA and translation of mature mRNA into proteins. This sequence of events represents the central dogma of flow of genetic information (Crick, 1970), involving several processes that are continuously coupled and highly regulated. Maintaining proper temporal and spatial control of gene expression is fundamental for all organisms.

1.1. Splicing of nuclear pre-mRNA

A process termed 'splicing' plays an extremely important role in gene regulation by expanding protein diversity through alternative splicing. This might partially explain the apparent discrepancy between gene number and organismal complexity (Graveley, 2001). There are four known mechanisms of splicing described for eukaryotes - two mechanisms of intron group I and II splicing, tRNA splicing, and pre-mRNA splicing governed by the spliceosome. The RNA splicing machinery involves a highly dynamic macromolecular complex, termed the spliceosome that assembles at the boundaries of exons and introns - the 'splice sites' in the precursor mRNA. The spliceosome catalyses the splicing reaction, in which introns are removed from the primary transcript, and exons are joined to yield the mature mRNA, which is exported to the cytoplasm and translated into protein. Therefore, the task of the mRNA splicing machinery is to recognise exons in the precursor RNA sequences and to catalyse the juxtaposition of correct exon-intron junctions, thus processing precursor mRNA into functional translatable mRNA.

Like other processes in eukaryotic gene expression, splicing is also coupled to other processes, such as capping, polyadenylation, nuclear export, and translation. RNA transcripts from mammalian genes are processed within seconds or minutes after their synthesis, which provides opportunity for a functional connection between transcription and splicing (Custódio & Carmo-Fonseca, 2016). Several links between splicing and transcription are known, and both transcription rate and chromatin structure can influence splicing outcomes

(Bentley, 2014). However, splicing also feeds back on transcription (Braunschweig *et al*., 2013). Apart from the enormous proteome expansion (Ule & Blencowe, 2019) by alternative splicing, the splicing machinery can also elaborately modulate when, where, and at what level a gene product should be expressed in an organism.

On an average, a human gene contains about 8 introns, with a median length of approximately 1 kb interspersed with exons that average <300 bp in length (Sakharkar *et al*., 2004). One extreme exception was reported for the human dystrophin gene (Pozzoli *et al*., 2002), where the pre-mRNA (> two million nt) is processed into a 14,000 nt mRNA by removal of 78 introns. To produce a functional protein, 78 pairs of intron-exon junctions need to be accurately recognised in this case. Therefore one of the major challenges in pre-mRNA splicing is reliable determination of these exon-intron boundaries.

1.1.1. Splicing pathway

The most common model of splicing mechanism relies on intron definition. Introns are defined by three important sites, the 5' splice site (5' SS), branch point (BP) adenosine, and 3' splice site (3' SS), all of which are defined by short conserved sequences. The consensus sequence elements important for intron recognition are depicted in **Figure 1.1A**.In budding yeasts, the 5' SS is followed by a highly conserved sequence, GUAUGU, and the YAG tri-nucleotide precedes the 3' SS, (Y is a pyrimidine). The BP adenosine is located 18–40 nucleotides upstream of the 3' SS in a highly conserved sequence, UACUA**A**C, in which the bold A denotes the BP adenosine (Spingola *et al*., 1999). The first and the last two nucleotides of the intron (GU- and AG-dinucleotide, respectively), as well as the branch point adenosine are invariable (**Figure 1.1B**). In humans, the sequences surrounding the 5' SS and the BP are less stringently conserved (Sheth *et al*., 2006); however the polypyrimidine tract is highly conserved.

Using radioactively labelled model substrates, the molecular mechanism of pre-mRNA splicing was first analysed *in vitro*, and the subsequent identification of splicing intermediates led to a mechanistic view of the splicing reaction (Padgett *et al*., 1984). In 1984, several groups independently carried out biochemical characterisation of splicing intermediates which established a two-step phosphoryl transfer mechanism of splicing (Padgett *et al*., 1984; Ruskin *et al*., 1984; Rodriguez *et al*., 1984; Domdey *et al*., 1984) similar to that of group II intron self-splicing. In the first reaction (termed branching), the 2' hydroxyl group of the BP adenosine targets the phosphodiester group at the 5' SS by a nucleophilic attack, producing a cleaved 5' exon and a lariat-intron–3' exon intermediate, in which the 5' phosphate of the first intron nucleotide (G) is linked to the 2' oxygen of the BP adenosine (**Figure 1.2**). In the second step (exon ligation), the newly exposed 3' hydroxyl group of the 5' exon makes the second nucleophilic attack on the phosphodiester group of the 3' SS, ligating the 5' and 3' exons to form mRNA and releasing the lariat intron.

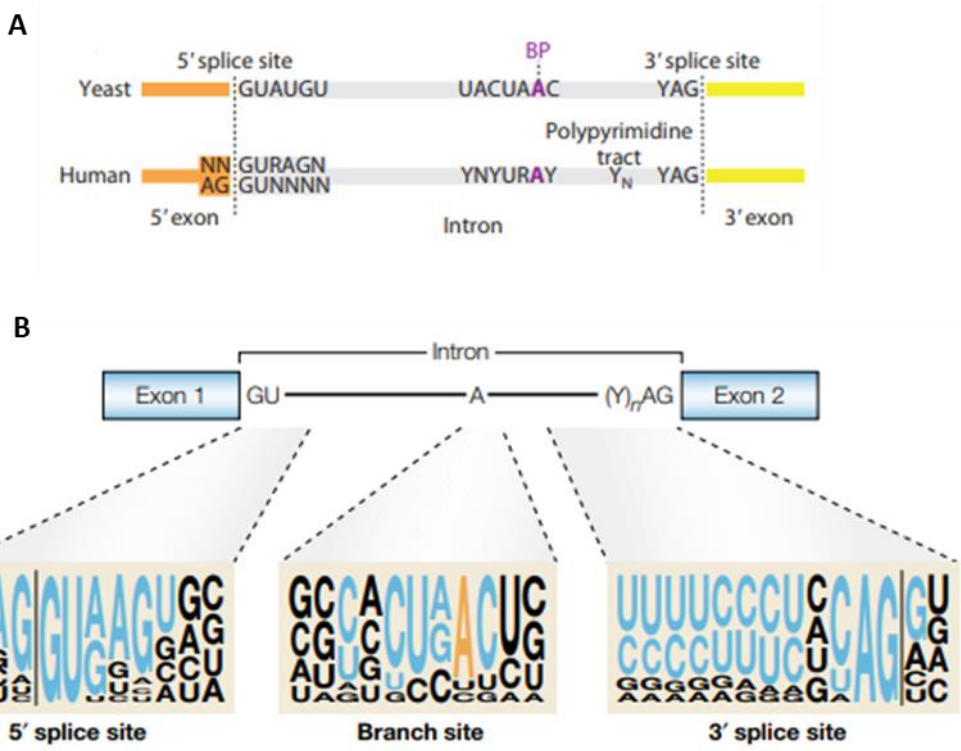

Figure 1.1 Yeast and human pre-mRNA substrates

(A) Introns are characterised by three short conserved sequences, the 5’ splice site (5’SS), Branch point sequence (BP), and 3’ splice site (3’SS). BP adenosine is indicate in purple. The sequences around splice sites are stringently conserved in yeast but more degenerate in humans. Human pre-mRNA is characterised by a polypyrimidine tract for binding of splice factors just before the YAG trinucleotide at the 3’ splice site. From Wilkinson *et al*. (2020), modified.

(B) Conserved motifs at or near the intron ends. BP adenosine is shown in orange, the invariant dinucleotide GU at the 5’SS and AG at the 3’SS respectively are shown in blue. From Cartegni *et al.* (2002).

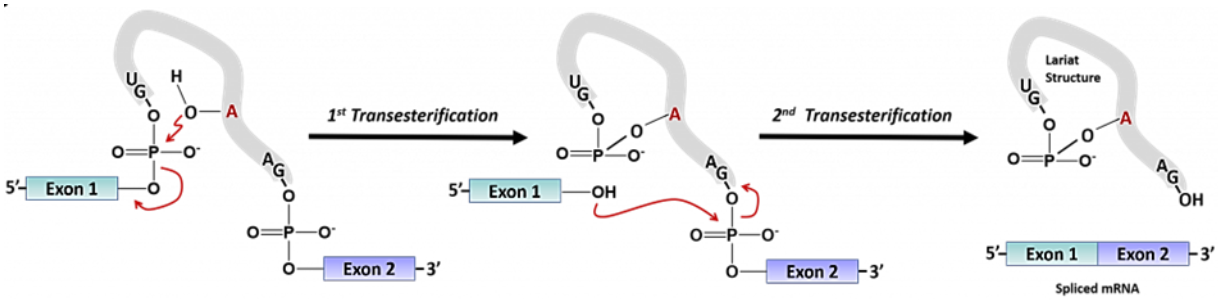

Figure 1.2 Chemistry of pre-mRNA splicing

Schematic diagram of the pre-mRNA with exons (in teal) and introns (in purple) depicting the two transesterification reactions required for intron removal. The branch point 2'-OH residue mediates attack on the 5'-phosphate of the intron guanosine residue located at the 5'-splice site. This releases the 3' hydroxyl of Exon 1, which subsequently mediates attack of the 5' phosphate of the first guanosine residue in Exon 2. The 3' hydroxyl of the intron guanosine residue is released forming the lariat structure (debranched and degraded) and Exon 1 is ligated to Exon 2 [adapted from Anon, 2021. RNA Processing].

[Available at: https://bio.libretexts.org/@go/page/15198 (Accessed September 2, 2021)].

Thus, by two transesterification reactions, exons are ligated and the intervening introns are spliced out. These two transesterification reactions are catalysed by a complex molecular machine termed the spliceosome (Brody & Abelson, 1985).

In 1993, Steitz & Steitz proposed a general two-metal-ion mechanism for catalytic RNA, in which the pentacovalent transition states of the two splicing transesterifications are stabilised by two metal ions in the active site of the spliceosome (Steitz & Steitz, 1993; see **Figure 1.3**). For the first phosphoryl transfer reaction, one of the two metal ions (M1) stabilises the leaving group, the 3' hydroxyl of the last 5' exon nucleotide, and the second metal ion (M2) activates the attacking nucleophile, the 2' hydroxyl group of the BP adenosine. During the first reaction (branching), the BP adenosine, must leave the active site to allow the binding of the 3' SS to the active site. In the second phosphoryl transfer reaction (exon ligation), M1 activates the 3' hydroxyl group of the 5' exon, and M2 stabilises the leaving group, the 3' hydroxyl group of the last intron nucleotide. Both M1 and M2 are coordinated by spliceosomal RNA (Fica *et al*., 2013); therefore the spliceosome is also considered a ribozyme (reviewed by Wilkinson *et al*., 2020).

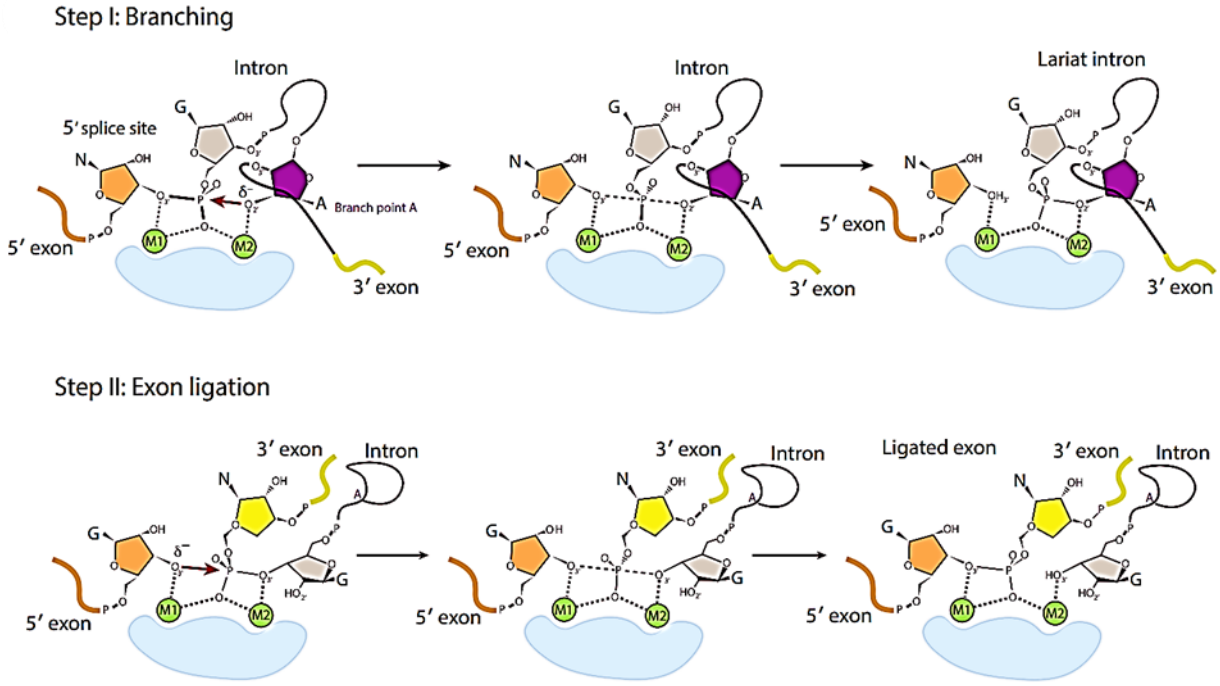

Figure 1.3 Two-metal-ion mechanism for splicing catalysis

Schematic representation of two-metal-ion mechanism, proposed by Steitz & Steitz (1993); 5'SS is depicted in orange and the 3' SS in yellow, intron is depicted with a black line. The two metal ions labelled as M1&M2 in green are located in the active sites of the spliceosome (in light blue). The BP adenosine is shown in purple and the two nucleophilic attacks are pointed out by red arrows (mechanism is explained in the text). From Wilkinson *et al*. (2020), modified.

1.1.2. Basic splicing factors and their assembly

The spliceosome is a highly dynamic macromolecular machinery that catalyses the removal of introns from eukaryotic pre-mRNA (Staley & Guthrie, 1998), and combines flexibility with accuracy. The spliceosome is composed of five small nuclear RNAs (snRNAs) namely – U1, U2, U4, U5 & U6, and approximately 100 proteins (Kastner *et al*., 2019). Each snRNA binds a specific set of seven common proteins, the Sm proteins, which assemble as a heptameric ring on the snRNA to form the small nuclear ribonucleoprotein (snRNP) particle (Lerner & Steitz, 1979). An exception is the U6 snRNA which is transcribed by RNAP III and associates with a preassembled ring of seven paralogous LSm proteins (LSm2-8) (Zhou *et al*., 2014). U1-U5 snRNAs are transcribed by the RNAP II. Several non-snRNP–associated proteins and protein complexes, including splicing factors and eight ATP-dependent helicases, are also involved in splicing. Initial mass spectrometric analysis of mixed population of affinity-purified spliceosomal complexes indicated that between 150 (Zhou *et al.,* 2002) and 300 different proteins (Rappsilber *et al.,* 2002) co-purified with the spliceosome. Within the spliceosome, the snRNAs perform the essential roles of substrate recognition and catalysis.

The spliceosome is not preassembled, but rather is formed anew on its substrate through sequential and highly coordinated interactions between the pre-mRNA, the snRNPs and numerous splicing factors. The sequence of events in an intron-defined spliceosome is schematically depicted in **Figure 1.4**. Briefly, The U1 snRNP is recruited to the 5’ splice site, where the 5’ end of the snRNA base-pairs with the 5’ splice site consensus sequence to form the early spliceosomal complex (E-complex). Alongside ATP-independent binding of U1 snRNP, SF1 (splicing factor 1) and U2AF65 – which recognise the branch site in the intron and the polypyrimidine tract, respectively, are also recruited.

U2AF65 together with U2AF35 constitute the U2AF (U2 auxiliary factor)-complex, which directly interacts with the 3’ splice site. In the second step, U2 snRNP is recruited to the branch point sequence displacing SF1, via base-pairing between the branch site and the U2 snRNA in an ATP-dependent manner. This complex containing pre-mRNA, U1 and U2 snRNP constitutes the pre-spliceosomal A-complex. U1 and U2 snRNP-associated proteins interact with each other, bringing the two splice sites in close proximity for recruitment of U4/U6.U5 tri-snRNP. The pre-spliceosome then associates with the preassembled U4/U6.U5 tri-snRNP to form the fully assembled spliceosome known as the pre-catalytic B complex. U6 snRNA, which ultimately folds to form the active site of the spliceosome, is extensively base-paired with U4 snRNA within the tri-snRNP (Plaschka *et al*., 2019). The complex is stabilised by direct interaction of U4/U6 snRNAs and the U5 snRNP, which is associated by protein-protein interactions.

At this stage, major rearrangements in RNA-RNA and RNA-protein interactions occur: The DEAD-box helicase Prp28 releases the 5’ SS from U1 snRNP and transfers it to the ACAGAGA box within U6 snRNA. The RNA helicase Brr2 then separates U4 snRNA from U6 snRNA and allows the U6 snRNA sequence adjacent to the 5’ SS-bound ACAGAGA box to fold and associate with part of U2 snRNA to yield the active site harbouring two catalytic metal ions. The U1 and U4 snRNPs leave the spliceosome, which is now termed the activated B complex (B^{act}). The B^{act} complex is then transformed into the catalytically active B* complex by Prp2

RNA helicase, and the first step of the splicing reaction takes place when the BP adenosine is docked into the active site – C* complex (compare to **Figure 1.3**).

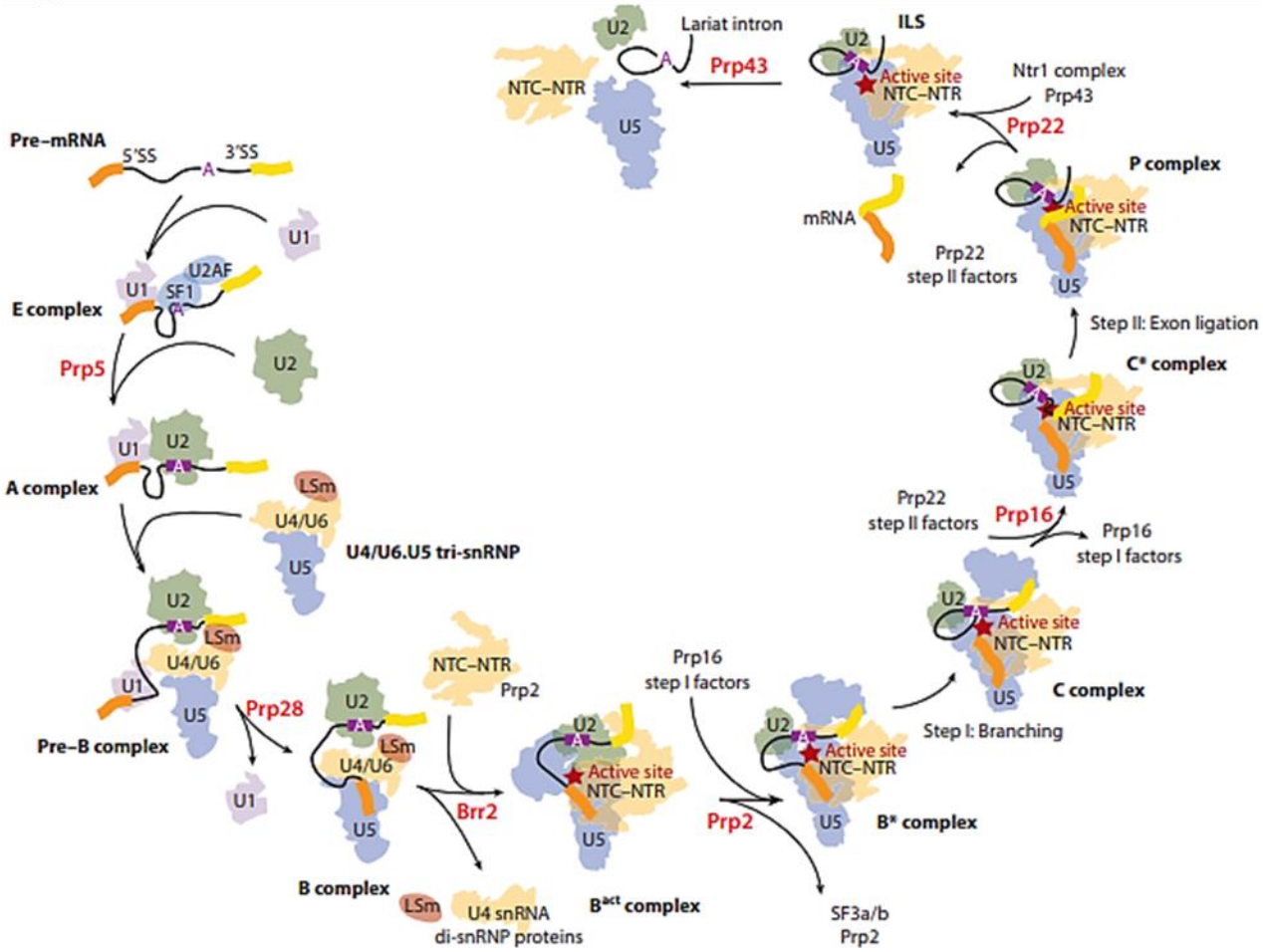

Figure 1.4 Assembly and disassembly cycle of the spliceosome

Canonical cross-intron assembly and disassembly pathway of the spliceosome is depicted. Exons and introns are indicated by boxes (orange and yellow) and black lines, respectively. Stages at which ATPases in the DEAD-box, DEAH box and Ski-2 families (red) play crucial roles is indicated. Abbreviations: 5'SS and 3'SS, splice sites; BP, branch point; U1-U6, snRNPs; ILS, intron-lariat spliceosome; mRNA, messenger RNA; NTC, Prp19-associated complex; NTR, Prp19-related complex; snRNP, small nuclear ribonucleoprotein. From Wilkinson *et al*. (2020), modified.

The two exons in the pre-mRNA are kept in close proximity throughout the B and C complexes through interactions with U5 snRNA and this facilitates joining of the two exons. The 5' exon remains in the active site, but the branch point adenosine must vacate the active site for the incoming 3' SS site for the second step in splicing: exon ligation reaction. Lastly, the 5' and 3' exons are joined, and the resulting mRNA (ligated exons) is released from the active site of the spliceosome. Following this, the spliceosome structure dissociates, snRNPs are recycled to participate in new rounds of splicing and the intron lariat is instantly removed by DBR1 debranching and degraded (Mohanta & Chakrabarti, 2020).

In case the intron length exceeds ~200 to 250 nt, the spliceosome cannot assemble across the intron (Fox-Walsh *et al*., 2005). Most mammalian introns are several hundred to several

thousand nucleotides in length (Deutsch & Long, 1999), therefore a mechanism of exon-definition has been suggested to be favoured in mammals (**Figure 1.5**).

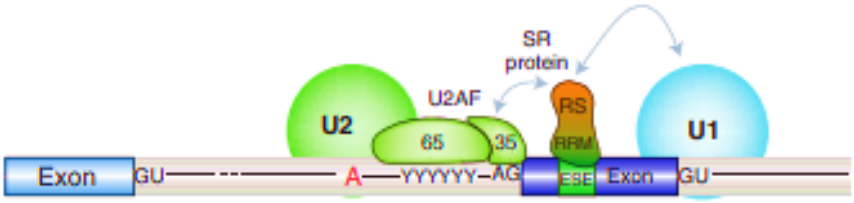

Figure 1.5 Model of interactions occurring during exon definition
Splicing enhancers like SR proteins bind to enhancer sequences of exon (ESE) and establish a network of protein-protein interactions across the exon thus stabilizing the exon-defined complex. From Will & Lührmann (2011).

During exon definition, the U1 snRNP binds to the 5' splice site downstream of an exon and promotes the association of U2AF to the 3' splice site upstream of it. This leads to the recruitment of the U2 snRNP. Splicing enhancer sequences within the exon (ESEs) recruit proteins of the SR protein family (see **section 1.2.2**), which establish a network of protein–protein interactions across the exon that stabilise the exon-defined complex (Reed, 2000). However, the subsequent switch from the exon-definition complex to the intron-defined spliceosome is poorly understood until now (reviewed by Will & Lührmann, 2011). In a recent article, Wan *et al*. (2021), proposed and validated a unified theoretical model to explain the general features of transcription and pervasive stochastic splice site selection, through nascent RNA profiling in single cells, transcriptome-wide footprinting of U2AF complex and lariat sequencing. This study presents a model of pervasive stochastic splice site selection rather than exclusive splicing at annotated splice sites.

1.2 Alternative splicing regulation mechanisms

According to the International Human Genome Sequencing Consortium, the human genome contains only about 20,000 protein-coding genes. This estimate is not large compared to the primitive nematode *C. elegans* (19,000 genes) and the common fruit fly *D. melanogaster* (~14,000 genes). However, the complexity of the proteome produced by this limited set of genetic information is quite large in each organism. The diversity of the human proteins is clearly more complex than that of invertebrates. Although the fruit fly has much fewer genes, it has a more complex proteome than the undoubtedly simpler organism *C. elegans*, and one reason for this is alternative splicing. Alternative splicing is a crucial mechanism, allowing individual genes to express multiple mRNAs that encode proteins with diverse and sometimes even antagonistic functions. This phenomenon was first observed when one gene could encode both a membrane-bound as well as a secreted antibody (Early *et al*., 1980). Genome-wide sequencing and bioinformatic analyses revealed that approximately 95% of multi-exon human genes are alternatively spliced (Pan *et al*., 2008; Wang *et al*., 2008). Alternative splicing can generate more transcripts from a single gene than the number of genes in an entire genome (Graveley, 2001). One exceptional example was the discovery of the *Drosophila Dscam* (Down syndrome cell adhesion molecule) gene, which codes for a cell surface protein

responsible for neuronal connectivity. The combinatorial alternative splicing of the exons can potentially generate up to 38,016 distinct mRNA isoforms, more than twice the number of genes in the entire *Drosophila* genome (Celotto & Graveley, 2001).

Constitutive splicing includes all the exons of a gene in the mature mRNA whereas in alternative splicing, alternative 5' or 3' splice sites may be used, resulting in exons of different lengths. Whole exons may be skipped or included, i.e. exon skipping or exon inclusion (cassette exons). Another modification to the cassette exon mechanism is mutually exclusive exons. Here, inclusion of one cassette exon dictates skipping of the adjacent exon, and *vice versa*. Finally, in some cases, introns are not spliced, they are rather retained in the mature mRNA (intron retention). **Figure 1.6** summarises the five basic patterns into which alternative splicing may be classified. Often, different modes of alternative splicing are employed in splicing of a single pre-mRNA species Furthermore, usage of alternative promoters and alternative poly(A) sites contribute to isoform diversity (Park *et al.,* 2018). Each mRNA isoform thus formed can have distinct properties in the cell, such as stability, localisation, and translational efficiency, and can be translated into stable protein isoforms with divergent structures and functions (Braunschweig *et al*., 2013). Additionally, around one-third of all alternative splicing events leads to the introduction of a premature termination codon (PTC) subjecting the mRNA to degradation by nonsense-mediated decay (NMD) and mRNA transcript elimination (McGlincy & Smith, 2008). NMD represents one RNA surveillance pathway to ensure the fidelity of gene expression.

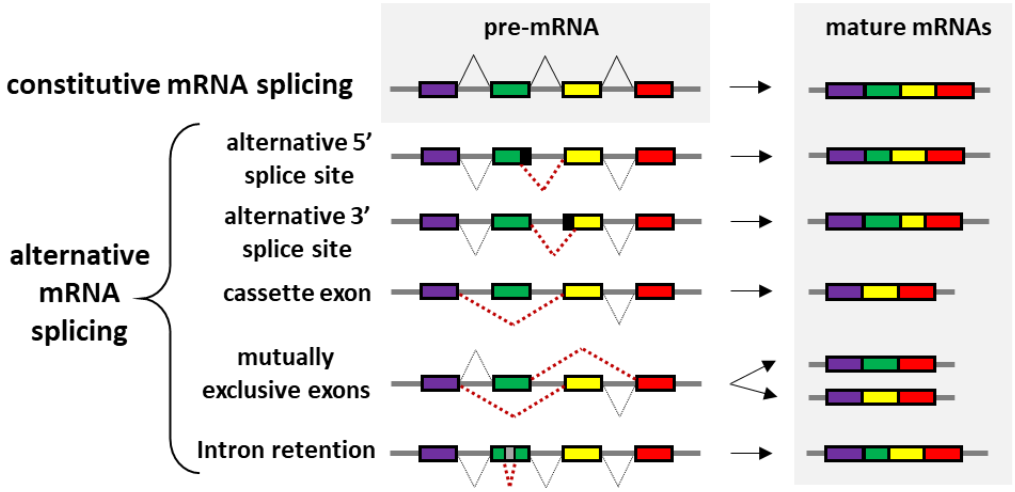

Figure 1.6 Modes of alternative splicing

Constitutive mRNA splicing is depicted schematically at the top. Exons are indicated by coloured boxes and introns with grey lines. Below, the five modes of alternative splicing are illustrated. Removal of introns by splicing are shown by triangles above and below the introns using black straight and dashed-lines. Alternative splicing event(s) are marked by red dashed-lines. The mature mRNA resulting from the respective splicing event is shown to the right. From Hui & Bindereif (2005), modified.

Recent advances in high-throughput sequencing (e.g., Nanopore sequencing) and the application of RNA-seq since 2008, have enhanced the detection of alternatively spliced mRNA isoforms. These technological advancements have overcome the limitations posed by

traditional RT-PCR (reverse transcription polymerase chain reaction) and ESTs (expressed sequence tags) for detection of mRNA isoforms and allowed for detailed analysis of the transcriptome.

1.2.1 Splicing enhancers and silencers

Alternative splicing is regulated in a spatial (cell-type and tissue type) and temporal (developmental-stage or differentiation) manner, orchestrated through external stimuli such as activation of signal transduction cascades. These external stimuli include extensive protein-RNA interactions between *cis* elements within the pre-mRNA and *trans*-acting factors that bind to these *cis* elements (Park *et al*., 2018).

Splice-site selection is a key mechanism in alternative splicing: Generally, splice sites are recognised based on their match to the consensus sequence (see **Figure 1.1**). While a 'strong' splice site shows a good match to the consensus sequence, a 'weak' splice site usually deviates from the consensus to a certain degree, and is less likely to be recognised by the splicing machinery. Spliceosome assembly is a common feature for both constitutive and alternative splicing, where numerous *trans*-acting factors interact with each other and with the *cis*-elements within the pre-mRNA to form active spliceosome. As described above (**section 1.1.1**), the most conserved *cis* splicing elements are 5' and 3' splice sites defining the boundary of the intron, the branch point (BP) adenosine and the polypyrimidine tract close to the 3' splice site. The spliceosome recognises these core components, which play an essential role in defining the identity of exon and intron (Wang & Burge, 2008). However, these classical splicing signals are not sufficient for recognition of all exons because cryptic splice sites which loosely match the consensus sequence, are very common in introns. In humans, canonical splicing signals provide less than 50% of the necessary information for accurate removal of introns (Lim & Burge, 2001). To correctly define the exon-intron boundary, additional splicing regulatory sequence elements (auxiliary signals) are required. Auxiliary splice signals play an important role in splice site recognition by enhancing accurate removal of introns. These auxiliary signals also known as splicing regulatory elements (SREs) are classified according to their location (exonic or intronic) and their functional effects on splicing (activation or repression) as exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS). Taken together, an important goal in current research is to arrive at the "splicing code" which comprises a set of regulation rules for splicing by studying SREs and their cognate factors (Barash *et* al., 2010). These elements are highly diverse in their sequence composition, and therefore bind a variety of splicing-regulatory proteins (**Figure 1.7**).

SREs are usually located near weak splice sites that they regulate and share degenerate sequence motifs, which makes identification of these SREs rather difficult. Therefore, several approaches were undertaken to identify splicing-regulatory sequences experimentally. ESEs are the best characterised splicing regulatory elements, which were identified by deletion and/or mutational analyses in minigene constructs transfected *in vivo* (for examples, see Preußner *et al*., 2012; Fairbrother *et al*., 2002). A randomised sequence stretch was inserted into an enhancer-dependent exon, and those elements that promoted inclusion of the exon

were amplified and analysed by RT-PCR (Coulter *et al*., 1997).

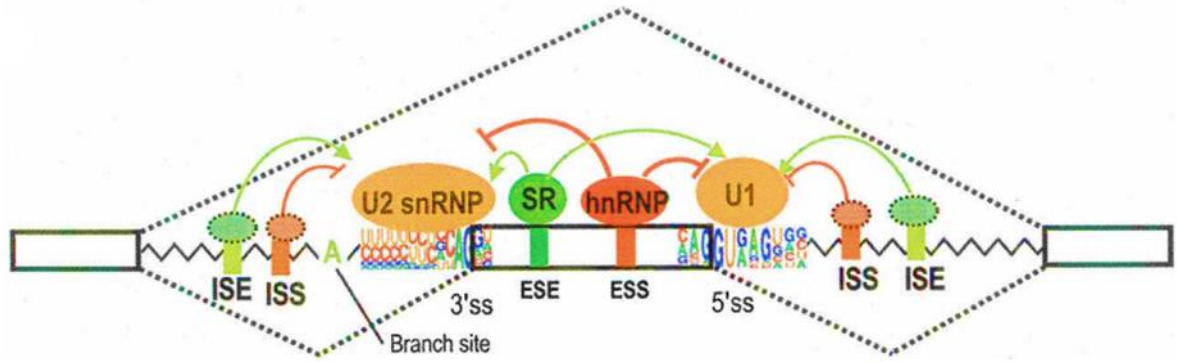

Figure 1.7 Regulated splicing

The schematic depicts splicing regulatory elements and *trans*-acting factors. Open boxes indicate exons, and jagged lines show introns. The consensus motifs of splice sites are shown in the pictogram. Two predicted alternative splicing pathways with the exon included or skipped, are marked with dashed lines. *Cis*-elements within the pre-mRNA include the canonical splicing elements 5'splice site (5'SS), branch point adenosine (A) in green, polypyrimidine tract [Y(n) – not shown], and 3' splice site (3'SS), as well as the auxiliary *cis* elements: exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs). 5'SS is recognised by the U1 snRNP, branch site – by the U2 snRNP complex, 3'SS and Y(n) are bound by the U2AF proteins. ISS and ESS elements recruit hnRNP proteins, which hinder binding of the spliceosomal components to the pre-mRNA. ISE and ESE elements are bound by SR proteins, which promote splicing. From Wang & Burge (2008), modified.

Sequences with enhancer activity were further identified from a pool of RNA sequences using functional *in vitro* and *in vivo* SELEX (systematic evolution of ligands by exponential enrichment) analysis (Woerfel & Bindereif, 2001). This method characterised two classes of ESEs, the purine-rich ESEs and the adenosine/cytosine-rich ESEs. These experimental and computational approaches provided a global picture that ESEs promote exon definition because of their enrichment and conservation in exons (Ke *et al*, 2018) whereas the ISSs are more enriched in introns and help define alternative splice sites by suppressing pseudoexons (Wang *et al*., 2006).

In comparison, only a few intronic splicing regulatory elements (ISREs) have been characterised (reviewed by Matlin *et al.,* 2005). Based on intronic sequence conservation, most predicted elements were enriched near alternatively spliced exons and resemble RNA motifs recognised by splicing factors such as Fox1, Nova and nPTB (Yeo *et al*., 2007). CA-rich and CA-repeat sequences were identified as one class of ISREs that could function both as intronic splicing enhancers and silencers depending on their relative location in pre-mRNA (Hui *et al*., 2005; Hung *et al*., 2008; Heiner *et al*., 2010). Ule and co-workers identified a tissue-specific ISRE known as the YCAY motifs (Y=C or U) using CLIP (crosslinking and immunoprecipitation) analysis. These motifs are recognised by the neuron-specific hnRNP-like proteins belonging to the neuro-oncological neural antigen (NOVA) family, regulating several splicing events in mouse brain (Ule *et al*., 2003). Another similar ISRE is the highly conserved UGCAUG motif recognised by the brain-specific factor Fox1 (Underwood *et al*.,

2005) and muscle-specific Fox-2 (Brudno *et al*., 2001). AU-rich motifs are strongly associated with constitutive splicing and may function as ISEs (Voelker & Berglund, 2007).

1.2.2 Trans-acting regulatory factors

Cis-regulatory elements mediate their function through recruitment of *trans*-acting factors to promote or suppress the use of adjacent splice sites (Matera & Wang, 2014). SR proteins and hnRNP family constitute the two major *trans*-acting protein classes that are conserved between plants and humans and have regulatory roles in alternative splicing, often functioning antagonistically (Raczynska *et al*., 2010; Reddy, 2007). In general, SR proteins play a key role in spliceosome assembly and regulate specific splicing events, mainly as activators bound to ESEs. In contrast, hnRNP proteins often bind to ESS and ISS elements and act mainly as splicing repressors (Wang & Brendel, 2004b). Both factors act on early stages of spliceosome assembly (on E and A complexes – see **Figure 1.4**) and cause splicing activation or repression. Typically, these splicing factors have a modular domain configuration, and contain one or many RNA binding domains (RBDs) to specifically recognise cognate *cis*-elements in pre-mRNA targets and functional domain(s) to affect splicing (Matera & Wang, 2014).

SR and SR-related proteins: Serine-arginine-rich (SR) proteins constitute a family of highly conserved non-snRNP splicing factors with diverse roles in constitutive and alternative splicing. They were the first splicing regulators described, which are characterised by a modular structure with a domain of variable-length arginine-serine-rich dipeptides (RS domain) at the C-terminus and one or two N-terminal RNA recognition motifs (RRMs). The RRM domain allows sequence-specific binding to RNA, whereas the RS domain is necessary for protein-protein interactions with other RS-domain containing proteins (Graveley, 2000) and include signals for cellular localisation (Long & Caceres, 2009). Other proteins, distinct from SR proteins, which contain an RS domain, but lack RRMs, are referred to as SR-related proteins. These proteins include the U2AF and U1 snRNP 70 kDa (U1 70K) protein (**Figure 1.8**).

One of the best characterised functions of SR proteins is recognition of ESE sequences in regulating alternative splicing. They can effect splice site choice in a concentration and phosphorylation-dependent manner (via serine residues in the RS domain), thus regulating tissue-specific and stress responsive alternative splicing in plants and animals (Duque, 2011). For example, through binding to an ESE close to a weak 5’ splice site during early spliceosome assembly, SR protein SRSF1 stimulates splicing by recruiting U1 snRNP in a phosphorylation-dependent manner mediated by interaction with U1 70K. (Cho *et al*., 2011). Activation of the 3’ splice site on the other hand is mediated by recruitment of U2AF65 to a weak pyrimidine tract (Zuo & Maniatis, 1996). Incorporation of the tri-snRNP complex (U4/U6.U5 snRNP) into the spliceosome is also mediated by SR proteins (Roscigno & Garcia-Blanco, 1995).

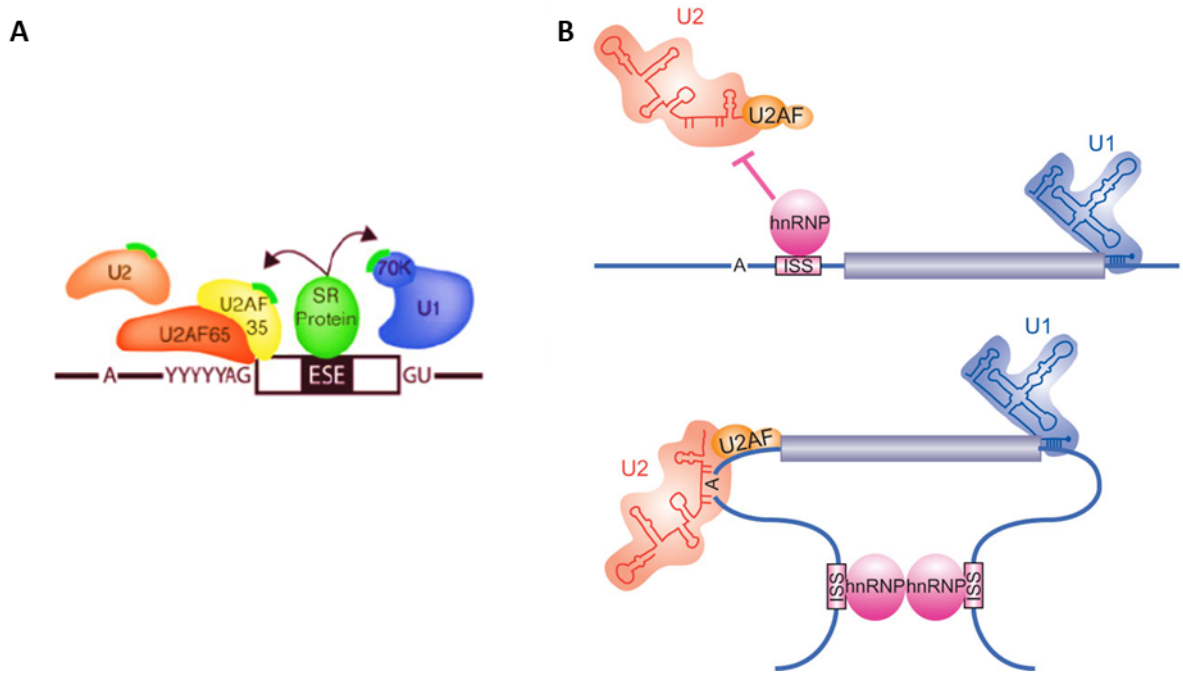

Figure 1.8 Splicing regulation through *trans*-acting factors

(A) Regulation of splicing by SR proteins: Exonic splicing enhancers (ESE) within the exon (open box) is the binding site for SR proteins (in green) which stimulate splicing by recruitment and interaction (indicated by arrows) with U1 snRNP (in blue) to the downstream 5' splice site and/or U2AF (65 and 35 kDa subunits – in orange and yellow respectively) to the upstream polypyrimidine tract and the 3' splice site on the intron (indicated by black line). U2AF thereafter, recruits U2 snRNP to the BP adenosine. SR proteins assist in the recruitment of spliceosomal complexes to the pre-mRNA through RRM interaction with the RNA and RS domain-mediated (indicated by bright green arcs) interaction with other proteins. From Mueller & Hertel (2012), modified.

(B) Splicing regulation by hnRNP proteins: Intronic splicing silencers (ISS) in intronic regions (blue line) close to the exons (in purple) act as binding sites for hnRNP proteins (in pink), which interfere with binding of U2AF (in orange) to the 3' splice site thereby causing splicing repression (top panel). Alternatively, hnRNP proteins can loop-out an intervening exon by binding to ISS sequences flanking the exon (bottom panel). From Graveley (2009).

ESE-bound SR proteins can also activate splicing by antagonizing repressive factors. For example, hnRNP A/B proteins and SRSF1 function antagonistically in alternative splice site selection. While elevated levels of hnRNP A1 favour distal 5' splice sites, increase in SRSFI promotes the use of proximal 5' splice site. This points to the importance of relative ratios of opposing splicing factors. Another antagonistic function is established between two SR proteins SRSF1 and SC35 in the regulation of β- tropomyosin mRNA (Wu & Maniatis, 1993).

Heterologous nuclear ribonucleoproteins (hnRNPs): Primary transcripts of protein-coding genes synthesised by RNAP II in the nuclei of eukaryotic cells undergo extensive posttranscriptional processing (Darnell, 1982). These transcripts were termed heterogeneous nuclear RNAs (hnRNAs) to describe their size heterogeneity and cellular location (Dreyfuss *et* al., 1993). Although used interchangeably, the terms pre-mRNA and hnRNA do not refer to the same RNA species, since only a subset of hnRNAs may actually be precursors to mRNAs,

and the rest turn over in the nucleus. Nascent hnRNAs are immediately bound by a family of proteins, termed hnRNPs, forming hnRNP complexes. These hnRNPs are associated with the hnRNAs from the time that they emerge from the transcription complex throughout the time they reside in the nucleus, and such hnRNPs are not stable components of other ribonucleoprotein (RNP) complexes like snRNPs.

The hnRNP proteins are among the most abundant nuclear proteins in higher eukaryotes, regulating several RNA-related biological processes such as – transcriptional regulation, pre-mRNA 3' end processing, splicing, mRNA stability and export to cytoplasm, telomere-length maintenance, mRNA translation and turnover (Kim *et* al., 2000). HnRNP proteins were first identified by immunopurifications with hnRNP A1- and hnRNP C-specific antibodies in combination with two-dimensional gel electrophoresis. A total of 24 hnRNP proteins (labelled hnRNP A through U) in the molecular mass range of 34 to 120 kDa, were distinguished with a diffuse nuclear localisation, in contrast to SR proteins, which concentrate in nuclear speckles (Dreyfuss *et* al., 2002). All hnRNP proteins share a common modular structure of multiple domains, connected by flexible linker regions. A common feature of hnRNP proteins are RRM domains, often present in tandem and other auxiliary domains which are composed of clusters of certain amino acids. These include the RGG boxes (Arg-Gly-Gly tripeptides), or acidic glycine- or proline-rich regions. Such domains can help to mediate protein-protein interaction or facilitate protein localisation (Dreyfuss *et* al., 1993). Post-translational modifications of hnRNP are also common - the arginine residues within the RGG box have the potential to be methylated, thereby regulating RNA-binding activity. In addition to arginine methylation, serine/threonine phosphorylation are common post-translational modifications of hnRNP proteins (Mayrand *et al*., 1993).

Initially, it was thought that hnRNP proteins associate non-specifically with hnRNA. However, SELEX and CLIP assays revealed the binding specificities of most hnRNP proteins, confirming previous observations of the preferential association of hnRNP proteins with homopolymeric RNAs (Swanson & Dreyfuss, 1988). For example, hnRNP L recognises and binds to CA-rich and CA-repeat sequences (Hui *et al*., 2005). All members of the hnRNP H protein family recognise GGGA sequence (Caputi & Zahler, 2001), and UCUU motifs are recognised by hnRNP I/PTB.

Apart from mediating important cellular functions such as telomere maintenance, chromatin remodelling and DNA repair; hnRNP proteins mainly function as repressors of splicing by binding to the pre-mRNA and prevent other splicing factors from accessing the binding site (Wang & Brendel, 2004b). Nonetheless, the type of action of the hnRNP proteins depends on the context where it is bound (Wachter *et al*., 2012). For example, hnRNP proteins stimulate splicing, predominantly from ISEs, by promoting interaction of U1 snRNP with weak 5' splice sites, as observed for the TIA proteins. Another striking example, is the dual functioning of hnRNP-like proteins in mice. The Nova family proteins in mice, recognise the YCAY motif that can act either as an ESS, if it is in an exon preceding an alternatively spliced exon, or as an ISS, if it is in an intron following an alternatively spliced exon (Ule *et al*., 2006).

A complete review of hnRNP proteins and their functions, was compiled by Dreyfuss *et al*., (2002). Briefly, hnRNP A/B protein family is involved in many aspects of RNA processing, especially in splicing. HnRNP A1 is required for efficient miRNA processing, and hnRNP A/B

proteins play a role in telomere biogenesis. HnRNP C is implicated in splicing and in enhancing translation mediated by internal ribosomal entry sites (IRES). HnRNP E/K family members contain KH (K Homology) domains instead of RRMs, differentiating them from the other hnRNPs. HnRNP K is necessary for splicing, transcription, mRNA stability and translation. HnRNP R1/R2 was observed to be necessary for retinal development in rats (Peng *et al*., 2009).

HnRNP A1 and I (PTB, polypyrimidine tract binding proteins) are probably the most intensively studied hnRNP proteins, with homologues between plants and mammals (Kaminski *et al*., 1995). PTB affects splicing of α-tropomyosin, by binding to pyrimidine tracts and inhibits exon inclusion, while U2AF antagonises this. SR proteins on the other hand, activate splicing, demonstrating how hnRNP and SR proteins work together to produce alternative splicing. PTB has been found to auto-regulate its own mRNA by alternative splicing (Wollerton *et al*., 2004) like hnRNP L (Rossbach *et al*., 2009, discussed in **section 1.4**). PTB also mediates long-range interactions between distant RNA regions flanking alternative exons by dimerization of the protein surrounding the exon, thus looping-out the intervening region of the pre-mRNA and preventing splicing of the excluded RNA region [Preußner *et al*., 2012; (**Figure 1.8B**)].

Recent advances in functional understanding of hnRNPs propose a role of hnRNPs to act as "RNA scaffolds" and recruit mRNA, long non-coding RNA (lncRNA), microRNA (miRNA) and circular RNA (circRNA – discussed in section 1.6) to affect mRNA splicing and processing. RNA scaffolds also regulate transcription from genes, post-transcriptional translation and change genome structure. Furthermore, hnRNPs are essential for self-renewal and differentiation of stem cells and play crucial roles in stem-cell functions (reviewed in Xie *et al*., 2021).

In addition to SR proteins and hnRNP proteins, several non-SR and non-hnRNP proteins, like the Nova-1 protein in mice, bind to specific intronic sequences and promote the inclusion of neuron-specific introns (Jensen *et al*., 2000), playing a key role in regulating splicing. Apoptosis promoting protein TIA-1 is also found to regulate human Fas receptor pre-mRNA splicing (Foerch & Valcarcel, 2001). These examples show how several non-canonical splicing factors may also function as splicing regulators on specific pre-mRNAs.

1.3 RNA binding domains

Fundamental biological processes of the cell such as transcription, mRNA processing, translation and several other developmental stages rely on interactions between nucleic acids and proteins. A variety of *in vitro* and *in vivo* experimental strategies including computational methods, have been developed in the last decade to study these interactions. Consequently, large amounts of experimental data have fuelled structural studies such as NMR (nuclear magnetic resonance), X-ray crystallography and cryo-EM analyses (Schlundt *et al*., 2017). Additional computational methods, such as neural networks (Alipanahi *et al*., 2015), were developed to predict protein binding sites in nucleic acid sequences (Tuvshinjargal *et al*., 2016).

'Conventional' RNA-binding proteins (RBPs) recognise specific sequences and/or structural motifs in RNA to form ribonucleoprotein (RNP) complexes that principally regulate gene expression (**Figure 1.9A**). The RBP binds to target genes on specific nucleotide sequences in the open reading frame (ORF), intron region or untranslated region (UTR) of the transcripts via modular combinations of structurally well-defined RNA-binding domains [RBDs, (Hentze *et al*., 2018)]. Several RBDs have been characterised and classified principally into: (a) RNA-recognition motifs [RRM, (Clery *et al*., 2008)], (b) hnRNP K homology (KH) domain, (c) DEAD/DEAH box helicase, (d) Zinc finger and (e) RGG (Ala-Gly-Gly) box. Briefly, hnRNP K homology (KH) domain is an RNA-binding domain of ~70 amino acids that folds into three α-helices packed against a three-stranded β-sheet; the two-core α-helices form a hydrophobic cleft with a GXXG loop that interconnects the helices, where RNA binds (Valverde *et al*., 2008). DEAD/DEAH box proteins are RNA helicases with two highly similar domains containing the conserved sequence Asp-Glu-Ala-Asp (DEAD); RNA binds across both helicase domains (Hentze *et al*., 2018). Zinc finger domain – a protein structural domain that is coordinated by one or more Zinc ions to stabilise the fold (Klug, 2010). RGG (Ala-Gly-Gly) box is a high-affinity RNA-binding domain, serving also as a substrate recognition site for protein arginine methyl transferases [PRMTs (Boisvert *et al*., 2005c; Lischwe *et al*., 1985)]. Other RBDs such as cold-shock domain (Y-box proteins), Pumilio/FBF (PUF) domain, double stranded RNA-binding domain (dsRBD), Piwi/Argonaute/Zwille (PAZ) domain, and Sm domain, also bind to RNA with sequence-specific interactions.

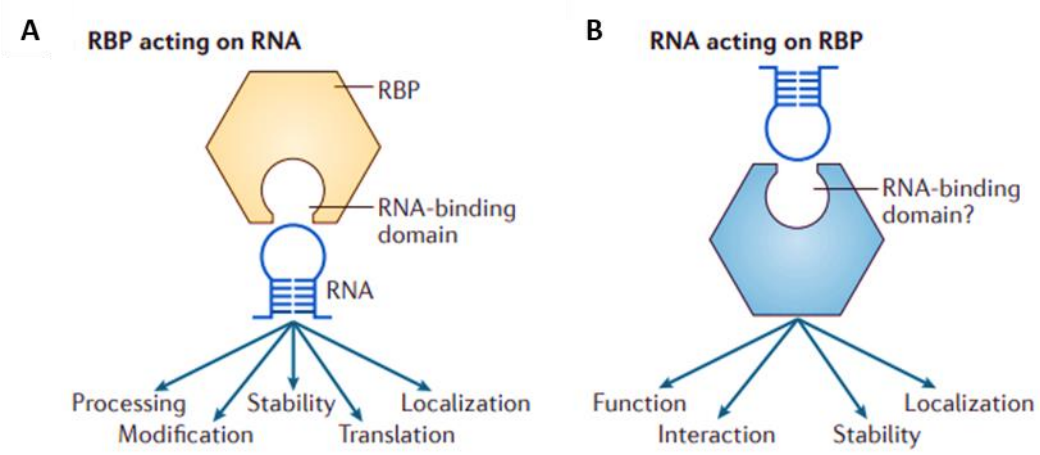

Figure 1.9 Functional interactions between protein and RNA

Schematic depiction of protein-RNA interactions. RNA-binding protein (RBP - in yellow) binds to RNA through the RNA-binding domain (RBD) and regulates its function and metabolism (A). Inversely, the RNA can also bind to the RBP (in blue) regulating its function and stability (B). From Hentze *et al.* (2018), modified.

One of the most abundant protein domains in eukaryotes is the RRM domain – around 14432 RRM domains have been identified in 8767 proteins in the SMART nrdb database, updated last year (Letunic *et al*., 2021). In eukaryotic proteins, RRMs are often present in multiple copies within a protein (44%, two to six RRMs). A classic example of this is the multidomain hnRNP L protein with 4 RRM domains (discussed in section 1.4). RRMs are also present together with other RBDs (21% RRM-containing proteins) like the KH domains in

IMP3/IGF2BP3 protein. Proteins containing RRMs regulate cellular processes such as mRNA and rRNA processing, RNA stability, export and translation (Dreyfuss *et al*., 2002). Typically, an RRM is approximately 90 amino acids in length with a characteristic βαββαβ topology that forms a four-stranded β-sheet packed against two α-helices. **Figure 1.10** depicts the secondary structures of the four RRM-domains of hnRNP L. The most conserved RRM signature sequence is an eight-residue motif called RNP1 (in β3-sheet), which has the consensus [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY]. A second six-residue region of homology, called RNP2 (in β1-sheet), is typically located 30 residues N-terminal to RNP1, and has the consensus [ILV]-[FY]-[ILV]-X-N-L. Additional conserved amino acids define an 80-residue domain that encompasses the RNA-binding function (Scherly et al., 1989). Both β-sheet surface as well as the loops connecting the β-strand and α-helices (like the loop connecting RRM3-4 in hnRNP L, **see Figure 1.10**) are crucial for RNA recognition.

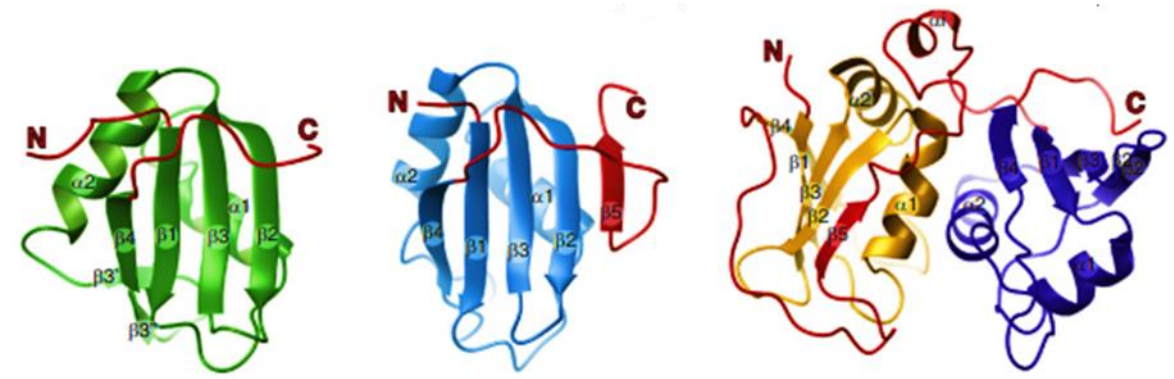

Figure 1.10 Ribbon diagrams of the four RRM domains of hnRNP L
Ribbon-diagrams of protein structure of RRM1 (green), RRM2 (light-blue), RRM3 (yellow) and RRM4 (blue) of hnRNP L is represented. Arrows and tubes illustrate β-sheets and α-helices in secondary structure respectively. Conserved and structured extensions are shown in red – at the N- and C-terminus of all four RRMs and the inter-domain linker of RRM3-4. RRM2 and RRM3 possess an additional fifth β-strand on the side of the RNA binding surface. From Blatter *et al.* (2015), modified.

Although several RBDs as above are now defined, structural analysis of complex RNP machines such as the ribosome (Behrmann *et al*., 2015; Amunts *et* al., 2014) and the spliceosome (Plaschka *et al*., 2017; Matera & Wang, 2014) point to a more complex protein-RNA interaction network that may not involve any canonical RBDs. While it is assumed that RBPs that bind to their targets with higher affinity are more likely to have ascertainable biological functions (Hentze *et al*., 2018), growing evidence for RBP modulation by RNA has also been documented (**Figure 1.9B**). The recent characterisations of multiple microscopically visible, membraneless RNP granules such as Cajal bodies and paraspeckles in the nucleus as well as stress granules and processing (P-) bodies in the cytoplasm (Buchan, 2014; Anderson & Kedersha, 2009), indicate the limited applicability of RBPs regulating RNA function. This is due to the dynamic composition and amorphous structure of these granules with functions not yet well-defined. These RNP bodies are composed of RBPs containing intrinsically disordered regions (IDRs) that drive their formation by liquid-liquid phase separation [LLPS (Alberti *et al*., 2019). In addition, a myriad of long non-coding RNAs (lncRNA) have now been

discovered and implicated in recruitment of transcription factors and/or chromatin-modifying complexes to chromatin, in organization, scaffolding or inhibition of protein- complexes (Cech & Steitz, 2014). These observations break with convention by indicating that RNAs may regulate RBP function (reviewed by Hentze *et al*., 2018). Taken together, the RBPs, ncRNAs and mRNAs form a dynamic environment termed the ribonome, which acts as one of the layers in transcriptome regulation (Morris *et al*., 2010).

1.4 hnRNP L – global alternative splicing regulator

Heterogeneous nuclear ribonucleoprotein L (hnRNP L) is a member of the hnRNP family of proteins, and was identified along with the set of 24 proteins labelled from hnRNP A to U, based on 2D-gel electrophoresis (Piňol-Roma *et al*., 1989). These protein complexes were further purified by single-stranded DNA (ssDNA) agarose chromatography. HnRNP L was characterised by raising a monoclonal antibody termed '4D11' (also used in this thesis) against hnRNP L (Piňol-Roma *et al*., 1989). Several isoelectric forms of hnRNP L were detected by 2D-gel electrophoresis, suggesting posttranslational modifications of hnRNP L. The hnRNP L open reading frame was revealed by screening cDNA-clone libraries for proteins that react with hnRNP L-specific antibodies, which helped in the prediction of its amino acid sequence. It was found that hnRNP L-protein consists of 558aa, with a calculated molecular mass of 60.4kDa. Sequence analysis revealed N-terminal glycine-rich region and four conserved RRM domains (Hahm *et al.,* 1998a). In addition, the linker sequence between RRM2 and RRM3 is rich in proline (see **Figure 1.11A**). HnRNP L shows 68% sequence identity with its closely related paralog, hnRNP L –like (hnRNP LL) and is also similar to three paralogous hnRNP proteins in its domain composition: polypyrimidine tract-binding protein (PTB), neural polypyrimidine tract-binding protein (nPTB), and polypyrimidine tract-binding protein 3 (Rod1) (Blatter *et al*., 2015).

The first functional implication of hnRNP L was observed in the processing of the Herpes simplex virus thymidine kinase gene (HSV-TK). HnRNP L was shown to bind its pre-mRNA in a sequence-specific manner (to a 49-nt core region characterised by CA-rich stretches), and nuclear export of the intronless chimeric construct was dependent on hnRNP L binding (Liu & Mertz, 1995). Another example in viral RNA processing by hnRNP L was reported by Sikora *et al.* (2009). HnRNP L was found to be associated with the Hepatitis delta virus (HDV) RNA. Hahm *et al* (1998b) also identified a cytoplasmic function of hnRNP L in activating IRES (internal ribosome entry site)-mediated translation by binding to the 3' end region of the hepatitis C virus (HCV). Activation of translation by hnRNP L is mediated by recruitment of canonical translational initiation factors, a mechanism evidenced by *in vitro* binding assays with competitor RNA (Hwang *et al*., 2009).

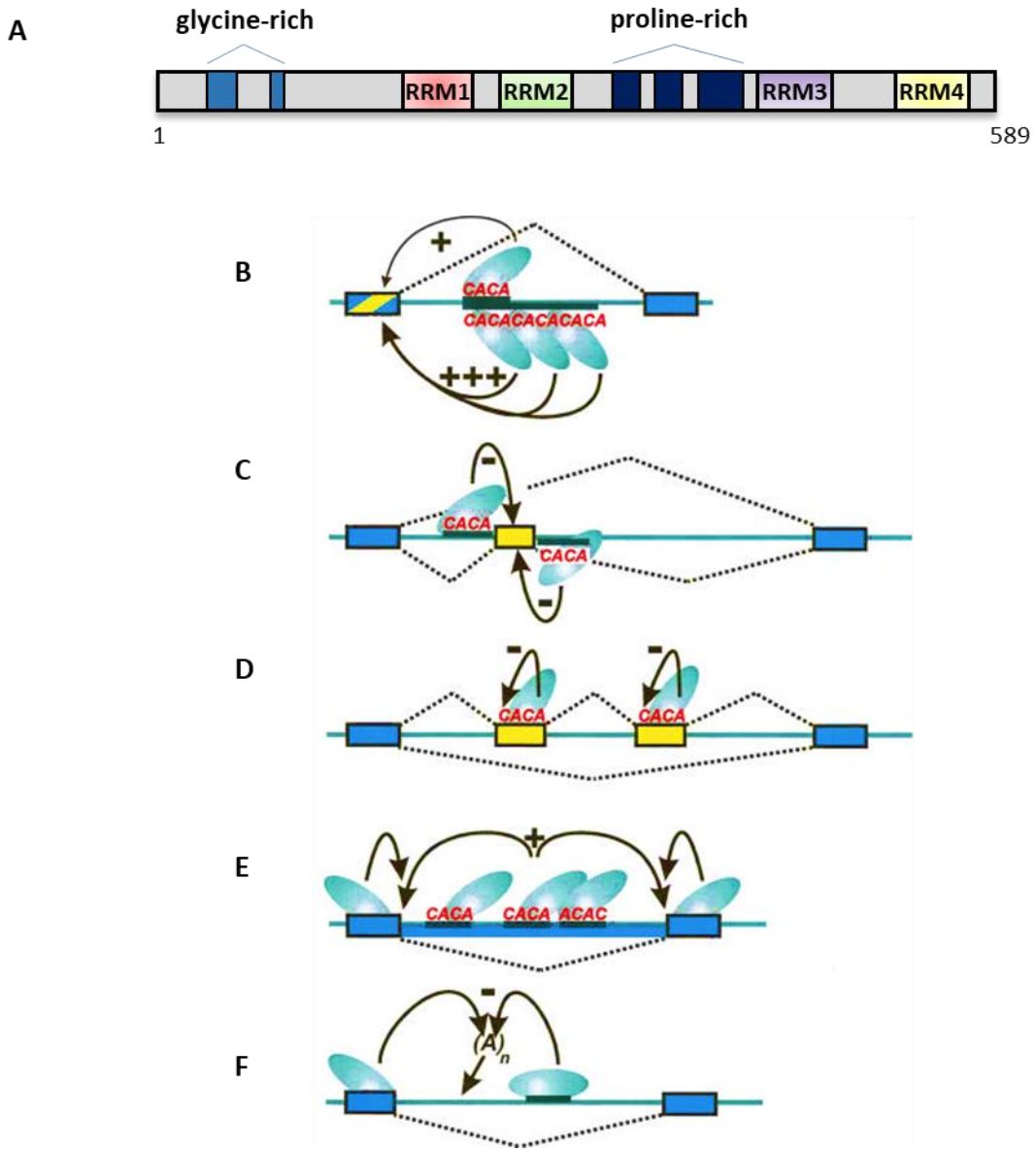

Figure 1.11 Domain organization of hnRNP L protein and summary of activities in regulating alternative splicing

(A) The domains of hnRNP L (NP_001524.2; 589 amino acids) are schematically represented. The four RRM domains are indicated – RRM1 (red), RRM2 (green), RRM3 (purple) and RRM4 (yellow); glycine-rich and proline-rich regions in light blue and blue, respectively. **(B-F)** Regulatory activities of hnRNP L are schematically represented, with hnRNP L functioning either as an activator (+) or repressor (-), and using intronic or exonic CA-rich elements (constitutive exons as blue boxes, regulated exons in yellow). **(B)** Splicing activation by binding to (length-dependent) intronic splicing enhancers **(C)** Repression of cassette-type exons by binding to either upstream or downstream intronic splicing silencers **(D)** Repression of (multiple) alternative exons or regulation of variable exons by binding to exonic splicing silencers **(E)** Suppression of intron retention by binding to intronic splicing enhancers **(F)** Repression of internal polyadenylation by binding to an exonic/intronic regulatory element. From Hung *et al.* (2008), modified.

IRES-mediated translation was also observed in some human endogenous genes when cap-mediated translation is compromised. For example, *SLC7A1* gene encodes the cationic amino acid transporter 1 (Cat-1). During amino acid starvation, hnRNP L promotes efficient translation of Cat-1 by IRES-mediated translation along with PTB (also known as hnRNP I) (Majumder *et al*., 2009; Komar & Hatzoglou, 2005).

Initial observations by immunofluorescence microscopy showed hnRNP L to be evenly distributed throughout the nucleoplasm (excluding the nucleoli) in human and mouse cells. However, hnRNP L shuttles from nucleus and cytoplasm and plays crucial roles in both (Kim *et al*., 2000). In the nucleus, hnRNP L mainly regulates the synthesis, transport and processing of mRNA. HnRNP L also functions in many biological functions such as DNA repair, alternative splicing, transcription factor activity, translation, signal transduction and gene expression (Hung *et al*., 2008). Numerous cross-linking and purification assays (Hui *et al*., 2003b; Shih & Claffey, 1999) and *in vitro* SELEX experiments revealed that full-length hnRNP L binds specifically to CA-rich sequences on the RNA (Hui *et al*., 2005). Thus, hnRNP L functions in RNA stability (Hui *et al*., 2003a; Hamilton *et al*., 1999; Shih & Claffey, 1999), as positive or negative regulator in IRES-mediated translation (Peddigari *et al*., 2013), miRNA (microRNA) silencing and riboswitch activity in the *VEGFA* (vascular endothelial growth factor A) mRNA (Jafarifar *et al*., 2011; Ray *et al*., 2009) and polyadenylation of mRNA (Hung *et al*., 2008; Guang *et al*., 2005). Furthermore, alternative splicing of *CD45* is regulated in a position-dependent manner by the action of *trans*-acting hnRNP L and *cis*-acting CA-rich elements (Chiou *et al*., 2013; Preußner *et al*., 2012). Several other genes such as mouse *integrin alpha2beta1* and carcinoembryonic antigen-related cell adhesion molecule-1 (*CEACAM1*) are regulated in a similar manner by hnRNP L (Dery *et al*., 2018; Heiner *et al*., 2010; Cheli & Kunicki, 2006; Hui *et al*., 2005). These examples including autoregulation of hnRNP L by its own mRNA through nonsense-mediated decay (NMD) (Rossbach *et al*., 2009) indicate the global regulatory function of hnRNP L (Blatter *et al*., 2015).

Highly versatile, hnRNP L also exhibits activities in the cytoplasm, such as nucleo-cytoplasmic transport (Liu & Mertz, 1995; Guang *et al*., 2005) and activation of *VEGFA* translation via differential mechanisms which is associated with tumorigenesis (Peddigari *et al*., 2013; Ray *et al*., 2009). Apart from these aspects of RNA metabolism, hnRNP L also plays a role in epigenetic methylation of histones. HnRNP L is a component of the Set2 complex in humans, and its knockdown by RNA interference (RNAi) down-regulates the trimethylation mark on lysine 36 of histone H3 exclusively (Yuan *et al*., 2009). Furthermore, hnRNP L plays diverse roles in the progression of several cancers (reviewed by Gu *et al*., 2020). The role of hnRNP L in several tumours and the possible mechanism of action is enlisted in *Supplementary information*, **Table 1.1**. A recent study also reports that hnRNP L is essential for myogenic differentiation and modulation of myotonic dystrophy pathologies (Alexander *et al*., 2021), which reinforces the versatility of hnRNP L's function in eukaryotic cells.

The most prominent role of the global regulator hnRNP L, however, is splicing regulation. Beginning with the first observations (by mass spectrometry analyses) of hnRNP L assembling on two different *in vitro* splicing substrates (Zhou *et al*., 2002) and H (heterogeneous) complex

formation at initial stages of spliceosome assembly (Black *et al*., 2003), hnRNP L was found to regulate global alternative splicing. Hui *et al*. (2003b) presented the first experimental evidence for hnRNP L in splicing of human endothelial nitric oxide synthase gene (*NOS3* or *eNOS*). HnRNP L activated splicing of an *NOS3* minigene construct by binding to CA-repeat sequences, in an *in vitro* splicing and complementation assay (Hui *et al*., 2003a&b). Interestingly, the effects of splicing by hnRNP L were dependent on the position of the CA-repeat regions in the *NOS3* minigene (Hui *et al*., 2005). HnRNP L's function in splicing in a genomewide context was addressed by a SELEX approach, where, in addition to CA-repeat RNA-elements, certain CA-rich sequences were identified as high-affinity hnRNP L binding targets; this sequence-specific binding of hnRNP L was further validated *in vitro* (Hui *et al*., 2005). Building upon these results, global splicing alterations caused by RNAi-mediated knockdown of hnRNP L were analysed via an exon-specific microarray platform. This study identified eleven hnRNP L-regulated splicing events that were experimentally validated (Hung *et al*., 2008). First, hnRNP L prevented intron retention in *CD55* and *STRA6*; second, exon repression was validated in *TJP1*, *BPTF* and *PARK7*; third, activation of exon inclusion was demonstrated in *MYL6*, *FAM48A* and *PAPOLA*; and fourth, suppression of multiple exons in long introns was verified in *SORBS2* and *LIFR*. Additionally, a novel function of hnRNP L was described: regulation of alternative polyadenylation site selection in *ASAH1*. A mechanistic view of exon repression mediated by hnRNP L was presented by Heiner *et al*. (2010): hnRNP L binds to an ISS near the 5' splice site of the *SLC2A2* pre-mRNA, preventing splice site recognition by U1 snRNP. A summary of hnRNP L- mediated splicing regulatory mechanisms is depicted in **Figure 1.11B-F**.

1.5 RBM24 – muscle specific alternative splicing regulator

Members of the ubiquitously expressed SR and hnRNP protein family regulate splicing by either facilitating or inhibiting splice site recognition. Since splice site recognition is crucial in determining the outcome of splicing, several mechanisms exist to control this process. One common mechanism is the activity of tissue-specific splicing factors, which allows more specific control of splice site selection. Although only a few tissue-specific splicing factors have been identified so far (Nilsen & Graveley, 2010), the first tissue where alternative splicing was observed in a tissue-specific manner was the muscle, where more than 1000 muscle-specific splicing events have been mapped (Llorian & Smith, 2011; Castle *et al*., 2008). Yang and co-workers characterised RBM24 (RNA-binding motif protein 24) as one such splicing factor, that plays an important role in regulating muscle-specific alternative splicing (Yang *et al*., 2014). RBM24 is a highly conserved RBP containing a single RRM domain at its N-terminus and two conserved domains at the C-terminus (Sun *et al*., 2016;). During vertebrate development, RBM24 exhibits strongly restricted tissue-specific expression in myoblasts and is required for myogenic differentiation (Grifone *et al*., 2014). Deficiency in the expression level of the RBM24 manifests in congenital disorders such as cardiomyopathy, myopathy or blindness in different animal models, although no mutations in the *RBM24* human gene have been linked to any disease so far. RBM24 is also crucial for embryonic cardiac development and regulates at least 68 alternative splicing events. RBM24 is sufficient to activate muscle-specific splicing

in non-muscle cells as well as in HeLa nuclear extracts *in vitro*. (Yang *et al*., 2014). RBM24 is also involved in cytoplasmic polyadenylation (CPA) during lens fiber cell terminal differentiation to ensure accumulation of crystalline proteins (Shao *et al*., 2020). It therefore appears that RBM24 is involved in most aspects of post-transcriptional regulation, especially as a key factor in regulating alternative splicing to establish the contractile function in developing cardiac and skeletal muscles (Lin *et al*., 2018; Weeland *et al*., 2015; Yang *et al*., 2014).

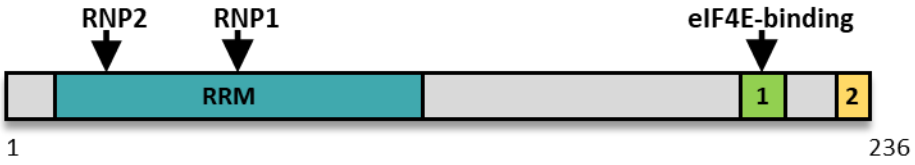

Figure 1.12 Domain structure of the protein RBM24

Schematic representation of human RBM24 protein domains (NP_001137414.1; 236 amino acids). The amino-terminal half contains a canonical RRM (in teal), with two consensus RNP (ribonucleoprotein) sequences – RNP1 and RNP2. The C-terminal regions contains two conserved domains -1 (in green, for eIF4E binding) and 2 (in yellow). From Grifone *et al.* (2020), modified.

RBM24 mediates muscle-specific alternative splicing and regulates cell-cycle progression and apoptosis, by binding to GU-rich sequences in target mRNAs in a way similar to its closely related paralog -- Rbm38 (RNPC1) (Qian *et al*., 2020; Ray *et al*., 2009). The RRM domain at the N-terminus in RBM24 and Rbm38 is almost identical; especially RNP1 and RNP2 are highly conserved from humans to nematodes (Afroz *et al*., 2015). Although sequences outside the RRM domain may be relatively divergent, the two domains at the C-terminus share sequence identity (Boy *et al*., 2004). One motif close to the extreme C-terminus (domain 1, see **Figure 1.12**), contains a serine residue (serine 181 in RBM24 and serine 195 in Rbm38) that can interact with eukaryotic initiation factor 4E (eIF4E) and prevent it from binding to 5' cap of mRNAs (Lucchesi *et al*., 2018; Zhang *et al*., 2011). In several cancer cell lines, glycogen synthase kinase 3 (GSK3) phosphorylates the serine residues in RBM24 and Rbm38, preventing their interaction with eIF4E, and this converts RBM24 into an activator of mRNA translation (Zhang *et al*., 2013). The exact function of the other conserved domain (domain 2) at the C-terminus of RBM24 is not clear, but both N- and C-termini are required for interactions with other protein partners (Liu, *et al*., 2019a). Recent developments in deciphering RBM24 function implicate an involvement of RBM24 in bladder cancer progression (Yin *et al*., 2021), maintenance of auditory and motor coordination (Zheng *et al*., 2021), adult skeletal muscle regeneration (Zhang *et al*., 2020) and dilated cardiomyopathy (Liu, *et al*., 2019a). Interestingly, all these cellular processes are regulated by alternative splicing mediated by RBM24. On the contrary, a circular RNA (SMARCA5) was found to suppress non-small cell lung cancer progression by regulating the miR-670-5p/RBM24 axis (Zhang *et al*., 2020).

RBM24 therefore acts as a multifaceted regulator in several cellular processes involved in development and disease. It regulates cell differentiation through distinct mechanisms varying in tissue-specific and stage-specific manners. However, many aspects related to RBM24 such as its dynamic subcellular organization, tissue-specific function and its modulation via interaction with other partners in the cell, remain to be explored. **Table 1.2** in *Supplementary information*, lists a complete set of potential RBM24 functions in development and disease.

1.6 Circular RNA

CircRNAs represent a novel class of covalently closed RNAs, among which exonic circRNAs derived from eukaryotic protein-coding genes are most prominent and relatively well-characterised. CircRNAs were first identified in 1976 by Sanger *et al*. as small single-stranded circular viroids, which are pathogens in higher plants. These viroids are uncoated RNA molecules (246-401nt) that do not encode any proteins. Unsuccessful end-labelling and electron microscopy of viroid RNA followed by subsequent sequencing of viroid nucleotide sequence confirmed that viroids are true circular RNAs (Gross *et al*., 1978). In addition, circular forms of RNA were observed in cytoplasmic fractions of eukaryotic cell lines (HeLa cells) by electron microscopy (Hsu & Coca-Prados, 1979). However they were considered products of aberrant splicing events with no functional relevance (Cocquerelle *et al*., 1993). Interestingly, hepatitis delta (δ) virus, a satellite virus of the hepatitis B virus, unlike plant viroids, was shown to be a circular RNA molecule (~1700nt) that encodes a protein required for viral replication (Kos *et al*., 1986; Weiner *et al*., 1988). In the following years, several studies identified circular forms of RNAs in various species including prokaryotes (Ford *et al*., 1994), unicellular eukaryotes (Grabowski *et al*., 1981), mammals (Capel *et al*., 1993) and viruses (Kos *et al*., 1986). CircRNAs in unicellular eukaryotes were found to be generated from rRNA (ribosomal RNA) type-I self-splicing introns (Grabowski *et al*., 1981), while circRNAs in archaea were formed as intermediates during rRNA processing (Kjems & Garrett, 1988). It was not until the 1990's that the first multicellular eukaryotic endogenous circRNAs were identified. Nigro *et al*. (1991), studied the exact order of exons in the DCC (Deleted in Colorectal Carcinoma – a tumor suppressor gene) mRNA in rodents and human cells, and found that the exons were arranged in the "wrong" order, i.e., although the canonical splice sites were conserved, the exons seemed to be scrambled in their arrangement. This study defined circRNAs as products of pre-mRNA splicing where the downstream splice-donor site, connects to the upstream splice acceptor site, thus forming a circRNA (Nigro *et al*., 1991). This process is also called "backsplicing", the mechanism, by which pre-mRNAs generate circRNAs through non-canonical splicing events (explained in **section 1.6.1**).

Additionally, as a direct evidence of circRNA production from nuclear pre-mRNA, scrambled exons at higher levels than DCC were identified from human *ETS-1* gene after transcription. Characterisation of ETS-1 transcripts revealed that they were circRNAs with predominant cytoplasmic localisation and high stability (Cocquerelle *et al*., 1993). The first mechanistic insight into how specific exons could be selected for circularisation was provided by the identification of the mouse *Sry* [Sex determining region (Y)] gene, which determines the sex in mammals (Capel *et al*., 1993). More than 90% of *Sry* transcripts in the adult mouse testis,

that accumulated in the cytoplasm and was not associated with polysomes, corresponded to the single-exon Sry circRNA. Several such endogenous circRNAs were identified, produced from protein-coding genes like rat cytochrome *P450 2C24* (Zaphiropoulos, 1996) and human *dystrophin* gene (Surono *et al*., 1999), without well-defined functions and characterised until early 2012 (Hansen *et al*., 2011; Burd *et al*., 2010; Houseley *et al*., 2006; Li & Lytton, 1999). With the development and availability of high-throughput RNA-sequencing and bioinformatics tools, circRNAs gathered much interest between 2012 and 2014 and were re-discovered as novel and ubiquitous non-coding RNAs (Salzman *et al*., 2012; Jeck *et al*., 2013). RNA-seq studies confirmed initial observations and characterised thousands of circRNAs produced from annotated protein-coding genes across various eukaryotes including metazoans, protists, fungi and plants (Conn *et al*., 2015, Ivanov *et al*., 2015; Lu *et al*., 2015; Guo *et al*., 2014; Zhang *et al*., 2014; Salzman *et al*., 2013). Surprisingly, thousands of circRNAs were also identified in cancer cells caused by chromosome rearrangements (Salzman *et al*., 2012).

One of the caveats in identifying circRNAs by RNA-seq is the relative low expression of circRNAs in cells and detection of potential artefacts of reverse transcription and/or other false-positive reads. It is therefore strongly recommended that at least two algorithms should be used to identify circRNAs: Five existing algorithms were used to identify circRNAs, total circRNAs identified ranged from 1532 to 4067. However, only 854 circRNAs were identified by all five algorithms (Hansen *et al*., 2016). This reinforces the need to use multiple independent pipelines to identify circRNA candidates, and newly identified circRNAs must be subsequently validated using RT-PCR across backspliced junction, Northern blotting, or using NanoString technology (Dahl *et al*., 2018; Hansen, 2018; Tatomer *et al*., 2017; Jeck & Sharpless, 2014). In this thesis, circRNAs were also characterised by RT-PCR and Northern blotting (see section 3.8). To overcome the limitation of low abundance of circRNA in cells in RNA-seq detection, in addition to using random priming in rRNA-depleted samples, exonucleases such as RNase R, or poly(A) selection steps are employed to deplete contaminating linear RNA and enhance sequencing coverage of circular RNAs (Panda *et al*., 2017; Jeck *et al*., 2013; Danan *et al*., 2012). Some common pipelines developed to detect circRNAs from RNA-seq datasets include CIRCexplorer3 (Ma *et al*., 2019), UROBORUS (Song *et al*., 2016), KNIFE (Szabo *et al*., 2015); CIRI (Gao *et al*., 2015); circRNA_finder (Westholm *et al*., 2014); CIRCexplorer (Zhang *et al*., 2014) and find_circ (Memczak *et al*., 2013). Many putative circular RNAs predicted through these tools have been assembled into searchable online databases like CIRCpedia2 (Dong *et al*., 2018); CircInteractome (Dudekula *et al*., 2016), CIRCpedia (Zhang *et al*., 2016) and circBase (Glažar *et al*., 2014).

1.6.1 CircRNA biogenesis

Circular RNAs depend on canonical splice sites and are generated by a special alternative splicing mechanism termed ‘backsplicing’, which has a different molecular mechanism than linear alternative splicing (Fu & Ares, 2014). When a pre-mRNA is backspliced, the 3’-end of an exon ligates to the 5’-end of the same or of an upstream exon through a 3’,5’–phosphodiester bond, forming a closed structure with the characteristic backsplice junction site (BSJ) (Chen, 2020; Zhou *et al*., 2020; Kristensen *et al*., 2019). Mutational analyses using circRNA expression vectors and blocking spliceosome assembly using the splicing inhibitor

isoginkgetin revealed that circRNA biogenesis depends on the canonical splicing machinery (Starke *et al*., 2015). When pre-mRNA processing events are slowed down, backsplicing is facilitated, as observed for *D. melanogaster,* where depletion of the components of U2 snRNP increased the ratio of circular to linear RNA products (Liang *et al*., 2017).

According to the order, in which splicing occurs and the intermediates generated, two models of circRNA biogenesis are proposed and validated (Jeck *et al*., 2013): the lariat model and the backsplicing model (Chen, 2015). **Figure 1.13** depicts the main hypothesis of backsplicing where flanking intron sequences of the downstream splice-donor site and upstream splice-acceptor site form a loop, thus bringing the two splice sites into close proximity. This looping is facilitated by base-pairing between Alu elements located in upstream and downstream introns; Alu elements are inverted highly repetitive elements composed of ~300 bases and are the most abundant primate transposable elements (Ivanov *et al*., 2015; Kelly *et al*., 2015; Zhang *et al*., 2014). Another mechanism in juxtaposing the splice sites, by the dimerization of RBPs that bind to specific motifs in the flanking introns. Common examples are the binding of quaking protein (HQK, encoded by *QKI*- Conn *et al*., 2015) and FUS (Errichelli *et al*., 2017). Taken together, it is now becoming evident that a combination of *cis*-acting elements (as demonstrated by the generation of a 490-nt circRNA from *D.melanogaster laccase gene2*) and *trans*-acting splicing factors such as hnRNP and SR-proteins, is required for circRNA biogenesis (Kramer *et al*., 2015). Nevertheless, just the presence of inverted repeats in the flanking introns does not necessarily trigger backsplicing, since several circRNAs expressed in humans, pigs, mice, *C.elegans* and *D.melanogaster* are not flanked by complementary repeats (Kristensen *et al*., 2017; Barrett *et al*., 2015; Ashwal-Fluss *et al*., 2014; Westholm *et al*., 2014). CircRNAs are usually expressed in a tissue- and developmental-stage-specific manner (Memczak *et al*., 2013). Interestingly, DHX9 (ATP-dependent RNA helicase A) and ADAR (adenosine deaminase) enzymes suppress the biogenesis of circRNAs that rely on base-pairing between inverted repeats (Eisenberg *et al*., 2018). Adenosine-to-inosine editing by dsRNA (double-stranded RNA)-specific ADAR enzymes preventing activation of innate immune system, and the unwinding of dsRNA helical structures by DHX9 prevent the looping of intron sequences (Aktaş *et al*., 2017). By contrast, base pairing between intronic RNA pairs is stabilised by the action of NF90 and NF110, protein products of ILF3 (interleukin enhancer-binding factor 3) involved in host antiviral mechanisms, thus promoting the production of circRNAs (Li *et al*., 2017).

Another interesting mechanism of circRNA biogenesis, proceeds from lariat precursors (Kristensen *et al*., 2019). CircRNA biogenesis was coupled to lariat formation during exon skipping events, thereby allowing a single pre-mRNA to generate both a linear mRNA and a circRNA (Eger *et al*., 2018; Kelly *et al*., 2015). Barrett *et al*. (2015) have demonstrated a mandatory coupling between exon skipping and circRNA formation for the *mrps16* gene in *S.pombe*, where splicing of exon 1 to exon 3 of *mrps16* releases an intron lariat containing exon 2, which is spliced again to form a circRNA. Also, intronic lariats that escape debranching by DBR1 enzyme lead to formation of ciRNAs (circular intronic RNAs).

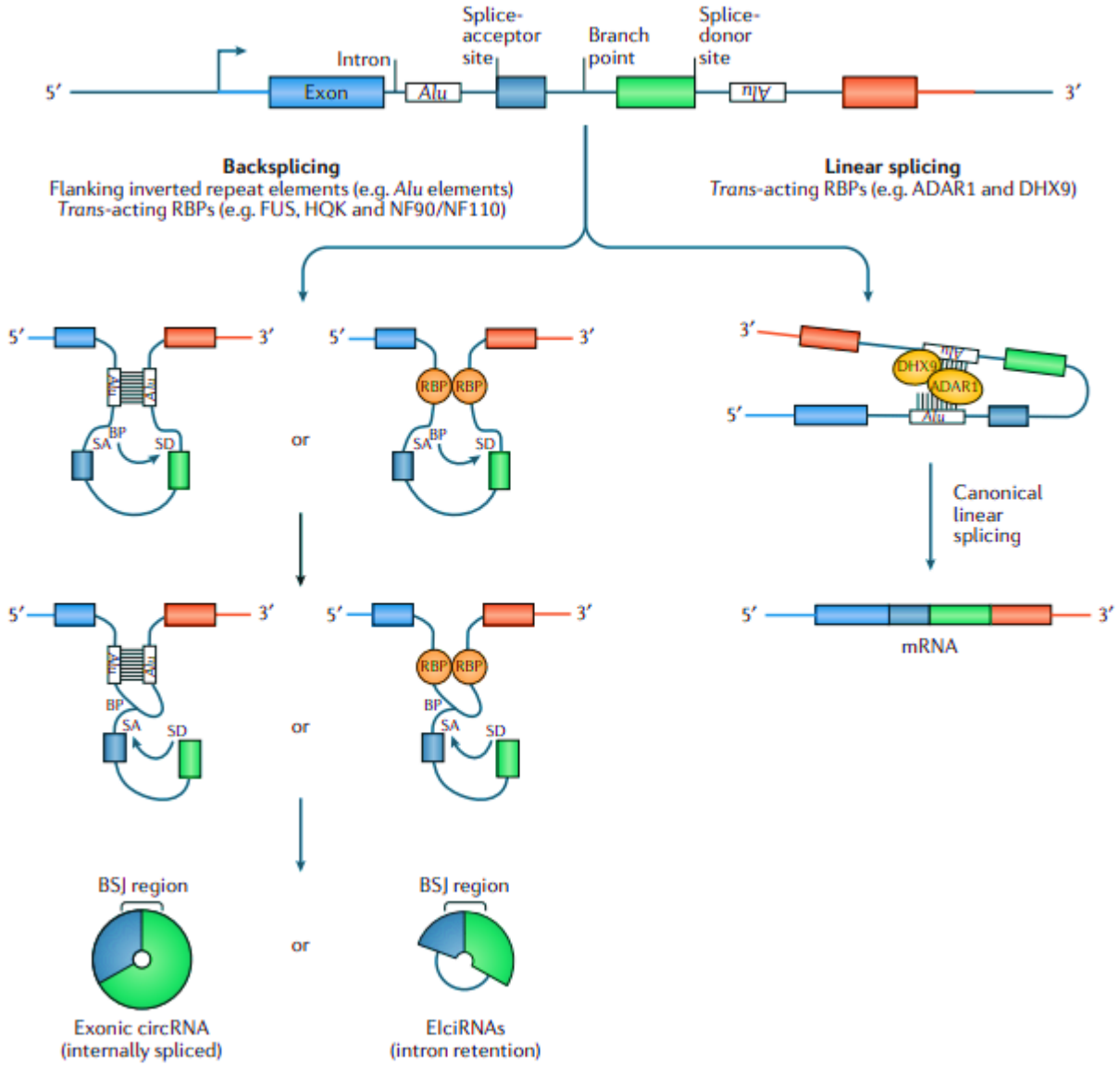

Figure 1.13 Biogenesis of circRNAs versus linear mRNA splicing

Schematic representation of competition between the linear splicing and backsplicing of exons (coloured boxes). Long flanking introns (in blue lines), inverted repeat elements (such as Alu elements – white boxes) and trans-acting RNA binding proteins (RBPs; for example, RNA-binding protein FUS, protein quaking (HQK) and NF90 and NF110) favour backsplicing (left). Canonical linear splicing (right) is favoured by exons surrounded by short flanking introns and by introns bound by the trans-acting RBPs double-stranded RNA-specific adenosine deaminase (ADAR1) and ATP-dependent RNA helicase A (DHX9). Base pairing between Alu-elements or dimerization of RBPs brings splice sites together forming exonic circRNAs or exon–intron circRNAs (EIcircRNAs). From Kristensen *et al.* (2019), modified.

Recently, an excellent study by Li *et al*. (2019) proposed an integral model for exon and intron definition, remodeling and, circRNA biogenesis mediated by backsplicing, through structural and biochemical analyses of the spliceosomal E complex assembled on the pre-mRNA (Zhou *et al*., 2020). Finally, circRNAs can also be produced during tRNA (transfer RNA) biogenesis. Splicing of pre-tRNA yields exons and introns with 2’,3’ cyclic-phosphate ends that are ligated

by endogenous RtcB ligase to generate mature tRNA and tricRNA (tRNA intronic circular RNA) respectively (Schmidt *et al*., 2019; Noto *et al*., 2017, reviewed in Popow *et al*., 2012). Functions of tricRNAs are still unclear, with just one exceptional example of its role in leucine tRNA methylation by the methyltransferase Trm4 in *S.pombe* (Müller *et al*., 2019).

Following biogenesis, most exonic circRNAs are localised to the cytoplasm (Salzman *et al*., 2012), however, little is known about circRNA nuclear export. Huang *et al*., (2018) proposed a length-dependent mechanism of circRNA export based on their observations that depleting spliceosomal helicases caused enrichment of circRNAs of different sizes. The study reports that while the spliceosome RNA helicase DDX39B (UAP56) drives the export of long (>1300 nt) circRNAs, ATP-dependent RNA helicase DDX39A (URH49) was found to export short (<400 nt) circRNAs to the nucleus. Some circRNAs have also been found to be localised in cell organelles such as the mitochondria (Zhao *et al*., 2020; Wu *et al*., 2020). In addition to regulating biogenesis of circRNA (Di Timoteo *et al*., 2020; Zhang *et al*., 2020; Tang *et al*., 2020), N6-methyladenosine (m6A) modification may also control circRNA translocation (Chen *et al*., 2019).

A general degradation pathway for circRNAs has not yet been explained, nevertheless, global reduction in circRNA abundance due to dilution by proliferation in highly proliferative tissues is reported (Bachmayr-Heyda *et al*., 2015). Latest research in this field identified some global mechanisms including – (a) degradation by cytoplasmic endonuclease RNase L upon viral infection (Liu *et al*., 2019b), (b) removal of circRNAs from cytoplasm via packaging into extracellular vesicles (EVs), detected in blood and urine (Preußer *et al*., 2018) and (c) a structure-mediated circRNA decay by UPF1 (RNA helicase and ATPase) and G3BP1 (stress granule assembly factor 1) (Fischer *et al*., 2020).

1.6.2 Functions of endogenous circRNAs

For a long time, circRNAs have been considered non-coding RNAs (ncRNAs) with their function remaining elusive (Guo *et al*., 2014; Memczak *et al*., 2015). Only for a very minor fraction of the circRNAs identified, miRNA sponging has been investigated as their biological function (Piwecka *et al*., 2017; Zheng *et al*., 2016; Hansen *et al*., 2013). CiRS7/CDR1as, perhaps the best-characterised circRNA to date, illustrates the best example for a miRNA sponge circRNA function. CiRS7 is a naturally expressed circRNA containing more than 70 binding sites for miR-7 and has the potential to regulate expression of miR-7 target genes (Hansen *et al*., 2013). Since ciRS7 is particularly abundant in neuronal tissues (Memczak *et al*., 2013), it may most likely function in neuronal function and differentiation, by protecting or inhibiting miR-7 (Kleaveland *et al*., 2018; Piwecka *et al*., 2017; Weng *et al*., 2017; Yu *et al*., 2016). Knockout of ciRS7 locus in mouse genome led to downregulation of miR-7 and a subsequent upregulation of miR7 target gene *Fos*, which affects the behavioral phenotype of knockout mice, displaying neuropsychiatric disorders (Piwecka *et al*., 2017).

Although several other circRNAs were reported to display miRNA sponging properties, like circHIPK3 (Zheng *et al*., 2016) and circBIRC6 (Yu *et al*., 2017), most circRNAs do not possess multiple miRNA binding sites. Nonetheless, circRNAs play a crucial role in development and disease through their miRNA sponging function, which is confirmed for the many circRNAs

involved in cancer progression or inhibition, such as circHIPK3 (Zheng *et al*., 2016), circPVT1 (Verduci *et al*., 2017), ciRS7 (Kristensen *et al*., 2018a) and circCCDC66 (Hsiao *et al*., 2017). circCCDC66 contains binding sites for many miRNAs targeting oncogenes, and therefore functions as a sponge for more than one miRNA such as miR-33b and miR-93 which target the *MYC* oncogene. circBIRC6 and circCORO1C inhibit miRNA-mediated suppression of pluripotency genes *SOX2*, *OCT4* and *NANOG* by sponging and thereby promote pluripotency in human embryonic stem cells (Yu *et al*., 2017).

In addition, circZNF91 which contains 24 binding sites for miR-23b-3p, induces differentiation in epidermal stem cells (Kristensen *et al*., 2018b). A complete list of proposed biological functions and mechanism of action for key circRNAs is enlisted in *Supplementary information* **Table 1.3**. Finally, there is expanding interest in recent years to identify circRNAs as cancer-specific biomarkers, emphasizing the need for good standards of circRNA characterisation and validation (Pfafenrot & Preußer, 2019).

Crosslinking immunoprecipitation datasets suggested several putative functions of circRNAs (reviewed by Kristensen *et al*., 2019; Hentze & Preiss, 2013) such as interactions with several RBPs (Dudekula *et al*., 2016), enhancing protein function (Li *et al*., 2015; Zhang *et al*., 2013), scaffolding function to mediate complex formation between specific substrates and enzymes (Zeng *et al*., 2017; Du *et al*., 2017) and recruitment of proteins to specific cellular locations (Chen *et al*., 2018). **Figure 1.14** schematically depicts potential functions of circRNAs. Furthermore, it has been proposed that circRNAs may also function as protein sponges (Ashwal-Fluss *et al*., 2014; Abdelmohsen *et al*., 2017). Notwithstanding, the first evidence for a protein sponging function was reported recently by our group for hnRNP L, based on designer circRNAs (Schreiner *et al*., 2020). A large portion of this thesis is dedicated to understanding the design and development of such designer-protein sponges for alternative splicing regulators, such as RBM24 and hnRNP L (see section 3).

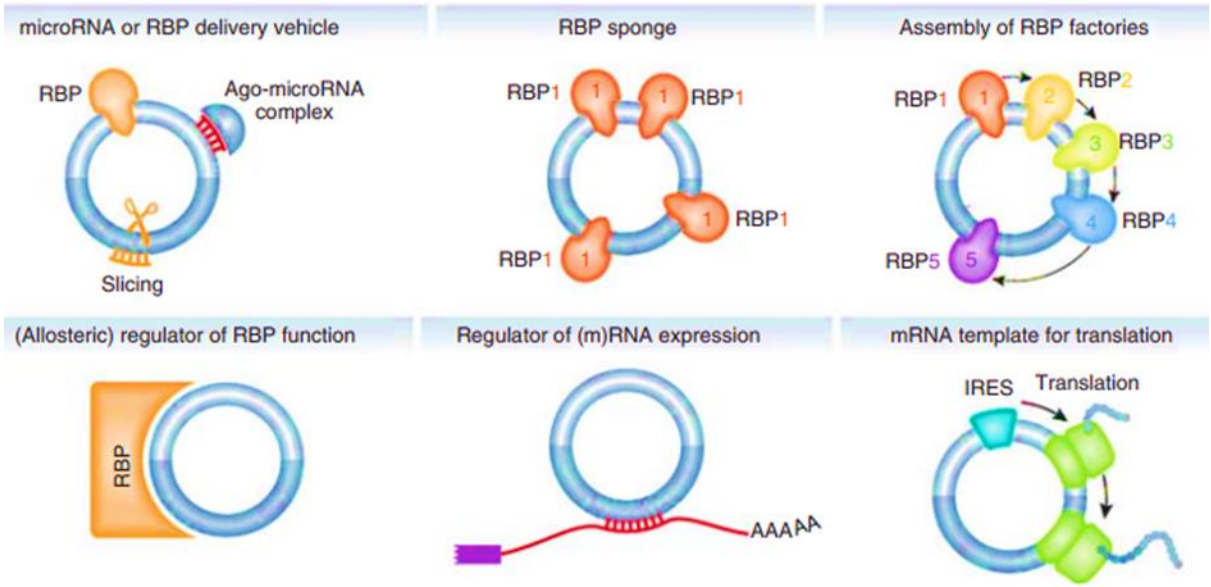

Figure 1.14 Plausible options for circRNA function

Pictogram represents plausible functions of circRNAs as delivery vehicles for microRNA and RBP (top-left), RBP sponge for one or several different RBPs (top-middle and right), regulator of RBP function (bottom-left) regulators of (m)RNA expression (bottom-middle) and mRNA template for translation of unique circRNA peptides (bottom-right). From Hentze & Preiss (2013), modified.

Although much debated, several reports also defined circRNAs to be translatable (Perriman *et al*., 1998; Chen *et al*., 1995), and more recent studies presented evidence for translation potency of a subset of circRNAs *in vivo* (Pamudurti *et al*., 2017) and some other endogenous circRNAs (Begum *et al*., 2018; Lei *et al*., 2020).

Possible mechanisms for circRNA translation involve Internal Ribosome Entry Site (IRES) and N^6 – methyladenosine (m^6 A)-mediated cap-independent translation (Lei *et al*., 2020). Interestingly, circZNF609, an endogenous circRNA, was reported to be associated with heavy polysomes, to be translated in a splicing-dependent and cap-independent manner, and to function in myogenesis (Legnini *et al*., 2017). Another study coupled the activity of hnRNP L in circularizing *ARHGAP35* (also known as P190-A, a tumour suppressor gene) to produce an oncogenic protein that promotes cancer cell progression (Li *et al*., 2021). In contrast, studies based on small subsets of highly abundant human circRNAs and polysome gradients (Schneider *et al*., 2016; Jeck *et al*., 2013), global analysis of ribosome-footprinting data from human U2OS cells (Guo *et al*., 2014) and polysome fractionation combined with RNA-seq in the mouse-brain (You *et al*., 2015) – all argue against endogenous circRNAs to recruit ribosomes and be translated at significant levels. Nonetheless, while it is thought that a vast majority of circRNAs are noncoding (Stagsted *et al*., 2019; Kristensen *et al*., 2019; You *et al*., 2015), there are open questions regarding the overall biological significance of circRNA translation and regarding the borderline to pervasive background translation. More work in this direction may help overcome this experimental and intellectual challenge in the future (Schneider & Bindereif, 2017; Hansen, 2021).

1.7 CircRNA expression systems in human cells

The functional and biological relevance of endogenous circRNAs can be studied by overexpressing the circRNA of interest. A reliable overexpression system is therefore essential to functionally study both endogenous and artificial circRNAs *in vivo*. The entire gene (Barrett & Salzman, 2016) or a mini-gene construct containing the circularising exon flanked by inverted complementary sequences, can be expressed in cells (Starke *et al*., 2015; Liang & Wilusz, 2014). Usually these inverted repeats are naturally occurring *Alu* repeat elements, or a short region of at least 30 nucleotides from the upstream intron inserted around the circularising exon in an inverted orientation (Liang & Wilusz, 2014; Hansen *et al*., 2013). Including binding motifs for specific RBPs in the flanking introns is another strategy to enhance circRNA expression using plasmids (Conn *et al*., 2015). Artificial circRNAs are usually generated *in vitro* by T7 transcription from synthetic DNA oligonucleotides followed by enzymatic ligation of the RNA transcribed (Müller & Appel, 2017; Breuer & Rossbach, 2020). Circularisation efficiency of RNAs *in vitro* is sequence-specific and conventionally *in vitro* circRNA production is more suitable for shorter circRNAs (in this thesis, RNA up to 120 nt in length was circularised *in vitro*, see **section 2.2.16**). A stem region can be inserted around the circularising region which stabilises the base-pair interactions across the exon and enhances circularisation by preventing additional secondary structure formation. CircRNAs thus produced can be directly transfected in desired cell systems.

Based on a vector with group I self-splicing introns [permuted intron-exon (PIE)] found in *Anabaena* pre-tRNA, Wesselhoeft and colleagues demonstrated a large-scale production strategy of circRNA *in vitro* (Wesselhoeft *et al*., 2018). To generate circRNA of interest, the sequence is inserted into the so-called PIE vector which contains homology arms and spacers flanking the circularising exonic region. First, the RNA is transcribed *in vitro.* After transcription, circRNA is formed by autocatalytic ligation, and circRNAs are further purified by RNase R treatment and HPLC purification. Using this strategy one can generate circRNAs up to 5 kb in length. However, a drawback of this technique is the formation of several contaminating circRNA concatemers, which are a result of intermolecular reactions between PIE RNA molecules instead of intramolecular reactions.

Conventionally, circRNAs were produced in eukaryotic cell systems by transfecting vectors containing the circRNA of interest. Hansen and colleagues first demonstrated transient circRNA expression *in vivo* using a vector containing splice sites from the circularizing exon of ciRS-7 pre-mRNA and the flanking intronic sequences surrounding the exon (Hansen *et al*., 2013). Another attempt in transient circRNA expression was made by developing a vector based on the human *ZKSCAN1* gene (Liang & Wilusz, 2013). This vector contains *ZKSCAN1* circRNA splice sites, which were surrounded by intronic regions with upstream AluSq2 and downstream AluJr & AluSz sequences. However, the major drawback of these vector-based expression systems was inefficient circularisation. The principal outcome of splicing appeared to be a linear RNA instead of a circular RNA. To overcome the limitations of exonic vector-based circRNA production systems, a vector based on a tRNA splicing mechanism was introduced for expressing and imaging of stable circRNAs *in vivo* (Schmidt *et al*., 2016; Lu *et al*., 2015). This vector is based on an intron from the *Drosophila* tRNA:Tyr_{GUA} gene, where excision of tRNA intron generates an intronic intermediate that is circularised by the endogenous RtcB ligase (2',3'-cyclic phosphate and 5'-OH ends). Transcription of this vector is under the influence of U6 promoter. This promotes 5' capping of the transcript by γ-monomethyl phosphate (Good *et al*., 1997) which improves the stability of the RNA expressed.

1.8 CircRNA overexpression by tornado system

The challenge in designing optimal expression systems for RNA aptamers has been the low expression (low nanomolar range) in mammalian cells and rapid degradation. With the introduction of Twister ribozymes a new level of circRNA expression with almost no detectable linear RNA by-products was achieved. The Tornado (Twister-optimised RNA for durable overexpression) expression system was designed to achieve rapid RNA circularisation, resulting in circRNA with high stability and expression levels (Litke & Jaffrey, 2019). The Tornado expression cassette includes the most effective promoter for small RNAs – the U6 promoter (Good *et al*., 1997), which is transcribed by RNAP III.

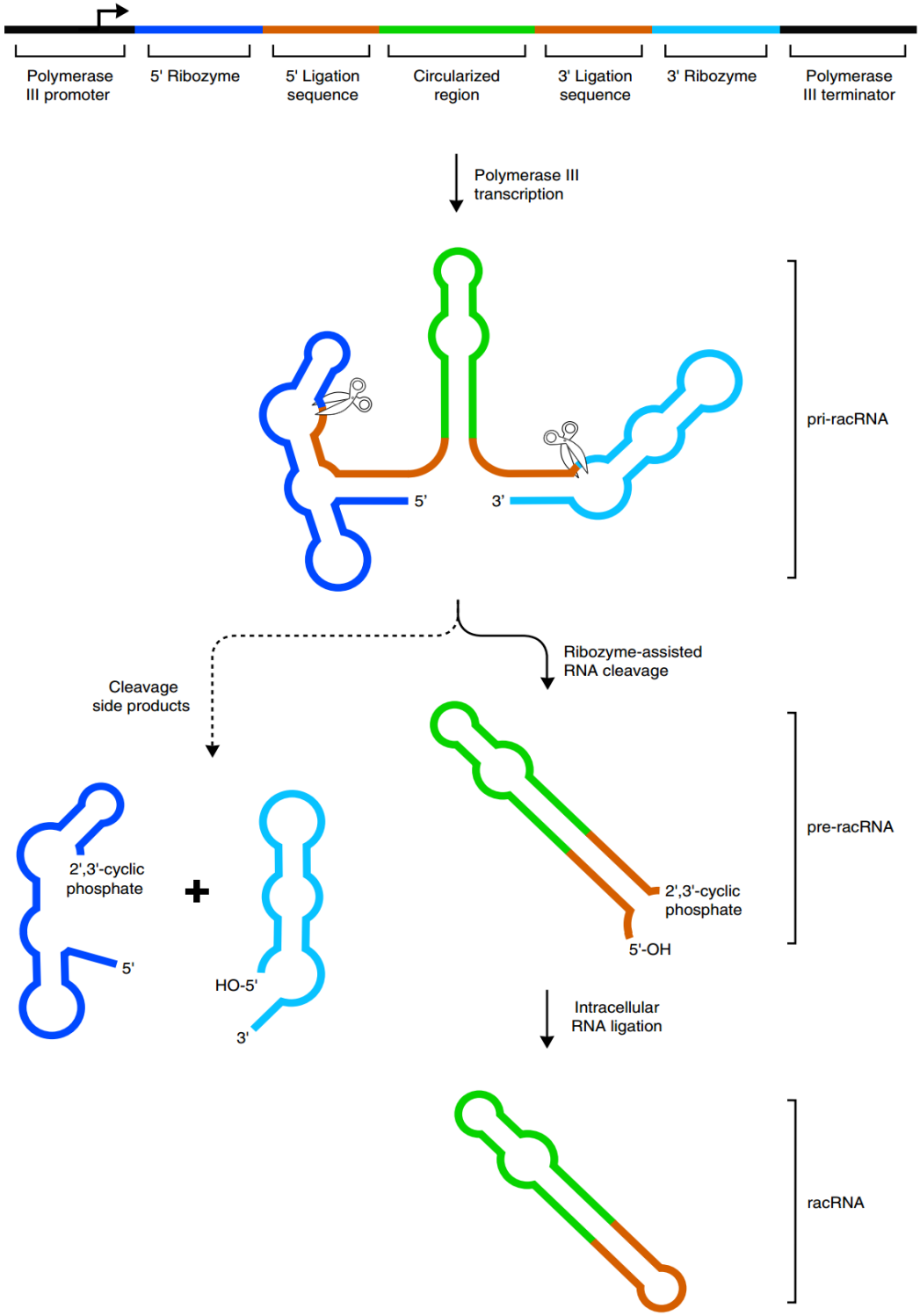

Figure 1.15 Tornado expression system.
Schematic representation of autocatalytically processed circRNA expression. circRNA sequence (green) is flanked by the 5′- and 3′-stem-forming sequences (orange), each of which is flanked by the 5′- and 3′-self-cleaving ribozymes (blue and cyan, respectively). After transcription, autocatalytic cleavage of ribozymes generates a 2′,3′-cyclic phosphate and 5′-OH on the new RNA ends, which are ligated by RtcB ligase. RacRNA, ribozyme-assisted circular RNA. From Litke & Jaffrey (2019).

The Tornado expression cassette also includes a Broccoli aptamer, which can be used for visualisation of the RNA by fluorescence imaging (see below). This Broccoli aptamer is flanked by combinations of the naturally occurring Twister ribozyme sequences – upstream Twister P3 U2A and downstream Twister P1. Twister ribozymes belong to the class of fastest-cleaving ribozymes (Roth *et al*., 2014), which undergo self-cleavage producing 5'-OH and 2',3'-cyclic phosphate ends in the intermediate RNA species. This intermediate complex resembles the TSEN-(tRNA splicing endonuclease) driven pre-tRNA intron excision complex in tRNA splicing. The 5'-OH and 2',3'-cyclic phosphate ends of the ribozyme assisted circRNA are ligated by the almost ubiquitous endogenous RNA ligase RtcB. The unligated fragments of the precursor circRNA are rapidly degraded by exoribonucleases. Since circRNAs resist endogenous exoribonucleases, this approach enables circRNA formation to exceptionally high concentrations. One of the biggest advantages of using this approach is the stability of the resulting circRNAs, which are circularised rapidly and efficiently. Secondly, circularisation is achieved without co-expression of any additional proteins or enzymes like in vector-based systems. Finally, produced circRNA accumulates to very high concentrations. The mechanism of circRNA formation through the Tornado system is schematically represented in **Figure 1.15**.

In addition, for direct visualisation of expressed circRNA in living cells or in total RNA sample on a polyacrylamide-gel, fluorogenic aptamers such as Spinach (Paige *et al*., 2011) and Broccoli (Lu *et al*., 2015) can be introduced. These aptamers fold into a highly stable RNA structure capable of binding and retaining fluorophore molecules. When activated by binding to an aptamer, these molecules can be visualised by fluorescence detection. Broccoli aptamer is a 49 nt long RNA which activates fluorescence of DFHBI fluorophore (Filonov *et al*., 2014). These aptamers are routinely used in mammalian and bacterial live-cell imaging to detect direct localisation of RNA molecules. In fact, in-gel Broccoli imaging may serve as an alternative to detection by Northern blotting (Filonov *et al*., 2015).

1.9 Splicing and disease

Since a vast majority (>95%) of protein-coding genes are alternatively spliced (Pan *et al*., 2008; Wang *et al*., 2008), it is not surprising that defects in alternative splicing are frequently associated with human diseases and cancer. Given the complexity of regulatory patterns governing alternative splicing regulation, there are many functional links between disruption of alternative splicing and disease (reviewed by Montes *et al*., 2018; Scotti & Swanson, 2016). Most diseases arise due to mutations in the splicing regulatory elements (SREs) – both the *cis*-acting RNA sequence elements as well as the *trans*-acting splicing factors. Diseases associated with mutations in the splice sites are described in detail further below. In addition to mutations in SREs on the pre-mRNA, overexpression or mutation of splicing factors and aggregation of splicing factors induced by mutations (particularly notable in neurological diseases) contribute to disease progression. It is estimated that out of roughly 80,000 mutations in the human gene mutation database (HGMD) about 10% affect splice sites (Tazi *et al*., 2009). However, this is likely an underestimate considering the mutations affecting splicing indirectly by associating with the SREs, *cis*-elements, or *trans*-acting factors, which were not tracked. Mutations affecting splicing can be broadly classified into two groups: type

I mutations that affect the invariant canonical splice sites resulting in disruption of core consensus sequences (5' splice site, BP-adenosine, polypyrimidine tract and 3' splice site). About 9-10% of genetic diseases are caused by canonical splice site mutations (Wang & Cooper, 2007). Type II mutations affect variant positions of splice sites or the regulatory elements that modulate spliceosome recruitment i.e., ESE and ESS elements in the exonic region and the ISE and ISS regions of the introns (Singh & Cooper, 2012). More than half of known disease-causing mutations are hypothesised to disrupt splicing (López-Bigas *et al*., 2005). For some genes it was reported that up to 50% of point mutations within the exon, affect splicing (Baralle & Buratti, 2017).

The first implication of splicing in disease was documented in the 1980s with the discovery of a point mutation in *HBB* gene encoding β-globin, which generated an alternative 3' splice site and promoted intron retention. This mutation resulted in β^{+}-thalassemia, a condition that is characterised by reduced β-globin protein levels and anaemia (Busslinger *et al*., 1981). Since then, several other examples have emerged that pointed out the role of mutations in splicing control such as: splice site mutation in dystrophin (*DMD*), resulting in loss of dystrophin function and Duchenne muscular dystrophy (Fletcher *et al*., 2013); polymorphic UG and U tracts near the 3' splice site of exon 9 in cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, which modulate the severity of cystic fibrosis (Tsui & Dorfman, 2013); and mutations in the 5' splice site, ESS and ESE of exon 10, in microtubule associated protein tau (*MAPT*), causing frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17, Niblock & Gallo, 2012). While these examples represent gene mutations causing a single type of disease, mutations in several types of sequence elements in the gene encoding lamins A, C, Δ10 and C2 result in multiple pathological phenotypes (Gruenbaum & Medalia, 2015). Lamins are type V intermediate filament proteins of the nucleus broadly classified into two groups based on function – nucleoplasmic lamin (gene expression) and peripheral lamins (differentiated cell nuclear architecture). Diseases caused due to lamin mutations are called laminopathies, and they comprise a heterogeneous group of over 14 diseases including, muscular dystrophies, cardiomyopathies, lipodystrophies, hereditary peripheral neuropathies and premature ageing (progeroid) syndromes (Luo *et al*., 2014).

Other major diseases caused due to splicing are: core spliceosome mutations causing retinitis pigmentosa (Liu & Zack, 2013); spliceosome dysregulation causing cancer (David and Manley, 2010); and a well-known example of spinal muscular atrophy (SMA) caused due to a single nucleotide substitution in exonic splicing signal in *SMN2* (survival of motor neuron 2) gene. SMA is a recessive autosomal disorder characterised by degeneration of spinal cord motor neurons leading to muscle atrophy (Cartegni & Krainer, 2002; Frugier *et al.,* 2002). In humans, SMN protein plays a critical role in snRNP biogenesis and is encoded by two genes, *SMN1* and *SMN2*, both of them encoding the same open reading frame. Most patients with SMA have deletions of the *SMN1* gene, but retain *SMN2*. A single nucleotide substitution C→T at position 6, in exon 7 of *SMN2* gene, causes frequent skipping of exon 7, thereby significantly altering the splicing pattern of SMN2 pre-mRNA. This effect is probably due to the disruption of an ESE element within exon 7. Thus, *SMN2* gene, producing a truncated and non-functional protein, is not able to completely compensate for the loss of *SMN1*, leading to SMA.

Interestingly, a very recent study also implicated a role of splicing in the novel SARS-CoV-2 virus life cycle, which causes the respiratory disease known as coronavirus disease 2019 (COVID-19), which developed into a pandemic in early 2020. According to this study, SARS-CoV-2 disrupts splicing, translation and protein trafficking to suppress host defences (Banerjee *et al.,* 2020). NSP16 (non-structural protein 16) of the virus binds to the mRNA recognition domains of the U1 and U2 splicing RNAs and acts to suppress global mRNA splicing upon SARS-CoV-2 infection. Disruption of the three non-overlapping steps of protein production – splicing, translation and protein trafficking, may represent a multi-pronged mechanism that synergistically acts to suppress the host antiviral response. These three cellular processes are disrupted before the formation of dsRNA (during viral genome replication) that could trigger type I IFN response in host defence, therefore allowing the virus to replicate its genome while the host innate immune response is minimised (Banerjee *et al.,* 2020).

1.10 Specific aims of this work

CircRNA functions are likely diverse and much debated. While, most studies focus on discovering endogenous functions of circRNAs; as an alternative approach, designing circRNAs to fulfil specific functions in the cell opens many possibilities to ascertain their functional relevance. Building on this idea, this thesis focuses on design, expression, and functional characterisation of artificial circular RNAs (circRNAs), to act as protein sponges for important regulators of alternative splicing such as hnRNP L and RBM24. We developed designer-made circRNAs, which, when transfected into the cell, function as a sponge specifically binding to these proteins. Sponging of these proteins interferes with their function in alternative splicing. Sequestration by designer circRNA protein-sponges expressed in cells, establishes conditions similar to RNAi-mediated knockdown of a targeted RBP. The basis for employing circRNAs over other linear forms lies in the chemical properties of a circRNA – its stability in the cell, which in some cases is up to four days post-transfection, and resistance to degradation by cellular RNases.

The rationale behind such an approach is to regulate alternative splicing in the cell by controlling the availability of the protein. Sponging of a splice regulator renders the protein temporarily unavailable for binding with the target genes, thereby causing variations in splice patterns. Our results strongly argue for protein-sponging as an alternative to RNAi-mediated knockdown of proteins because, sponging allows for broader control of the protein-availability. By sponging, one can precisely control the amount and time for which a protein can be sponged, since protein-sponging is a reversible reaction and protein function is not completely abolished.

The overall aim of this work is two-fold – (a) to establish a functional role for circRNAs in protein sponging (b) to develop optimal circRNA protein sponges and validate their functionality in the context of hnRNP L-mediated alternative splicing regulation. With this general agenda, this thesis has three specific aims which are outlined below.

First, to understand the combinatorial recognition of RNA elements that bind to multidomain proteins like hnRNP L, and to characterise RNA elements that bind to single domain RBM24. The primary step in designing optimal circRNA sponges was to decipher what is the exact binding sequence of hnRNP L and RBM24. To answer this question, we characterised the RNA elements that would bind to these proteins via SELEX-based selection and subsequent high-throughput sequencing.

Second, to design optimal circRNA sponges and to express them in cellular systems. SELEX-derived consensus sequences were analysed bioinformatically for motif enrichment over four rounds of SELEX selection and preferential spacing of RNA-binding domains. This formed the basis for circRNA design. In order to validate protein-sponging function, the experimental circRNA sponge was expressed *in vivo* using tRNA-based Tornado overexpression system.

Lastly, the final step in demonstrating functionality of circRNA acting as a protein sponge, was to validate its effects *in vivo* on cellular processes such as alternative splicing. As a model system, we used circRNA targeting hnRNP L, a global alternative splicing regulator. Direct *in vivo* hnRNP L/circRNA interactions were captured by RNA immunoprecipitation (RIP) coupled to cell fractionation and Northern blotting. Changes in alternative splicing patterns of the hnRNP L target genes were documented by RT-PCR. The results in this thesis establish a novel protein-sponging function of designer circRNAs discovered by our group recently. We argue for the high potential of designer circRNAs as a novel class of therapeutic RNAs to be applied in diseases caused due to RBPs. Protein sponging may add an additional layer of understanding gene regulatory networks within the cell by allowing precision and control of clinically- and biotechnologically-relevant RBPs.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

2- mercaptoethanol	Roth
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside(X-Gal)	Roche
Acetic acid	Roth
Acetone	Roth
Acrylamide	Bio-Rad
Acrylamide/bisacrylamide 30, 30%, 37.5:1	Roth
Acrylamide/bisacrylamide 40, 40%, 19:1	Roth
Agarose ultra-pure	Roth
Ammonium persulfate (APS)	Bio-Rad
Ampicilin	Roth
Bacto-agar	Roth
Bacto-tryptone	Roth
Bacto-yeast extract	Roth
Bisacrylamide	Bio-Rad
Blocking reagent	Roche
Boric acid	Roth
Bovine serum albumin (BSA)	Roche
Bromophenol blue	Merck
Calcium chloride	Merck
CDP-Star chemiluminiscent substrate	Sigma-Aldrich
Chloroform	Roth
Creatine phosphate	Roche
Deoxynucleotide triphosphates (dNTPs)	Peqlab
DFHBI	Sigma-Aldrich
Dimethyl pyrocarbonate (DMPC)	Sigma-Aldrich
Dithiothreitol (DTT)	Roche
Dulbecco's Modified Eagle's Medium (DMEM)	Gibco
Ethanol (EtOH)	Roth
Ethidium bromide	Roth
Ethylendiaminetetraacetic acid (EDTA)	Roth
Ethylene glycol	Roth
Fetal bovine serum (FBS)	Gibco
Formamide	Roth
Gentamycin	Sigma-Aldrich
Glucose	Sigma-Aldrich
Glycerol	Roth
Glycine	Roth
Glycogen	PeqLab
Heparin	Sigma-Aldrich
Hydrochloric acid	Roth
N-2-hydroxyethylpiperazine (HEPES)	Sigma-Aldrich

Imidazole	Roth
InstantBlue Protein Stain	CBS Scientific
Isoamyl alcohol	Roth
Isopropanol	Roth
Lipofectamine 2000 transfection reagent	Thermo Scientific
Lumi-Light Western blotting substrate	Roche
Isopropyl-1-thio-β-D-galactoside (IPTG)	Roche
Kanamycin	Sigma-Aldrich
Lysozyme	Sigma-Aldrich
Magnesium chloride ($MgCl_2$)	Merck
Magnesium sulphate	Sigma-Aldrich
Maleic acid	Roth
Methanol	Roth
Milk powder (fat-free)	Roth
Nonidet P-40 /Igepal CA-630	Sigma-Aldrich
NorthernMax hybridization buffer	Ambion
OptiMEM	Gibco
Phenol/chloroform/isoamylalcohol (25:42:1)	Roth
Phenylmethylsulfonyl fluoride (PMSF)	Roth
Phosphate-buffered saline (PBS), X10	Gibco
Polyoxyethyleneorbiten monolaurate (Tween20)	Sigma-Aldrich
Polyvinylalcohol	Merck
Potassium chloride (KCl)	Roth
Roti-phenol	Roth
Roti-phenol/chloroform	Roth
Rotiphorese gel 30% (37.5:1)	Roth
Sodium acetic acid (NaAc)	Merck
Sodium chloride (NaCl)	Roth
Sodium citrate	Roth
Sodium dihydrogen phosphate monohydrate($NaH_2PO_4 \cdot 2H_20$)	Merck
Sodium dodecyl sulfate (SDS)	Roth
Sodium Hydroxide(NaOH)	Merck
Super Optimal Broth (SOC) medium	Invitrogen
SYBR Gold solution	Roche
Tetracycline	Sigma-Aldrich
N,N,N´,N´-tetramethylenediamine (TEMED)	Bio-Rad
Tris-hydroxymethylaminomethane (Tris)	Roth
Triton X-100	Merck
Trizol reagent	Thermo Scientific
TRIzol-LC	Ambion
tRNA from yeast	Roche
Trypsin EDTA solution	Gibco
TurboFect in vitro transfection reagent	Thermo Scientific
Urea	Roth
Xylene cyanole FF	Sigma-Aldrich

2.1.2 Nucleotides

Deoxynucleoside triphosphate set (dNTP, 100mM)	Roche
[γ-^{32}P]ATP (3,000Ci/mmol, 5µCi/µl)	Hartmann Analytic
[α-^{32}P]CTP (800Ci/mmol, 10µCi/µl)	Hartmann Analytic
[α-^{32}P]UTP (800Ci/mmol, 10µCi/µl)	Hartmann Analytic

2.1.3 Enzymes and Enzyme inhibitors

AcTEV Protease (10U/ µl)	Invitrogen
Protease inhibitor cocktail tablets	Roth
Poteinase K (PK, 10 µg/µl)	Roche
Restriction endonucleases	New England Biolabs
RNase R	Epicenter
RNase inhibitor (RNaseOUT, 40U/µl)	Invitrogen
RQ1 RNase free DNase (1U/µl)	Promega
Shrimp alkaline phosphatase (SAP, 1U/µl)	Roche
Taq DNA polymerase (5U/µl)	Self made
T4 DNA ligase (400U/µl)	New England Biolabs
T4 RNA Ligase	New England Biolabs
T7 RNA Polymerase(20U/µl)	Fermentas
qScript reverse transcriptase	Quanta Biosciences

2.1.4 Reaction Buffers

10× restriction enzyme buffer	New England Biolabs
10× RQ1 DNase buffer	Promega
10× SAP buffer	Roche
10× Taq polymerase buffer	Promega
5× T4 DNA ligase buffer	New England Biolabs
10×T4 PNK Buffer	New England Biolabs
10× T4 RNA ligase buffer	Thermo Scientific
5× qScript RT buffer	Invitrogen

2.1.5 Molecular weight markers

DNA DIG-labeled Molecular Weight Marker VIII	Roche
GeneRuler DNA Ladder Mix	Thermo Scientific
GeneRuler Low Range (LR) DNA Ladder	Thermo Scientific
Riboruler High/Low Range (HR/LR) RNA Ladder	Thermo Scientific
PeqGOLD Protein-Marker IV	Peqlab

2.1.6 Kits

DIG-detection system	Roche
Gel extraction kit	QIAGEN
HiScribe T7 High Yield RNA synthesis kit	New England Biolabs
Monarch RNA cleanup kit	New England Biolabs

NE-PER nuclear and cytoplasmic extraction reagents	Thermo Scientific
Norgen total RNA purification kit	Norgen Biotek
Plasmid Maxi kit	QIAGEN
QIAprep Spin Miniprep kit	QIAGEN
QIAquick gel extraction kit	QIAGEN
qScript cDNA synthesis kit	Quantabio
qScript Flex cDNA kit	Quantabio
Qubit quantification assay kit, RNA BR	Thermo Scientific
Quick Spin RNA columns	Roche
RNeasy Mini kit	QIAGEN
TOPO TA cloning kit Invitrogen	Invitrogen

2.1.7 Plasmids

pGEX-6P2-hnRNP L	
pGEX-5x-2	Amersham Bioscience
pFAST-Bac-Htb-hnRNP L	described in Hui et al., 2005
pAV-U6+27-Tornado-Broccoli	Addgene
pAV-U6+27-Tornado-F30-Broccoli- GSE1	This study
pAV-U6+27-Tornado-F30-Broccoli- Plt_circR4	This study
pAV-U6+27-Tornado-F30-Broccoli-cirS7	This study
Tornado- (CA)20	This study
Tornado CA-SELEX X4	This study
Tornado-L9/10	This study
Tornado-L9/15	This study
Tornado-L12/10	This study
Tornado-L12/15	This study
Tornado-L12/10_mut	This study

2.1.8 Antibodies

Anti-Digoxigenin-AP Fab coupled to alkaline phosphatase	Roche
Anti-FLAG monoclonal antibody	Sigma-Aldrich
Anti-GAPDH monoclonal antibody 71.1	Sigma-Aldrich
Anti-GST monoclonal antibody	Sigma-Aldrich
Anti-hnRNP L monoclonal antibody 4D11	Sigma-Aldrich
Anti-Mouse IgG peroxidase antibody	Sigma-Aldrich
Anti-Rabbit IgG peroxidase antibody	Sigma-Aldrich

2.1.9 *E.coli* strains and cell lines

BL21 Star (DE3) F- ompT hsdSB (rB-mB-) gal dcm rne131 (DE3)	Invitrogen
DH10Bac Competant cells	Thermo Scientific
HeLa (human cervix carcinoma cells)	ATCC No. CCL-2
HEK293 (human embryonic kidney cell)	ATCC No. CRL-1573

JM109 {genotype: endA1, gyrA96, thi, hsdR17 (rk-, mk+), relA1, supE44, ≅(lacproAB), [F´, traD36, proAB, laqlqZ≅M15]}	Promega
Sf21 (*Spodoptera frugiperda* cell)	Invitrogen
TOP 10 high-competent cells	Invitrogen

2.1.10 Laboratory equipment

Agfa Curix 60 processing machine (developing machine)	AGFA
BLX 254 UV-crosslinker	Biolink
Centrifuge 5424	Eppendorf
Excella Eco-170 CO2 incubator	New Brunswick Scientific
G:Box gel documentation	Syngene
HB-1000 hybridization oven	UVP
InLab Expert Pro-ISM pH meter	Mettler Toledo
INTAS Imager Transilluminator ChemiDoc MP	Bio-Rad
Micropipettors (0.5 µl to 1000 µl)	Eppendorf, Gilson
Mini PROTEAN electrophoresis system	Biorad
Moxi Z Mini automated cell counter	ORFLO Technologies
NanoDrop 1000 spectrophotometer	Thermo Scientific
Pipetus pipet filler	Hirschmann
Qubit 2.0 fluorometer	Invitrogen
Realplex Mastercycler (thermocycler)	Eppendorf
Scintillation Counter	Packard
Subcell GT agarose gel system	BioRad
Systec DB-23 autoclave	Systec
Thermo Scientific Safe 2020 (laminar flow)	Thermo Scientific
Thermomixer L057	Kisker
Trans-Blot semi-dry transfer cell	BioRad
Trans-Blot Turbo transfer system	BioRad
Typhoon	GE Healthcare
Veriti thermal cycler	Applied Biosystems
Vortex-Genie2	Scientific Industries

2.1.11 Other consumables

Autoradiography Detection films	GE Healthcare
Culture bottles with filter (75 cm2)	Greiner Bio-One
Cell culture dishes 100 X 20 mm	Sarstedt
Cell culture plates (6, 12, 24 wells)	Sarstedt
Chemiluminescent detection films	GE Healthcare
Eppendorf tubes (1.5 ml, 2 ml; Safe Lock)	Sarstedt
Falcon tubes (15 ml, 50 ml)	Sarstedt
Multiply PCR strips and lids	Sarstedt
Pipette tips (10, 20, 2000, 1250 ml)	Sarstedt
qPCR Seal optical clear films	VWR
qPCR semi-skirted plates	VWR
Sterile serological pipettes (5, 10, 25, 50 ml)	Greiner

2.1.12 Oligonucleotides

The siRNAs were obtained from Sigma-Aldrich and all synthetic oligonucleotides from Merck.

siRNA oligonucleotides	
human hnRNPL 3' UTR 1581	GACAUUUCUCUUUCCUUUATT
luciferase GL2	CGUACGCGGAAUACUUCGATT
hnRNP L alternative splicing targets	
BPTF (FALZ) fwd	TCATCAAACCTTTGCTACATGG
BPTF (FALZ) rev	CTGACTGGTACCTGTACTTGATG
CC2D2A fwd	CCAGGGAAAGAGGTAGAAAGGAC
CC2D2A rev	GTCCTCGGCATCATCACCAT
HMMR fwd	AGATACTACCTTGCCTGCTTCA
HMMR rev	GCCTTGCTTCCATCTTTTCCA
GPBP1 fwd	TCCTGAGTATGAGAGAGAACCAA
GPBP1 rev	TCTTAACTGGCTGTGACGGA
PPP3CB fwd	AGTGGAGTGTTAGCTGGAGGA
PPP3CB rev	CCGAGGTGGCATTCTCTCAT
RIF1 fwd	TCCATACCATGCCCAACAGA
RIF1 rev	GAGTTGTCCCAGGCCTCTTG
TJP1 fwd	ATATCCTCCTTACTCACCACAAG
TJP2 rev	TTCAAAACATGGTTCTGCCTC
Housekeeping/ Reference genes	
ACTB fwd	GTGGAACGGTGAAGGTGACA
ACTB rev	GGGACTTCCTGTAACAACGCA
Tornado	
(CA)20 circ fwd	TCGGCGTGGACTGTAGAAC
(CA)20 circ rev	TGTGTGTGTGTGTGGCGG
CA-SELEX X4 Broccoli circ fwd	GGTCGGCGTGGACTGTAG
CA-SELEX X4 Broccoli circ rev	GATACGAATATCTGGACCCGACCGTC
T-L_12/10_fwd_circPrimer1.1	ACCCTCCACTACATCTTATGAAA
T-L_12/10_rev_circPrimer1.2	GTGGAGGGTGCGTGTCAT
T-L_12/10_mut_ fwd_circPrimer3.1	GCACCCTCCACTTGTTCTTA
T-L_12/10_mut_ rev_circPrimer3.2	TGCCACACATAAGCCCTTG
hnRNP L cloning	
hnRNPL_RRM1-2_fwd	CAGGAATTCTGGGTGAAAATTATGATGATCCG
hnRNPL_RRM1_rev	CCGCTCGAGTTAGTGATGGTGATGATGATGACCCTGG AAATACAGATTTTCGCTATTAACGCTACGGCTATC
hnRNPL_RRM2_fwd	CAGGAATTCTGCAGAAAATTAGCCGTCCG
hnRNPL_RRM1-2_rev	CCGCTCGAGTTAGTGATGGTGATGATGATGACCCTGG AAATACAGATTTTCACCCTGACCGCTCAGATT
hnRNPL_RRM3-4_fwd	CAGGAATTCTGTATGGTCCACATGCAGATAGTCC
hnRNPL_RRM3-4_rev	CCGCTCGAGTTAGTGATGGT

Endogenous circRNA cloning	
F30-CDR1as-KFlI_fwd	GATGGGTCCCGGTTTCCGATGGCACCTGTG
F30- CDR1as -KFlI_rev	GATGGGACCCCTGGATATTGCAGACACTGG
F30-GSE1-KFlI_fwd	GATGGGTCCCGCATGAGCCATGAGCCCAAG
F30-GSE1-KFlI_rev	GATGGGACCCCTCTGGGCTCCTCCGCCT
F30-NA-KFlI_fwd	GATGGGTCCCACTATGGAAATGGTGTCAGA
F30-NA-KFlI_rev	GATGGGACCCCTGCAAAAAGTTGTGGAGAT

hnRNP L cloning in baculovirus	
EcoRI_fwd	CCGGAATTCAAAGCCGTCGTCTGCTGCC
XhoI_rev	GTTAGCTGGCATGCTGTGCGG
Fwd_L_pFAST_Seq	ACCATCACCATCACCATCAC
Rev_L_pFAST_Seq	TCTACAAATGTGGTATGGCT
SeqPrimer_L_CO_fwd	CCGGATTATTCATACCGTCCCA
SeqPrimer_L_CO_rev	ACCACGAATATGAACAACCGG
SeqPrimer_L_rev	TTCCACCACACCGTCAATCA
hnRNP L_Bac_Seq_Fwd	GCAAGCCTGAATGGTGCAGA
hnRNP L _Bac_Seq_Rev	ATGATCACCCAGCAGCGGA
pUC/M13_fwd	CCCAGTCACGACGTTGTAAAACG
pUC/M13_rev	AGCGGATAACAATTTCACACAGG
U6_termination seq	AAAAGCGGACCGAAGTCCGCTCTAG

hnRNP L oligonucleotides

L_9/10_s	TAATACGACTCACTATAGGGCTTATGACACTCCACTTGGACACCTCC ACTTGGACATTCCACTTGGACATCTTATG
L_9/10_as	CATAAGATGTCCAAGTGGAATGTCCAAGTGGAGGTGTCCAAGTGGA GTGTCATAAGCCCTATAGTGAGTCGTATTA
L_9/15_s	TAATACGACTCACTATAGGGCTTATGACACTCCACTTGGACACGCAC CCTCCACTTGGACATTCCACTTGGACATCTTATG
L_9/15_as	CATAAGATGTCCAAGTGGAATGTCCAAGTGGAGGGTGCGTGTCCAA GTGGAGTGTCATAAGCCCTATAGTGAGTCGTATTA
L_12/10_s	TAATACGACTCACTATAGGGCTTATGACACGCACCCTCCACTACACC TCCACTTGGACATGCACCCTCCACTACATCTTATG
L_12/10_as	CATAAGATGTAGTGGAGGGTGCATGTCCAAGTGGAGGTGTAGTGG AGGGTGCGTGTCATAAGCCCTATAGTGAGTCGTATTA
L_12/15_s	TAATACGACTCACTATAGGGCTTATGACACGCACCCTCCACTACACG CACCCTCCACTTGGACATGCACCCTCCACTACATCTTATG
L_12/15_as	CATAAGATGTAGTGGAGGGTGCATGTCCAAGTGGAGGGTGCGTGT AGTGGAGGGTGCGTGTCATAAGCCCTATAGTGAGTCGTATTA
L_9/10_mut_s	TAATACGACTCACTATAGGGCTTATGTGTGTCCACTTGGTGTGCTCC ACTTGGTGTTTCCACTTGGTGTTCTTATG
L_9/10_mut_as	CATAAGAACACCAAGTGGAAACACCAAGTGGAGCACACCAAGTGG ACACACATAAGCCCTATAGTGAGTCGTATTA

L_9/15_mut_s	TAATACGACTCACTATAGGGCTTATGTGTGTCCACTTGGTGTGGCACCCTCCACTTGGTGTTTCCACTTGGTGTTCTTATG
L_9/15_mut_as	CATAAGAACACCAAGTGGAAACACCAAGTGGAGGGTGCCACACCAAGTGGACACACATAAGCCCTATAGTGAGTCGTATTA
L_12/10_mut_s	TAATACGACTCACTATAGGGCTTATGTGTGGCACCCTCCACTTGTGCTCCACTTGGTGTTGCACCCTCCACTTGTTCTTATG
L_12/10_mut_as	CATAAGAACAAGTGGAGGGTGCAACACCAAGTGGAGCACAAGTGGAGGGTGCCACACATAAGCCCTATAGTGAGTCGTATTA
L_12/15_mut_s	TAATACGACTCACTATAGGGCTTATGTGTGGCACCCTCCACTTGTGGCACCCTCCACTTGGTGTTGCACCCTCCACTTGTTCTTATG
L_12/15_mut_as	CATAAGAACAAGTGGAGGGTGCAACACCAAGTGGAGGGTGCCACAAGTGGAGGGTGCCACACATAAGCCCTATAGTGAGTCGTATTA
L_12/10+SL_s	TAATACGACTCACTATAGGGAGTAAGCCTTATGACACGCACCCTCCACTACACCTCCACTTGGACATGCACCCTCCACTACATCTTATGGCTTACAGTA
L_12/10+SL_as	TACTGTAAGCCATAAGATGTAGTGGAGGGTGCATGTCCAAGTGGAGGTGTAGTGGAGGGTGCGTGTCATAAGGCTTACTCCCTATAGTGAGTCGTATTA
L_12/10_mut+SL_s	TAATACGACTCACTATAGGGAGTAAGCCTTATGTGTGGCACCCTCCACTTGTGCTCCACTTGGTGTTGCACCCTCCACTTGTTCTTATGGCTTACAGTA
L_12/10_mut+SL_as	TACTGTAAGCCATAAGAACAAGTGGAGGGTGCAACACCAAGTGGAGCACAAGTGGAGGGTGCCACACATAAGGCTTACTCCCTATAGTGAGTCGTATTA
CA#51_fwd (20nt)	TAATACGACTCACTATAGGGATACATGACACACACACGCA
CA#51_rev (20nt)	TGCGTGTGTGTGTCATGTATCCCTATAGTGAGTCGTATTA
$(CA)_{10}$_fwd	TAATACGACTCACTATAGGGCACACACACACACACACACA
$(CA)_{10}$_rev	TGTGTGTGTGTGTGTGTGTGCCCTATAGTGAGTCGTATTA
T7_OR_CA#51_fwd	TAATACGACTCACTATAGGGAGTAAGCATACATGACACACACACGCAGCTTACAGTA
T7_OR_CA#51_rev	TACTGTAAGCTGCGTGTGTGTGTCATGTATGCTTACTCCCTATAGTGAGTCGTATTA
T7_OR_$(CA)_{20}$_fwd	TAATACGACTCACTATAGGGAGTAAGCCACACACACACACACACACACACACACACACACACACAGCTTACAGTA
T7_OR_$(CA)_{20}$_rev	TACTGTAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGGCTTACTCCCTATAGTGAGTCGTATTA

hnRNP L Tornado oligonucleotides

T_L9/10_fwd	GGCCGCAAGGGCTTATGACACTCCACTTGGACACCTCCACTTGGACATTCCACTTGGACATCTTATGAAAAACCGC
T_L9/10_rev	GGTTTTTCATAAGATGTCCAAGTGGAATGTCCAAGTGGAGGTGTCCAAGTGGAGTGTCATAAGCCCTTGC
T_L9/15_fwd	GGCCGCAAGGGCTTATGACACTCCACTTGGACACGCACCCTCCACTTGGACATTCCACTTGGACATCTTATGAAAAACCGC
T_L9/15_rev	GGTTTTTCATAAGATGTCCAAGTGGAATGTCCAAGTGGAGGGTGCGTGTCCAAGTGGAGTGTCATAAGCCCTTGC

T_L12/10_fwd	GGCCGCAAGGGCTTATGACACGCACCCTCCACTACACCTCCACTTGG ACATGCACCCTCCACTACATCTTATGAAAAACCGC
T_L12/10_rev	GGTTTTTCATAAGATGTAGTGGAGGGTGCATGTCCAAGTGGAGGTG TAGTGGAGGGTGCGTGTCATAAGCCCTTGC
T_L12/15_fwd	GGCCGCAAGGGCTTATGACACGCACCCTCCACTACACGCACCCTCCA CTTGGACATGCACCCTCCACTACATCTTATGAAAAACCGC
T_L12/15_rev	GGTTTTTCATAAGATGTAGTGGAGGGTGCATGTCCAAGTGGAGGGT GCGTGTAGTGGAGGGTGCGTGTCATAAGCCCTTGC
T_L12/10_mut_fwd	GGCCGCAAGGGCTTATGTGTGGCACCCTCCACTTGTGCTCCACTTGG TGTTGCACCCTCCACTTGTTCTTATGAAAAACCGC
T_L12/10_mut_rev	GGTTTTTCATAAGAACAAGTGGAGGGTGCAACACCAAGTGGAGCAC AAGTGGAGGGTGCCACACATAAGCCCTTGC

Oligonucleotides for Northern blot riboprobe

Riboprobe_L12-10_fwd	TAATACGACTCACTATAGGGAGGGTGCATGTCCAAGTGGAGGTGT AGTGGAGGGTGCGTGTTTTTTTTTTTTTTTTTTTTTT
Riboprobe_L12-10_rev	AAAAAAAAAAAAAAAAAAAAAAAACACGCACCCTCCACTACACCTC CACTTGGACATGCACCCTCCCTATAGTGAGTCGTATTA
Riboprobe_mut_fwd	TAATACGACTCACTATAGGGAAGAACAAGTGGAGGGTGCAACACC AAGTGGAGCACAAGTGGAGTTTTTTTTTTTTTTTTTTTT
Riboprobe_mut_rev	AAAAAAAAAAAAAAAAAAAAAAACTCCACTTGTGCTCCACTTGGTG TTGCACCCTCCACTTGTTCTTCCCTATAGTGAGTCGTATTA
Riboprobe_CA20_fwd	TAATACGACTCACTATAGGGGTGTGTGTGTGTGTGTGTGTGTGTGT GTGTGGCTTACTTTTTTTTTTTTTTTTTTTTTT
Riboprobe_CA20_rev	AAAAAAAAAAAAAAAAAAAAAAAAGTAAGCCACACACACACACAC ACACACACACACACACCCCTATAGTGAGTCGTATTA

2.2 Methods

2.2.1 DNA cloning in E. coli

2.2.1.1 Preparation of plasmid DNA

Plasmid DNA for transformation was extracted from an overnight bacterial culture using QIAprep spin miniprep kit or QIAGEN plasmid maxi kit. Plasmid DNA was prepared according to manufacturer's instructions. Concentration of the plasmid extracted was determined by UV light absorption at 260nm using a spectrophotometer (Nanodrop1000, Thermo Scientific). Pure DNA should have an $OD_{260/280}$ ratio of approximately 1.8.

2.2.1.2 Agarose gel electrophoresis

0.8-2% agarose gels were routinely used for analysing DNA. Agarose was melted in 1xTBE buffer (100mM boric acid, 100mM Tris and 2mM EDTA pH8.8) in a microwave oven and cooled to 55°C. Before pouring into the casting tray, ethidium bromide was added to the gel in a 1:20,000 dilution. DNA samples were mixed with appropriate amount of 6x loading buffer (30% (v/v) glycerol, 0.025% (w/v) bromophenol blue) for loading into the wells. Typically, the voltage for electrophoresis was set at 10V/cm from a power supply (electrophoresis apparatus, BioRad). DNA was visualised over a UV transilluminator, using a gel documentation system (SynGene).

2.2.1.3 DNA extraction from agarose gels

Required DNA bands were excised from agarose gels with a scalpel, transferred to Eppendorf vials, weighed and subsequently purified using QIAprep / MinElute gel extraction kit (QIAGEN) as described by the manufacturer.

2.2.1.4 DNA cleavage using restriction enzymes

DNA was digested with appropriate restriction endonucleases (New England Biolabs) in a reaction mixture consisting of DNA sample, 10x restriction buffer, restriction enzyme and distilled deionised water (ddH_2O). The reaction mixture was incubated at the recommended temperature for 1-2h. Usually, 1 to 5 units of enzyme was used to digest 1 µg of DNA. Digested DNA was purified by gel extraction or phenolisation.

2.2.1.5 Dephosphorylation of DNA

In order to prevent self-ligation of the linearised vector DNA, the 5'-termini of the vector DNA were dephosphorylated using SAP. DNA sample was mixed with 1µl of SAP (1U/µl) and 2µl of 10x SAP buffer in a 20 µl reaction. The reaction mixture was incubated at 37°C for 1h. SAP can be inactivated either by heat treatment at 65°C for 15 min or by phenolisation.

2.2.1.6 DNA ligation

Linearised DNA vector and purified DNA fragment were ligated in molar ratios of 1:5 and 1:7, respectively, ensuring a molar excess of insert over vector. The reaction mixture consisted of 1x T4 DNA ligation buffer, 1µl T4 DNA ligase (400U/µl, New England Biolabs) and distilled deionised water (ddH_2O) in a total volume of 20µl. The reaction was incubated at 16°C overnight.

2.2.1.7 Transformation of *E.coli* cells

10ng of plasmid DNA or 2-5µl of the ligation mixture was mixed with 50µl rapidly thawed high-competent Top10/BL21 *E.coli* cells by gently tapping on the tube and incubated on ice for 30min. The cells were then subjected to a quick heat shock at 42°C for 30sec and placed on ice immediately. 200µl of SOC medium (Invitrogen) was added to the cells to recover and grown at 37°C for 1h on a shaker. Afterwards, the cells were plated on pre-warmed LB-agar plates supplemented with 100µg/ml ampicillin. The plates were incubated overnight at 37°C. Single colonies were picked and grown in 3 ml of LB medium (1% tryptone, 0.5% yeast extract, and 1% NaCl) with 100µg/ml ampicillin. Plasmids isolated from these colonies were analysed by colony PCR and digestion with restriction endonucleases and verified by running on agarose gel. Finally, the positive clones were confirmed by Sanger Sequencing (SeqLab, Microsynth). If blue/white screening was carried out for selecting recombinants, 40µl of X-gal (40mg/ml) and 40µl of IPTG solution (1M) was spread on LB plates containing appropriate antibiotics before the transformation cultures were plated.

2.2.2 Generation of hnRNP L subdomains

The full-length hnRNP L gene sequence (1829 bp) was ordered as codon-optimised DNA in a pMA-RQ-(AmpR) plasmid backbone (ThermoFisher), with additional 6xHis-tag and TEV-cleavage site. Full-length hnRNP L was re-cloned (*Eco*RI, *Xho*I) into the pGEX-6P2 expression vector (GE Healthcare). Similarly, truncations of hnRNP L - RRM1, RRM2, RRM1-2 and RRM3-4 (with 6xHis-tag and TEV-cleavage site) were generated by PCR from full-length plasmid template, purified and cloned (*Eco*RI, *Xho*I) next to GST-tag in the pGEX-6P2 expression plasmid. (See **Supplementary Figure S1.A-D**)

2.2.2.1 PCR amplification with specific primers

The hnRNP L open reading frame from pGEX-6P2 plasmid served as template to generate truncation versions of hnRNP L protein –RRM1, RRM2, RRM1-2 and RRM3-4 (See **Supplementary Figure S1.E-F**). Primers specific to each domain were used to amplify each domain by PCR according to the following conditions.

10.0 µl	5x buffer (HF)
1.0 µl	dNTPs
2.5 µl	Primer mix (10µM)
0.5 µl	Phusion Polymerase
2.0 µl	1ng template DNA
34.0 µl	deionised water
50.0 µl	total volume

PCR temperature profile:

initial denaturation	98°C	30s	
denaturation	98°C	15s	30 cycles
annealing	61-65°C	30s	
elongation	72°C	20s	
final elongation	72°C	7min	
storage	4°C		

Primer combinations used for amplification of individual domains are as follows (see *Materials* **section 2.1.12**);

hnRNP L	Primer combination
RRM1	hnRNPL_RRM1-2_fwd (001) hnRNPL_RRM1_rev (001')
RRM2	hnRNPL_RRM2_fwd (002) hnRNPL_RRM1-2_rev (002')
RRM1-2	hnRNPL_RRM1-2_fwd (001) hnRNPL_RRM1-2_rev (002')
RRM3-4	hnRNPL_RRM3-4_fwd (003) hnRNPL_RRM3-4_rev (003')

2.2.2.2 Gel extraction and purification

PCR amplified DNA was run on a 0.8% agarose gel and stained with ethidium bromide. DNA bands corresponding to the respective subdomains of hnRNP L were excised from the gel under low UV-trans illumination with a clean scalpel and transferred to sterile eppendorf vials. The weight of the gel slice is determined and 3-fold volume of QC buffer (Qiagen) is added to the gel slice and incubated at 55°C until the gel dissolves completely. DNA thus obtained was purified using MinElute columns (Qiagen) according to manufacturer's instructions.

2.2.3 Expression and purification of proteins in *E.coli*

Expressions constructs of GST-RBM24, GST-hnRNP L-TEV-His fusion protein and its truncations, were transformed into *E. coli BL21* strain. Single colony from each transformation was then inoculated into 3-5ml LB medium and grown overnight at 37°C to make the pre-culture. 200ml fresh LB medium containing ampicillin (100 µg/ml) was inoculated with 1% (2mL) overnight pre-culture, and incubated at 37°C with shaking until OD_{600} = 0.6 is reached. For protein expression, cells were induced with 1mM IPTG and incubated on a shaker at

25/16°C. After induction, cells were harvested by centrifugation at 400rpm for 15min. Recombinant proteins thus expressed were purified either using GST-tag or the 6x-His tag.

2.2.3.1 GST Tagged RBM24 and hnRNP L

Harvested bacterial cell-pellet was subjected to one freeze/thaw cycle at -196°C in liquid nitrogen and then resuspended in 2-5ml/g wet-weight GST-lysis buffer (1xPBS pH7.4; 0.5mM DTT; 0.05% NP40; protease inhibitor cocktail, Roche). Cells were lysed by sonication after incubation with lysozyme (1mg/ml) on ice for 30min with gentle vortex every 10min. Cell suspension was sonicated on ice, thrice by pulses of 30s at 50-60% amplitude (Branson sonifier model B-15). RNAse A (10µg/ml) and DNase I (5µg/ml) were added to the cell suspension to remove contaminating nucleic acids, incubated on ice for 15min and the reaction was stopped by adding 25mM $MgCl_2$. Lysate was separated from cellular debris by centrifugation at 14000rpm for 30min at 4°C. Additionally, lysate was filtered through a 0.45 µm syringe filter. The lysate was incubated for 30min at RT on a wheel with rotation, with 500µl pbv (packed bead volume) glutathione beads (Gluthathione-Sepharose 4B, GE Healthcare) pre-incubated in 1 ml cold GST-lysis buffer. Beads were washed three times with 1ml cold GST-lysis buffer and once with GST-wash buffer (1x PBS pH7.4; 2M NaCl; 5mM DTT; 2% Triton X-100; 10% Glycerol; protease inhibitor cocktail, Roche) on a rotating wheel at RT. Recombinant GST-proteins were eluted thrice with 100µl GST-elution buffer (50mM Tris-HCl, pH 8.0; 40mM reduced glutathione; 120mM NaCl; 0.5mM DTT; 0.05% NP40; protease inhibitor cocktail, Roche). Fractions containing recombinant GST-proteins were pooled, dialysed against Dialysis buffer (20mM Tris-HCl, pH7.9; 20% glycerol; 100mM KCl; 0.2mM EDTA, pH 8.0; 1mM DTT) and aliquots of each elution fraction were separated on 10% SDS-PAGE and stained by Coomassie Blue (see **section 2.2.5**). Protein concentration was estimated on Coomassie blue stained SDS-PAGE using BSA (Biolabs) as a standard.

In order to remove GST-tag for binding experiments, upto 2µg of fusion protein was incubated with 1 unit AcTEV Protease (stored in 50mM Tris-HCl, pH7.5; 1mM EDTA; 5mM DTT; 50% glycerol; 0.1% Triton X-100, Invitrogen), 0.1 M DTT and 20x TEV buffer (1M Tris-HCl, pH 8.0; 10mM EDTA, Invitrogen) at 30°C for 30min as follows:

20 µg	Fusion protein
7.5 µl	20x TEV Buffer
1.5 µl	0.1M DTT
1.0 µl	AcTEV Protease, (10 units)
to 150 µl	deionised water
150.0 µl	total volume

2.2.3.2 His Tagged hnRNP L subdomains

Recombinant hnRNP L_RRM1, hnRNP L_RRM2, hnRNP L_RRM1-2 and hnRNP L_RRM3-4 were purified through His-tag purification in a manner similar to GST-tag purification mentioned above. Induced bacterial cell pellet was freeze-thawed, once in liquid N_2, followed by lysis of the cells by lysozyme and sonication in 6x-His-lysis and wash buffer (50mM NaH_2PO_4; 2M

NaCl; 50mM Imidazole; 10mM β-mercaptoethanol; 10% glycerin; 0.05% Tween; pH adjusted to 8.0 with NaOH; protease inhibitor cocktail, Roche). Lysate separated from cell debris is incubated with 1 ml pbv Ni-NTA agarose beads equilibrated in 6xHis lysis and wash buffer. The mixture was incubated for 1 h at 4°C to allow binding of the protein to the beads followed by 5 washes each with 10 ml 6xHis lysis and wash buffer. The beads were transferred to a chromatography column, 1 ml elution buffer (50 mM NaH_2PO_4; 300 mM NaCl; 250 mM Imidazole; pH adjusted to 8.0 with NaOH) was added and incubated for 30 min at 4°C. Fractions containing recombinant protein were collected in sterile vials and pooled for dialysis against Dialysis buffer (20 mM Tris-HCl, pH 7.9; 20% glycerol; 100 mM KCl; 0.2 mM EDTA, pH 8.0; 1 mM DTT). Purified recombinant proteins were stored at -80°C and aliquots of each elution fraction were separated on 10% SDS-PAGE and stained by Coomassie Blue (see *Methods* **section 2.2.5**). Protein concentration was estimated on Coomassie blue-stained SDS-PAGE using BSA (Biolabs) as a standard.

2.2.4 Generation and purification of recombinant baculovirus expressed His-tagged hnRNP L

Recombinant baculoviruses have been widely used as vectors to express heterologous genes in cultured insect cells. Heterologous genes placed under the control of the strong polyhedron promoter of the *Autographa californica* nuclear polyhedrosis virus (AcNPV) are often expressed at high levels of properly post-translationally modified, biologically active and functional recombinant proteins during the late stages of infection. Baculovirus expression system is based upon the ability to propagate AcMNPV (*Autographa californica* multiple nuclear polyhedrosis virus) in insect cells. Baculoviruses are safer to work with, than other mammalian viruses because they are non-infectious to vertebrates (Inceoglu *et al*., 2001). Recombinant baculovirus expressing His-tagged hnRNP L codon-optimised sequence was constructed with the Bac-to-Bac baculovirus expression system (Thermo Fisher Scientific, USA) according to the manufacturer's protocol. It is a rapid and efficient system that relies on site-specific transposition.

Fig. 2.1 summarises the principle behind the Bac-to-Bac baculovirus expression system. In order to generate recombinant baculovirus expressing hnRNP L, a full-length codon-optimised cDNA encoding hnRNP L was cloned into pFastBac HTb donor plasmid, which contains the mini-Tn7 elements. The recombinant plasmid pFast-hnRNP L-His was transformed into DH10Bac competent cells containing a bacmid with a mini-*att*Tn7 target site and a helper plasmid. The mini-Tn7 elements on the pFast-hnRNP L-His were transposed to the mini-*att*Tn7 target site on the bacmid by the transposase encoded by the helper plasmid. Transposition of the cloned gene into the bacmid disrupted the expression of the *lac*Zα gene. White colonies containing recombinant bacmids were selected by antibiotic selection and blue/white screening and amplified. High molecular weight DNA prepared from these *E. coli* clones was used to transfect insect cells. Recombinant viruses were harvested from the transfected insect cells.

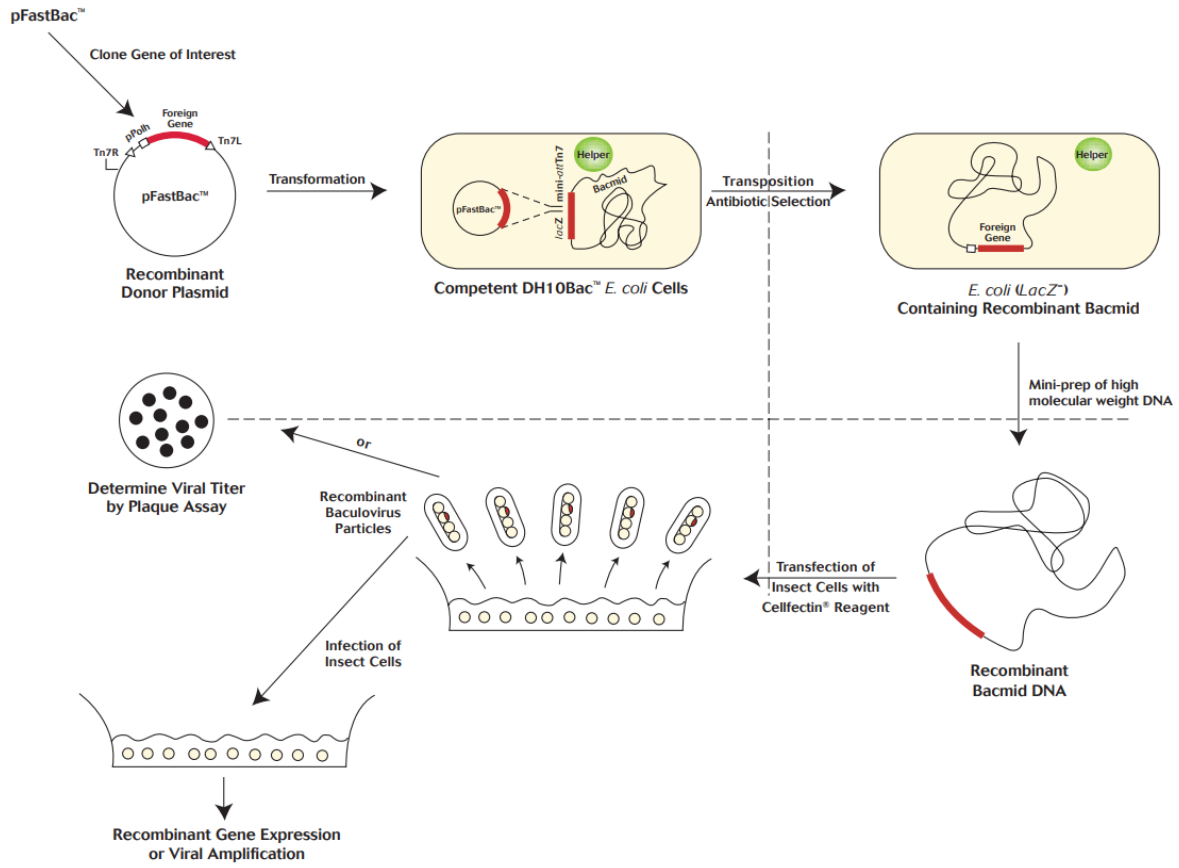

Fig. 2.1 Generation of recombinant baculovirus. pPolh represents polyhedron promoter of AcNPV, Tn7R and Tn7L indicate transposon 7 elements, and mini-*att*Tn7 refers to a short segment containing the attachment site for Tn7 (adapted from Invitrogen instruction manual).

2.2.4.1 Infection of insect cells

1x10^7 Sf21 insect cells were seeded in a 75cm^2 flask containing SF-900 II medium with 10% FBS and incubated at 27°C for 1h. After the cells were attached to the flask, they were infected with 25µl of viral stock [1x10^7 plaque forming units (pfu)/ml] at a multiplicity of infection (MOI) of 0.1. Four days after infection, cells were harvested by centrifugation for 10 min at 500x g at 4°C.

2.2.4.2 Purification of His-tagged hnRNP L from baculovirus infected insect cells

Cell pellet from one 75 cm^2 flask (1 × 10^7 cells) was resuspended in 4 ml of lysis buffer (50 mM NaH_2PO_4; 2 M NaCl; 50 mM Imidazole; 10 mM β-mercaptoethanol; 10% Glycerin; 0.05% Tween; pH adjusted to 8.0 with NaOH; protease inhibitor cocktail, Roche) after freeze-thawing on dry-ice and incubated on ice for 10 min. Cell lysate was passed through a 25 G needle connected to a 1 ml syringe 5 times, followed by centrifugation at 14000 rpm at 4°C for 10 min to pellet cellular debris and DNA. 50 µl of packed Ni-NTA beads was added to cleared lasyte and incubated at 4°C for 2 h with rotating. The bound Ni-NTA beads were pelleted by centrifugation at 6,000 rpm for 1 min and washed with 1 ml of wash buffer (50 mM NaH_2PO_4; 2 M NaCl; 50 mM Imidazole; 10 mM β-mercaptoethanol; 10% Glycerin; 0.05%

Tween; pH adjusted to 8.0 with NaOH; protease inhibitor cocktail, Roche) four times. His-tagged protein was eluted with 50 µl of elution buffer (50 mM NaH_2PO_4; 300 mM NaCl; 250 mM Imidazole; pH adjusted to 8.0 with NaOH) four times. The eluates were collected and analysed by SDS-PAGE followed by Western blot to detect hnRNP L using anti-hnRNP L monoclonal antibody (4D11).

2.2.5 SDS polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining

Proteins were resolved on denaturing SDS polyacrylamide gels. Two buffers were used: 4x stacking gel buffer (0.5 M Tris-HCl pH 6.8; 0.4% (w/v) SDS) and 4x separating buffer (1.5 M Tris-HCl pH 8.8; 0.4% (w/v) SDS) to prepare either stacking gel (5% acrylamide/bisacrylamide 37.5:1; 1x stacking gel buffer; 80 µl APS; 8 µl TEMED in 8 ml) or separating gels (10-12% Acrylamide/bisacrylamide 37.5:1; 1x separating gel buffer; 40 µl APS; 4 µl TEMED in 4 ml) respectively. The proteins were mixed with 2x Laemmli loading buffer (4% SDS; 20% glycerol; 100 mM Tris-HCl pH 6.8; 0.2% BPB, 5% β-mercaptoethanol), denatured at 100°C for 10 min and loaded on the gel. Electrophoresis was carried out at 300 V, 35 Amp in a BioRad gel running tank containing SDS-gel running buffer (25 mM Tris; 250 mM glycine pH 8.3; 0.1% (w/v) SDS). Gels were subjected to either Western blot analysis or Coomassie blue staining. For Coomassie staining, gels were incubated with InstaBlue coomassie stain (Expedeon) overnight at 4°C or at room temperature for 15min with shaking after heating up the solution for 10-15s in a microwave oven. The gel was dried onto a Whatmann filter paper using a gel dryer (SE1160, Hoefer Scientific instruments, USA) which is connected to a vacuum pump (Vacuubrand, Germany) and at 60°C for 1h.

2.2.6 Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

2.2.6.1 SELEX selection of binding motifs for RBM24 and hnRNP L

For a detailed explanation of the steps in SELEX selection, see Schneider *et al*., 2016.

An RNA pool with a degenerate sequence of 40 nucleotides was generated by T7 transcription, using an annealed template of SLX-N_{40} and T7-fwd oligonucleotides (T7 High-Yield Kit, NEB). This random pool comprises the 'sequence space' within which binding preferences of RBM24, hnRNP L and variants was tested. 40pmol of RBM24 and GST-hnRNP L/truncated variants were incubated with 4nmol of SLX-N_{40} transcript over four rounds of selection. Selection of RNAs using GST alone, served as negative control. SELEX selections were carried out with the fusion proteins bound to pre-blocked glutathione-sepharose beads (GE healthcare) in total volume of 200µl buffer (10mM Tris-HCl, pH 7.5; 100mM KCl; 2.5mM $MgCl_2$; 0.1% Triton X-100; Roche protease inhibitor). After 30min incubation at room temperature, samples were washed thrice with 1ml washing buffer (10mM Tris-HCl, pH 7.5; 100/300/600mM KCl; 2.5mM $MgCl_2$; 0.1% Triton X-100). Samples were then treated with proteinase K (Roth), RNA was extracted using phenol/chloroform extraction and ethanol precipitation. The stringency of washing steps increased gradually during the rounds of selection (R1: 3x 100mM; R2: 2x 100mM, 1x 300mM; R3: 1x 100mM, 2x 300mM; R4: 1x 300mM and 2x 600mM KCl washing buffer).

2.2.6.2 Library preparation

Selected RNAs were reverse transcribed (qScript Flex cDNA Synthesis Kit, Quanta) with SLX_RT reverse primer. cDNA libraries were amplified by PCR (16 cycles; SLX_Sol-5xN_fwd and SLX_Sol_rev). Transcripts for the next round of selection were produced by *in vitro* transcription. After the four rounds of selection RNA aliquots from each round and from the fourth round of GST selection were used for barcoding by reverse transcription with the SLX_RX reverse primers. The final library pool was subjected to high-throughput sequencing on a Miseq platform (single-read 150 bp, Illumina). PhiX control library was added to increase the sample complexity (Illumina). Primer sequences are listed in *Materials* **section 2.1.12**.

2.2.6.3 SELEX-seq data analysis (motif enrichment and spacing analysis)

SELEX-seq data analysis was done by Lee-Hsueh Hung, a member of AG Prof. Bindereif's lab. Methods are described in detail by Schneider *et al.* (2019), and applied here for the analysis of RBM24 and hnRNP L.

2.2.7 RNA design based on SELEX-derived data

Highly specific SELEX-derived RNAs were designed based on motif enrichment studies, spacing analysis and secondary structure prediction (see *Methods* **sections 2.2.6** and **2.2.14.2**). For RBM 24, the RNA was designed such that it contains the binding motif UGUU. On the other hand, four RNAs with binding motifs in the order- ACAC, ACAC, ACAU and ACAU for binding to the four respective RRM domains were designed for hnRNP L. Binding motifs of these RNAs differed from one another in their spacing between the individual domains i.e., each of the 4 RRM domains (9-12nt) and between the di-domains i.e., between RRM1-2 and RRM3-4 (10-15nt). The RNAs were labelled L_9/10, L_9/15, L_12/10 and L_12/15 where the first number refers to spacing between motifs for individual domains and the second number represents spacing between motifs for the di-domains of hnRNP L (represented by L). In each case, for mutant control, ACAU was mutated to UGUG and ACAT to UGUU. Mutants were labelled as L_9/10_mut, L_9/15_mut, L_12/10_mut and L_12/15_mut. These served as controls for binding specificity in band-shift assays and pulldown assays.

2.2.8 *In vitro* RNA Transcription

2.2.8.1 Annealing of oligonucleotides

Commercially obtained (Sigma-Aldrich/Merck) complementary oligonucleotides were mixed at the same molar concentrations in 2x Annealing Buffer (20mM Tris-HCl, pH 7.5; 50mM NaCl, 1mM EDTA) to a final volume of 50µl in a PCR tube. The tube was placed in a thermocycler, which was programmed to initial denaturation at 95°C for 2-5min, and then gradually cooled down to room temperature at the rate of 1°C per minute. Successful annealing can be verified by running about 200ng of non-annealed and annealed oligonucleotides on a 2% agarose gel. When compared to the individual oligonucleotide pairs, the annealed oligonucleotides slightly shift upward on the gel, indicating successful annealing.

2.2.8.2 Transcription of ^{32}P-labelled RNA

RNAs were radioactively-labelled internally by T7 *in vitro* transcription. Annealed oligos served as templates for transcription. 5µl of template DNA (200ng/µl) was mixed with 5µl 5x transcription buffer (Thermo Scientific), 2.5µl 100mM DTT; 1.25µl 10mM ATP; 1.25µl 10mM CTP; 0.5µl 2mM UTP; 1.25µl 10mM GTP; 1µl RNase inhibitor (RNase OUT); 1 µl [α-^{32}P]UTP (800Ci/mmol) and 1µl T7 RNA polymerase (20U/µl). DMPC-H_2O was added to the final volume of 25µl. In case of high-yield transcriptions the respective 10x reaction buffer and polymerase were used (NEB). Additionally, to facilitate *in vitro* circularisation, the 5' end of the transcripts were primed with 40mM GMP. Transcriptions were carried out at 37°C for 2h. 2µl RQ1 DNase was added to each reaction and incubation continued for 30min at 37°C.

2.2.8.3 Removal of unincorporated nucleotides by gel filtration

Unincorporated nucleotides were removed from the transcription reactions using RNA spin columns following the manufacturer's instructions (Roche). Transcribed RNAs were precipitated with 600 µl 100% ethanol; 20µl 3M NaAc pH 5.2; and 1µl (20mg/ml) glycogen. RNA was precipitated at -80°C for at least 20 min and then centrifuged at >14000x g for 30 min at 4°C.

2.2.8.4 Determination of RNA concentration

After pelleting, washing twice with 70% ethanol and drying, the amount of RNA transcribed was measured using a scintillation counter (1600TR, Packard) and calculated using the following formula:

RNA [ng] = CTP cold [µM] x volume of reaction [µl] x %incorporation x 0.0132

The transcripts were dissolved in an appropriate volume of DMPC-H_2O.

2.2.8.5 Gel extraction of radiolabelled RNA

RNA transcripts were separated by denaturing polyacrylamide gel (10% (v/v) Acrylamide/Bis (1x TBE, 50% urea); 10% (v/v) APS; 0.001 % (v/v) TEMED) electrophoresis. The RNA bands were cut out of the gel and shredded into tiny pieces by forcing through a tiny hole pierced at the bottom of the tube and collecting by centrifugation into another tube. RNAs were eluted from the gel by adding 300 µl of 2x PK buffer (200mM Tris-HCl pH8.0, 300mM NaCl, 25mM EDTA, 2% SDS) overnight at room temperature with rotating. The eluate was phenolised with Roti-phenol/chloroform to remove the enzymes and buffers and precipitated with 1µl of glycogen (20mg/ml), 900µl of 100% ethanol and 30µl 3M NaAc pH5.2. After washing and air-drying, the purified RNAs were dissolved in DMPC-treated H_2O.

2.2.8.6 Transcription without ^{32}P labelling

For transcription reactions without the ^{32}P-label, 100mM UTP was added to the transcription reaction described above instead of [α- ^{32}P] UTP. RNAs were synthesised by *in vitro* transcription, using double-stranded DNA-oligonucleotide templates and HiScribe T7 High Yield RNA synthesis kit (NEB), followed by RQ1 DNase treatment (10U per reaction). Transcribed RNAs were extracted with 200µl phenol/chloroform/isoamyl alcohol (25:24:1)

and then precipitated. After pelleting, washing, and drying, transcripts were dissolved in 20µl of DMPC-H_2O. The concentration of RNA was estimated spectroscopically by Qubit measurement (Invitrogen).

2.2.9 Electrophoretic mobility shift assay (Band shift)

50pmol of ^{32}P-labelled substrate RNAs were incubated with different concentrations of recombinant RBM24 and hnRNP L protein (20, 40, 80, 160, 320 or 640nM) in a 10x binding buffer containing 20mM Tris-HCl pH 7.9; 100mM KCl; 1mM DTT; 0.2mM EDTA pH8.0; 20% glycerol; 1µg tRNA; 1µg BSA and RNase inhibitor (RNase OUT). The total reaction volume was 10µl. The mixtures were incubated at 25°C for 30min. Samples are kept on ice until they are fractionated on a native RNA gel. 1µl of native RNA gel loading buffer (0.025% (w/v) bromophenol blue; 30% (v/v) glycerol) was added and the samples were fractionated on a 5% native RNA gel (30% acrylamide/bisacrylamide 37.5:1; 1xTBE; 400µl 10% APS; 40µl TEMED filled up with H_2O to 50ml). Gel is pre-run at 300V for 30min. After samples are loaded, gel is run for 1h40min at 300V. Pre-run, loading and gel-run are all carried out in a 4°C cooling chamber. Samples were visualised by collecting the radioactive signals using a phosphorescence imaging plate (Fujifilm, Japan) placed on the gel for 3h. The phosphor screen is then placed in a phosphorescence imaging (PI) scanner (Typhoon FLA9500, GE, USA) which detects the phosphorescence and images are analysed using ImageQuant software.

2.2.10 HnRNP L sponging assays

2.2.10.1 *In vitro* pulldown assay with biotinylated RNAs

SELEX-derived RNAs for hnRNP L sponging were produced by T7-transcription (T7 High-Yield Kit, NEB) from annealed oligo cassettes and chemically modified at the 3'-biotinylation (Wilkomm & Hartmann, 2005). For pulldown of endogenous hnRNP L from HeLa cell lysate and nuclear extract, 1x10^6 cells (and equivalent amount of nuclear extract) were lysed in RIPA lysis buffer [50mM Tris-HCl, pH 7.4; 150mM NaCl; 5mM EDTA; 1% NP40 (v/v), 0.1% SDS] and incubated with 50pmol of 3' biotinylated RNA bound to MyOne Streptavidin C1 Dynabeads (binding capacity ~5000pmol/ml packed beads; Thermo Fisher Scientific) in a total volume of 200µl for 30min at room temperature. After three washing steps with wash buffer (1x WB100, 2x WB 300; 10mM Tris-HCl pH7.4; 100-600 KCl; 2.5mM $MgCl_2$; 0.1% Triton X-100), bound protein was released in SDS-sample buffer (50mM Tris-HCl pH 6.8; 2% SDS; 10% Glycerol; 2.5% 2-mercaptoethanol; 0.05% bromophenol blue) and heat denaturation at 95°C for 10 min. Samples were analysed by SDS-PAGE followed by Western blotting with hnRNP L monoclonal antibody [(4D11), see *Methods* **section 2.2.11**]

2.2.10.2 RNA-immunoprecipitation (RIP)

To demonstrate sponging of hnRNP L *in vivo*, HeLa cell lysate was extracted after transfection with respective sponges. 2 µg of each antibody (anti-hnRNP L monoclonal and anti-Flag, Sigma Aldrich) was added to pre-cleared lysate. 240 µl of pre-cleared cell lysate corresponding to 4.5x10^6 HeLa cells was incubated with the antibodies overnight at 4°C. 30 µl (1 mg) Protein - A - Dynabeads (Thermo Fisher Scientific) was added to the mixture and incubated for 2 h at 4°C with rotation. Protein-RNA complexes were washed with increasing stringency of salt concentration (1x 150mM; 2x 600mM; 1x 150mM NaCl). RNA from input and

immunoprecipitated fractions was extracted by TRIzol (Ambion) followed by reverse transcription (qScript cDNA Synthesis Kit, Quanta) and (q)PCR with gene and circular-specific primers. *In vitro* circularised L_12/10, L_12/10_mut and $(CA)_{20}$ RNAs were detected by Northern blotting. The fraction of the bound target RNAs was calculated for each target relative to the corresponding input fraction.

2.2.11 Western Blotting

Proteins were resolved on 10 % SDS-PAGE and transferred to PVDF nitrocellulose membrane (GE Healthcare) in transfer buffer (50mM Tris; 380mM glycine; 20% (v/v) methanol; 0,02% (w/v) SDS) for 30min at 300mA using a semi-dry transfer cell (Bio-Rad). The membrane was blocked for 1h in blocking buffer (5% Milk powder (Roth) in 1x PBS and 0.1% (v/v) Tween) at RT. The primary antibodies were diluted in fresh blocking buffer as follows: monoclonal anti-hnRNP L (4D11) 1:10000 (Sigma-Aldrich); mouse monoclonal anti-GAPDH 1:5000 (Sigma-Aldrich); anti-GST 1:5000 (Pharmacia biotech) and anti-hnRNP A1 1:1000 (Sigma-Aldrich) and added to the membrane and incubated overnight with shaking at 4°C. The membrane was washed three times for 10min in 1 x PBST [1xPBS (Gibco) and 0.1% (v/v) Tween] and incubated for 1h at RT with peroxidase conjugated secondary antibodies diluted 1:10000 (Sigma-Aldrich) in blocking buffer. The membrane was subsequently washed three times for 10min in PBST and the bound antibodies were detected by the Lumi-Light ECL system (Roche) according to manufacturer's protocol. The membrane was then exposed to an X-ray film (GE Healthcare) and developed.

2.2.12 Northern Blotting

For a more detailed view of RNA detection by digoxigenin (DIG) Northern blot, see Schneider *et al.*, (2018). HeLa total RNA was denatured by boiling at 95°C for 3min in loading buffer containing Formamide. Total RNA (500ng) and DIG-labelled DNA ladder (DIG VIII, Roche) were separated on a 10% denaturing polyacrylamide gel then transferred to a nylon membrane (Amersham Hybond-N+, GE Healthcare) for 1h, at 3mA/cm^2 using Trans-Blot semi-dry transfer cell (Biorad). Nucleic acids transferred to the membrane were cross-linked at 120mJ/cm^2 at 254nm in BLX 254 UV-crosslinker (Biolink) and pre-hybridised with the NorthernMax hybridization buffer (Ambion) at 60°C for 1 h. Hybridization with the riboprobe was carried out overnight at 60°C with constant rotation using HB-1000 hybridization oven (UVP). The DIG-labelled riboprobe directed against the circ-junction (for SELEX-derived RNAs and CA_{20}) or the Tornado stem region including circ-junction (for tornado generated circRNAs) was prepared *in vitro* by T7- transcription with DIG RNA labeling mix (Roche) according to the manufacturer's instructions. The riboprobe was treated with RQ1 DNase to remove unspecific DNA molecules. After hybridization, the membrane was washed with 2X SSC [30mM sodium citrate pH7.0, 0.1% (w/v) SDS, and 300 mM NaCl] and 0.5X SSC. The membrane was blocked for 1h with a DIG-blocking solution [2% (w/v) blocking reagent (Roche) in 100mM maleic acid buffer pH7.5]. The membrane was then incubated with anti-DIG-Fab (Roche) with a final dilution of 1:10000 for 1h. The excess antibody was washed with DIG-washing buffer [0.3% (v/v) Tween-20 in 100mM maleic acid buffer pH 7.5]. Northern blots were developed using the DIG-detection system, namely 0.5% CDP-star substrate solution (Roche) in DIG-detection buffer (100mM Tris-HCl pH9.5 and 100mM NaCl). The membranes were exposed to

Amersham Hyperfilm ECL (GE Healthcare) films for various time intervals. The films were developed on Agfa Curix 60 Processing Machine (AGFA).

2.2.13 *In vivo* splicing analysis

2.2.13.1 Cell culture

HeLa and HEK293 cells were cultured and maintained in DMEM supplemented with 10% FBS at 37°C in an incubator with 5% CO_2. When cells were 100% confluent (every 2-3 days), they were washed with 1x PBS, detached from the plate by 1x trypsin-EDTA, and reseeded into new dishes in a 1:10 ratio.

2.2.13.2 Transfection of circular expression constructs and circRNA in HeLa cells

One day before transfection, $1x10^6$ cells were seeded on a 10 cm dish such that cells were not more than 80-90% confluent at the time of transfection. 10 µg of plasmid DNA was used with appropriate amount (3µl/µg of DNA) of Lipofectamine 2000 transfection reagent (Invitrogen) according to the manufacturer's instructions. Plasmid DNA and transfection reagent were taken in 500 µl each of optiMEM (Gibco) medium in transfection tubes. They are mixed and incubated at room temperature for 20-30 min. After incubation the DNA-transfection mix is added drop-wise to the cells. Mock transfection was performed without adding the plasmid DNA. The medium was changed 4h after transfection. After 48h, cells were harvested with trypsin, collected in DMEM, and centrifuged at 500xg and 4°C for 5min. Cells were washed in 1x PBS and centrifuged as before. Cell pellets were subsequently used for RNA isolation, RNA immunoprecipitation, or cell fractionation as described below.

For *in vitro* circularised RNA transfection, $5x10^4$ cells were seeded per well in a 24-well plate one day before transfection. Gel-purified circRNA was transfected using Lipofectamine 2000 in optiMEM medium. CircRNA concentration ranged from 500ng to 2µg. Cells were harvested 24h after transfection and RNA was extracted.

2.2.13.3 Isolation of total RNA from HeLa cells

Cells were first washed with 1xPBS, collected and lysed with 450µl of TRIzol (Ambion). For tornado expressed RNA, total RNA was isolated using RNEasy kit (Qiagen). Additionally, the RNA was treated for removal of plasmid and DNA contaminants by adding 5 µl of RNase-free RQ1 DNase. For *in vitro* circularised RNA and smaller RNAs (<200nt) and after DNase treatment, total RNA was extracted by adding one volume of Phenol/Chloroform/Isoamyl alcohol (25:24:1) mix, vortex and centrifuge at 14,000xg for 5min. Transfer the aqueous upper layer into a new tube and add 3 volumes of 100% ethanol, 0.1 volume 3M sodium acetate (pH5.2) and 1µl glycogen. RNA is precipitated at -80°C for 20 min. Total RNAs were pelleted by centrifugation at 14,000rpm for 30min at 4°C, washed once with 1ml of 70% ethanol, dried and resuspended in 20 µl of DMPC-H_2O. The concentration of total RNA was determined by UV light absorption at 260nm (Nanodrop).

2.2.13.4 RNAi knockdown of hnRNP L in HeLa cells

HnRNP L protein was knocked down *in vivo* using synthetic siRNA against hnRNP L mRNA. SiRNA duplexes (hnRNP L and luciferase, see *Materials* **Table 2.1.12**) were mixed with

RNAiMAX (Invitrogen) at a final concentration of 20 µM and added to a 10 cm culture dish and incubated at room temperature for 15-25min in OptiMEM medium (Gibco). Seeding density of HeLa cells was ~$8.8x10^5$ cells/dish. Three days after transfection, total RNA was isolated using TRIzol and RNeasy columns (QIAGEN). Total RNA was reverse-transcribed using qScript Flex kit (Quantabio) with oligo(dT) primers, followed by PCR with the primers against hnRNP L alternative splicing targets.

2.2.13.5 Analysis of *in vivo* splicing by RT-PCR

Reverse transcriptions were carried out by qScript Flex cDNA synthesis kit (Quanta Biosciences). 1µg of total RNA was incubated at 65°C for 5min with 2µl each of random hexamers and oligo(dT) primers. Reaction was immediately chilled on ice. The RT reaction containing 4µl of 5x qScript reaction mix and 1µl qScript reverse transcriptase to a final volume of 20µl was added to each reaction. Reverse transcription reaction was carried out as follows 10min at 25°C for annealing of random primers, 65-90min at 42°C, 5min at 85°C and hold at 4°C.

PCR was done in 50µl reaction with 2µl RT reaction (from above), 1xPCR buffer, 400µl dNTPs, 1mM $MgCl_2$, 10µM each of gene specific forward and reverse primers, 0.2U/µl *Taq* polymerase. The following amplification profile was applied: 2min denaturation at 95°C, 30 cycles of amplification (15s at 95°C, 20s at 58°C and 30s at 72°C), and a final elongation step at 72°C for 7 min. 10µl of PCR reaction was mixed with 6XDNA loading buffer and analysed on a 2% agarose gel by ethidium bromide staining. Stained bands were quantified using ImageJ software.

2.2.14 Design and expression of circRNA

For *in vitro* circularisation, a stem sequence was incorporated into all SELEX-derived RNAs (L_9/10, L_9/15, L_12/15, L_12/10 and mutant) with an additional flex region of 5-7nt to enhance secondary structure of circRNA. T7-OR-$(CA)_{20}$ and T7-OR- CA-SELEX x4 contain the stem sequence as described in Jost *et al.* (2018).

2.2.14.1 Overexpression of circRNA from Tornado constructs

The Tornado overexpression vector system was used for overexpression of endogenous circRNAs (GSE1, Plt-circ-R4 and CiRS7/Cdr1as) and SELEX-derived RNAs. The primary Tornado vectors pAV-U6+27-Tornado-Broccoli and pAV-U6+27-Tornado-F30- Broccoli described by Litke and Jaffrey (2019) were ordered from Addgene. Both vectors are based on pAV vector with the U6+27 promoter [that includes the first 27 nucleotides of U6 RNA described previously (Lu *et al*., 2015)]. The Tornado circRNA expression cassette includes 49 bp sequence of Broccoli aptamer for fluorescent imaging with DFHBI (Filonov *et al*., 2014) flanked by ribozyme combinations of upstream Twister P1 and downstream Twister P3 U2A, and tRNA:Tyr intronic sequences that are cleaved by tRNA-specific endonuclease and subsequent ligation of RNA by RtcB ligase. pAV-U6+27-Tornado-F30-Broccoli vector was designed identically, but also contains a F30 three-way junction with Broccoli on one arm as described by Litke & Jaffrey (2019). Plasmids based on pAV-U6+27-Tornado-Broccoli vector were obtained by cloning directly into the Tornado expression cassette at NotI and SacII restriction sites, replacing the Broccoli aptamer sequence. Inserts comprising $(CA)_{20}$, CA-SELEX X4,

L_9/10, L_9/15, L_12/15, L_12/10 and mutant with the overhangs corresponding to restriction sites were chemically synthesised as individual oligonucleotides (Table 2.2) and annealed prior to ligation with linearised by *Not*I and *Sac*II vector. Obtained plasmids were named Tornado (CA) $_{20}$, Tornado CA-SELEX X4, T-L_9/10, T-L_9/15, T-L_12/15, T-L_12/10 and T-L_12/10_mut respectively. pAV-U6+27-Tornado-F30-Broccoli vector was used for the generation of Tornado CA-SELEX X4 Broccoli construct by Anna Didio (Schreiner *et al*., 2020). Chemically synthesised CA-SELEX X4 oligonucleotides flanked with *Kfl*I overhangs were annealed and cloned into *Kfl*I-linearised plasmid.

2.2.14.2 RNA secondary structure prediction

Secondary structure of RNA was predicted using the ViennaRNA package from RNAfold webserver (*http://rna.tbi.univie.ac.at*) implementing minimum free energy fold algorithm. For circRNA, the circular configuration of the RNA was taken into consideration for secondary structure prediction.

2.2.15 Cell fractionation

1x10^6 HeLa cells, corresponding to 10µl packed cell volume (PCV), was used to extract the cytoplasmic and nuclear fractions after transfection with circRNA expression constructs and *in vitro* circularised RNA. Extraction was performed according to manufacturer's instructions using NE-PER cytoplasmic and nuclear extraction reagents (Thermo Scientific) with protease inhibitor cocktail (Thermo Scientific). Cytoplasmic and nuclear extracts were lysed with TRIzol-LC (Ambion) for RNA isolation. Total RNA was isolated using RNeasy kit (Qiagen). Tornado expressed circRNAs were detected in both fractions with a vector specific probe by Northern blotting. To check protein distribution, equivalent lysate amounts of total cells, cytoplasmic and nuclear fractions were analysed by Western blotting for hnRNP L, and, as controls and for normalisation, for GAPDH and hnRNP A1. Band intensities in Western blot were quantified by densitometry using ImageJ software. HnRNP L content in each fraction was quantified as the percent of total: nuclear and cytoplasmic hnRNP L.

2.2.16 *In vitro* transcription and circularisation of RNA

To demonstrate alternative splicing changes due to protein sponging, designer sponge RNAs were synthesised by *in vitro* transcription using HiScribe T7 High Yield RNA synthesis kit (NEB) and DNA-oligonucleotides as template (see *Materials* **Table 2.2.12**). Transcription reaction mix contains ATP, CTP, GTP and UTP to a final concentration of 10mM in 20µl reaction. Additionally, to enhance circularisation, GMP is added to a final concentration of 40mM. Transcription was carried out at 37°C for 2h. Transcribed RNA was treated with RQ1 DNase (10U per reaction) for 30 min at 37°C to digest the DNA template. After DNase treatment RNA was purified using Monarch RNA cleanup kit. For circularisation, T4 RNA ligase (1.5U/µg RNA, Thermo Scientific) was used. To eliminate the linear precursor, RNA was treated with RNase R (10U/µg RNA). After each step, RNA was purified with Monarch RNA cleanup kit and visualised by denaturing polyacrylamide electrophoresis.

2.2.17 Testing circRNA integrity by RNase R treatment

HeLa total RNA or *in vitro* circularised RNA (500 ng) was treated with RNase R (2.5U/µg, Epicentre) for 30 min at 37°C and digested RNA was analysed on a 10% polyacrylamide gel.

As control for RNase R digestion, linear counterpart of the respective circular RNA was also tested.

3. Results

The following sections explain experiments that determine the designing and development of optimal circRNA sponges for two RBPs that regulate alternative splicing – hnRNP L and RBM24. Furthermore, designer sponges for hnRNP L have been characterised and functionally validated. However, functional characterisation of RBM24 sponges has not been documented here, since it is out of the scope of this thesis.

3.1 Analysis of RNA binding for hnRNP L and its subdomains

3.1.1 SELEX-based analysis of RNA binding for hnRNP L and subdomains

Our primary challenge in designing optimal circRNA sponges was, first, to understand the exact binding sequence of hnRNP L. This was addressed using SELEX-based selection of RNA elements that bind to the multidomain RNA-binding protein, hnRNP L. Secondly, we designed circRNA sponges based on binding specificity for hnRNP L and finally, we demonstrated alternative splicing variations due to hnRNP L sponging.

The multidomain RNA-binding protein hnRNP L recognises its cognate RNA through its four RRM domains (see **Figure 3.1** and **Supplementary Figure S1.F**). All four RRMs are involved in recognition and binding. Full-length (FL) and truncated hnRNP L were expressed and purified by 6xHis-tag pulldown from overnight *E. coli* (BL21) culture. After dialysis, protein concentration was estimated using a BSA-standard for hnRNP L (FL) and RRM-deletion variants and were analysed by SDS-PAGE (See **Supplementary Figure S2**).

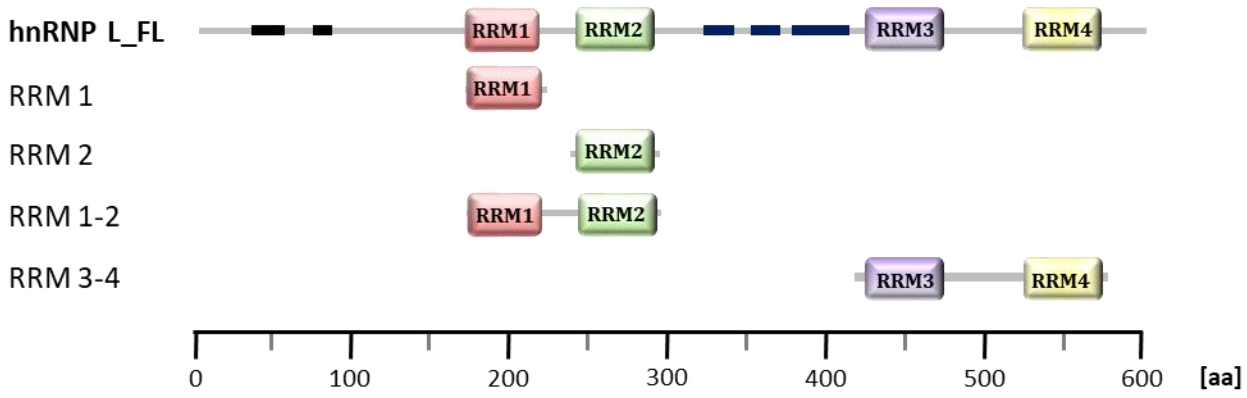

Fig. 3.1: Domain structure of hnRNP L and its truncations

Schematic representation of full-length (FL) hnRNP L and truncation variants (RRM1, RRM2, RRM1-2 and RRM3-4), used for SELEX experiments. RNA-binding domains are drawn as boxes (RRM1-red, RRM2-green, RRM3-purple and RRM4-yellow) and low complexity regions are shown as thick bars (black bars: glycine-rich and blue bars: proline-rich regions).

Since its discovery in 1990, SELEX has developed into a robust experimental strategy relying on combinatorial chemistry to identify specific sequences (single stranded DNA or RNA) that bind to certain ligands with high specificity and affinity [e.g. RNA-binding proteins (Tuerk & Gold, 1990; reviewed by Kohlberger & Gadermaier, 2021)]. Using SELEX-based selection, the specificity of RNA-protein interactions and their affinities can be effectively studied. Therefore, to dissect the complex RNA-binding properties of hnRNP L, we used individual GST-tagged subdomains of hnRNP L and performed an *in vitro* SELEX procedure. This procedure included selection by GST-pulldown of sequences bound and enriched from a random N_{40}-RNA pool. After each round of selection, bound sequences were identified by RNA-seq analysis (**Figure 3.2**). Bound RNA sequences after each round were amplified by PCR, reverse transcribed and fed as input into the next round, thus ensuring exponential enrichment. Note that as opposed to the standard protocol of using short degenerate region, we used an N_{40}-RNA pool. This enabled us to dissect and analyse arrays of several motifs, including their spacing. Additionally, enriched motifs were sequenced after each round of selection, allowing close monitoring of sequence enrichment throughout the SELEX procedure. In parallel, recombinant GST protein was also analysed for binding, and served as our negative control and for background correction.

Motif-enrichment analysis was carried out for all possible 4-mers, 5-mers and 6-mers by z-score calculation and was corrected at each round with the corresponding GST SELEX round (top-10 enriched 4-mer motifs in **Figure 3.2A**; top-20 enriched 4-mer, 5-mer and 6-mer motifs are enlisted for hnRNP L_FL in **Figure 3.2B**). For full-length hnRNP L protein, this SELEX analysis revealed an expected population of enriched CA-rich motifs. The individual RRM deletions (RRM1 & RRM2) as well as the di-domains (RRM1-2 & RRM3-4) did not greatly differ in their preference for binding motifs and showed a general affinity towards CA-rich and CA-repeat motifs. This observation is consistent with previous reports (Rossbach *et al*., 2014), including a previous SELEX with only hnRNP L_FL and a short degenerate RNA-pool carried out in our lab (Hui *et al*., 2015). A comparison of motif enrichment for full-length hnRNP L and the truncations indicates a high correlation between elements bound by full-length hnRNP L and RRM deletions. The most favoured tetramer for both hnRNP L_FL and subdomains was ACAC/T. Although poly(A) sequence seems to have been recognised, these sequences were mostly accompanied by a terminal C, especially for hnRNP L_FL, 5-mer and 6-mers (**Figure 3.2B**). In comparison to the di-domains, the single domains (especially RRM1) seemed to show an overall weak recognition and binding, which is reflected in the low z-score values for the selected motifs. This could partly be due to the very stringent washing conditions employed in higher SELEX rounds, which could inactivate the domains and hinder recognition or binding to the motifs.

Interestingly, we observed that the RRM3-4 tandem domains (which occur together as one folding unit) showed very high affinity for a CA-tetramer with a terminal T (or U in the RNA). Binding to ACAT outcompetes the typical CA-repeat for the tandem domain RRM3-4. This observation is also true for the full-length hnRNP L, which prefers an ACAT tetramer sequence over ACAC. This crucial observation distinguishes this SELEX experiment from the previous one, and was taken into consideration, while designing an optimal circRNA sponge.

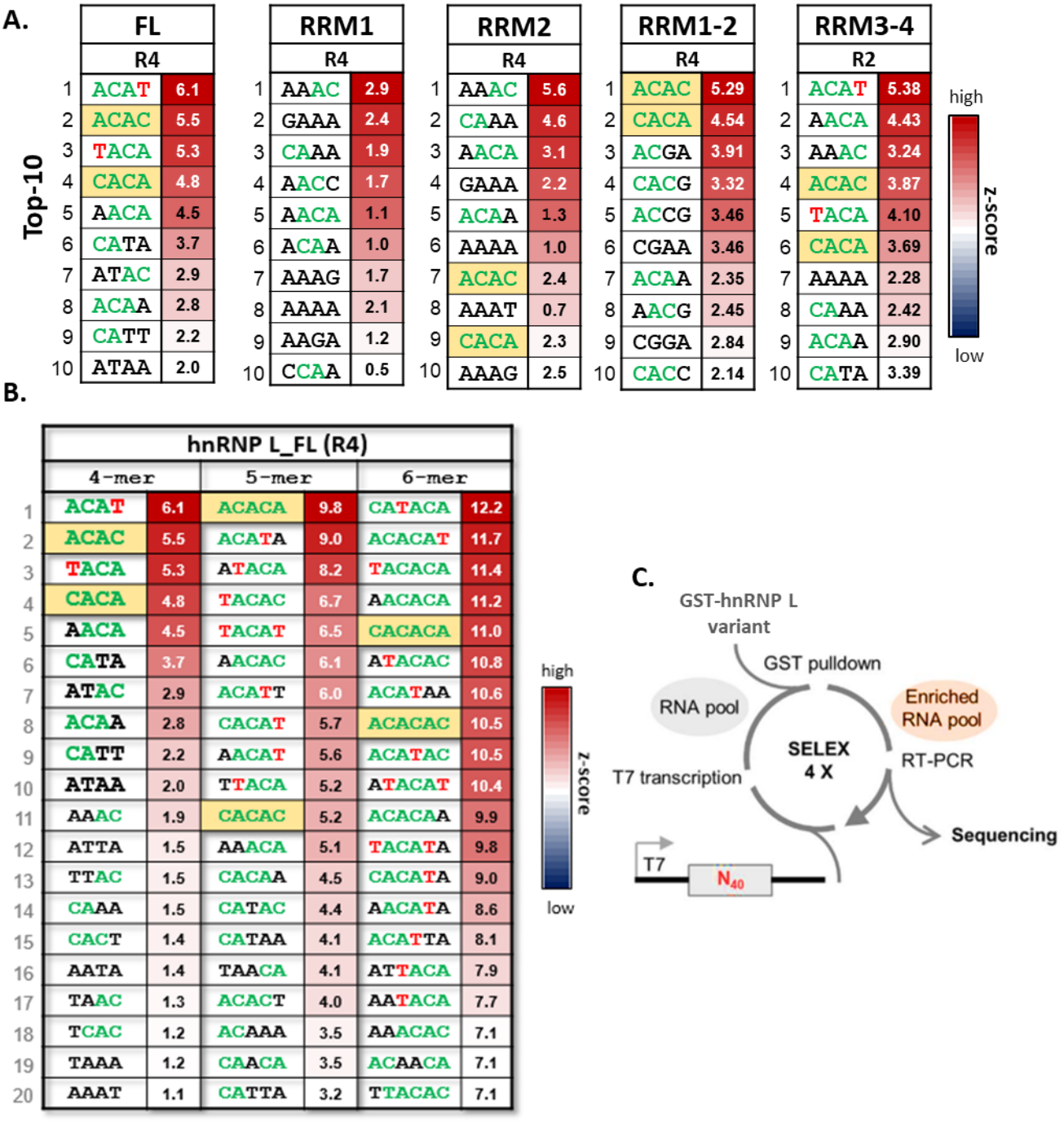

Fig. 3.2: SELEX-seq analysis of hnRNP L RNA-binding motifs.

(A) Top-10 enriched 4-mer motifs for full-length (FL) hnRNP L and all variants (RRM1, RRM2, RRM1-2 and RRM3-4) measured by z-score after the fourth round of selection (R4), except RRM3-4 where the second round (R2) is depicted.

(B) Top-20 table of SELEX-seq enriched 4-mer, 5-mer and 6-mer motifs with corresponding z-scores for full-length hnRNP L (FL) and the truncated protein variants (RRM1, RRM2, RRM1-2 and RRM3-4). CA-rich motifs are highlighted in green, elements with a common ACAT consensus in green with terminal T in red. CA-repeat sequences are highlighted in red with background in gold.

(C) Schematic representation of the SELEX-seq procedure**.** From a degenerate N_{40}-RNA pool, sequences bound by GST-tagged hnRNP L_FL and truncated variants respectively, were enriched and analysed by sequencing through the four SELEX rounds. From Schneider *et al.* (2018), modified.

However, binding specificity by the tandem domains RRM3-4 was lost after the second SELEX round due to stringency in washing. Therefore, only the first two SELEX rounds were analysed for RRM3-4. Furthermore, by comparing all SELEX rounds between the complete set of protein variants, we observe, as expected, a high overlap between the di-domains, but a low overlap between the individual domains in the motifs that were recognised. The tandem domain RRM3-4 showed the highest overlap with the full-length protein and the single RRM1 domain showed the least overlap in motif recognition (**Figure 3.2A** and **Supplementary Figure S3**). SELEX-strategy used for identifying binding-motifs is schematically represented in **Figure 3.2C**.

Taken together, our findings strongly argue for a coordinated combinatorial recognition of a specific array of CA-rich motifs by hnRNP L and subdomains. Our results further corroborate that the tandem domains RRM3-4 fold together as one unit because of their specific recognition pattern and binding to ACAT which, we presume, would be abolished if the domains were split. To identify how the four RRM domains of hnRNP L recognise consecutive elements on a single RNA, we analysed spacing between enriched minimal 4-mer motif combinations within a window of 0-25nts, using our SELEX-seq data. Enriched combinations of two predominant types of motifs (ACAC and ACAT) and their spacing were measured by z-score analysis (see **Figure 3.3A**).

Spacing analysis revealed that the most enriched motif for full-length hnRNP L was either the ACAC motif with another ACAC (ACAC-N_{0-25}-ACAC) or an ACAC in combination with ACAT (ACAT-N_{0-25}-ACAC). HnRNP L_FL showed such high specificity to CA-rich sequences that it bound to these sequences regardless of spacing. Interestingly, the orientation of the motif ACAT also played a role in recognition by hnRNP L. Spacing preference for the reverse orientation (TACA-ACAC) is different (TACA-N_{1-4}-ACAC) compared to ACAT-ACAC. Also, slight differences can be observed in the ACAC or CACA orientations respectively. But, given the general preference by hnRNP L for CA-rich and CA-repeat sequences, as reported by a previous SELEX experiment (Hui *et al*., 2015), we believe that these differences do not significantly contribute to RNA recognition by hnRNP L.

In contrast, analysis for the truncated versions (only di-domains) revealed that, surprisingly, RRM1-2 seems to show no spacing preference at all, although it is involved in recognition and binding to RNA (see **Figure 3.3A** middle panel). A maximum spacing of N_{7-8} was observed for the motif ACAC, which seems to tolerate other motifs containing at least one AC (ACGA, AACG and CACG). However, since these motif combinations were not enriched in the full-length analysis, we conclude that spacing may not be as crucial for RRM1-2 as long as the other domains are properly spaced.

Finally, for RRM3-4 we detected enrichment for the motif containing a terminal uracil in the RNA (i.e., ACAT). Nine out of ten motif combinations showed the enrichment of ACAT with an overall short to medium spacing (N_{1-12}) within the 25nts window. The most enriched combinations for RRM3-4 were two ACAT elements with short spacing [ACAT- N_{1-9}-ACAT (see **Figure 3.3A** bottom panel)]. In both cases (RRM1-2 and RRM3-4), it seems that long spacing between the individual domains is undesirable.

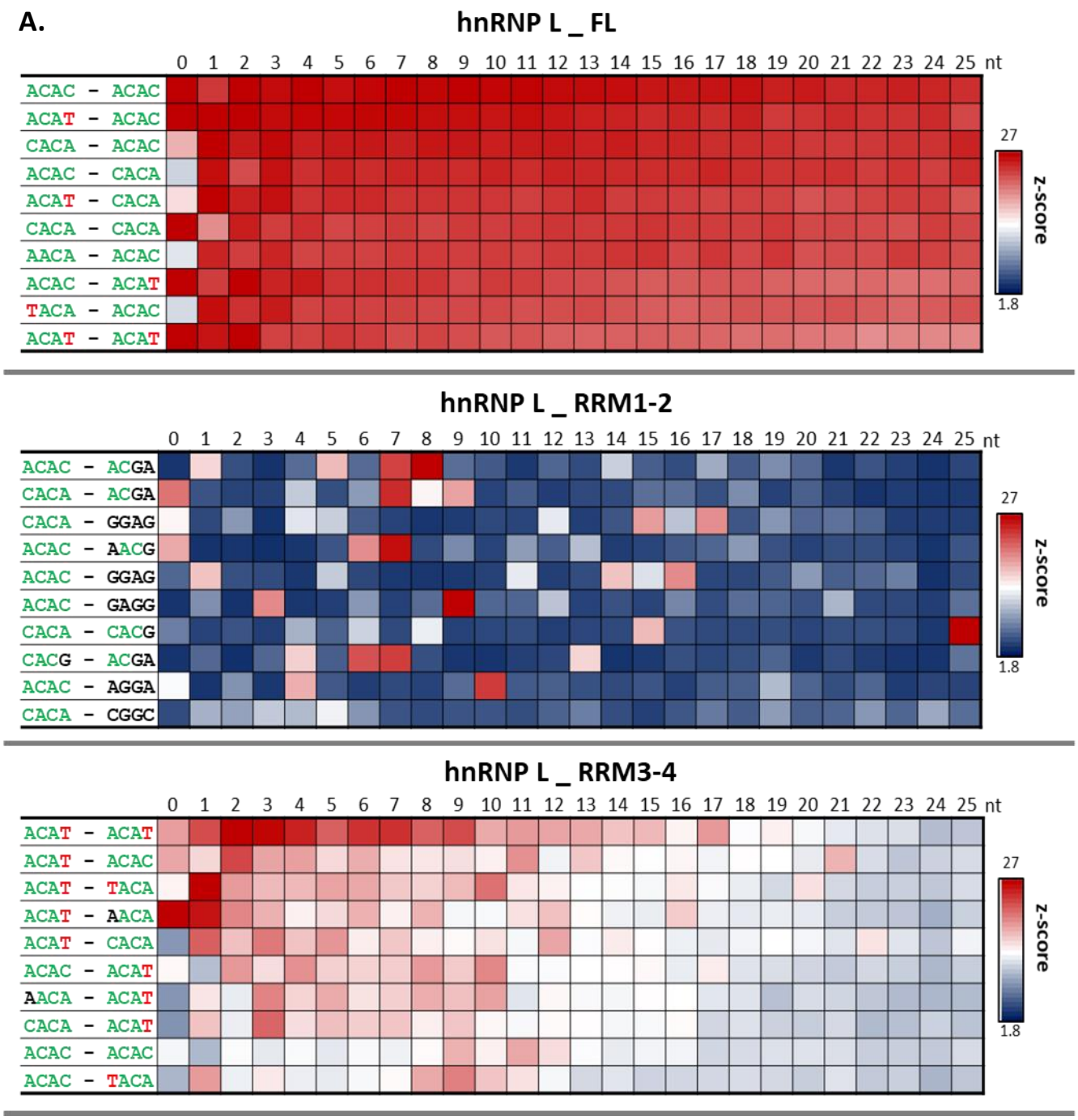

Fig. 3.3A: Spacing analysis reveals consensus array of hnRNP L-binding pattern.

Enrichment of motif combinations with spacing between 0-25nts for full-length hnRNP L (top), truncated RRM1-2 (middle) and RRM3-4 (bottom) di-domain variants, measured by z-score shown as heat-map. Spacing between combinations of CA-rich motifs and CA-repeat elements are shown on the left for full-length hnRNP L as well as RRM1-2 and RRM3-4. Individual z-score scales are given on the right.

The frequency with which ACAT is enriched for RRM3-4 binding indicates that this motif is strongly recognised by the tandem domain, and this is consistent with the binding property of the full-length hnRNP L. Therefore, we deduce that probably RRM3-4 first determines the binding by recognising ACAT within an array of AC-rich sequences in an RNA, and the other

domains subsequently bind in a combinatorial manner. Based on these datasets of full-length hnRNP L and its subdomains, we assembled a model of RNA recognition by hnRNP L (summarised in **Figure 3.3B**). It should be noted however, that this model would partially support both polarities of hnRNP L binding to its target RNA, at least for the first two domains owing to the symmetry of the sequence elements.

B.

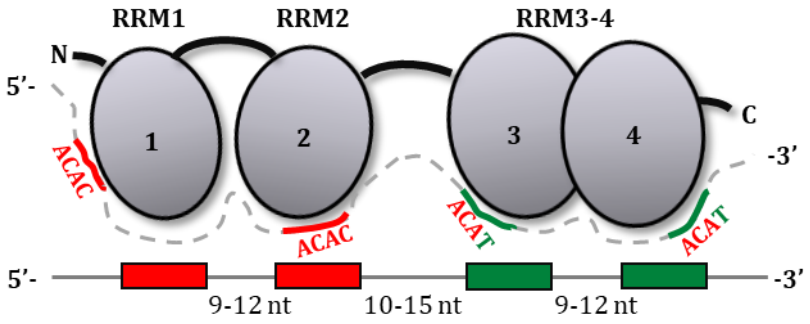

Fig. 3.3B: A model for RNA recognition by hnRNP L, based on SELEX-seq analysis.

Schematic representation of RNA recognition model for multidomain hnRNP L.

In addition to spacing analysis based on our SELEX-seq data, we also examined the binding behaviour of each of the RRMs of hnRNP L based on available crystal structures (Blatter *et al*., 2015). Using the predicted molecular structures of each RRM in hnRNP L, we calculated the size between individual domains and between the di-domains **(Figure 3.4).** In this prediction the authors used a long stretch of CA-repeat RNA to show interaction of the RRMs with the CA-dinucleotide in the RNA. The molecular structure of the protein-RNA interactions were depicted as 3D-ribbon diagrams (Richardson's diagrams), which are quite useful in predicting distances between the protein and the interacting RNA. Using PyMOL open-source software (https://pymol.org/2/), distances between the individual domains (RRM1 vs. RRM2 and RRM3 vs. RRM4) were calculated at nucleotide resolution. The program predicted the distance between the individual domains to be about ~11nt or 3.8nm and between the di-domains to be about ~13nt or 4.5nm and these predictions are in agreement with the spacing data we obtained from SELEX-seq analysis for hnRNP L and its subdomains.

Finally, based on these critical observations we were able to design short RNAs for the development of an optimal protein sponge for hnRNP L. Our design incorporated important information from motif enrichment data, which identified tetramer sequences bound by each of the four RRMs. These enriched motifs were arranged such that they were appropriately spaced, taking into consideration the differential spacing preference of the individual domains and the di-domains. We designed four RNAs containing a uniform sequence but only varying in spacing of the motifs (see **Supplementary Figure S4** and *Methods* **section 2.2.7**).

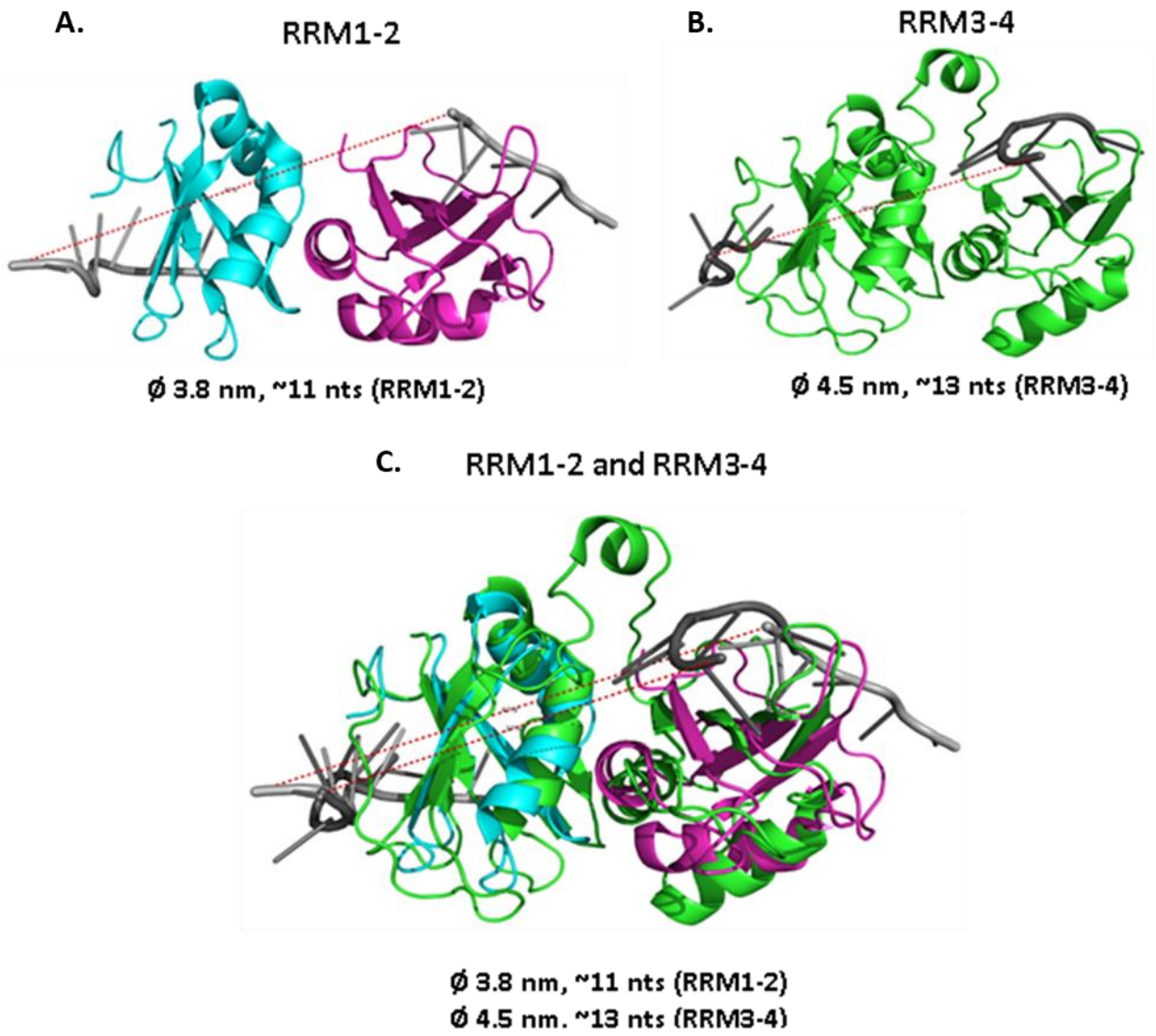

Fig. 3.4: Spacing predictions from available crystal structures of hnRNP L domains corroborates spacing analysis of hnRNP L based on SELEX-seq.

Based on crystal structures for each RRM domain (Blatter *et al*., 2015), spacing between each domain **(A & B)** and the di-domains **(C)** was calculated. Secondary structures of each domain is represented as ribbon-structures and the interacting RNA (CA-rich sequence) as line-structure in grey. The distance between the respective domains (represented by red dotted lines) is measured in nm (nanometer). Colour code: turquoise = RRM1, purple = RRM2, green = RRM3-4.

In summary, using SELEX-seq and molecular structure prediction tools, we demonstrate our ability to decipher the complex process of RNA recognition by RBPs such as hnRNP L, which usually occurs in a concerted manner, involving all four RRMs. We also provide insights into the design of RNA that can bind these RBPs with specificity and high affinity.

3.1.2 Analysis of hnRNP L binding by band shift assays

SELEX-seq data and spacing analysis for hnRNP L binding allowed us to predict a model which describes binding specificity of hnRNP L (**Figure 3.3B**). To validate this model, we designed synthetic RNA sequences containing domain-specific minimal 4-mer sequence elements, appropriately spaced by unrelated sequences and extending to a maximum total length of ~70nts (L_12/15, **Supplementary Figure S4**): ACAC-$N_{9\text{-}12}$-ACAC-$N_{10\text{-}15}$-ACAT- $N_{9\text{-}12}$-ACAT. Electromobility shift assays (EMSAs) with each *in vitro* transcribed ^{32}P-labelled RNA revealed that hnRNP L recognises each RNA with different affinities (see **Figure 3.5**). The strongest binding affinity was observed for L_12/10 RNA (K_D: ~60nM), while L_9/10 and L_9/15 seems to be bound with relatively low affinity (K_D >160nM). This was the first indication that the protein in general preferred medium spacing (N_{12}) over small spacing (N_9) between the individual domains. On the contrary, a slight preference toward small spacing (N_{10}) over medium spacing (N_{15}) was seen for the di-domains i.e., L_12/**10** versus L_12/**15** (here, spacing between di-domains highlighted, see **Figure 3.5B**). These effects were much more pronounced when these RNAs were used to pulldown endogenous hnRNP L from HeLa cell lysate (see *Results* **section 3.5**).

Next, to test for motif contribution of each RNA, we substituted the AC-motifs by mutating AC to UG and tested binding to hnRNP L. As a result, we were able to almost completely abolish binding of hnRNP L to the mutant RNA. Regardless of spacing variations, no binding was observed with increasing concentrations of hnRNP L. This observation indicated, firstly, that the binding motifs are necessary for high-affinity RNA recognition, and secondly, that sequences surrounding the binding motifs show little to no contribution in recognition and binding.

We consistently observed that the protein-RNA complex formed in these EMSAs seemed to be very large and fail to be resolved on the gel. As a result, large protein-RNA complexes accumulated at the wells of the gel (**Figure 3. 5**). Ideally, a one-to-one interaction is assumed for protein-RNA complexes; however we observe large complexes, suggesting that the protein forms oligomers with the RNA. Oligomer formation of hnRNP L with RNA was also observed previously in our lab (data not shown).

Furthermore, we argued that hnRNP L-RNA interaction is specific and independent of the GST-tag. To demonstrate this, we used recombinant GST-tagged hnRNP L and cleaved off the GST-tag by protease cleavage and observed binding behaviour with CA-rich RNAs by EMSA. We used shorter RNAs – $(CA)_{10}$ and CA#51 (20nts each), both of which have been shown to be bound by hnRNP L (Hui *et al*., 2005). $(CA)_{10}$ is a short RNA with 10 CA-repeats and CA#51 is a SELEX-derived short RNA, a high affinity target of hnRNP L (K_D 7.2nM, Hui *et al*., 2005). As predicted, no difference was seen in binding property of hnRNP L to the RNA even when the GST tag was removed (**Figure 3.6**). Interestingly, RNA-protein complexes formed in this EMSA were smaller and well resolved, both in the presence or absence of GST-tag, which further rules out binding mediated non-specifically by GST-dimerization. GST-removal by protease cleavage was tested and demonstrated by SDS-PAGE analysis (**Supplementary Figure S5**).

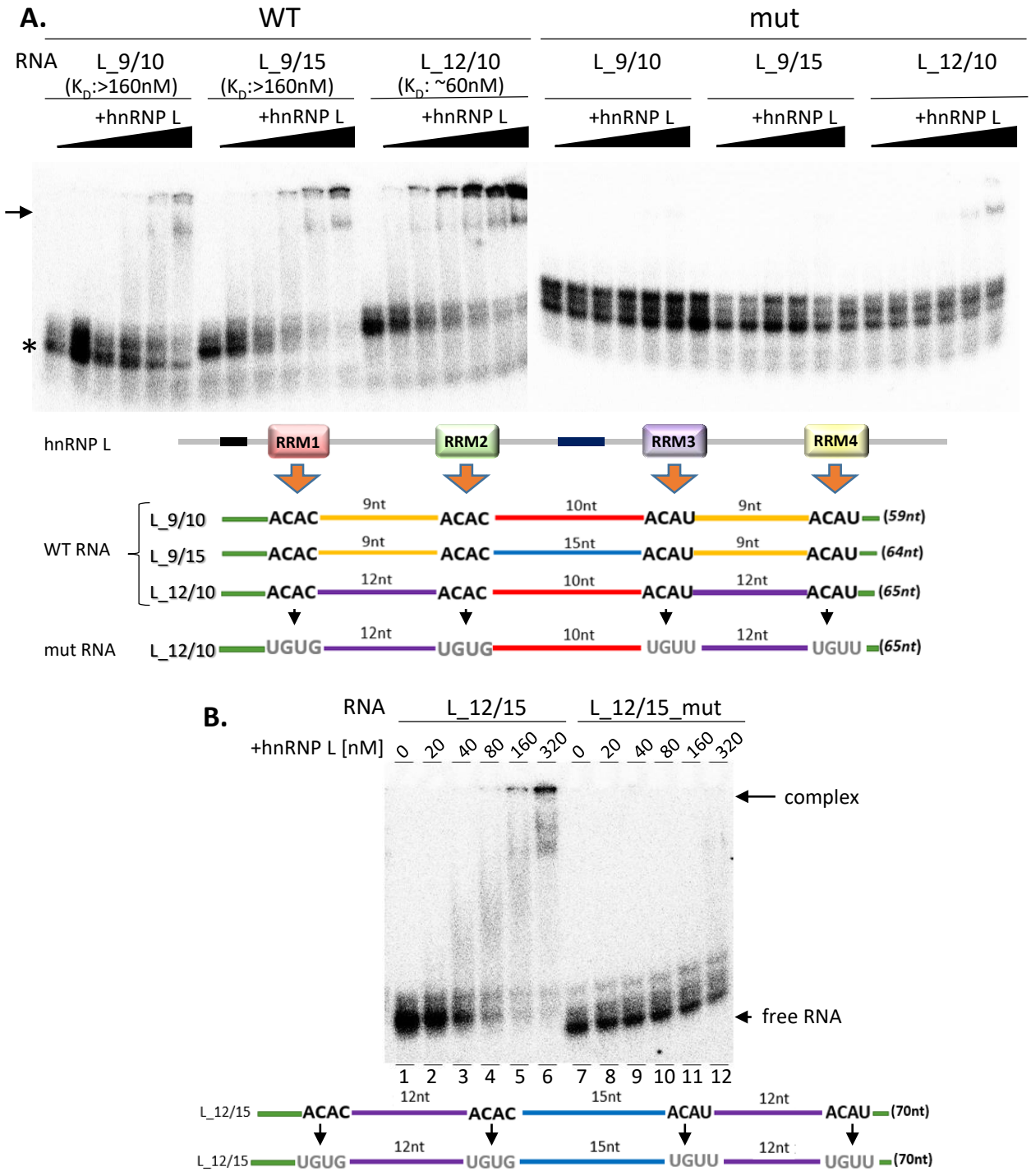

Fig. 3.5: Validation of SELEX-derived array of hnRNP L binding motifs.

Equimolar amounts of ^{32}P-labeled SELEX-derived RNAs (50nM) **(A)** Wild type (WT) - L_9/10 (K_D>160nM), L_9/15 (K_D>160nM), L_12/10 (K_D: ~60nM) and respective mutants (mut) were incubated with increasing concentration of recombinant hnRNP L. Complex formation (indicated by arrow) was analysed by a 5% native-PAGE and visualised by autoradiography. Free RNA is indicated by an asterisk. Estimated K_D values for WT RNAs are shown.

(B) Autoradiogram of hnRNP L band -shift assay with L_12/15 WT (lanes1-6) and mut (lanes 7-12) under similar conditions. The binding motifs, length and appropriate spacing between the motifs on the RNA is represented schematically under each autoradiogram. Bound RNA-protein complexes are shown with arrows and the unbound free-RNA fraction is indicated with an arrow head. *nM*, nanoMolar

Additionally, binding of hnRNP L to classical targets (CA-repeats) was also tested as described in Hui *et al*. (2003b) for $(CA)_5$, $(CA)_{10}$, $(CA)_{15}$, $(CA)_{25}$ and $(CA)_{32}$. In each case, with EMSAs we showed protein-RNA complex formation depending on the RNA size. In general, longer RNAs like $(CA)_{32}$ (64nt in length) formed complexes more efficiently with recombinant hnRNP L (data not shown).

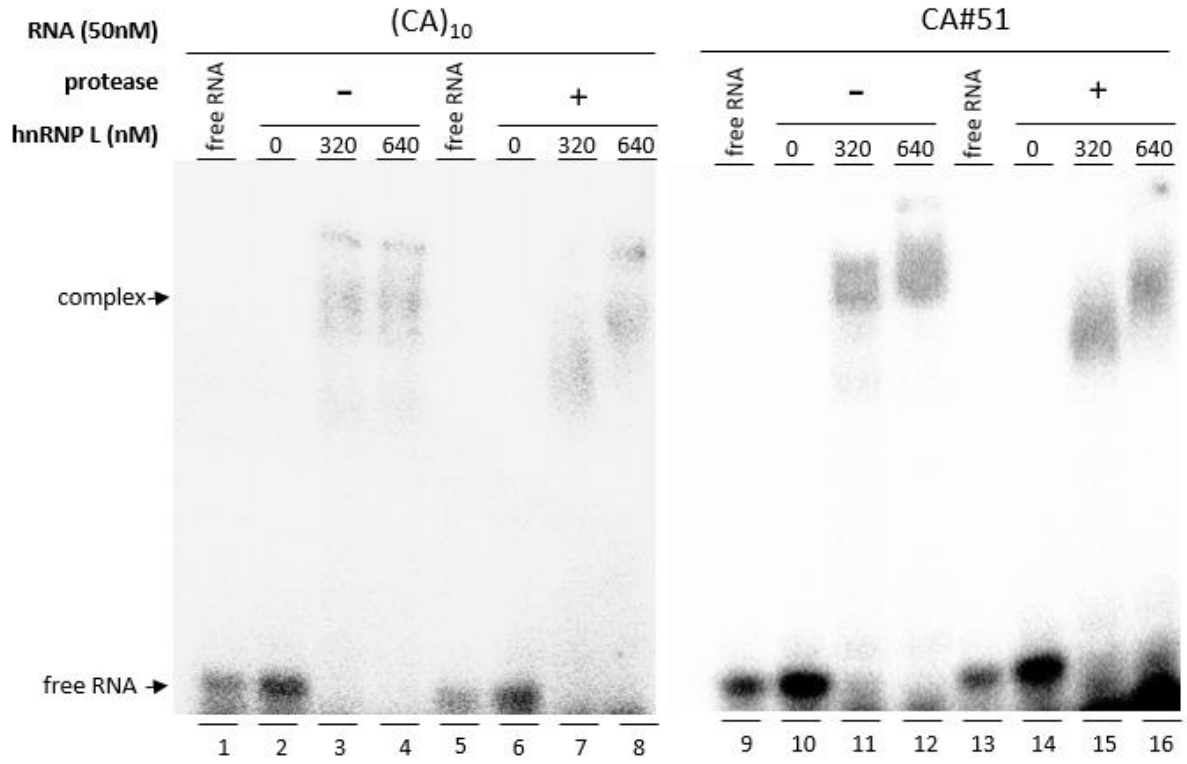

Fig. 3.6: Removal of GST-tag does not affect RNA-binding ability of hnRNP L.

Equimolar amounts of ^{32}P-labeled short RNAs (20nt) were incubated with GST-tagged hnRNP L in the presence (+) or absence (-) of TEV-protease. Lanes 1-8: hnRNP L binding to $(CA)_{10}$ RNA-containing 10 CA-repeats**.** Lanes 9-16: hnRNP L binding to CA#51 -- a previously reported high-affinity SELEX-sequence (Hui *et al.,* 2005). Unbound RNA transcript was loaded in each case as a reference for linear RNA mobility, complexes and unbound free RNA are indicated with an arrow. Increasing concentration of hnRNP L was titrated with a constant amount of RNA (50nM).

While screening for protein-RNA interactions, we encountered an unexpected problem with hnRNP L stability. Recombinant, purified hnRNP L was unstable after dialysis and upon longer storage. We observed degradation and precipitation of protein, which hindered EMSA assays. Therefore, we sought to express the protein by alternative expression systems and downstream purification. High concentrations of stable hnRNP L were obtained by purification steps involving size-exclusion chromatography. Purified protein was provided to us by Dr. Andreas Schlundt's lab, University of Frankfurt. However, the protein obtained was very sensitive to salt concentration and seemed to precipitate, when salt concentrations were lower than 1M. Since hnRNP L is a multidomain protein, the RRMs appear to be unstable in solution and are degraded. We speculate that high-salt concentration provides the right ionic condition to allow proper folding of the protein, so that weak areas that could be targeted by

proteases are not exposed. Therefore the protein was subsequently stored in a buffer containing 1M NaCl at -80°C.

Since all EMSAs until now were conducted at low salt concentrations (comparable to physiological levels), we questioned whether the influence of salt may affect binding ability of hnRNP L. Therefore, we tested hnRNP L binding using our RNA sequence which previously showed the highest binding affinity to hnRNP L (L_12/10) under low-salt (150mM) and high-salt (1M) conditions (see **Figure 3.7**).

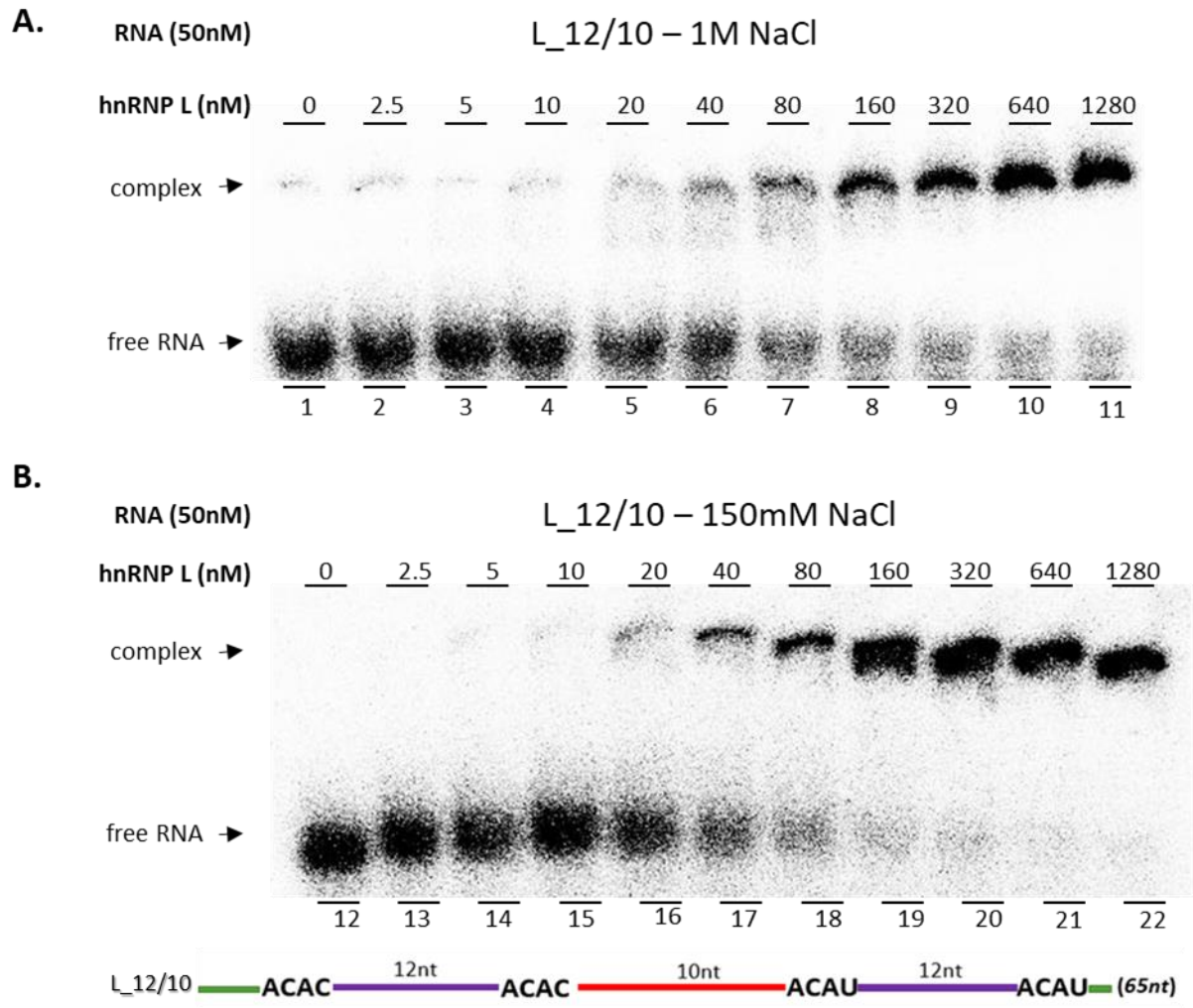

Fig. 3.7: High-salt conditions do not significantly alter hnRNP L affinity for SELEX-derived RNA.

Equimolar amounts of ^{32}P-labeled L_12/10 WT RNA was incubated with recombinant hnRNP L stored in buffer containing 1M NaCl. HnRNP L binding under **(A)** 1M NaCl concentration (lanes 1-11, K_D: ~80nM) and **(B)** normal salt concentration (150mM, lanes 12-22, K_D: ~60nM), was tested. Complexes and unbound free RNA are indicated with an arrow. Increasing concentration of hnRNP L was titrated with a constant amount of RNA (50nM).

EMSA assays under different salt conditions revealed that hnRNP L specificity to SELEX-derived RNA is not affected by high-salt concentration used for storage of hnRNP L. This further emphasises specific recognition and binding affinity of hnRNP L to designed RNA sequences.

In sum, by quantitative EMSAs we validated the interactions between hnRNP L and SELEX-based designer RNAs. Our results argue for a combinatorial recognition of RNA by the multidomain hnRNP L and provide insight into the specificity and affinity of RNA-protein interaction, both at physiological conditions as well as under high-salt concentration.

3.2 SELEX-based analysis of RNA-binding for RBM24

To determine the RNA binding specificity of the muscle-specific alternative splicing regulator RBM24, we performed an *in vitro* SELEX-based selection of RNA elements that bind to RBM24. RBM24 binds to its cognate RNA sequence via a single N-terminal RRM domain. SELEX-selection was carried out in a manner similar to hnRNP L described above. Recombinant RBM24 was purified by bacterial overexpression in *E. coli* (BL21 strain) and GST-pulldown using glutathione Sepharose beads. Recombinant GST-tagged RBM24 protein has a molecular weight of about ~51 kDa (24 kDa RBM24 +27 kDa GST), when analysed by SDS-PAGE with a 10% polyacrylamide gel (see **Supplementary Figure S2**).

Consistent with previous reports (Ray *et al*., 2009; Yang *et* al., 2014), our RNA-seq data, after three rounds of SELEX-selection, confirmed preferential binding of RBM24 to GT-rich sequences (GU in the RNA). Furthermore, sequencing data normalised to GST revealed that the most favourable tetramer (4-mer) sequence for RBM24 binding is GUGU (**Figure 3.8A**). Motif enrichment analysis for tetramers by z-score calculation showed the tendency of RBM24 to be generally associated with uridine stretches. However, enriched pentamer (5-mer) and hexamer (6-mer) motifs consisted of GU-rich sequences (**Figure 3.8B**) and poly(U) sequences were associated with at least one guanosine. Considering the relatively low enrichment of poly(U) sequences interspersed with adenosines, it is unlikely that RBM24 shows high specificity for 4-mers, 5-mers and 6-mers containing uridine alone or UA sequences without an accompanying guanosine. The observed GU-specificity was lost with highly stringent washing conditions in the fourth round of SELEX-selection for RBM24. This indicates less robust RNA interactions of the RRM. Since the single RRM with highly conserved RNP1/2 motifs in RBM24 is involved in RNA recognition (**Figure 1.12**), increasing salt concentrations for washing presumably weakened this interaction. Linear correlation (Pearson's r) of 4-mer motif enrichment for each SELEX round is shown in **Supplementary Figure S3**.

In addition, spacing analysis between enriched minimal 4-mer motif combinations showed minimal to no preference in spacing for binding of RBM24. The most enriched 4-mer motif consisted of GUGU with N_{0-3} spacing (**Figure 3.8C**), which may indicate that each RBM24 protein occupies a GU-rich motif independent of another RBM24, within a window of N_{0-25} nucleotides. Unlike hnRNP L, where four domains are involved in recognition and binding in a combinatorial manner, RBM24 relies on its single RRM and its binding occurs independently. To our knowledge, these results demonstrate the first SELEX-based analysis for RBM24, a single-RRM-domain RBP with crucial functions in sarcomerogenesis and muscle-specific splicing.

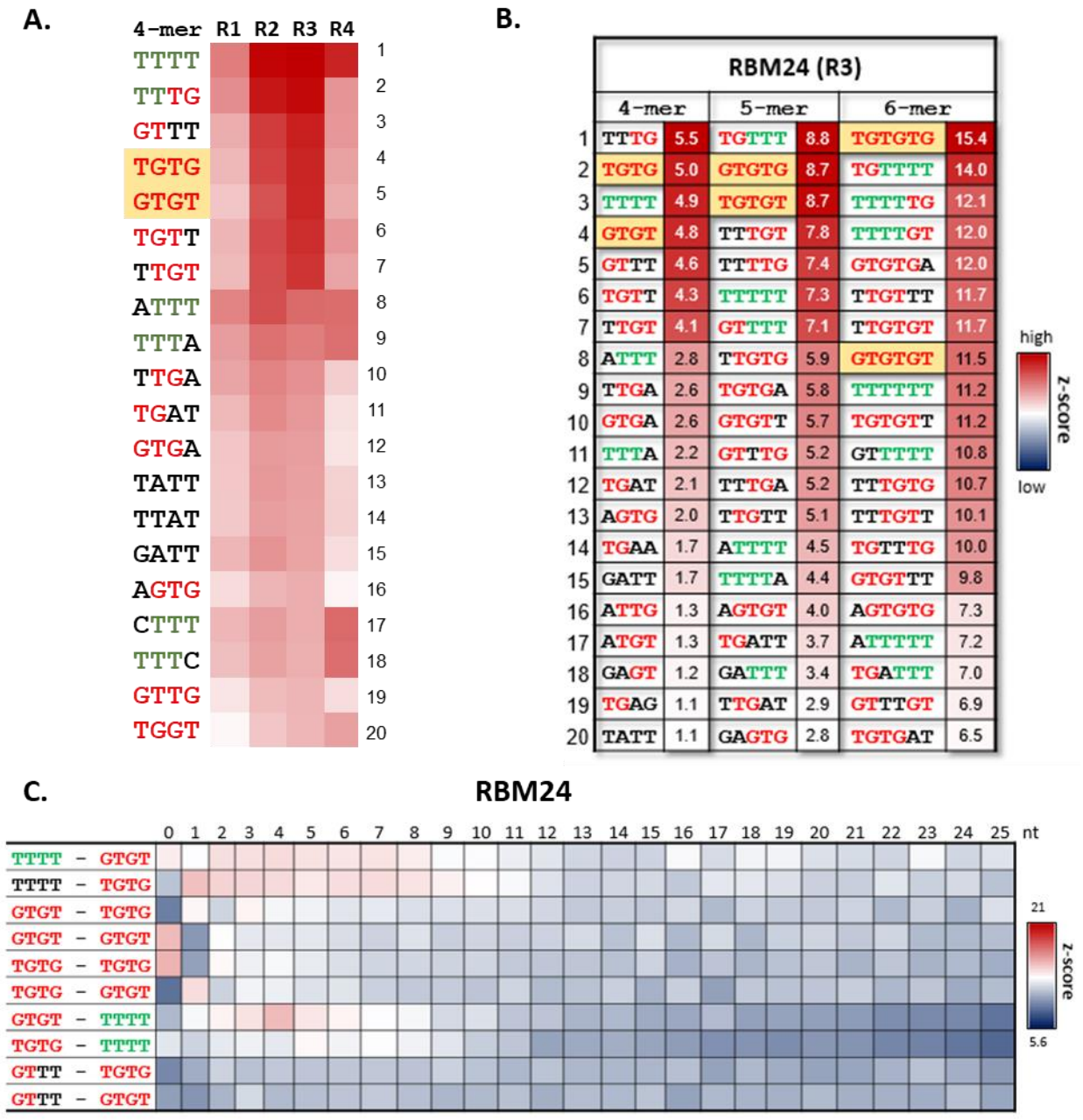

Fig. 3.8: RBM24 binding motif derived from SELEX-seq and spacing analysis.

(A) RBM24-specific motif enrichment from SELEX-seq analysis. Tetramer enrichment for the four SELEX rounds with GST-RBM24 (R1-R4) is illustrated. Top 20 tetramers are represented as colour-coded heat map, showing z-scores of the motif frequencies for each round. Motifs are ordered according to their cumulative z-score.

(**B)** Top-20 table of SELEX-seq enriched 4-mer, 5-mer and 6-mer motifs with corresponding z-scores for RBM24. GU-rich motifs are highlighted in red, poly-U regions in green and GU-repeat consensus sequences in red with background in gold.

(C) Enrichment of motif combinations with spacing between 0-25nts for RBM24. The combinations of poly-U and GU-rich elements are shown on the left. Individual z-score scales are given on the right.

3.3 Comparing binding of hnRNP L to linear and circular RNA

Due to its circular nature, circRNAs possess different properties compared to their linear counterparts. We speculated if the circular structure adds any advantage in binding to hnRNP L, compared to the linear RNA. We therefore intended to compare directly hnRNP L binding *in vitro* to linear versus circular RNA, using EMSA assays (**Figure 3.9**).

While both linear and circRNA bind to hnRNP L with similar affinities, circular configuration seems to bind hnRNP L with a slightly higher affinity (K_D: ~160nM) than the linear RNA (K_D: ~320nM). Noticeably, the complexes formed by the different forms of the RNA (linear vs. circular) with hnRNP L are seemingly different. HnRNP L forms two different complexes – larger complex which runs higher on the gel corresponding to linear RNA indicated with a straight line and a smaller complex which runs just below the linear RNA-protein complex corresponding to circular RNA-protein complex (indicated with a circle) (see **Figure 3.9**).

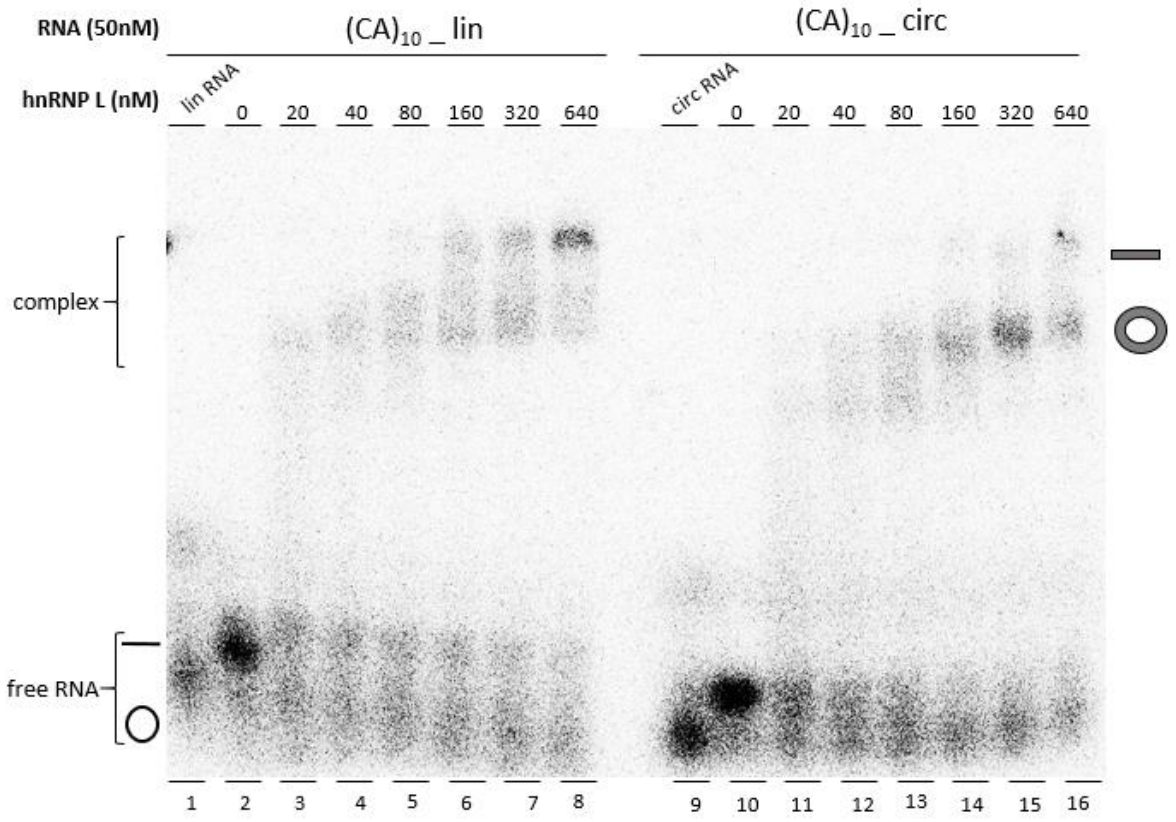

Fig.3.9: HnRNP L binds to linear and circular RNA with similar affinities.

Equimolar amounts of ^{32}P-labeled short $(CA)_{10}$ linear (lin) and circular (circ) RNA were incubated with recombinant hnRNP L. HnRNP L binding with linear $(CA)_{10}$ RNA (lanes 1-8) and *In vitro* circularised $(CA)_{10}$ RNA (lanes 9-16). RNA-protein complexes formed with linear and circRNA are indicated with a grey line and circle, respectively. Increasing concentration of hnRNP L was titrated with a constant amount of RNA (50nM).Unbound, free RNA corresponding to the respective linear and circRNA configurations are also indicated.

However, from this assay, it is difficult to estimate binding efficiencies based on different configurations of RNA, due to the overall poor signal from the autoradiogram. To show competitive binding of hnRNP L to either linear or circRNA, hnRNP L was incubated with equimolar concentrations of a mixture of linear and circRNA. After binding, we performed a GST-pulldown assay and looked for which RNA species was bound more efficiently, by hnRNP L. Unfortunately, we observed no clear preference in hnRNP L binding to circRNA over linear RNA (**data not shown**). This may be due to insufficient radioactive signal after circularisation and re-linearisation of ^{32}P-labelled circRNA during GST-pulldown assay. Nevertheless, we are confident that, by varying binding conditions and more efficient circularisation of labelled RNA, it should be possible to show circRNA binding preference over linear RNA.
Collectively, our results indicate that hnRNP L binding- depends not only on sequence specificity, but also, although to a small extent only, on RNA configuration.

3.4 Baculoviral expression and purification of recombinant hnRNP L

Prokaryotic expression systems like bacterial overexpression of proteins, though robust and easy-to-handle, have certain deficiencies, such as the lack of eukaryotic posttranslational modifications in expressed proteins. Multidomain RNA-binding proteins such as hnRNP L rely on posttranslational modifications for optimal stability and solubility. We encountered a similar problem with hnRNP L, which, when codon-optimised and bacterially overexpressed, was mostly insoluble upon storage and sensitive to degradation (see *Results* **section 3.1.2**). Therefore we decided to use a eukaryotic expression system that relies on baculovirus infection in Sf21 insect cells. Proteins expressed that way should contain proper posttranslational modifications and should be correctly folded, necessary for integrity and full activity of the protein. In contrast to bacterial overexpression, baculo expression results in lower yields, which is an additional advantage because slower expression ensures more time for correct folding.

The baculovirus expression system used here (Thermo Fischer Scientific), relies on site-specific transposition of an expression cassette into a baculovirus shuttle vector (bacmid) propagated in *E. coli* (Luckow *et al.,* 1993). HnRNP L with a 6xHis tag was cloned into the pFastBac™ donor plasmid to generate an expression construct, with expression controlled by a baculovirus-specific strong polyhedron promoter. This plasmid was transformed in *E. coli* host strain, DH10Bac™, that contains a baculovirus shuttle vector (bacmid) and a helper plasmid, and allows generation of a recombinant bacmid following transposition of the pFastBac™ hnRNP L construct. Recombinant baculovirus thus generated was then used to infect Sf21 cells. Infected Sf21 cells produced His-tagged hnRNP L. Different viral titres were used to infect cells, with the aim to maximise protein yield.

After hnRNP L was produced by the insect cells (2-4 days), cells were harvested and lysed. We performed a 6xHis-pulldown from cell lysate and purified the overexpressed hnRNP L by binding to Ni-NTA beads. Purified protein was washed, eluted and analysed by SDS-PAGE and Western blotting (see *Methods* **section 2.2.4.2**). Purified hnRNP L was detected by Western blotting using anti-hnRNP L antibody. When eluate and bead-fraction, after His-pulldown were analysed, we observed a high levels of His-tagged hnRNP L in both fractions (see **Figure 3.10**). As negative control, we infected insect cells with an empty bacmid (vector without any insert), where no hnRNP L was pulled down. Additionally, we tested uninfected cells for expression of hnRNP L, again resulting in no signal for hnRNP L. We also loaded GST-hnRNP L (93kDa) on the gel to serve as a marker and control for recombinant hnRNP L versus overexpressed hnRNP L (64kDa).

In sum, we have demonstrated overexpression of hnRNP L, based on a baculovirus-expression system.

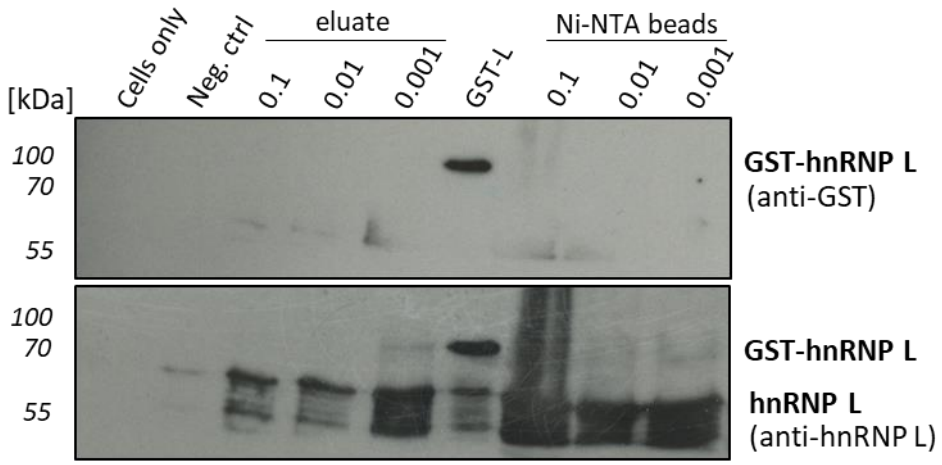

Fig. 3.10: His-pulldown of hnRNP L demonstrates baculovirus-based overexpression of hnRNP L

His-tagged hnRNP L was cloned into a baculovirus expression system and protein was overexpressed by viral infection for 72hrs, following which cells were harvested and lysed in RIPA buffer containing protease inhibitors followed by His-tag pulldown of hnRNP L using Ni-NTA beads. For viral titer estimation, 3-fold serially diluted baculovirus (0.1, 0.01 and 0.001) was used for infections. Protein expression was assayed by Western blot detection of hnRNP L. 25ng of GST-tagged hnRNP L was used as a positive control for Western blot and as a marker. Uninfected Sf21 insect cells were used as control (Cells only) for protein expression. Additionally, empty virus was infected to serve as negative control (Neg. ctrl). Anti-GST antibody was used to detect GST-hnRNP L (93kDa, top panel), and anti-hnRNP L monoclonal antibody (4D11) was used to detect overexpressed hnRNP L-alone (64kDa) as well as GST-hnRNP L (bottom panel).

3.5 *In vitro* pulldown assays for hnRNP L with SELEX-derived RNAs

RNA-binding specificity of RNA-binding proteins such as hnRNP L is highly dependent on the sequence context and the size of the RNA (see *Results* **section 3.1.2**). Based on our SELEX characterisation and spacing analysis, we identified four RNAs that showed binding specificity to hnRNP L, each with different affinity. After showing RNA-protein interactions by EMSAs, we attempted to replicate these observations using endogenous hnRNP L instead of recombinant hnRNP L. All four RNAs, along with their corresponding mutant forms were 3' end-labelled with biotin and used to capture endogenous hnRNP L from HeLa cell lysate. Pulldown of hnRNP L from cell lysate was analysed by Western blotting with anti-hnRNP L (4D11). Although we expected similar binding affinities for each of the four RNAs comparable to the EMSAs, strikingly, we observed that not all RNAs could capture endogenous hnRNP L. While it was clear that the RNAs with mutated binding motifs for hnRNP L did not interact with endogenous hnRNP L, the two shorter RNAs, L_9/10 & L_9/15, showed almost no interaction with hnRNP L. Compared to the input, pulldown efficiency of hnRNP L was maximal for L_12/10 and L_12/15 (see **Figure 3.11A**). This led us to conclude that the spacing between the binding-motifs likely plays an important role in determining recognition and binding efficiency.

To further test this assumption, we performed pulldown of hnRNP L from HeLa nuclear extract (**Figure 3.11B**). Since hnRNP L is predominantly nuclear and shuttling between the nucleus and the cytoplasm, there may be higher chances of successful hnRNP L pulldown from the nuclear fraction than from the total cellular extract. As expected, pulldown efficiency was much higher from HeLa nuclear extracts and we observed similar hnRNP L pulldown pattern for the four RNAs, comparable to HeLa cell lysate. In addition, and as a specificity control, hnRNP A1 protein, a member of the hnRNP family, was also tested for binding to the RNAs. We observed no pulldown of hnRNP A1 with any RNA, further validating the specificity of these RNAs for hnRNP L. Although one may predict such differences in binding relying on spacing differences in motifs, to our knowledge, this is the first experimental evidence for differential recognition of RNA by hnRNP L, where RNAs harbouring binding motifs differ only in their spacing.

In sum, the consistent results from both biochemical approaches, quantitative EMSAs (see *Results* **section 3.1.2**) and semi-quantitative pulldown, strongly support our binding model presented in **Figure 3.3B**. It appears that, since all four domains are involved in binding to RNA, each domain orients itself on the RNA with respect to the other domains. Furthermore, from our pulldown assays it is evident that binding specificity of hnRNP L not only depends on the sequence context and size of the RNA, but also on the spacing between the individual motifs within the RNA.

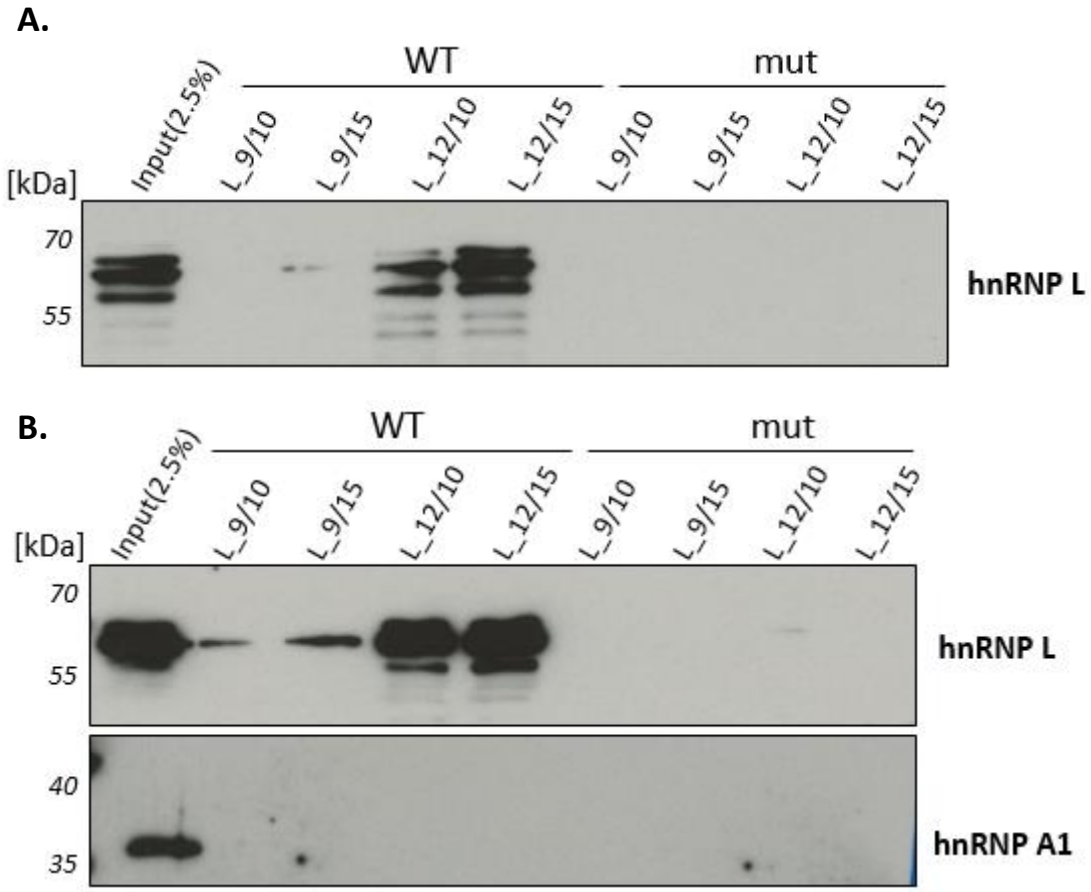

Fig. 3.11: HnRNP L binding-preference depends on spacing between individual motifs.

SELEX-derived RNAs that differ in motif spacing were synthesised *in vitro* by T7 transcription and subsequently biotinylated at the 3' end.
(A) After incubation with HeLa cell lysate, binding of endogenous hnRNP L was assayed by pulldown with streptavidin beads, followed by Western blot analysis, comparing in each case wild type (WT) and mutant (mut) RNAs. Pulldown efficiency for each variant was compared to input fraction (2.5%).
(B) Pulldown of hnRNP L from HeLa nuclear extract with biotinylated RNAs. Pulldown efficiency for each variant was compared to input fraction (2.5%). hnRNP A1 served as an additional negative control for pulldown from nuclear extract.

3.6 Overexpression of endogenous circRNAs using tornado system

The major challenge in expressing RNA aptamers in mammalian cells is their low expression and quick degradation. Therefore, although several expression systems have been developed to overexpress circRNAs in particular, based on minigene transfections and plasmid constructs with inverted repeats, none of them appears to sufficiently overexpress circRNAs. Wesselhoeft *et al.* (2018), developed an *in vitro* circRNA production strategy based on type I self-splicing introns. However, this system is also limited due to formation of several linear by-products and contaminating concatemers. To overcome these limitations, we applied another vector-based circRNA expression system developed by Litke and Jaffrey (2019). This system is based on the so-called Tornado expression cassette and relies on ribozyme activity of the transcript and ligation of tRNA intron by the ubiquitous endogenous RtcB ligase. The Tornado system has been shown to generate higher levels of circRNA than any other system developed so far, with almost no contaminating side-products. We therefore applied this strategy to overexpress our designer sponges in HeLa cells.

As proof of principle, we first attempted to express some endogenous circRNAs with different sizes using the Tornado system. Tornado constructs of *GSE1* (219nt), platelet-specific *Plt-circR4* (878nt) and the miRNA-7 sponge *CDR1as* (1485nt) were generated by cloning the sequences contained in these circRNAs into one of the arms of the F30, three-way junction in the Tornado vector. Each Tornado construct contains an F30-arm for proper folding of the Broccoli aptamer, as represented schematically in **Figure 3.12**. DFHBI is a fluorophore that becomes fluorescent upon binding to Broccoli; using this property, direct detection of RNA was possible by fluorescence imaging. HEK293 cells were transfected with each of the three Tornado constructs, and total RNA was analysed for circRNA. In addition, SPECC1 circRNA (1580nt) – with F30 (F30SPE) and without F30 sequence (SPE), were also cloned and expressed. The Tornado construct containing NFκB aptamer produces a circRNA of about 500nt**,** which served as a positive control for circularisation. For Predicted Tornado circRNA secondary structures, see **Supplementary Figures S6** & **S7**.

Total RNA obtained from HEK293 cells after Tornado transfections was analysed by gel electrophoresis, using staining with DFHBI fluorophore, followed by SYBR Gold detection (**Figure 3.13**). While SYBR Gold staining detected abundant endogenous RNAs such as tRNAs, rRNAs, 5S and 5.8S RNA, DFHBI staining detected only circularised RNA. Tornado-driven circularisation ensures proper folding of Broccoli aptamer, which is bound by DFHBI and visualised by fluorescent imaging. Noticeably, apart from the positive control NFκB aptamer, only GSE1 was directly detectable; suggesting that Tornado system may be optimal for production of relatively short circRNAs. Medium-length (Plt-circR4) to long circRNAs (SPECC1 & CDR1as) are expressed at lower levels compared to short RNAs. Another reason why these long circRNAs were undetectable may be caused by poly(U) stretches [$(U)_{>4}$] occurring in those genes. RNA polymerase III recognises oligo(U) regions as strong termination signals, so that transcription is halted, and only part of the circRNA may be overexpressed. However, these long circRNAs were detectable by RT-PCR (described below).

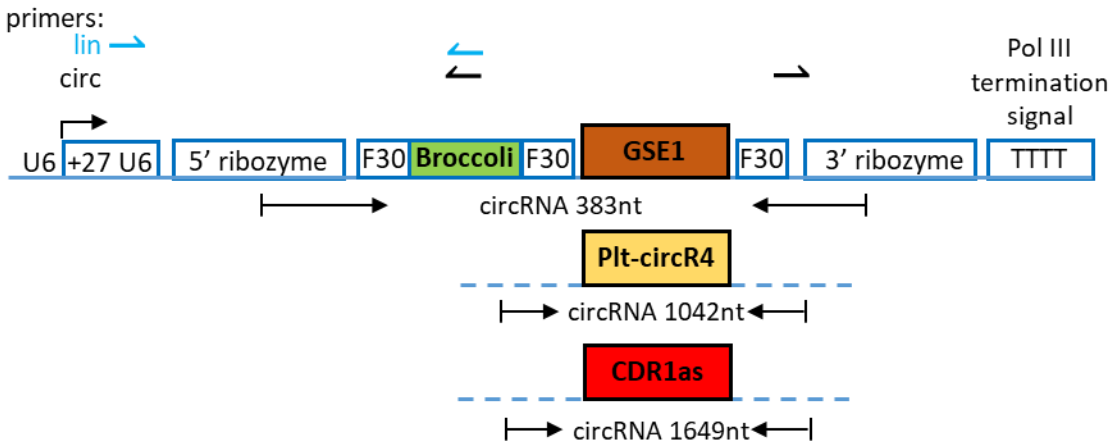

Fig. 3.12: Tornado constructs of endogenous circRNAs.
Schematic of GSE1, Plt-circR4 and CDR1as-Broccoli constructs. Pol III transcription is driven by U6 promoter with the first 27 nucleotides of U6. The circularizing sequences are flanked by Twister ribozymes (5' and 3' ribozymes). The Tornado circRNA expression cassette ends with the Pol III termination site (T-stretch). Each construct contains the Broccoli aptamer sequence for detection using DFHBI and three short sequences named F30, which allow proper folding of Broccoli. Primers used to detect linear and circular RNA are denoted in blue and black respectively, and the expected size (nt) of the circRNA is shown under each construct.

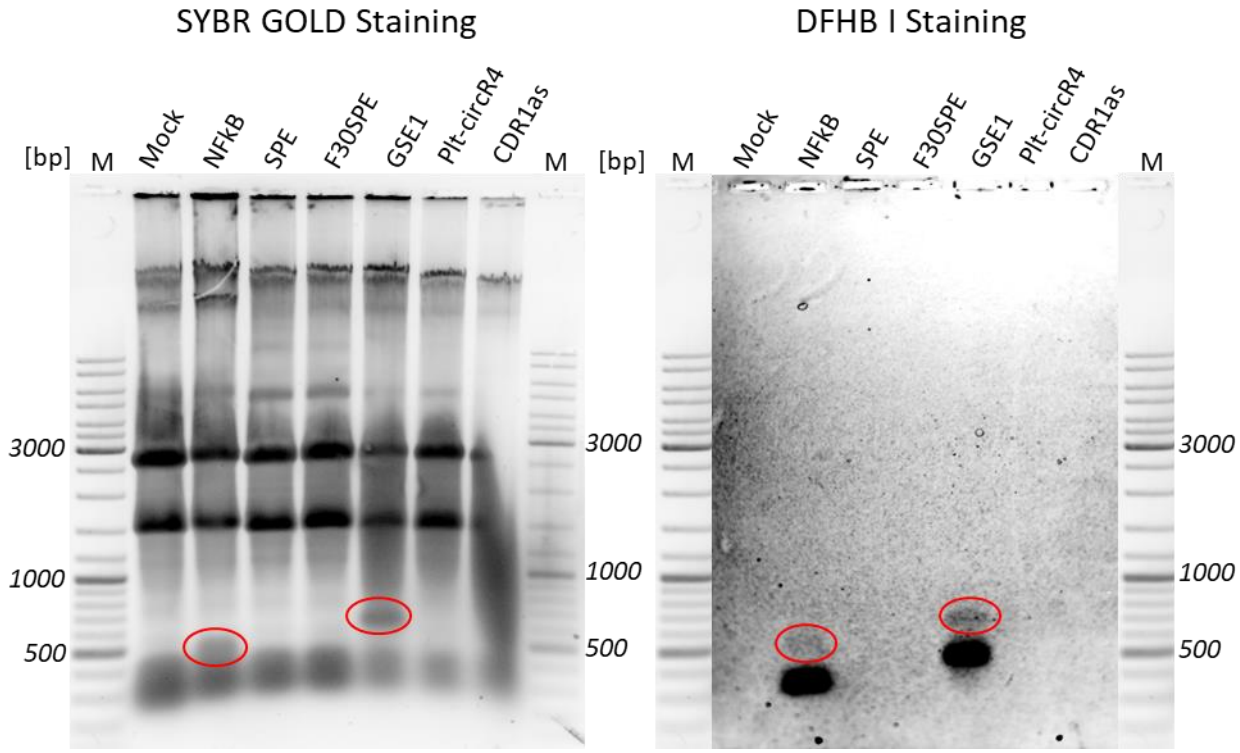

Fig. 3.13: Endogenous circRNAs are efficiently overexpressed by Tornado system.
HEK293 cells were transiently transfected with Tornado circRNA expression constructs- SPE, F30 SPE, GSE1, Plt-circR4 and CDR1as, each containing the Broccoli aptamer. Total RNA was extracted two days post-transfection. NFκB served as positive control for overexpression. Total RNA was analysed by 0.8% agarose gel electrophoresis. For direct in-gel detection of Broccoli-tagged circRNA, the agarose gel was stained with DFHBI (right panel), and then with SYBR Gold to stain total RNA (left panel). Bands corresponding to circRNA are marked with a red oval. *M*, marker.

In addition to total RNA staining, circRNA expression was also detected by RT-PCR using primers that span the circ-junction of the circRNA. HEK293 cells were harvested post-transfection with Tornado-GSE1, Plt-circR4 and CDR1as. Total RNA was extracted and reverse-transcribed into cDNA using random hexamers, followed by PCR with gene-specific and circRNA-specific primers (see **Figure 3.14**).

Strikingly, by RT-PCR we show the overexpression of all three circRNAs and not just GSE1. Out-facing (circ) primers were used to detect the circRNA, while in-facing primers (lin) were used to detect linear RNA (see **Figure 3.12**). The out-facing primers cover the length of nucleotides around the backsplice junction, thereby confirming its circular nature.

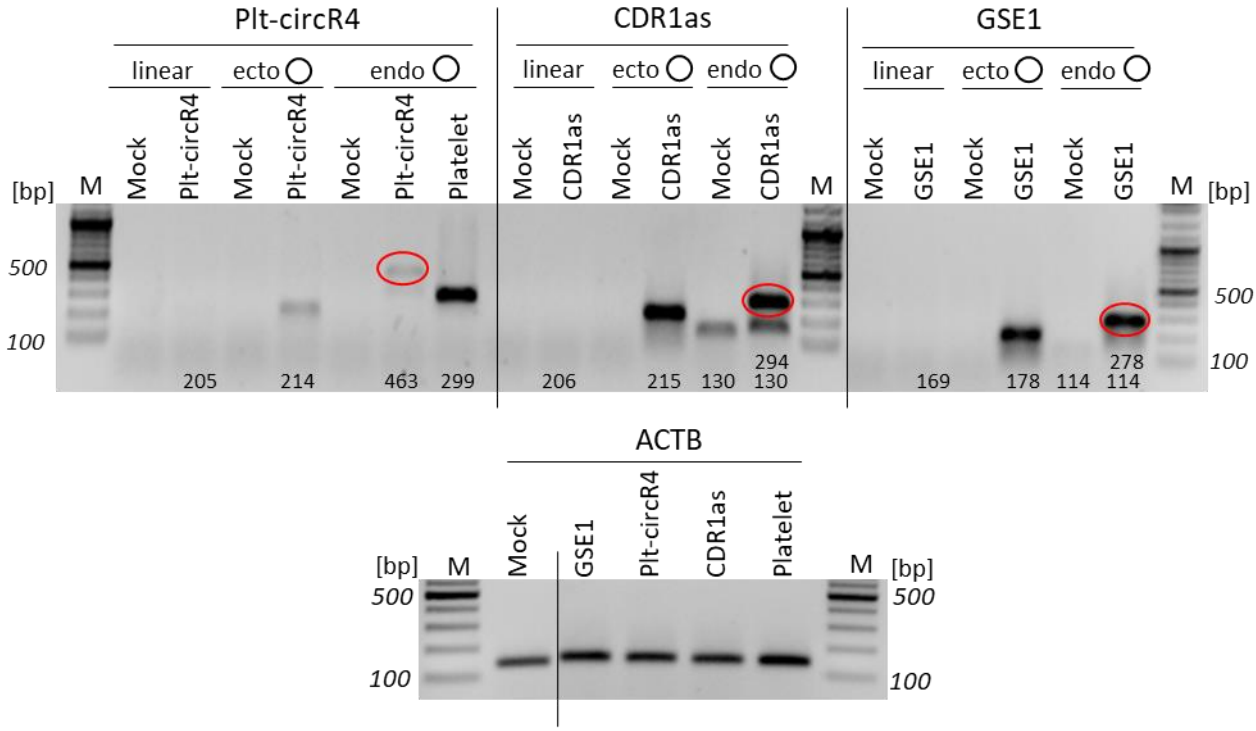

Fig. 3.14: Endogenous circRNA is distinguishable from ectopically overexpressed circRNA. Three endogenous circRNAs - GSE1, Plt-circR4 and CDR1as were overexpressed in HEK293 cells by Tornado-system. Total RNA was harvested two days after transfection and reverse-transcribed using mixed primers. RT-PCR was performed using linear and circRNA-specific primers (indicated with a line and circle, respectively). Additionally, to distinguish between endogenous and ectopically overexpressed circRNA, gene-specific and Tornado circRNA-specific primers were used, respectively. In addition to mock transfection, β-actin served as loading control for RT-PCR and overexpression. In each case, the expected band length is indicated under each lane for Plt-circR4, CDR1as and GSE1. Since endogenous circRNA-specific primers also bind to ectopically overexpressed circRNA, two PCR products are expected in this case, which are indicated one below the other for CDR1as and GSE1. For Plt-circR4, which is specific to platelets, no endogenous circRNA is detectable in HEK293 cells, therefore RNA extracted from platelets was used to detect endogenous Plt-circR4. Ectopically expressed circRNAs are distinguished from endogenous circRNAs, with red ovals. *M*, Marker

Tornado-expressed circRNAs possess the Broccoli-aptamer for in-gel detection and additional few nucleotides from stem region around the circRNA. Therefore, ectopically expressed circRNAs are slightly larger than endogenous circRNAs. Since endogenous circRNAs are slightly smaller than ectopically overexpressed circRNA, it is possible to distinguish the two circRNAs by their variable mobility in gel electrophoresis. Primers targeting the Tornado-expressed circRNAs ('ecto O' primers) recognise a region specific to the Tornado vector, therefore they are unable to bind to the endogenous circRNA. On the contrary, endogenous circ-primers ('endo O') bind to a region specific to the circRNA. Since this region is also common to the ectopically overexpressed circRNA, 'endo O' primers can also detect them. We observed good overexpression of GSE1 and CDR1as (see **Figure 3.14**).

When endogenous primers were used for Plt-circR4, only overexpressed Plt-circR4 was detected. This is because Plt-circR4 is platelet-specific, and when RNA from platelet was tested, we were able to detect endogenous Plt-circR4. This may be a reason why we do not observe very high levels of Tornado-Plt-circR4. It should be noted that neither 'ecto O' primers nor 'endo O' primers detected any linear precursor RNAs, although they have the same sequence as the circRNA. This strongly suggests that Tornado-driven overexpression produces circRNA without linear contaminants or concatemers, which was the main drawback of other systems. Detection of β-actin mRNA was used as loading control for PCR, and mock transfections as negative control.

In conclusion, we demonstrate the overexpression of circRNAs using the Tornado system developed by Litke and Jaffrey (2019). By considering three different examples, we show that RNAs can be efficiently and effectively circularised *in vivo.* Although circularisation efficiency is still higher for smaller RNAs, we were able to successfully overexpress also large circRNAs such as CDR1as (1485nt) and SPECC1 (1580nt). When producing large circRNAs it is necessary to bear in mind the limitation of oligo(U) mediated termination of transcription. By removing such potential termination sites, it should be possible to produce larger circRNAs. Since most endogenous circRNAs do not have well-defined functions, except for a couple of cases such as CDR1as (as miRNA-7 sponge, Hansen *et al*., 2013) and Sry (Capel *et al*., 1993), we also wanted to test if overexpression has any functional effects. Both GSE1 and especially Plt-circR4 are highly abundant in cells and platelets, respectively, but their function is still unknown. By overexpression using Tornado-system we expected to see phenotypic effects such as growth defects. In this context we observed no difference in growth patterns in normal versus transfected HEK293 cells (data not shown). This suggests that a more systematic approach with RNA-seq, metabolomics or proteomics studies, for example, may be required to detect changes occurring due to overexpression. Nonetheless, for the first time we now have a system that actually overexpresses any circRNA of interest without contaminants, and we believe that this system can be used to understand underlying functional mechanisms of circRNAs in the cell.

3.7 Overexpression of circRNA for hnRNP L sponging

After successfully demonstrating our ability to overexpress endogenous circRNAs using the Tornado system, we further utilised this strategy to overexpress circRNA sponges. SELEX-derived RNAs have previously been shown to bind to hnRNP L with high specificity and affinity (see *Results* **Sections 3.1.2** and **3.5**), therefore these sequences were used to design circRNA sponges against hnRNP L. Both wild type (WT) and mutant (mut) sequences of L_12/10 RNA, were cloned into the Tornado expression vector to overexpress these sponges in HeLa cells. In addition to our SELEX-derived sequences, Tornado-expression constructs for $(CA)_{20}$ sequence and a recently described hnRNP L sponge sequence - SLX4x, were also made. While, $(CA)_{20}$ sequence has been reported to sponge hnRNP L *in vivo* (Schreiner *et al*., 2020), SLX4x is another highly-specific hnRNP L sponge which is made up of four copies of a SELEX-derived, high-affinity sequence and shown to efficiently sponge hnRNP L (Schreiner *et al*., 2020). $(CA)_{20}$ and SLX4x therefore served as positive controls for hnRNP L sponging. All four Tornado-expression constructs (labelled T-L_12/10, T-mut, T-$(CA)_{20}$ and T-SLX4x) were made by replacing the Broccoli aptamer in the Tornado vector. These constructs also rely on the U6 promoter for transcription and twister ribozymes for circularisation *in vivo*, like the other Tornado vectors (see **Figure 3.15**). Since, the Tornado-expression system is ideal for small circRNAs, we had an added advantage in expressing these circRNA sponges, which range from 87nt [$(CA)_{20}$] to 149nt (SLX4x).

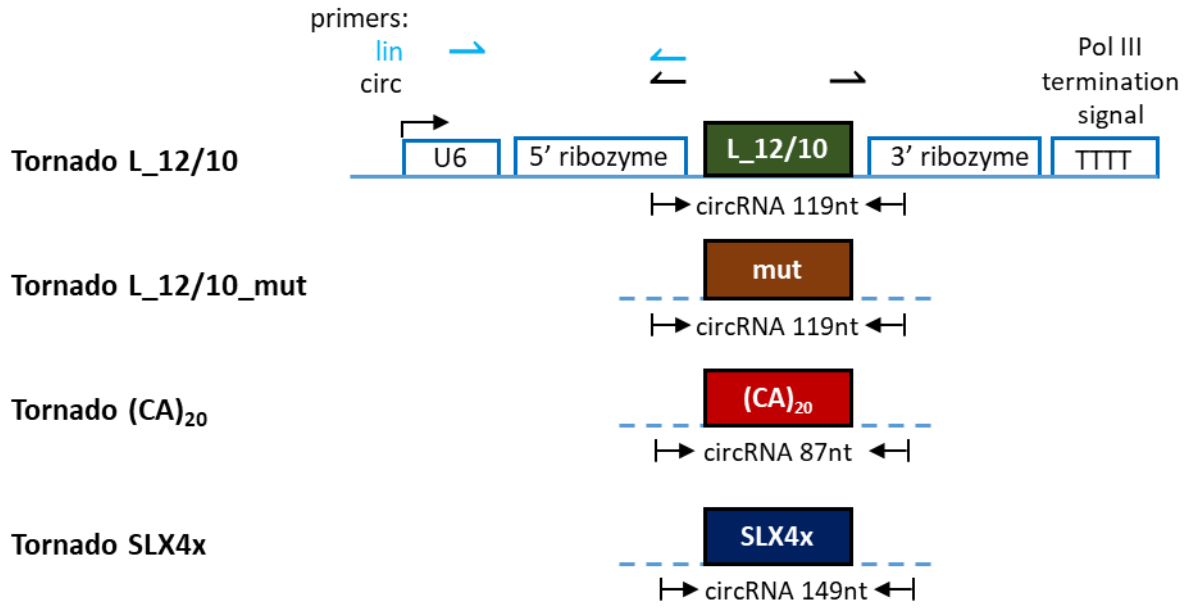

Fig. 3.15: Tornado expression constructs for hnRNP L circRNA sponges.
Schematic representation of Tornado constructs for hnRNP L sponges: T-L_12/10, T-L_12/10_mut, T-$(CA)_{20}$ and T-SLX4x. Pol III transcription is driven by U6 promoter. The circularizing sequences are flanked by Twister ribozymes (5' and 3' ribozymes). The Tornado circRNA expression cassette ends with the Pol III termination site (T-stretch). Primers used to detect linear and circular RNA are denoted in blue and black, respectively, with expected sizes (nt) of the circRNAs shown under each construct.

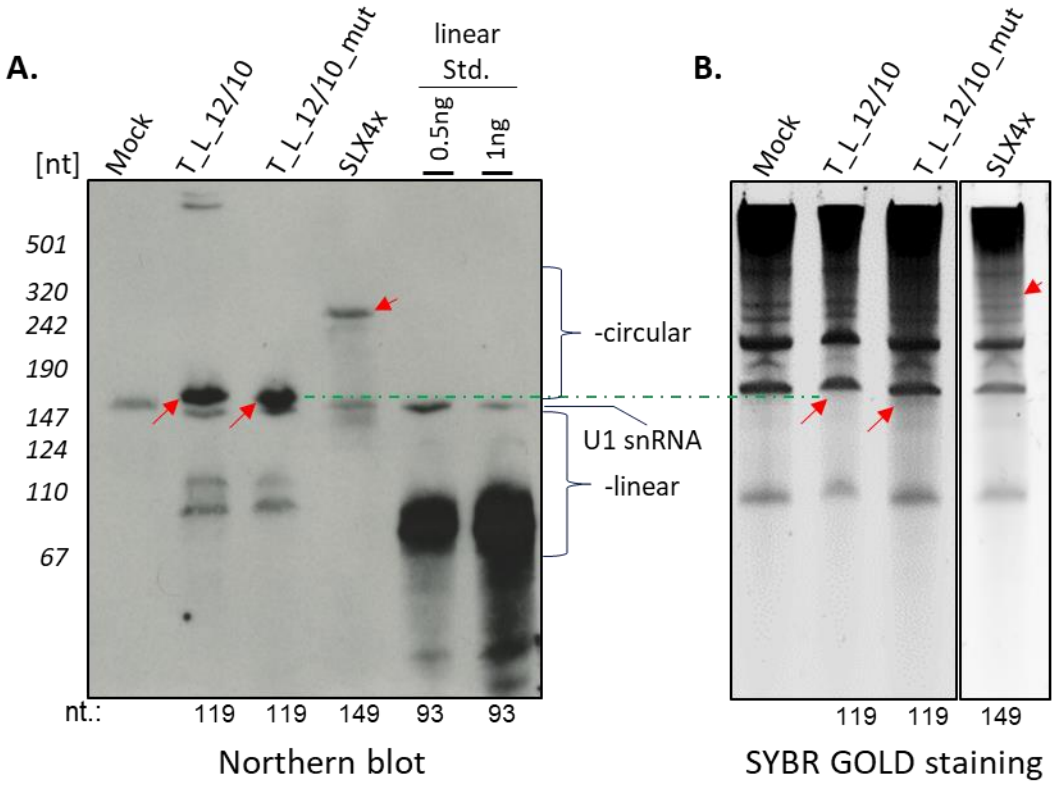

Fig. 3.16: Transfected circRNA sponge is overexpressed.
Total RNA was extracted from HeLa cells transfected with Tornado-expressed sponges for SELEX-derived sequences (T-L_12/10, T-L12/10_mut and SLX4x).
(A) Using a DIG-labeled riboprobe spanning the circ-junction and specific to tornado stem region, all transfected sponges were detected by Northern blot. 0.5ng and 1ng of a 93nt standard was used for comparison. U1 snRNA was detected using U1snRNA-riboprobe and this served as a loading control.
(B) 3µg of total RNA was run on a 10% denaturing polyacrylamide gel under the same conditions as for Northern blot and stained with SYBR Gold. The length of each circRNA is indicated below the gel. A green-line joining the Northern blot and stained gel indicates the same length of circRNA by two different detection techniques. Red arrows indicate overexpressed circRNA sponge.

These sponges were transfected into HeLa cells and high expression of circular sponges were observed both by direct RNA analysis as well as Northern blot detection of circular RNA (**Figure 3.16**). Tornado-expressed circRNA sponges were detected by Northern blotting using a DIG-labelled riboprobe specific for the circ-junction of the RNA and a portion of the Tornado-stem region. The riboprobe was successful in binding to WT, mut and SLX4x sponge (see **Figure 3.16A**), however $(CA)_{20}$ was undetectable (see **Figure 3.17B**). This is probably due to the secondary structure of the CA-repeat sequence, making it difficult for binding of the riboprobe. Compared to the mock transfections, it is clear that specific bands arise due to Tornado overexpression. Interestingly, both WT and mut sponges are overexpressed, compared to SLX4x. U1 snRNA is the most abundant snRNA with about $1x10^6$ copies per cell (Baserga & Steitz, 1993). In comparison, it appears that the Tornado-expressed sponges are

much more abundant than U1 snRNA, which has similar mobility on a 10% polyacrylamide gel. Also, comparing the linear standard, it seems that the sponges have relatively high molar concentration in the cells after transfection. Total RNA analysis by denaturing polyacrylamide gel electrophoresis showed many abundant RNAs, when stained with SYBR-GOLD (**Figure 3.16B**). By aligning the RNA bands on the gel to the Northern blot, we observed specific bands that correspond to Tornado overexpression. This is a key observation because, until now several expression systems have been developed to express specific circRNAs using different strategies. Unfortunately, none of these systems succeeded in circRNA expression at such high levels as to be directly visualised by total-RNA staining. Our results therefore provide strong evidence for high circRNA expression *in vivo* using the Tornado system developed by Litke and Jaffrey (2019). Although unexpected, we observed some linear byproducts along with circRNA and we think that these could be a result of incomplete ligation by RtcB, or circRNAs that were nicked. Linear and circular RNA are distinguishable by gel electrophoresis, based on their differential mobility.

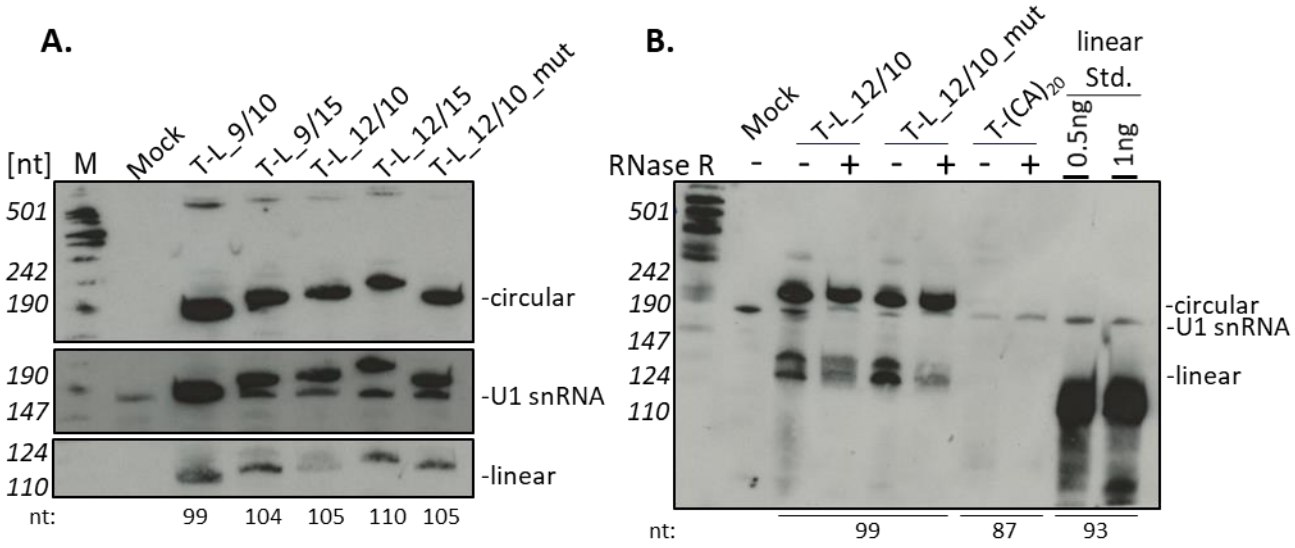

Fig. 3.17: Northern blot analysis of Tornado-expressed sponge confirms RNase R stability and high expression of circRNA.

Total RNA was extracted from HeLa cells transfected with Tornado-expression constructs of sponges for all SELEX-derived sequences (T-L_9/10, T-L_9/15, T-L_12/10, T-L_12/15 and T-L12/10_mut) and detected by Northern blot.

(A) Using a DIG-labeled riboprobe spanning the circ-junction and specific to tornado stem region, all transfected sponges were detected.

(B) In addition, circular configuration of sponges was confirmed by resistance to RNase R digestion. 500ng of total RNA for each variant was run on a 10% polyacrylamide gel. 0.5ng and 1ng of a 93nt linear standard was used for comparison. U1 snRNA was detected using U1snRNA-riboprobe and this served as a loading control. Size of RNA (nt.) is indicated below each blot.

In like manner, all SELEX-derived RNAs were cloned into Tornado plasmid for overexpression in HeLa cells. Since, each of the RNAs has similar sequence and vary only in their length and spacing of binding-motifs, all RNAs were successfully detected by the same riboprobe (see

Figure 3.17A). Stably expressed sponges T-L_12/10, T-mut and T-$(CA)_{20}$ were tested for their expression and resistance to RNase R digestion. T-$(CA)_{20}$ sponge was undetectable by Northern blot, but both WT and mut sponge were detectable. While the circRNA was resistant to digestion by RNase R, the linear RNAs were digested. This confirms the circular configuration of the sponges (see **Figure 3.17B**).

Based on the above observations, it is now evident that Tornado-expressed sponges are RNase R resistant, highly expressed and detectable. Since these sponges were designed to bind to hnRNP L, we tested binding ability of stably expressed sponges to hnRNP L *in vivo.* CircRNA-sponge binding to hnRNP L protein was assayed by RNA immunoprecipitation (RIP) (**Figure 3.18**). HeLa cells were transfected with the Tornado-expression constructs, and 48h after transfection, cell lysates were collected and probed with anti-hnRNP L and anti-FLAG antibodies. Immunoprecipitated protein was identified as hnRNP L by Western blot; GAPDH was detected as input and RIP-negative control (**Figure 3.18A**). Direct interaction of endogenous hnRNP L with expressed Tornado-sponges was detected by RT-PCR and quantitated by RT-(q)PCR (**Figure 3.18 B&C**), using primers that cover the circ-junction. In addition to the SELEX-derived sponges, T-L_12/10 (WT), T-mut and T-SLX4x; T-$(CA)_{20}$ sponge-interaction with hnRNP L was shown by RT-(q)PCR, as a positive control.

By RT-PCR detection T-L_12/10 demonstrated a 50% RIP efficiency, whereas T-SLX4x was shown to bind hnRNP L with 57% efficiency, compared to input. HnRNP L was not expected to interact with T-mut sponge, however we observe 40% efficiency in binding. This may be due to mis-priming of the DNA during PCR or background signal because, by RT-(q)PCR RIP efficiency for mut sponge was 0% (see **Figure 3.18C**). Nevertheless, the WT sponge still shows a 10% higher efficiency in binding to hnRNP L. Quantitative analysis of hnRNP L binding *in vivo*, by RT-(q)PCR showed a 6% RIP efficiency for T-$(CA)_{20}$ whereas WT sponge demonstrated a mere 3% RIP efficiency. While there was absolutely no interaction of hnRNP L with mut sponge, SLX4x seemed to interact with hnRNP L most efficiently (57% RIP efficiency). Surprisingly, compared to previous report, our results show a two-fold increase in RIP efficiency of hnRNP L with T-SLX4x and a slight decrease for T-$(CA)_{20}$, which showed a 10.7% RIP efficiency (Schreiner *et al*., 2020). Although WT sponge shows low RIP efficiencies, it is no cause for concern because a plausible explanation would be that, the very low RIP efficiency is a result of very high overexpression of the sponge. As observed from the Northern blot (**Figure 3.16A**), WT and mut sponges were expressed in almost three-fold higher amounts than T-SLX4x. Consequently, we speculate if all the hnRNP L in the cell is sponged, no protein is available for immunoprecipitation, therefore yielding low RIP-efficiencies. In any case, these observations confirm specific interaction of hnRNP L with sponges *in vivo*.

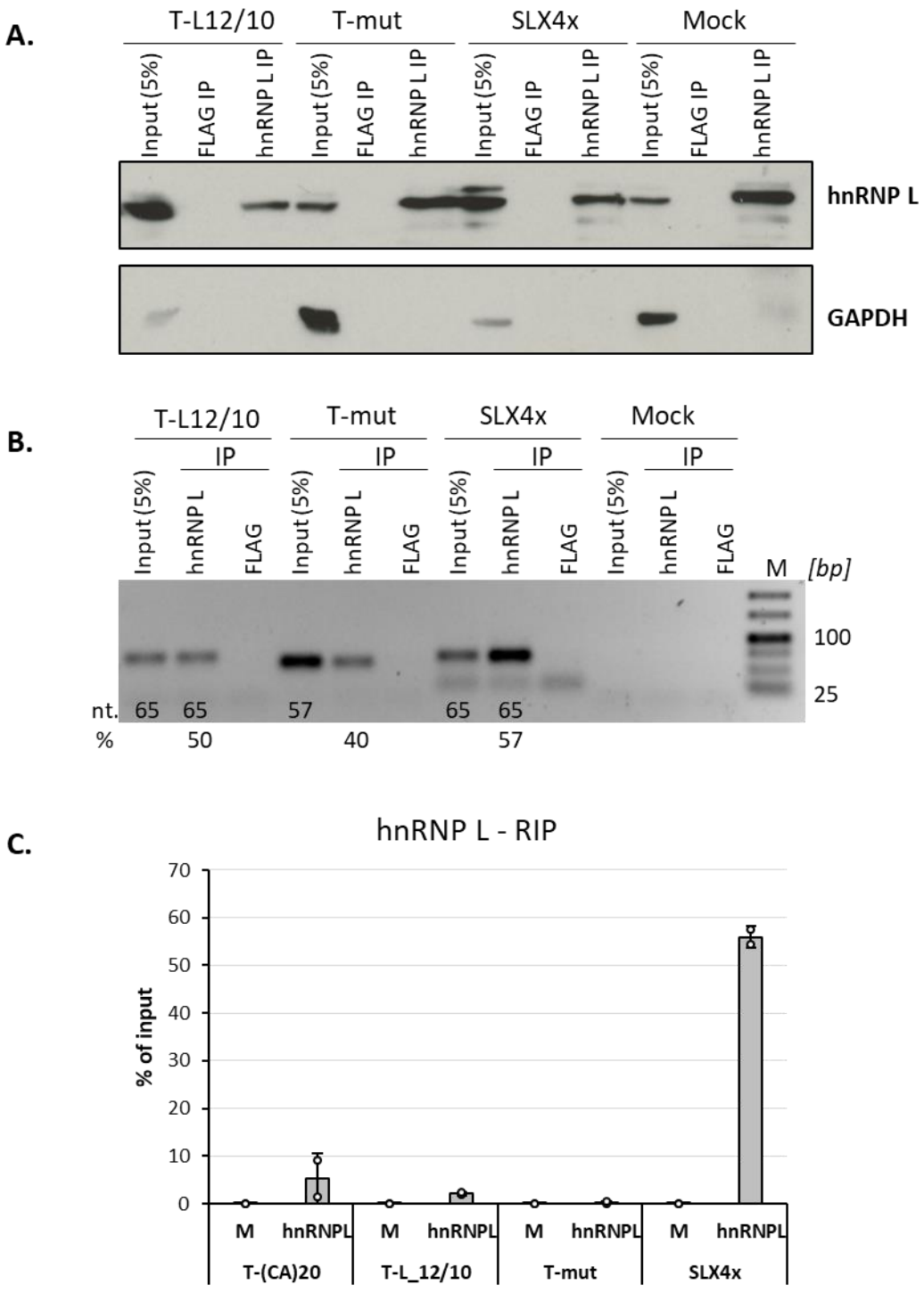

Fig. 3.18: Tornado-overexpressed circRNAs interact with hnRNP L *in vivo*.

HeLa cells were transfected with the T-$(CA)_{20}$, T-L12/10, T_L12/10_mut and SLX4x Tornado constructs. Cell lysates were prepared two days after transfection and immunoprecipitated using anti-hnRNP L or –FLAG (negative control) antibodies.

(A) Western blot demonstrating hnRNP L binding in RIP experiment. Input and IP samples were analysed by Western blot with antibodies against hnRNP L and GAPDH control.

(B,C) After RNA isolation from IP and input samples, transfected circRNAs were detected by **(B)** RT-PCR and **(C)** RT-qPCR. The fraction of bound target RNAs was calculated for each target relative to the corresponding input fraction – %RIP efficiency. Error bars represent standard deviations (n=3). *M*, Mock IP.

Taken together, this section establishes a framework for the overexpression of circRNA protein sponges using the Tornado system. Sponges designed to bind hnRNP L were overexpressed and resistant to RNase R digestion. The sponges also interacted with endogenous hnRNP L *in vivo*, giving us the first hints of hnRNP L sponging by designer circRNAs. The functional effects and relevance of such protein-circRNA sponge interactions will be discussed in the following sections.

3.8 Functional analysis of hnRNP L sponges

3.8.1 Design of optimal circRNA sponge

As a final step in establishing protein-sponging as a plausible function of circular RNAs, we tested artificially-made circRNA sponges for hnRNP L-sponging. Specific-binding of hnRNP L to SELEX-derived RNAs was demonstrated for *in vitro* transcribed linear RNA (see *Results* **section 3.1 & 3.5**), as well as for Tornado-expressed circular RNA *in vivo* (see *Results* **section 3.6**). These results indicated binding of both recombinant and endogenous hnRNP L to the SELEX-derived sequences, both *in vivo* and *in vitro*. However, we had not tested yet *in vitro* circularised RNA for binding to hnRNP L and differences, if any, to linear RNA-binding. Therefore, we designed three short RNAs and tested binding to endogenous hnRNP L. SELEX-derived circRNA sponges L_12/10 (WT) and L_12/10_mut, along with CA-sponge - $(CA)_{20}$, were modified by adding a stem sequence to enhance circularisation *in vitro* (see *Methods* **section 2.2.14**).

Secondary structure prediction of sponge sequences revealed accessibility of hnRNP L binding motifs both for the L_12/10 WT-sponge as well as for CA-sponge (see **Figure 3.19**). Due to redundancy of the CA-repeats, secondary structure of $(CA)_{20}$ CA-sponge reveals an open structure easily accessible to binding by hnRNP L. On the other hand, WT-sponge structure also indicates that the binding-motifs are fairly accessible to the four RRM-domains of hnRNP L; however, binding motifs in mut sponge are not so easily accessible to hnRNP L.

Several sponges were generated *in vitro* by T7-transcription and *in vitro* circularisation by T4-RNA ligase, namely- SELEX-derived sponges, L_12/10, L_12/10_mut, CA#51 and CA-sponges $(CA)_{10}$ and $(CA)_{20}$. Linear RNA-sponge and circRNA-sponge were distinguished from each other by their differential running behaviour on denaturing polyacrylamide gel. Due to their circular configuration circRNAs either shift upward (usually >50nt) or downward (<50nt) based on size of the RNA and percentage of PAA (see **Figure 3.20A**).

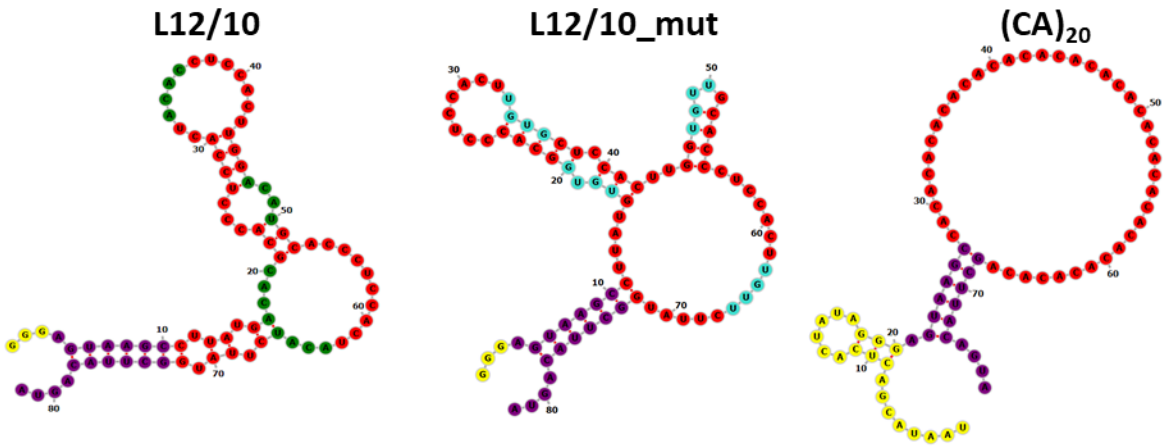

Fig. 3.19: Secondary structures of hnRNP L circRNA sponges.

Predicted circRNA secondary structures for L_12/10, L_12/10_mut and $(CA)_{20}$. HnRNP L-binding motifs in L_12/10 are marked by green. Turquoise represents mutated binding sites in L_12/10_mut and the twenty CA-repeats in $(CA)_{20}$ are shown in red. The stem sequence, which enhances circularisation *in vitro,* is shown in purple.

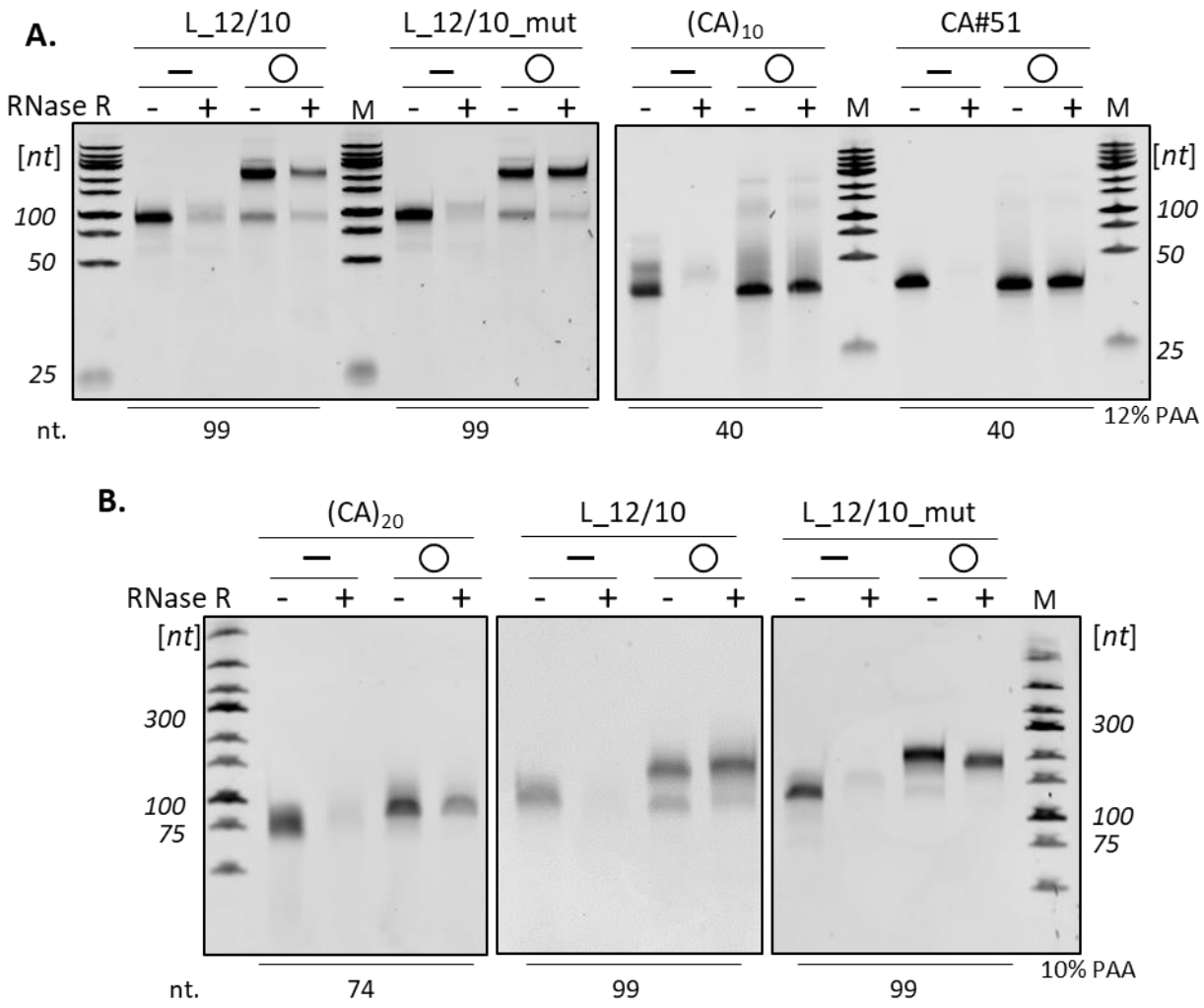

Fig. 3.20: *In vitro* generated hnRNP L circRNA sponges are stable and resistant to RNase R digestion.

CircRNA sponges for hnRNP L sponging were synthesised *in vitro* by T7 transcription, followed by RNA ligation. Five different RNAs were synthesised based on experimental design and requirement:- sponges with CA-repeats $(CA)_{10}$ and $(CA)_{20}$, SELEX-derived sponges L_12/10 and CA#51, and L_12/10_mut.

(A) Both short circRNAs $(CA)_{10}$ and CA#51 and longer circRNAs $(CA)_{20}$, L_12/10 and L_12/10_mut show differential mobility compared to the linear configuration on a 12% polyacrylamide gel. Furthermore, circularity was validated by integrity of circRNA in the presence (+) or absence (-) of RNase R.

(B) circRNA mobility on a 10% polyacrylamide gel. Linear RNA is marked with a horizontal line and circRNA is indicated by a circle. Sizes of each RNA are indicated at the bottom of the gel (in nt). *M*, marker (Gene ruler low range).

$(CA)_{10}$ and CA#51 (both 40nt) circRNAs shift slightly downward on a 12% PAA while $(CA)_{20}$, L_12/10 and L_12/10_mut (>50nt) circRNAs shift upward both on a 10% and 12%PAA. Circular configuration of each sponge was further tested by its resistance to digestion by RNase R. Since circRNAs lack 3' and 5' ends like their linear counterparts, they exhibit resistance to digestion by RNase R, in contrast to linear RNAs (see **Figure 3.20**).

Interaction of *in vitro* generated circRNA sponges with endogenous hnRNP L was captured by RIP assay, in a manner similar to Tornado-expressed sponges (see *Results* **section 3.7**). HeLa cells were transfected with equimolar amounts of *in vitro* synthesised circRNA sponges – WT, mut and $(CA)_{20}$, which served as positive control for hnRNP L binding. Cell lysates from each transfection were probed with anti-hnRNP L and anti-FLAG, 48h post-transfection. FLAG antibody was used as non-specific, negative control for RNA-binding. Immunoprecipitated hnRNP L was detected by Western blot, GAPDH was detected as input control and RIP negative control. (**Figure 3.21A**). RT-PCR analysis of sponge RNA after immunoprecipitation with hnRNP L showed a selective enrichment of WT circRNA sponge, whereas no enrichment of mut sponge was detectable (**Figure 3.21B**). We used out-facing primers on the transcript (in-facing on the circRNA), such that it spans the circ-junction for detection of circRNA. These primers can distinguish circRNA from linear RNA (**Figure 3.21B** schematic).

Furthermore, by quantitative RT-(q)PCR, we observed >20% interaction of WT sponge with hnRNP L while mut sponge was unable to bind hnRNP L (**Figure 3.21C**) . Unfortunately, due to its small size and repetitive CA-repeat sequences, we were unable to design compatible primers for the detection of $(CA)_{20}$ sponge. Primers that were designed targeting the circ-junction were unsuccessful in detecting circRNA. RT-(q)PCR analysis showed very high C_T-values, therefore data for $(CA)_{20}$ has not been shown. However, we showed specific interaction of hnRNP L to our SELEX-derived circRNA sponge *in vivo*. This specificity can be attributed to the binding motifs on the RNA since, mut sponge with a similar sequence, but differing in the binding motifs shows almost no binding at all. Additionally, we observed in general, that mut circRNA sponge is less stable than the WT sponge. This is one of the reasons why we detect only a very faint signal for the input fraction of mut sponge by RT-PCR analysis. Although no binding of hnRNP L with mut sponge is expected, presumably any interaction of mut sponge with hnRNP L causes degradation of the circRNA, further adding to instability. Instability of mut circRNA was also tested by a time-course experiment and Northern analysis (see **Figure 3.23**). Interestingly, by RT-(q)PCR, we were unable to identify hnRNP L interaction with linear sponge. Linear RNA-sponges seem to bind hnRNP L less-efficiently compared to circRNA. This observation was true not only for RIP assays but also for *in vivo* splicing assays – a key finding which will be discussed in detail in the following sections.

To overcome the limitation posed by RT-PCR approaches to detect small circRNAs, as an alternative, we identified circRNA-hnRNP L interaction by Northern analysis after RIP experiment, (**Figure 3.22**). HeLa cells were transfected with circRNA sponges, hnRNP L-RNA interactions were assayed by RIP and the extracted RNA after RIP assay was tested by a Northern blot. Enriched RNA from input (5% of total RNA), hnRNP L IP and FLAG IP were probed with a DIG-labelled riboprobe specific to the circ-junction of each circRNA sponge. The RNA-probes were specific to each circRNA, and could bind to both linear, as well as the circular configuration of RNA. The two configurations were distinguished on gel by their differential mobility. Usually, due to very little material bound by the protein in RIP assay, Northern detection after RIP is uncommon; however, we successfully detected all three sponges after RIP assay by Northern blotting (see **Figure 3.22**).

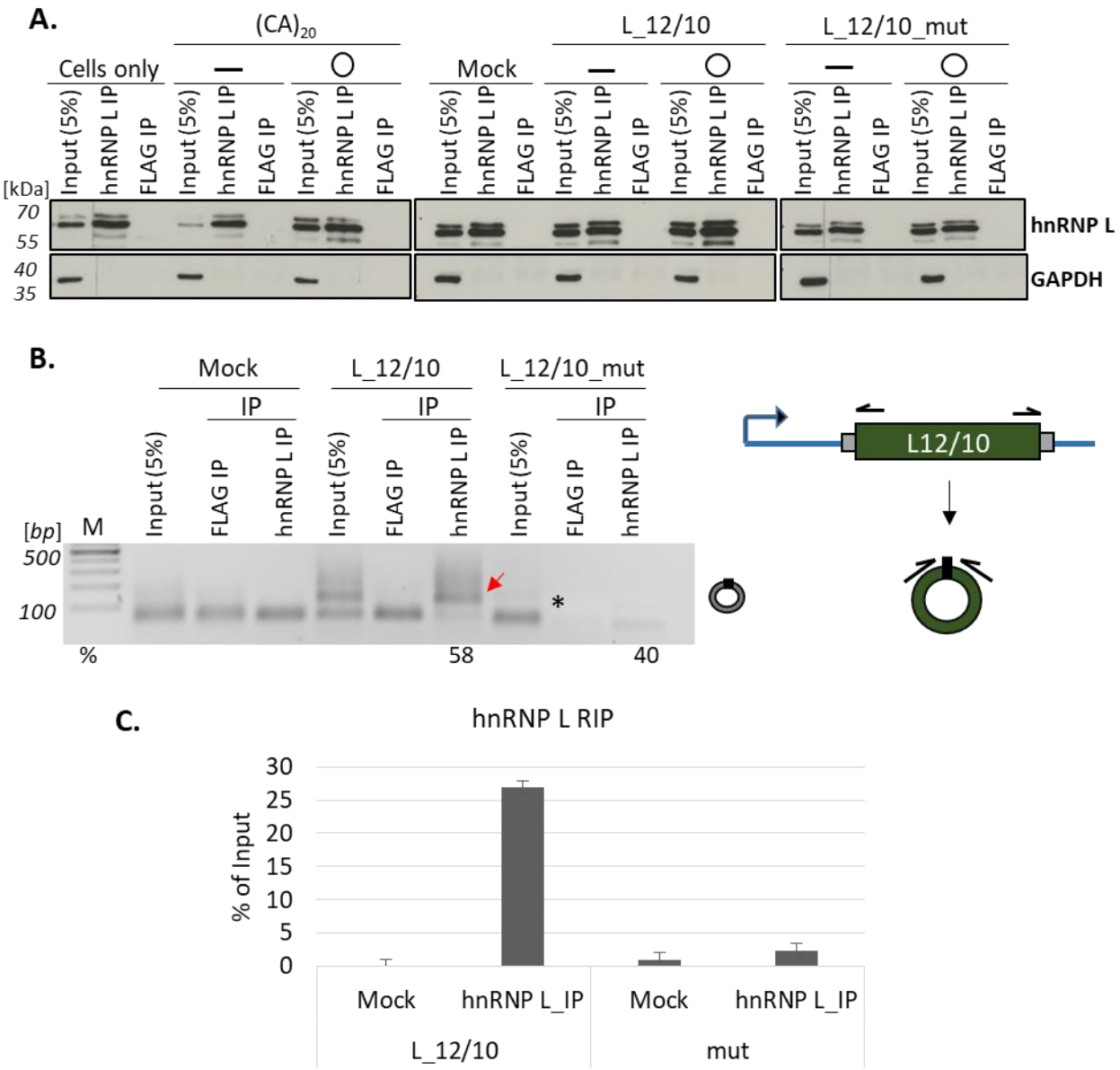

Fig. 3.21: *in vitro* generated circRNA sponge interacts with hnRNP L *in vivo*.

HeLa cells were transfected with the $(CA)_{20}$, L_12/10 and L_12/10_mut circRNA sponges. Cell lysates were prepared two days after transfection and immunoprecipitated using anti-hnRNP L or anti–FLAG (negative control) antibodies.

(A) Western blot demonstrating hnRNP L binding with linear(-) and circRNA (O) in RIP experiment. Input and IP samples were analysed by Western blot with antibodies against hnRNP L and GAPDH control.

(B,C) After RNA isolation from IP and input samples, transfected circRNA sponges were detected by RT-PCR **(B)** and RT-qPCR **(C)**. The fraction of bound target RNAs was calculated for each target relative to the corresponding input fraction – %RIP efficiency. Error bars represent standard deviations (n=3). Primers used for RT-PCR are shown schematically in panel (B). Red arrow indicates immunoprecipitation of hnRNP L with WT sponge, detected by RT-PCR. Asterisk denotes expected band size in input fraction of mut sponge. %RIF efficiency by RT-PCR, was calculated for WT and mut sponge relative to corresponding input fraction, indicated below the gel image.

Northern analysis of RIP samples revealed that WT sponge was most efficient in binding to endogenous hnRNP L (see **Figure 3.22**). Both CA-sponge as well as WT SELEX-sponge in their circular configuration bound hnRNP L specifically and efficiently, whereas mut sponge did not show any interaction with hnRNP L. Strikingly, interactions with linear RNA were completely abolished both for CA-sponge and WT sponge. Although linear RNA is detected, the major product is a circRNA and as expected, linear products arise out of re-linearisation of circRNA and also degradation after interaction with hnRNP L. A mock transfection control was also maintained and U1 snRNA served as loading control. On a relative basis, compared to the linear and circular standards (1ng each), it appears that there is almost a 100% RIP efficiency with the WT sponge for hnRNP L. This observation supports our conclusion of preferential binding of hnRNP L to circRNA over linear RNA. This observation has also been validated in a functional context for hnRNP L (see *Results* **section 3.8.1**).

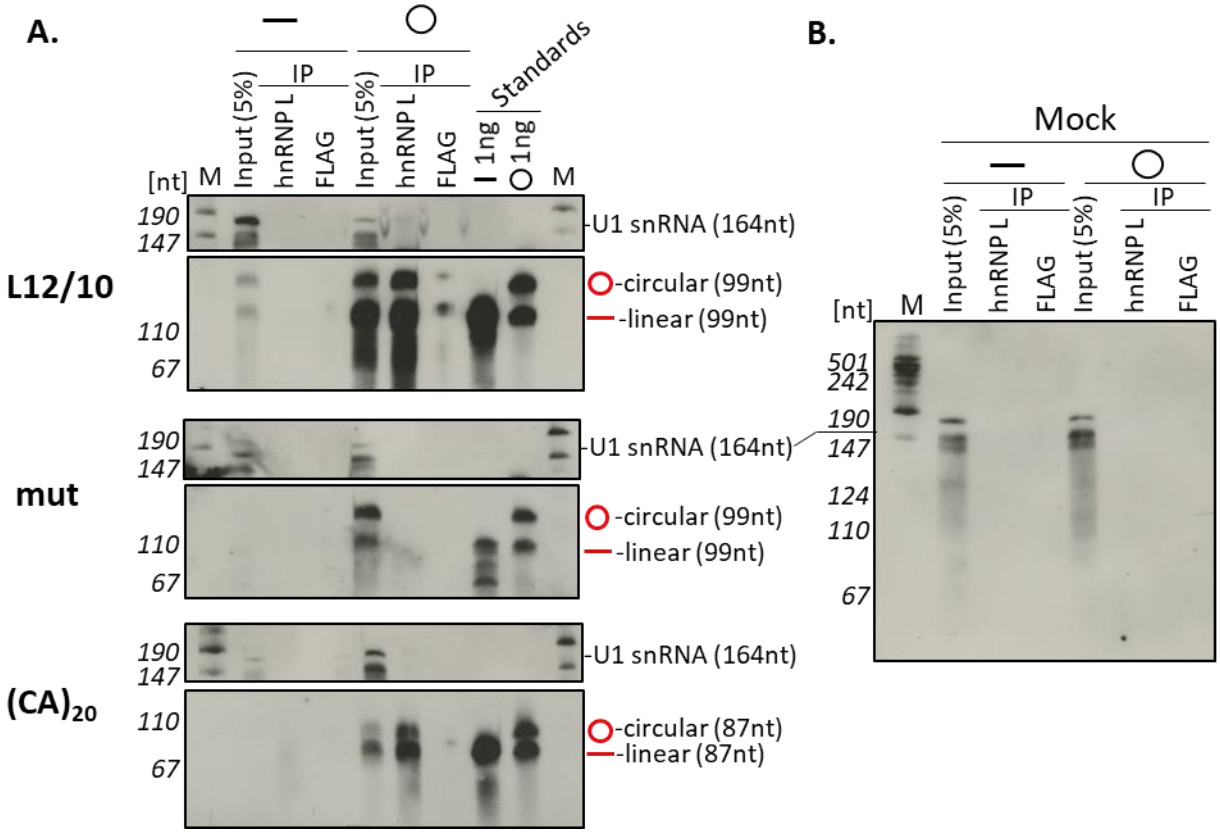

Fig. 3.22: Northern analysis confirms circRNA sponge interaction with hnRNP L *in vivo*.
HeLa cells were transfected with $(CA)_{20}$, L_12/10 and L_12/10_mut, linear and circRNA sponges. Cell lysates were prepared two days after transfection and immunoprecipitated using anti-hnRNP L or anti–FLAG (negative control) antibodies. RNA extracted after immunoprecipitation was analysed by Northern blot.
(A) CircRNA sponge was detected using a specific DIG-labeled riboprobe spanning the circ-junction of each variant. Total RNA after immunoprecipitation with hnRNP L was analysed on a 10% denaturing polyacrylamide gel. U1 snRNA served as loading control. Linear RNA is indicated with a horizontal straight line and circRNA with a circle.. 1ng each of linear and circular RNA was loaded as standards for Northern blot for each variant.
(B) Northern analysis of mock transfected cells after RIP assay.

Finally, before testing the functional implications of hnRNP L sponging, we decided to first test the stability of circRNA after transfected. HeLa cells were transfected with linear and circular WT, mut and CA-sponge; total-RNA was extracted at different time points (1, 2, 3 and 4 days post-transfection) and analysed by a Northern blot. Using sponge-specific DIG-labelled riboprobes, we identified all transfected sponges (**Figure 3.23**). Linear and circRNA sponges were distinguishable due to differential running behaviour, and compared to linear and circRNA standards.

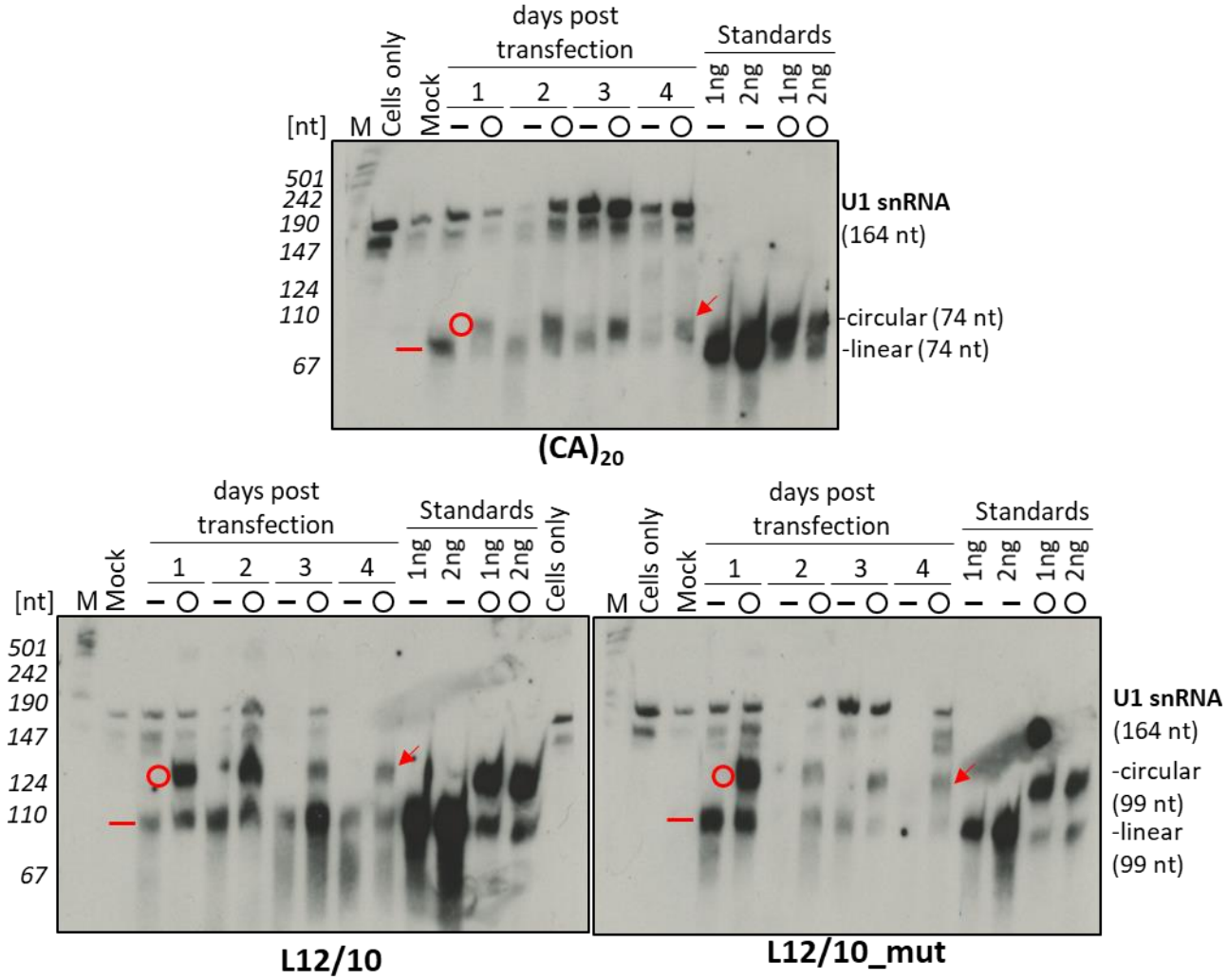

Fig. 3.23: Time-course experiment demonstrates circRNA stability by Northern blot detection.

HeLa cells were transfected with *in vitro* generated circRNA sponges - $(CA)_{20}$ (top panel) L_12/10 and L12/10_mut (bottom panels). RNA was extracted after 1, 2, 3 and 4 days post transfection. Using a specific DIG-labeled riboprobe spanning the circ-junction of each variant, all transfected sponges were detected. 500ng of total RNA for each variant was analysed on a 10% denaturing polyacrylamide gel. U1 snRNA served as loading control. Linear RNA is indicated with a horizontal straight-line and circRNA with a circle. For each transfection a mock control and cells-only control were maintained. 1ng and 2ng each of linear and circular RNA was loaded as standards for Northern blot for each variant. Linear and circular RNA on the blot are indicated in red, and circRNA stability four days post-transfection is indicated with a red arrow.

Time-course experiment by Northern analysis of transfected hnRNP L-sponges revealed that circular configuration of sponges is more stable to *in vivo* degradation by cellular nucleases compared to their linear counterparts. It must be noted that, since both linear and circRNA essentially have the same sequence, the stability of circRNA is a result of different configuration of the RNA. Interestingly, WT circRNA-sponge appears to be most stable, even four days after transfection compared to CA-sponge and mut sponge, which is least stable (see **Figure 3.23**). Linear RNA in all three cases degrades very quickly, creating smears on the blot. It appears that >50% circRNA also re-linearises, from third day post transfection. However, compared to the standards, a small amount circRNA is still available for sponging.

In summary, we have designed and tested artificial circRNA sponges for their ability to bind to hnRNP L, and our results indicate successful protein-binding by designer circRNAs. By RIP assays and Northern analysis of transfected sponges we demonstrate superior stability of circRNAs over linear RNAs and their specificity for hnRNP L-binding. An interesting observation was the preferential binding of hnRNP L to circRNAs over linear RNAs. The next step in establishing protein-sponging function of designer circRNAs would be to test the effects of hnRNP L in a functional context. Since, hnRNP L is an alternative splicing regulator, we next tested the effects of hnRNP L-sponging on alternative splice patterns of hnRNP L target genes. This will be discussed in the following section.

3.8.2 *In vivo* splicing assays

In addition to mediating important cellular functions such as telomere maintenance, chromatin remodelling and DNA repair. HnRNP proteins mainly function as repressors of splicing by binding to the pre-mRNA and prevent other splicing factors from accessing the binding site (Wang & Brendel, 2004b). HnRNP L is a global regulator of alternative splicing (Zhou *et al*., 2002; Hui *et al*., 2003a) and can function both either as activator or repressor of splicing. By siRNA-mediated knockdown of hnRNP L, we demonstrate loss of repressor-function of hnRNP L in alternative splicing of target genes (**Figure 3.24**). HnRNP L was specifically knocked-down using commercially available siRNA (human hnRNP L 3' UTR – see *Materials* **Table 2.1**) in HeLa cells, and changes in alternative splicing patterns were observed by RT—PCR analysis of hnRNP L target genes. Knockdown of hnRNP L was confirmed by Western blot analysis of hnRNP L (**Figure 3.24A**), where mock transfection and control luciferase GL2 (see *Materials* **Table 2.1**), did not affect levels of hnRNP L. Alternative splicing was assayed for four known targets of hnRNP L – *TJP1*, *BPTF* (*FALZ*), *RIF1* and *GPBP1*, relative to luciferase control (**Figure 3.24B**). In each case hnRNP L functions as a repressor of an alternatively spliced exon, promoting exon skipping in the targets (Hung *et al*., 2008). By knockdown of hnRNP L this function is compromised and therefore we observe more exon inclusion. For *TJP1* and *BPTF* (*FALZ*), an increase in exon inclusion was observed from 22% to 32% and from 24% to 48%, respectively. *RIF1* and *GPBP1* showed slightly more exon inclusion from 28% to 40% and from 28% to 41%, respectively.

Using this model system, we tested if our designer circRNA sponges could mimic hnRNP L knockdown by sponging the protein instead of knocking-down the mRNA. Both Tornado-expressed sponges (see **Supplementary Figure S8**) as well as *in vitro* synthesised circRNA successfully sponged hnRNP L (**Figure 3.25**).

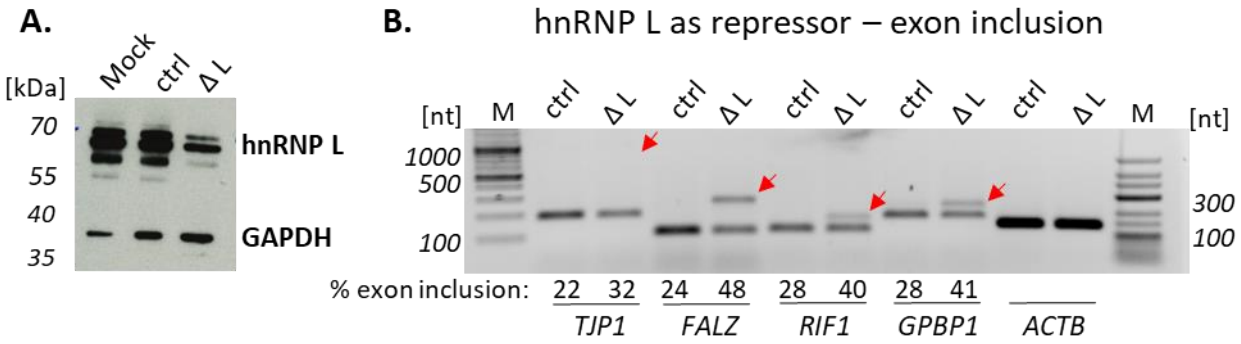

Fig. 3.24: Alternative splicing is affected by siRNA-mediated knockdown of hnRNP L.

HeLa cells were transfected each with equimolar concentrations of siRNA against either hnRNP L (Δ L) or luciferase (ctrl), and total RNA was extracted one day post-transfection.

(A) Western blot analysis of hnRNP L knockdown efficiency by siRNA, using anti-hnRNP L and anti-GAPDH antibodies.

(B) Alternative splicing of endogenous mRNA targets of hnRNP L – *TJP1, BPTF* (*FALZ*), *RIF1* and *GPBP 1* (exon inclusion) was assayed by semi-quantitative RT-PCR. Percentage of exon inclusion is shown below. *ACTB* served as control for RT-PCR. Δ L and Ctrl designate knockdown of hnRNP L and luciferase respectively. *M*, marker.

Differences in splicing *in vivo*, as a result of hnRNP L sponging is specific to WT-sponge and CA-sponge RNA. Since mut sponge cannot bind to hnRNP L, no effect was observed on exon inclusion. Moreover, splicing effects were dose-dependent, i.e., with increasing sponge-RNA transfection there was increased exon inclusion for WT-sponge and CA-sponge RNA. Around $1x10^6$ HeLa cells were transfected with increasing amounts of RNA (0.5, 1 & 2µg), and 1µg sponge RNA per-million cells was found to be the most effective dosage (see **Figure 3.25**).

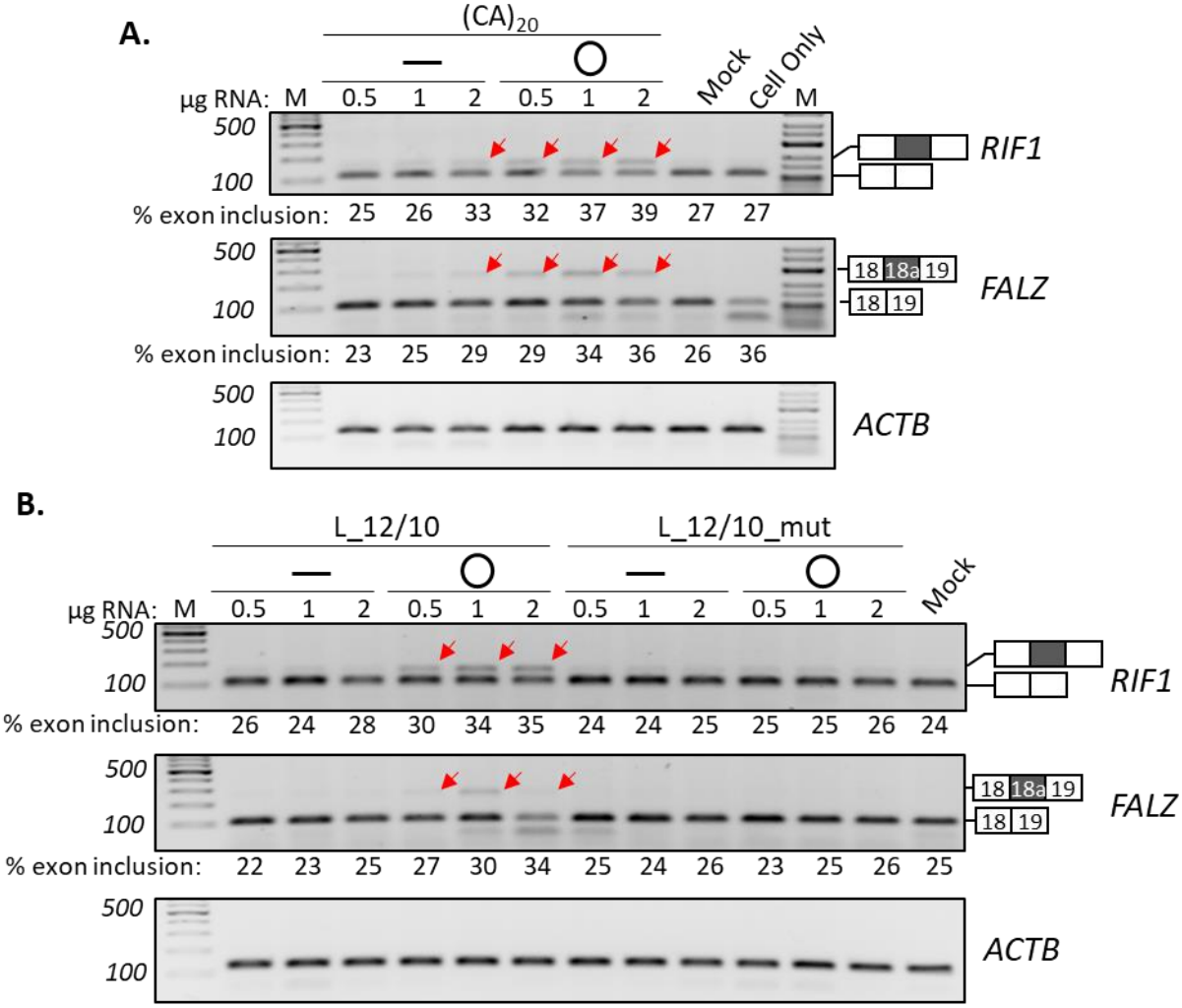

Fig. 3.25: Dose-dependent modulation of alternative splicing by hnRNP L sponging.
(A,B) HeLa cells were transfected with 0.5µg, 1 µg or 2 µg of *in vitro* generated linear and circRNA sponges - $(CA)_{20}$ **(A)** L_12/10 and L12/10_mut **(B)**. RNA was extracted two days post-transfection. Alternative splicing of endogenous targets of hnRNP L – *BPTF* (*FALZ*) and *RIF1* (exon inclusion) was assayed by semi-quantitative RT-PCR. Percentage of exon inclusion is shown below. *ACTB* served as loading control for RT-PCR. *M*, marker.

Interestingly, the maximal sponging effect on hnRNP L was observed in each case for circRNA sponge when compared to linear sponges. This further validates the efficiency of binding of hnRNP L to circRNA over linear RNA. In addition to *RIF1* and *FALZ*, we tested two more targets (*TJP1* and *GPBP1*) where hnRNP L acts as repressor of splicing (**Figure 3.26**). In all genes tested,

repressor-activity of hnRNP L was compromised due to sponging and the circRNA sponge was superior to the linear sponge in binding to hnRNP L.

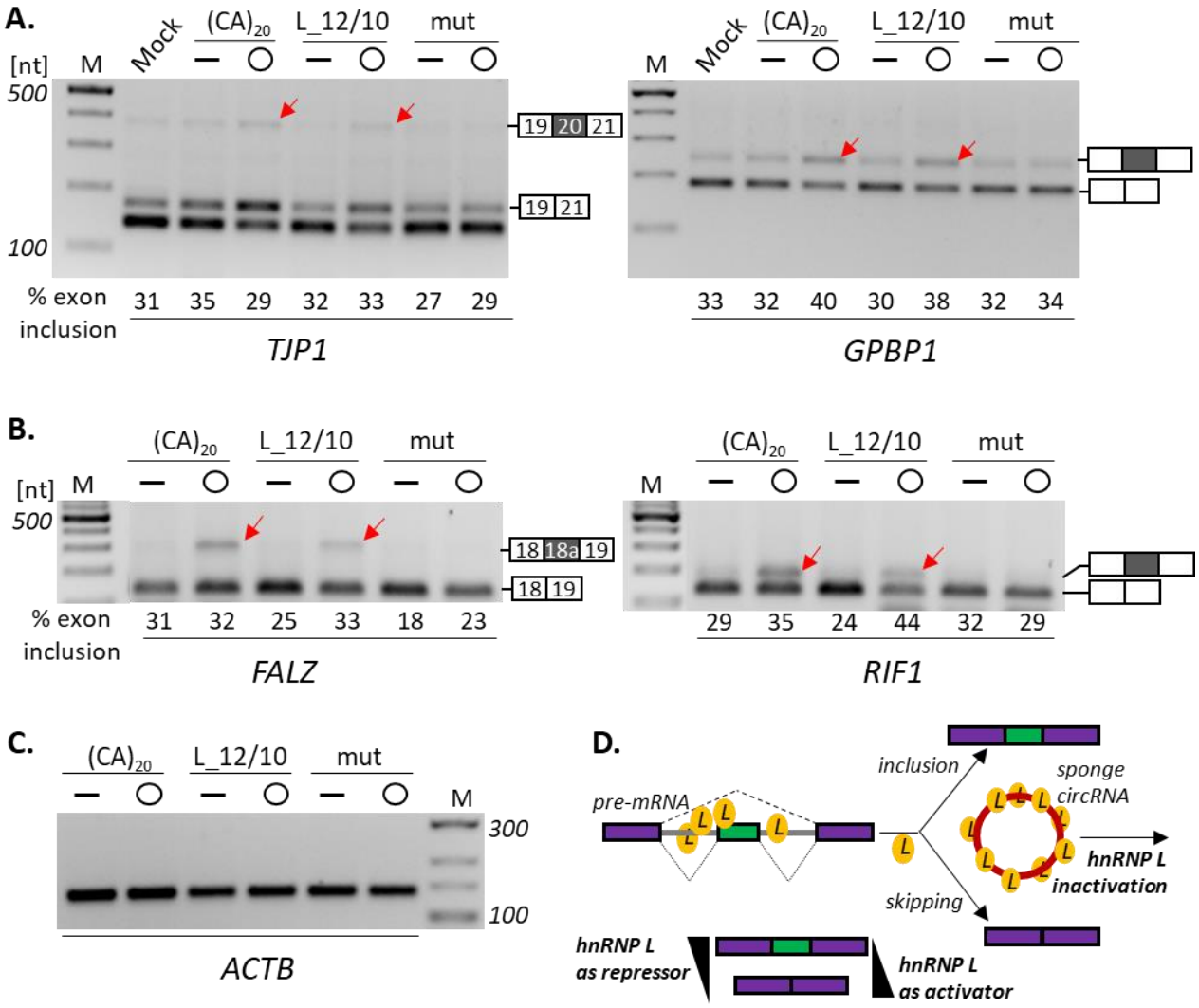

Fig. 3.26: Transfected circRNA sponges affect hnRNP L alternative splicing targets in HeLa cells (exon inclusion).

(A,B,C) Total RNA was extracted from HeLa cells transfected with *in vitro* generated linear and circRNA sponges - L_12/10, L12/10_mut and $(CA)_{20}$ two days post-transfection. Isolated RNA was used for RT-PCR with primers against hnRNP L alternative splicing targets - *TJP1, GPBP1* **(A)**, *FALZ* and *RIF1* **(B)**. Percentage of exon inclusion is shown below the corresponding lanes. Positions of PCR products corresponding to alternatively spliced mRNAs are indicated with red arrows. ACTB served as loading control for RT-PCR **(C)**. Linear RNA is indicated with a horizontal straight line and circRNA with a circle. Exon inclusion is schematically represented on the right of each panel. *M*, marker.
(D) Schematic of alternative splicing modulation by circRNA sponges. HnRNP L regulates exon skipping and inclusion, acting either as splicing activator or repressor. Depending on whether hnRNP L acts as a repressor or activator, splicing decisions can be modulated by a sponge circRNA, which inactivates hnRNP L by sponging, resulting in a shift in the ratio of splice isoforms (skipping / inclusion). Exons represented as bars, skipped exon in green; circRNA is shown as a red circle and hnRNP L in yellow. From Schreiner *et al*. (2020), modified.

HnRNP L acts as a splicing repressor when bound to CA-rich ISS or ESS sites, hindering the recruitment of spliceosomal components, which in turn leads to exon skipping. When hnRNP L is limiting, its repressor function is minimised, resulting in partial exon inclusion. Splicing activator function of hnRNP L assumes hnRNP L-driven exon inclusion or intron removal, which is diminished when hnRNP L is bound by circRNA sponges. When hnRNP L regulates exon inclusion, hnRNP L sequestration promotes exon skipping (see **Figure 3.27**). HnRNP L acts as activator of splicing for *CC2D2A*, *HMMR* and *PPP3CB* and mediates exon skipping. Our results strongly indicate exon-skipping as a result of hnRNP L sponging by sponge RNAs.

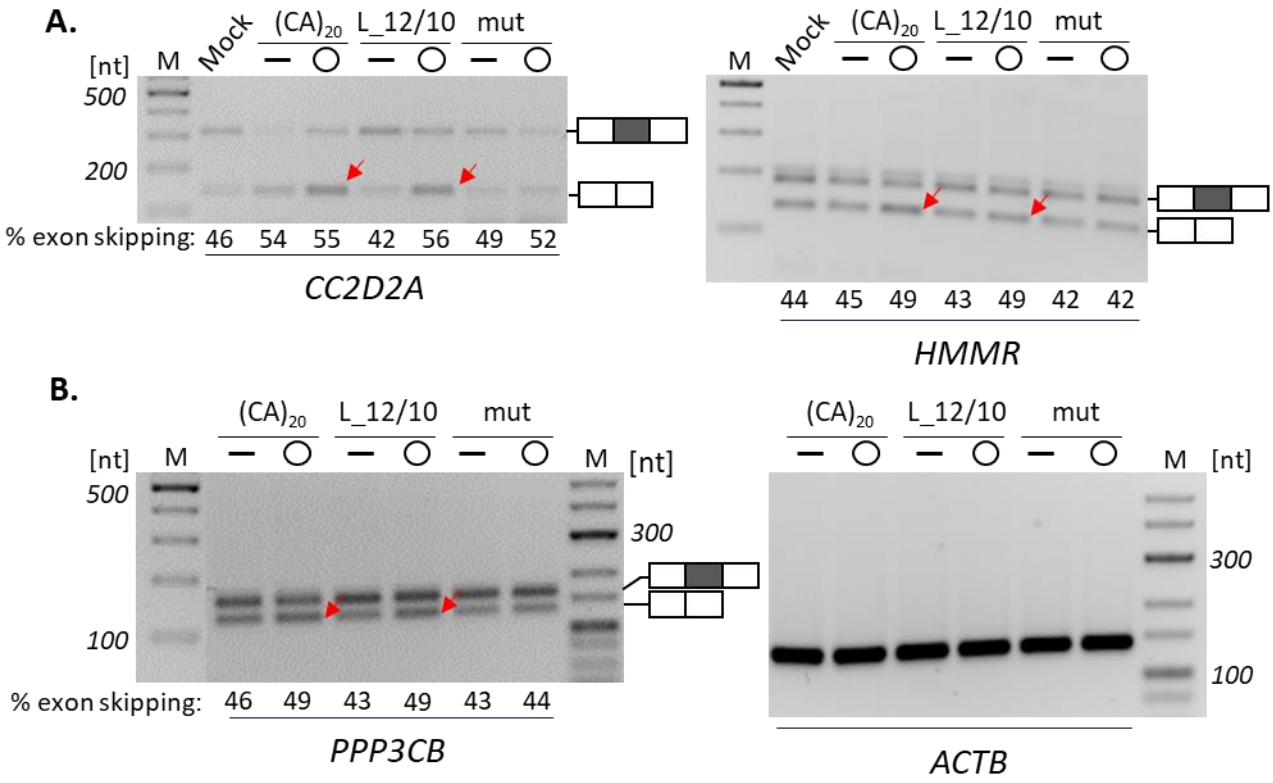

Fig. 3.27: Transfected circRNA sponge affect hnRNP L alternative splicing targets in HeLa cells (exon skipping).

(A,B) Total RNA was extracted from HeLa cells transfected with *in vitro* generated linear and circRNA sponges - L_12/10, L12/10_mut and $(CA)_{20}$ two days post transfection. Isolated RNA was used for RT-PCR with primers against hnRNP L alternative splicing targets – *CC2D2A, HMMR* **(A)** and *PPP3CB* **(B)**. Percentage of exon skipping is shown below the corresponding lanes. Positions of PCR products corresponding to alternatively spliced mRNAs are indicated with red arrows. *ACTB* served as loading control for RT-PCR. Linear RNA is indicated with a horizontal straight line and circRNA with a circle. Exon skipping is schematically represented on the right of each panel. *M*, marker.

The two-fold function of hnRNP L in acting as splicing activator or repressor, promoting exon skipping or exon inclusion, respectively, is schematically represented in **Figure 3.26D**. Sponging of hnRNP L by designer circRNAs inactivates hnRNP L as described, thereby causing changes in splicing of target genes.

These observations led us to conclude that designer circRNA sponges reproducibly and strongly affect alternative splicing in a variety of hnRNP L target genes where hnRNP L causes exon inclusion or skipping, based on different mechanisms of action. In general, hnRNP L sponging by designer circRNAs outperformed siRNA-mediated knockdown of hnRNP L in observed alternative splicing effects (compare **Figures 3.24** and **3.26**) and showed a dose-dependent effect (**Figure 3.25**). Tornado-expressed circRNAs showed maximal effects two days post-transfection, presumably due to the compromise between endogenous RNA turnover and Tornado-circRNA accumulation (data not shown). Besides the report from our own group (Schreiner *et al*., 2020), to our knowledge, this is the only other evidence for protein-sponge function of designer circRNAs. In our opinion, this observation is pivotal in understanding circRNA function and design, in the context of RNA-binding proteins.

3.8.3 Cell fractionation

Since hnRNP L shuttles between nucleus and cytoplasm with a predominant nuclear localisation (Pinol-Roma *et al*., 1989), we assayed the effects of hnRNP L sponging by designer circRNAs, on nuclear-cytoplasmic distribution of hnRNP L in cells expressing circRNA sponges (**Figure 3.28**). HeLa cells were transfected with designer sponges (Tornado-expressed, as well as *in vitro* generated circRNA sponges - $(CA)_{20}$, L_12/10 and L_12/10_mut) and mock control. Each sponge was transfected in linear as well as circular configurations. Twenty-four hours post–transfection, cells were fractionated into cytoplasmic and nuclear fractions. Equivalent lysate amounts of total cells, nuclear and cytoplasmic fractions were analysed for hnRNP L by Western blotting. GAPDH and hnRNP A1 were also identified as controls and normalisation for cytoplasmic and nuclear fractions, respectively. The cytoplasmic:nuclear ratio of hnRNP L for mock transfection is 36:64, based on Western signals. This ratio is comparable to the distribution seen in mut sponge transfection (36:64 for linear and 28:72 for circRNA) and CA-sponge (31:69 for circRNA). Contrastingly, with WT circRNA-sponge transfection this ratio dramatically shifts to 61:39, comparable to shifting of cytoplasmic-nuclear ratio observed by $(CA)_{100}$ sponge transfection (Schreiner *et al*., 2020).

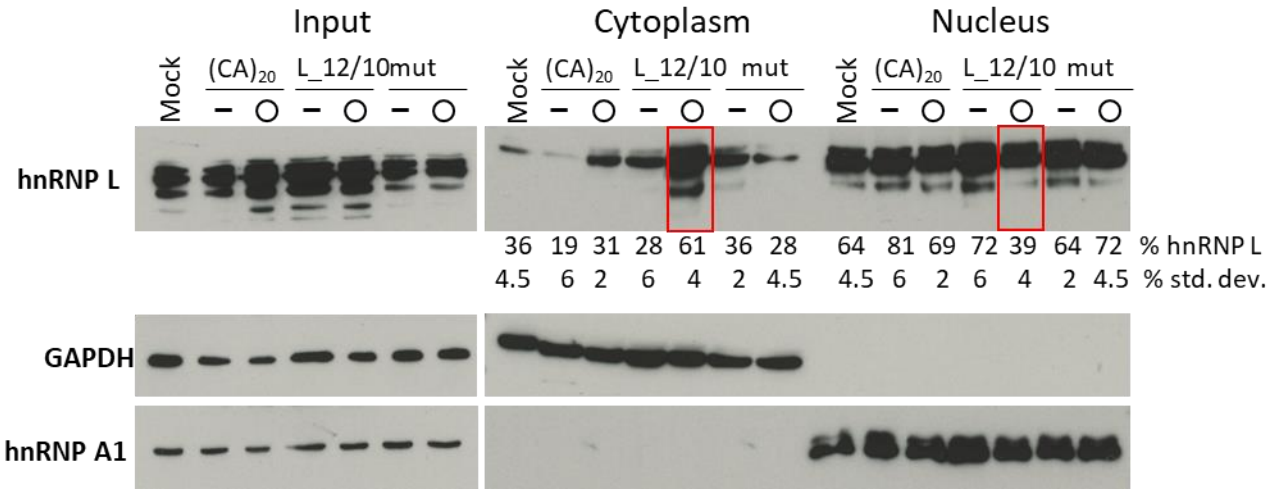

Fig. 3.28: Transfected circRNA sponge shifts nuclear-cytoplasmic distribution of hnRNP L. Small hnRNP L-sponge circRNA shifts nuclear-cytoplasmic distribution of hnRNP L. $(CA)_{20}$, L_12/10 and L_12/10_mut , each in linear or circular configuration (−, O), were transfected in HeLa cells, followed by cell fractionation after 24h. Equivalent lysate amounts of total cells, cytoplasmic and nuclear fractions were analysed by Western blotting for hnRNP L, GAPDH, and hnRNP A1. The distribution of hnRNP L between nuclear and cytoplasmic fractions was quantitated, based on Western signals (mean values and standard deviations given below the respective lanes; *n* = 3). HnRNP L distribution shifted by L_12/10 sponge, is marked with a red box.

Our results strongly argue for translocation of hnRNP L from nucleus to cytoplasm by the L_12/10 circRNA sponge. This strong effect is specific for the circular configuration of WT sponge and was not observed for mut sponge. However, total steady-state levels of hnRNP L did not change significantly. The circRNA-specific effect of WT sponge may be attributed to the differential stabilities of circular versus linear configurations which, most likely also differ between nuclear and cytoplasmic residence. Nuclear-cytoplasmic shift of hnRNP L was also observed for cells transfected with Tornado-expressed circRNA sponge (see **Supplementary Figure S9**).

In parallel, subcellular localisation of transfected circRNA sponges was also tested by Northern analysis (**Figure 3.29**). Tornado-expression constructs of all SELEX-derived sponges, namely – T-L_9/10, T-L_9/15, T-L_12/10, T-L_12/15 and T-L_12/10_mut, along with two positive controls T-$(CA)_{20}$ and T-SLX4x were transfected into HeLa cells. 48 hours after transfection, cells were harvested and fractionated into cytoplasmic and nuclear fractions. RNA was extracted from the respective fractions and analysed by Northern blotting using a riboprobe spanning the circ-junction and the Tornado-stem region. For each sponge, all three fractions, whole-cell (**Figure 3.29A**), cytoplasmic and nuclear fractions (**Figure 3.29B**) were analysed. All circRNA sponges were detected in the whole-cell fraction with varying intensities. This differential expression of Tornado-expressed circRNA sponges is sequence and size-dependent. With the exception of T-SLX4x, a predominant cytoplasmic localisation was observed for all SELEX-derived linear sponges. Circular sponges were either equally distributed between nucleus and cytoplasm, or showed a slightly higher cytoplasmic localisation. Surprisingly, T-SLX4x and T-$(CA)_{20}$ circRNAs showed a predominant nuclear localisation. However, lower expression and faint signals on Northern blot may suggest that sponges localised in the cytoplasm were undetected for T-SLX4x and T-$(CA)_{20}$. In contrast, T-L_9/15 sponge showed an exclusive cytoplasmic localisation. These differences in localisation patterns, we assume, are due to variable sequence-context and their ability to sponge hnRNP L.

Although the exact mechanism of nucleo-cytoplasmic transport of circRNA sponges is unknown, we think that these circRNAs after binding to hnRNP L passively diffuse through the nuclear pores into the cytoplasm. Based on our understanding so far, RNAL Pol III transcribed Tornado-sponges are synthesised in the nucleus and transported to the cytoplasm, and in the process they bind to hnRNP L, which is co-transported with the circRNA. However, this proposed mechanism needs more experimental validation. On the other hand, transfected circRNAs, due to their high-affinity to nuclear hnRNP L, are found to be equally distributed in the nucleus and cytoplasm. Interestingly, *in vitro* synthesised linear sponges show a predominant cytoplasmic localisation (data not shown), which is in line with the previous observation (Schreiner *et al*., 2020). This is most-likely due to the lower affinity of linear sponges to hnRNP L.

We conclude that designer circRNAs efficiently sponge hnRNP L and cause translocation of hnRNP L from nucleus to cytoplasm. This translocation may also have effects on the functional properties of hnRNP L such as alternative splicing, transport and/or localisation of mRNAs (Krecic & Swanson, 1999; Schreiner *et al*., 2020).

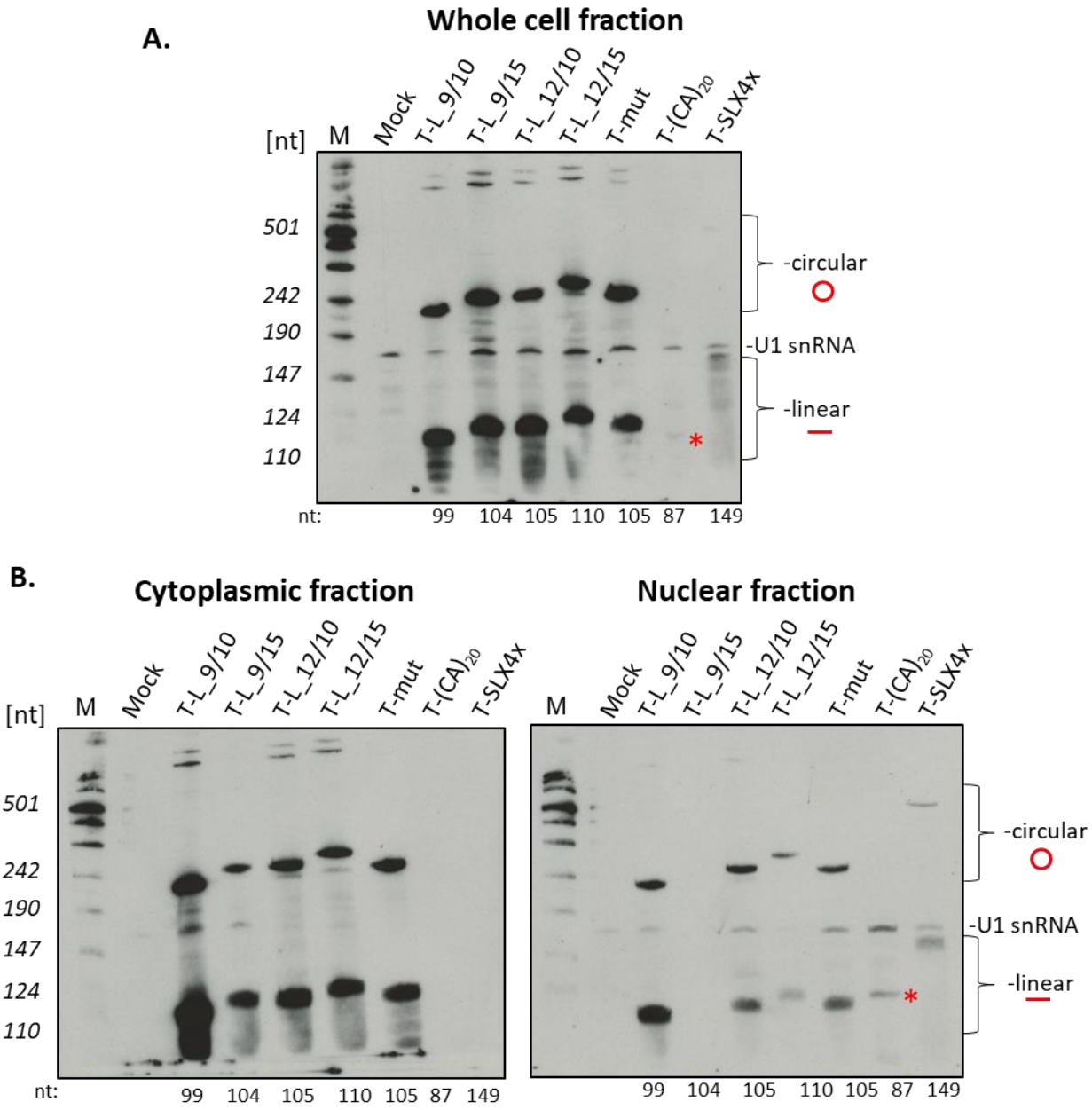

Fig. 3.29: Northern blot analysis after cell fractionation reveals predominant cytoplasmic localisation of transfected sponges

Total RNA was extracted from HeLa cells transfected with Tornado-expressed sponges for all SELEX-derived sequences (T-L_9/10, T-L_9/15, T-L_12/10, T-L_12/15, T-L12/10_mut and SLX4x) and T-$(CA)_{20}$ sponge.

(A) Using a DIG-labeled riboprobe spanning the circ-junction and specific to tornado stem region, all transfected sponges were detected.

(B) Equivalent amounts of cytoplasmic (left) and nuclear (right) RNA fractions were loaded on a 10% polyacrylamide gel. U1 snRNA served as loading control. Size of each RNA is shown below each blot. T-$(CA)_{20}$ circRNA is represented by a red asterisk.

4. Discussion

4.1 Determination of exact binding motifs for RBPs by SELEX

Systematic evolution of ligands by exponential enrichment, SELEX in short, is a widely known combinatorial method in molecular biology, for the directed evolution of oligonucleotide sequences that bind to selected biomolecules (Tuerk *et al*., 1990; Ellington & Szostak, 1990). SELEX has been described as an unbiased method for the determination of sequences that optimally bind to biomolecules in general, fulfilling specific functions. Therefore it can be applied to several examples where biomolecules interact, like DNA/RNA interaction with proteins, as ligands. The biggest advantage of this technique is the ability to determine specificity of binding to both small and large molecules. As a result, SELEX has been frequently used to determine *de novo* the RNA-binding specificity of RBPs (Schneider *et al.,* 2019). In this study, we applied this effective approach in determining the exact binding sequence of two important proteins – hnRNP L and RBM24, both of which regulate alternative splicing. RBM24, a single-domain RBP, regulates muscle-specific alternative splicing, while hnRNP L is a classical multidomain RBP, regulating splicing. How multidomain proteins exactly recognise their respective targets is still an open question in the field. What advantage in target recognition does a multidomain protein like hnRNP L possess over a single domain protein, such as RBM24? What roles do the individual protein domains play in binding? These are some questions that can be addressed by SELEX.

Owing to its versatility and specificity, *in vitro* SELEX followed by RNA-seq, has become a powerful tool in understanding how several elements are recognised at the same time by RBPs. Multidomain proteins have an obvious advantage in target recognition and binding compared to single-domain proteins, due to the multiple domains involved. HnRNP L has four RRM domains, all of which are involved in recognition and binding. Although the kinetics of motif recognition are not fully understood, it is most likely that the fidelity of recognition is maintained by combinatorial binding as opposed to a perfect 'lock-and-key' model of binding. Based on our understanding, we presume that hnRNP L recognises its target by means of one RRM binding specifically to one motif, as a result the other three domains simultaneously, occupy the adjacent motifs. This model of binding reduces the overall 'burden' on each domain diffusing through space to find their exact binding-motif, since the domains are tethered to each other. It also adds a probable proof-reading function to domain recognition i.e., when one domain binds to its respective motif in the RNA, if the tethered second or third domain does not find its exact binding motif, it disassociates and finds another target. Single-domain proteins like RBM24 on the other hand, need to rely on just one domain that recognises its target, which may be more challenging based on the kinetics of binding.

In view of this complex procedure required for RNA binding, SELEX in our opinion is the ideal method to decipher these interactions. We modified the SELEX approach by using a 40nt (N_{40}) random sequence instead of 20-nt long, frequently used random-RNA sequences. The rationale behind the approach is to accommodate as many binding motifs as possible for recognition by the multidomain hnRNP L protein. By introducing a long degenerate sequence for binding, one can expect that the interaction will be most-specific because at least two to three motifs will be bound by the protein. This approach has been thoroughly validated by

our group previously for another RBP, IMP3- a multidomain protein involved in posttranscriptional gene regulation (Schneider *et al.,* 2019; Bell *et al*., 2013).

Another layer of complexity in understanding protein-RNA interactions is the secondary structure of RNA and the tertiary/quaternary structure of the protein. Why and how multidomain proteins identify their targets can be explained through crystal structures, to understand the binding in a 3D context. Unfortunately, full-length hnRNP L has not been crystallised. However, individual domains of hnRNP L have been crystallised (Blatter *et al*., 2015), and whatever information we have concerning binding come from these partial structures.

Our work demonstrates a systematic experimental approach to study combinatorial RNA recognition by the multidomain RBP, hnRNP L and single domain RBM24. For hnRNP L, both affinity and specificity are achieved through simultaneous engagement of multiple domains with their respective RNA elements. Bioinformatic analyses can predict certain features of RNA recognition by multidomain proteins, but may not explain combinatorial recognition (Dominguez *et al*., 2018; Ray *et al*., 2013). In addition, commonly used global approaches to map protein–RNA interactions, such as CLIP and RIP, have been analysed with the aim to reveal short consensus sequences. By employing SELEX, a systematic analysis of multidomain RBPs, as well as rational searches for high-confidence and functional target sequences was made possible.

4.2 A new and efficient role of circRNAs in protein sponging

CircRNAs were re-discovered as a large class of non-coding RNAs in the last decade, and are present in almost all eukaryotes investigated so far (Jeck *et al*., 2013; Salzman *et al*., 2012). Several naturally occurring circRNAs have also been identified and characterised based on high-throughput sequencing and computational analyses (Chen, 2020; Kristensen *et al*., 2019). However, most of these findings have been unable to comprehensively determine the functions of circRNAs. Although some functions have been established, such as miRNA-sponging for the naturally occurring CDR1as sponge (Hansen *et al*., 2013) and the Sry sponge (Capel *et al*., 1993), functions of a vast majority of circRNAs identified remains elusive (Wilusz, 2018). Several hypothetical functions have been proposed for endogenous circRNAs such as protein-sponging, allostery, templates for translation (Hentze & Preiss, 2013), but upon critical analysis of the available data, sufficient evidence is lacking to prove these ideas. This is especially true for circRNA translation, which is highly debated (Hansen, 2021).

In view of this predicament, we used an alternative approach to study circRNA function. Instead of characterising each endogenous circRNA for a putative function in the cell, we designed circRNAs to fulfil a specific function. By this approach we identified circRNAs that can be categorised under a certain family of functions, such as protein-sponging. We designed and developed optimal circRNAs for RBPs, to function as protein sponges. Protein-sponging by circRNAs was previously only proposed for the circMbl circRNA, encoded by the *muscleblind* (*mbl*/*MBNL1*) gene in *D. melanogaster* (Ashwal-Fluss *et al*., 2014). However, this study did not satisfactorily demonstrate protein sponging of MBL protein by the circRNA. In contrast, artificial circRNAs have been shown to sponge proteins and this sequestration also

affects the functionality of the protein sponged (Schreiner *et al*., 2020). This thesis further establishes the principle of efficient protein-sponging by circRNAs for alternative splicing regulators such as hnRNP L and RBM24. Our contribution to understanding protein-sponging suggests that, since proteins can be specifically and efficiently sponged by designer circRNAs, it is plausible that such an interaction also exists in nature. It is quite possible that circMbl may be one of those naturally occurring protein-sponges capable to functionally affecting MBL by sponging, but this has to be investigated further.

It is therefore clear that a careful, systematic approach needs to be applied when defining circRNA sponge functions. Aspects such as stoichiometry of protein-circRNA interaction and size, must be considered while designing circRNAs. In general, for circRNA functions (like miRNA sponging) and specifically for protein sponging, it is necessary to understand the competition between circRNAs available for binding and the respective protein (or miRNA) levels in the cell. Our colleagues have demonstrated that designer circRNAs can indeed outcompete hnRNP L for binding (Schreiner *et al*., 2020). Our results consistently showed a high accumulation of Tornado-expressed circRNA, which was greater than 10^6 molecules per cell (comparable to U1 snRNA) (**Figure 3.16**). This further corroborates previous observations where Tornado circRNAs accumulated to levels of up to ~10^6-10^7 molecules per cell (Schreiner *et al*., 2020). Such high expression is clearly sufficient to quantitatively sponge endogenous hnRNP L, which is usually found at ~10^6 molecules per cell (Schreiner *et al*., 2020).

Another aspect to be considered is the accessibility of the circRNA to hnRNP L binding. Since circRNAs usually accumulate in the cytoplasm, they may not be accessible to hnRNP L, which is predominantly nuclear. However, sponge transfections revealed a dramatic shift in the cytoplasmic-to-nuclear ratio of hnRNP L distribution from 36:64 to 61:39 (**Figure 3.28**). This suggests molecular interaction between circRNA sponge and hnRNP L and translocation of hnRNP L as a result of sequestration. The exact mechanism of this translocation is unclear, and probably occurs by diffusion through nuclear pores.

However, an open question is, can increased circRNA expression or increased binding sites on the circRNA, affect its sponging function? CDR1as is an efficient miRNA sponge, partly because it possesses >70 highly specific and conserved binding sites for miR-7 (Hansen *et al*., 2013). Our hnRNP L-sponge SLX4x, is one such example of increased binding-efficiency with more binding sites (**Figure 3.18**). A SELEX-derived, high-affinity hnRNP L-binding sequence (CA# 51, 5’-ATACATGACACACACACGCA-3’; Hui *et al*., 2005), was used in four copies to increase the binding sites for hnRNP L binding. Using PIE (permuted intron–exon) overexpression vector (Wesselhoeft *et al*., 2018), up to sixteen-copies of CA#51 were incorporated into the circRNA by Anna Didio (data not shown). On the other hand, with the introduction of the Tornado-overexpression system, it is now possible to express circRNAs at high concentrations. Currently, in our opinion, the Tornado system is the most efficient overexpression system for small circRNAs. Although Pol III-driven transcription from tRNA-based vectors limits circRNA size to ~800nt (Noto *et al*., 2017) and also requires the absence of poly(T) stretches, circRNA expression by Tornado system is efficient and robust. In comparison, circRNA expression by vector systems based on backsplicing (Wesselhoeft *et al*., 2018) are limited in terms of low yield and contaminating linear RNA and RNA concatemers. The choice of circRNA expression

system therefore depends on several factors such as length, secondary structure and composition of sequence of interest. Notwithstanding, future experiments incorporating more protein binding-sites in the circRNA and/or consequent high-expression, may be necessary to fully understand and further optimise protein sequestration by circRNA sponges.

4.3 Global gene regulation through protein sponging

Natural biological processes like embryonic development, organ formation and disease pathogenesis are governed by highly coordinated genetic /gene regulatory networks (Singh *et al*., 2018). Differential expression of regulatory genes induce differential expression of their respective target genes, by varying mRNA and protein levels, which in turn determine cellular functions. Differential expression of these genes is brought about by regulatory molecules which include DNA, RNA and proteins. Interactions between these molecules, or complexes of these, determine gene expression. Quantitative transcriptome profiling by microarray hybridization and high-throughput RNA-seq (Mortazavi *et al*., 2008), have enabled the systematic mapping of genes whose expression levels are correlated. Such a correlation in expression levels across biological samples established gene expression correlation networks, which provide genome-scale views of gene regulation in the context of events like haematopoiesis, oncogenesis, embryogenesis and inflammation (Singh *et al*., 2018). However, early research on these networks only considered gene-gene interactions because eukaryotic expression is thought to be hierarchically organised i.e., protein products of regulated genes in turn regulate expression of other genes, often involving feedback mechanisms. Network analysis has since been modified to include interactions among cellular molecular constituents of other types such as protein-metabolite networks as well as protein-protein interaction networks, to arrive at a more global understanding of these complex networks – a branch of science now known as systems biology.

Alternative splicing networks are key players in global gene expression and are embedded into these gene regulatory networks. Hung *et al.* (2008) demonstrated specific changes in alternative splicing patterns on a global scale by microarray analysis and knockdown of hnRNP L, a splice regulator. In this study, the authors have convincingly shown that knockdown of hnRNP L promoted exon skipping (in *CC2D2A* and *HMMR* etc.), exon inclusion (in *TJP1* and *FALZ* etc.) and intron retention (in *DAF1*), thus altering splicing patterns in hnRNP L target genes. These results have been reproduced, in part, also in this thesis using a novel protein-sponging mechanism (Schreiner *et al*., 2020) instead of hnRNP L knockdown (see *Results* **section 3.8.1**). CircRNA protein-sponges were designed to sequester hnRNP L and diminish its functions in alternative splicing regulation. Such a sequestration establishes conditions comparable to RNAi-mediated knockdown of hnRNP L. Both transfected circRNAs as well as Tornado-expressed circRNAs sponged hnRNP L specifically, such that alternative splicing of hnRNP L target genes was altered. Another example of modulating gene regulatory networks by sponging was shown for Hepatitis C Virus (HCV) propagation in cell culture system. CircRNA-mediated sponging of endogenous miRNA-122, inhibited HCV, since miR-122 is necessary for virus propagation (Jost *et al*., 2018).

Taken together, these observations along with our results strongly suggest that designer circRNAs can affect gene regulatory networks. Global gene regulation is a complex process

involving several interconnected networks, and altering one part of these networks could have bearing on gene expression on the whole. This is clearly demonstrated by the changes in alternative splicing networks, which was caused by the sequestration of one alternative splice regulator- hnRNP L, using designer circRNAs.

4.4 Designer circRNAs for various clinical/ biotechnological applications

The work presented in this thesis was aimed at demonstrating a functional role for designer circRNAs in protein-sponging. Although the design of circRNAs in this work was limited to two known RBPs – hnRNP L and RBM24, both involved in alternative splicing regulation. This novel technology can be applied to any RBP of clinical and biotechnological relevance. RBPs are multi-faceted proteins regulating almost all steps in RNA metabolism, such as splicing, translation, stability, localisation and degradation. Therefore, proper regulation of RNA-RBP networks is crucial for good health (Kelaini *et al.,* 2021). RBP-regulation relies on binding-affinities to RNA, micro-environment and events such as metabolism, and stress-response. Misregulation or dysfunction of RBPs, or a disruption in RNA homeostasis often leads to several diseases including diabetes, cardiovascular disease and other disorders such as neurodegenerative disorders and cancer (reviewed in Kelaini *et al.,* 2021). RBPs recognise hundreds of transcripts and form extensive regulatory networks that maintain cell homeostasis. From the formation of membraneless organelles to Mendelian and somatic genetic disorders, RBPs play an active role in various processes of the cell (Gebauer *et al*., 2021).

Antisense oligonucleotides (ASO), RNAi-mediated knockdowns and pharmacological inhibitions are currently the first-choice of treatment for most of these pathologies. Therapies targeting alternative splicing disorders are limited to ASOs and siRNA-mediated knockdowns. Although several ASOs have been clinically approved by the FDA (Food and Drug Administration, USA) for treatment (Quemener *et al.,* 2020), most of these therapies are limited due to toxic side effects. Since, our work demonstrating protein-sponging through designer circRNAs mimics RNAi-mediated knockdown, we propose their use as a sustainable alternative. CircRNA-based interventions offer several advantages over ASOs. *Firstly*, most ASOs and siRNAs need to be chemically modified to overcome endogenous degradation pathways. These modifications, upon degradation in the cell are often metabolised inefficiently causing toxic side effects. However, substantial chemical developments have been introduced in the field to diminish toxicity of ASOs, making them more tolerable. But when administered chronically, toxicity may still be observed (Quemener *et al.,* 2020). On the contrary, circRNAs are naturally occurring compounds and widespread in diverse cell types (Jeck *et al*., 2013; Salzman *et al*., 2012). They are degraded within the cell by natural degradation pathways by cellular nucleases (endonucleases and exonucleases), without any toxic side products. CircRNA degradation is not fully understood, however a few mechanisms such as Ago-mediated cleavage of CDR1as initiated by miRNA (Hansen *et al*., 2011) and RNase L-mediated degradation (Liu, *et al*., 2019b), have been reported. *Secondly*, due to their circular configuration and lack of 5’ and 3’ ends, circular RNAs are more stable and resistant to cellular endonucleases compared to linear RNA (Suzuki *et al*., 2006). We also demonstrated that designer circRNAs were more stable compared to linear RNA, and detectable up to four

days post-transfection (see *Results* **section 3.8**). *Finally,* miRNA-sponging studies have shown that circRNAs were more efficient than ASOs in preventing HCV propagation in cells (Jost *et al*., 2018). Considering these observations, we strongly believe that circRNAs could serve as an efficient alternative to ASOs and siRNA-mediated knockdown of RBPs, in several diseases caused due to disruption of RBP-RNA networks. Since they occur naturally, if unmodified, it is less likely that circRNA administration could elicit an innate immune response (Wesselhoeft *et al*., 2019). Interestingly, in a recent report our colleagues have demonstrated the successful inhibition of the novel SARS-CoV-2 coronavirus proliferation by designer antisense-circRNAs (Pfafenrot *et al.,* 2021).

Numerous RBPs are known to be overexpressed or their genes upregulated, in various types of cancers and therefore becoming targets of new cancer therapeutics. Interestingly, hnRNP L was identified as one such therapeutic target in the progression of pancreatic cancer (Fei *et al*., 2017). HnRNP L directly regulates the alternative splicing of the androgen receptor- a key lineage-specific oncogene in prostate cancer. HnRNP L also regulates circRNA formation by backsplicing in these cells (Fei *et al*., 2017). By another mechanism involving nonsense mediated decay (NMD), hnRNP L has also been linked to B-cell lymphoma progression (Kishor *et al*., 2018). HnRNP L protects mRNAs with NMD-inducing features such as long 3'-UTRs, which, in the case of B cell lymphoma leads to cancer. Protection by hnRNP L allows aberrant BCL2 mRNAs to evade NMD, thus promoting BCL2 overexpression and neoplasia (Kishor *et al*., 2018). In both these cases, it is clear that targeting hnRNP L as a therapeutic target may alleviate cancer progression. Blocking the protein using ASOs binding to specific 3' UTRs has been proposed as potential chemotherapy option. However, developing these therapies may need more careful examination and time. An alternative strategy could be the proposed use of our designer circRNAs. Designer circRNAs with high-affinity for hnRNP L have been shown to have dramatic effects on hnRNP L distribution (see *Results* **section 3.8.2**) and alternative splicing of target genes (see *Results* **section 3.8.2** and Schreiner *et al*., 2020). We are confident that applying designer circRNAs in these disease contexts offers a new paradigm in cancer therapy. A recent example of circRNA-protein sponging has been described for the HuR protein, which is bound by circPABPN1, derived from *PABPN1* gene (Abdelmohsen *et al*., 2017). CircPABPN1 sequesters HuR, and prevents it from binding to *PABPN1* mRNA, which in turn lowers translation of the mRNA. Although, circPABPN1 has not been shown to have any role in cancer, since HuR targets many mRNAs from tumour-suppressor and cancer related genes like *MYC*, *HIF1A* and *BCL2*, it would be interesting to find if circPABPN1 also promotes cancer progression (Abdelmohsen *et al*., 2017).

In addition to clinical applications via protein-sponging function, circRNAs and SELEX-derived aptamers in general, have been implicated in various biotechnological applications such as conditional gene regulation, biosensing, riboswitch for gene regulation, tags to visualise RNA or protein distribution in living cells (molecular probes), and for diagnostic purposes (Weigand & Suess, 2009). Latest developments in the field have applied aptamers for immunoassays and aptamer-based magnetic cell sorting (Nimjee *et al*., 2017). The first breakthrough in the therapeutic application of an RNA aptamer came with the approval of Pegaptanib (PEGylated aptamer angiogenesis inhibitor), by the FDA in 2004 (Ulrich *et al*., 2006). Pegaptanib (also known as Macugen), is an anti-vascular endothelial growth factor (anti-$VEGF_{165}$) RNA aptamer

developed by Ruckman *et al.* (1998) using the SELEX technique. It was approved for the treatment of pathological choroidal neovascularisation, associated with neovascular age-related macular degeneration (AMD). Following this, several SELEX-based aptamers were developed for various applications as antivirals, anticoagulants, anti-inflammatory compounds, aptamers against cell-surface antigens and whole organisms and aptamers in cancer treatment (reviewed in Ulrich *et al*., 2006 and Nimjee *et al*., 2017). These examples reiterate the applicability of the SELEX-technique in designing functionally relevant and biotechnologically important compounds. Building on this idea, we propose the development of circRNA as aptamers that bind with high specificity to small molecules such as adenosine and ATP or to hormones like insulin or other growth factors. Targeting these molecules would first require the design of circRNA-aptamer with a high affinity binding-site. The binding site for these molecules may be SELEX-derived (as obtained for hnRNP L protein-sponges, see Figure 4.1) or an existing high-affinity site, which can be implemented into a circRNA. Iaboni *et al.* (2016), developed a novel internalising aptamer, GL56, which targets the insulin receptor (IR). This nuclease–resistant, SELEX-derived RNA aptamer, specifically recognises the IR and rapidly internalises into target cells. The aptamer is found to specifically inhibit IR-dependent signalling, when applied to cancer cells expressing IR (Iaboni *et al.* 2016). By incorporating oligomers or multimers of GL56, and expressing them as a circRNA, it is very likely that binding-affinity can be enhanced. Furthermore, circRNA may be coupled to a degradation signal and administered directly to target cells and tissues, or into serum, such that after its function it may be removed safely through the secretory machinery.

In summary, the unique features of nucleic acid aptamers including high-affinity binding to target molecules and specificity, provide vast potential for future applications. Properties such as diversity (aptamers can be developed for any molecule), uniform activity regardless of batch synthesis, easily modifiable pharmacokinetic properties, no/low immunogenicity and unlimited shelf life, make RNA aptamers a better choice of drugs over antibodies (reviewed in Nimjee *et al*., 2017). A unique feature of aptamers is the reversal of inhibitory activity of the drug by the administration of an antidote which disrupts binding of the drug to its target. This property has been proposed for the development of reversal agents for new oral anticoagulant drugs (Nimjee *et al*., 2017). Such reversal is not possible with antibody-based therapies, thereby making the future of aptamer-based therapies very promising.

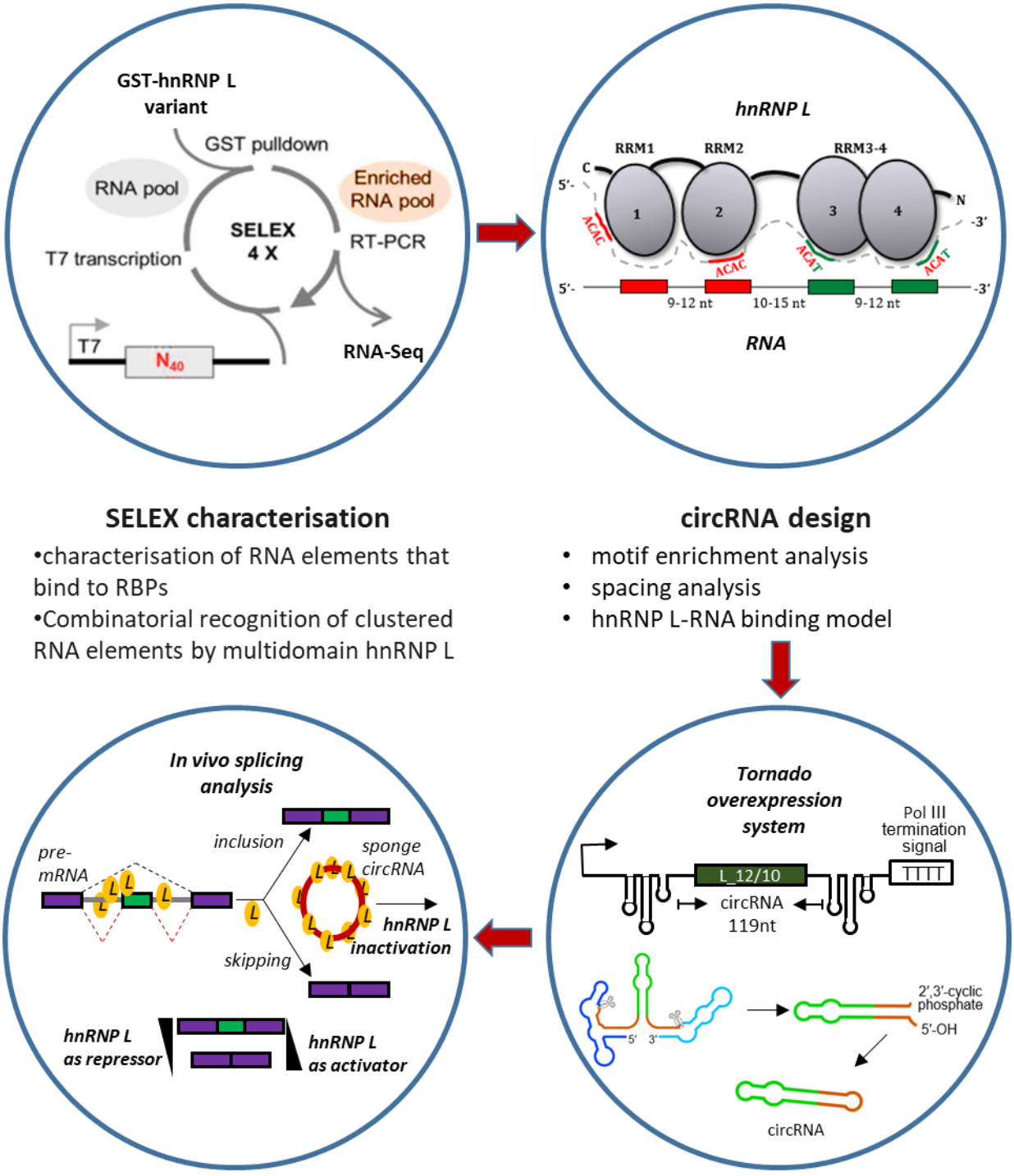

Fig. 4.1: Summary of optimal circRNA protein-sponge design

Schematic representation of steps involved in the design and development of optimal protein sponge for RBPs (hnRNP L is shown as an example).

4.5 Perspectives

The fascinating field of circular RNAs has drawn much attention in recent years due to their special properties such as high stability in cells owing to their closed covalent structure (Kristensen *et al*., 2019). CircRNAs are ubiquitous, and have been identified in all eukaryotes studied; however, their expression is highly tissue-specific (Rybak-Wolf *et al*., 2015). Although the biogenesis and expression profiles of circRNAs have been well examined, except for a couple of well-established examples, functional relevance of circRNAs is still largely unsettled (Hansen, 2021). With the introduction of the novel protein-sponging function using designer circRNAs, circRNA function can be now viewed from a new perspective (Schreiner *et al*., 2020).

The primary objective of this thesis was aimed at establishing this functional role of circRNAs in protein sponging, using artificial designer circRNAs. We have demonstrated that designer circRNAs were capable of sequestering RBPs specifically (protein-sponging) and modulate their function. Observed effects on alternative splicing due to sponging of hnRNP L, correlated with RNAi-mediated downregulation of hnRNP L. Although very promising, this strategy is limited to those RBPs with a known and highly specific binding-sequence. Nonetheless, this novel technique seems to be a promising alternative to siRNA-mediated knockdown of hnRNP L, in therapy. HnRNP L is essential for cell-survival, and complete knockdown of hnRNP L is lethal to the cell (Gaudreau *et al.,* 2016). In light of this, circRNA-mediated sponging offers target-protein regulation on a broader range, without completely abolishing its activity. Since in theory, sponging is also reversible, partial sequestration of hnRNP L by circRNA is a sustainable alternative compared to complete removal by siRNA. CircRNA sponging, like other aptamer-based therapies, allows controlled-regulation with the possibility to reverse inhibitory activity of the sponge. **Figure 4.1** schematically outlines the basic steps involved in optimal circRNA design.

It must be noted that therapeutic circRNAs are a new class of clinically relevant molecules, which need to be investigated further for approval as therapeutic agents. We discuss current limitations of therapeutic RNAs in general and potential strategies to overcome these limitations, here.

Firstly, according to the 73rd Ensembl version, out of ~20,300 protein-coding genes annotated, only about 4479 (22%) are estimated to be drugged/druggable by well-established small molecules or antibodies (Finan *et al.,* 2017). Several factors influence design and development of drugs, the major factor being target-binding site. Since small-molecules and antibodies can effectively bind to a limited set of molecules (like proteins), their effectiveness for the 'undruggable' targets is compromised. RNA aptamers on the other hand, can potentially bind to a wide-range of targets including non-coding RNAs. This dramatically expands their therapeutic possibility. ASOs and siRNAs were described as therapeutic agents in 1978 and 1998, respectively (Deprey *et al.,* 2020), and since then have been most widely applied in drug development and for personalised immunotherapy. Until now, seven ASO drugs have been approved for use in humans, while many others are undergoing clinical trials (Setten *et al.,* 2019). A recent and well publicised example was the use of Milasen – a personalised ASO drug

used for a child suffering from a rare neurodegenerative disorder (Batten's disease), caused by a genetic mutation leading to an alternative splicing error (Kim *et al.,* 2019).

Interestingly, from our SELEX experiments we observed that designing circRNA sponges is easier for multidomain RBPs (hnRNP L) compared to single-domain RBPs such as RBM24. This is because hnRNP L binds to its target RNA via all four RRM domains. Due to the recognition of an array of binding elements in the target RNA by respective RRM domains, interaction between protein and RNA is more favourable and the protein binds tightly to its target. In contrast, RBM24 relies on a single domain for recognition and binding to RNA. This interaction may be relatively less compact compared to hnRNP L, where multiple domains bind in a combinatorial manner. Combinatorial recognition of RNA-binding elements is an important aspect of protein-RNA interactions, which was also recently established by our group for a multidomain RBP, IMP3 (Schneider *et al.,* 2019).

Secondly, a major challenge in RNA therapy is stability of RNA compounds in blood. Therapeutic RNAs are highly charged and have a large molecular mass compared to small molecules. This may pose a significant hindrance for delivery through the plasma membrane. Double-stranded siRNAs cannot pass through cell membranes as easily as single stranded ASOs, and upon entering the cell, unmodified RNAs are highly accessible to cellular nucleases. This problem can be overcome by using circRNAs which show high stability *in vivo*. But in general, efficacy of targeted RNA therapy *in vitro* does not always correlate with efficacy *in vivo*; and commonly known targeted delivery is only well-established for the liver.

Finally, delivery strategies for RNA therapeutics are limited. There are two main reasons why targeted delivery, in most cases is unsuccessful: susceptibility of oligonucleotides to degradation by cellular nucleases and inefficiency of cellular uptake. Both these challenges have been overcome in part, by chemical modifications of the ASOs and siRNAs. RNA therapeutics can be modified by diverse chemical modifications of the nucleobases, phosphate backbone and sugars (reviewed in Wan & Seth, 2016). The most commonly used modifications include – using Phosphorothioate (PS) nucleotide backbone and 2'-*O*-methyl (2'-OMe) ribose sugars instead of phosphate backbone and ribose sugar, respectively. These modifications promote higher cellular uptake and increase resistance to nucleases (Matsukura *et al.,* 1987; Inoue *et al.,* 1987). Therapeutic RNAs also interact with proteins on cell membrane and are effectively internalised by endocytosis. However, most of these RNA molecules remain trapped in the endosomes and cannot fulfil their therapeutic roles, a phenomenon known as non-productive uptake (Crooke *et al.,* 2017). Additional modifications such as 2'-*O*-methoxyethyl (2'-MOE), locked nucleic acid (LNA) and cationic guanidinium groups for backbone modification have been shown to enhance cellular uptake. Since siRNAs are larger and require separate sense and antisense strands, 2'-OMe and 2'-fluoro (2'-F) modifications for the sugars and PS modifications at the 5' and 3' ends of both strands are introduced (reviewed in Deprey *et al.,* 2020). Although these modifications ensure efficient uptake of RNA drugs and resistance to degradation, gymnosis (the process of cellular uptake of therapeutic RNAs unaided by chemical or physical means of drug delivery) remains highly inefficient (Deprey *et al.,* 2020). An effective way to overcome gymnosis is to employ physical delivery strategies such as microinjection or electroporation. Also, chemical delivery using

cationic lipids such as lipofectamine are routinely applied for the transfection of RNA therapeutics into cells (also applied for circRNA transfection in this thesis). Furthermore, formulating RNA therapeutics into lipid nanoparticles or liposomes containing neutral and cationic lipids such as cholesterol, promotes delivery across plasma membrane, while avoiding toxicity (Prakash *et al*., 2013).

To sum it up, if unmodified, therapeutic circRNAs correspond to natural RNAs in their safety and are a promising alternative to other therapeutic RNAs. However, circRNA-based therapy is still largely unexplored. Delivery of certain circRNAs may be hindered due to large size, and efficiency of circRNA-production is cell-type and tissue-specific. Furthermore, addition of specific localisation signals by end modification is not possible due to their closed configuration. Regardless, many circRNAs have been proposed as putative regulators of miRNA (by sponging) in spite of stoichiometric challenges and lack of miRNA response element enrichment in circRNA sequences (Guo *et al*., 2014; Stagsted *et al*., 2019). Although the prospects of identifying a non-canonical game-changer in gene regulation is very high, it is our opinion that scientific data must be evaluated critically before studying noise and artefacts generated from circRNA functional studies (Hansen, 2021). Scrutinising the functional properties of circRNAs with increased stringency would be beneficial to the circRNA field as well as the scientific community to address future challenges and applications of therapeutic circRNAs.

In conclusion, this study expands the functional possibilities of circular RNAs and establishes for the first time, a novel application in protein sponging. Our data strongly suggests that this approach may be extended to any RBP of clinical importance. Designer circRNAs can be developed into a novel and highly specific new class of therapeutic RNAs, to be applied in cases where overexpressed (or mislocalised) RNA-binding proteins cause human disease, such as in many tumour tissues.

Supplementary information

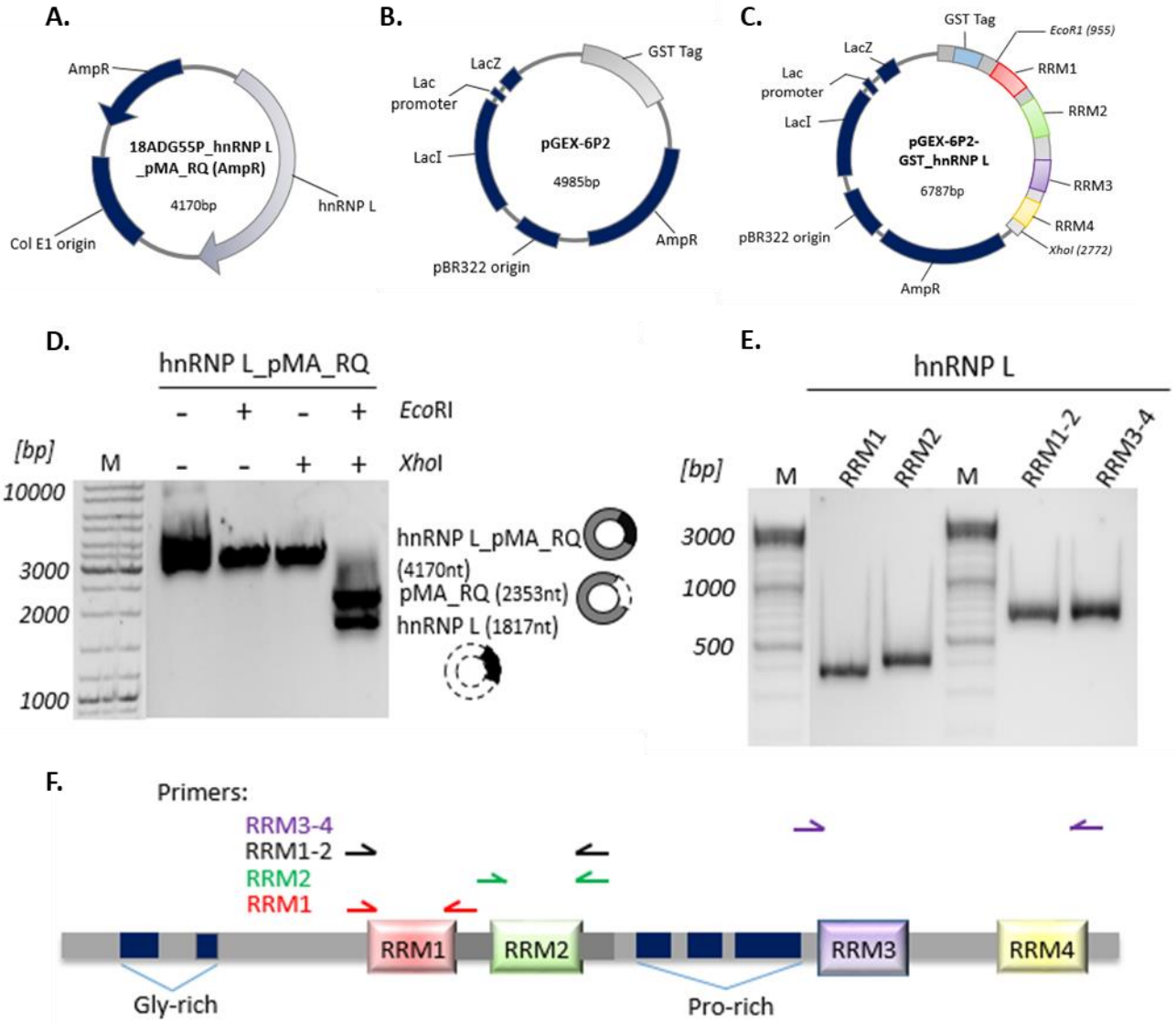

Fig. S1: Cloning strategy of codon-optimised hnRNP L in pGEX-6P2 expression vector.

(A-C) Vector map of donor plasmid - 18ADG55P _hnRNP L_pMA_RQ **(A)**, empty pGEX-6P2 expression vector **(B)** and pGEX-6P2-GST_hnRNP L **(C)**. HnRNP L sequence from the donor plasmid was cloned into pGEX-6P2 vector between *Eco*RI and *Xho*I restriction sites. **(D)** Restriction digestion of hnRNP L sequence out of donor plasmid using *Eco*RI and *Xho*I enzymes. Double digestion yields hnRNP L insert and digested vector showing differential mobility on 0.8% agarose gel. **(E)** HnRNP L truncations obtained from pGEX-6P2-GST_hnRNP L using domain-specific primers and PCR amplification on a 2% agarose gel. **(F)** Schematic representation of hnRNP L domain organisation and PCR amplification of hnRNP L domains with domain-specific primers indicated by colour code: red = RRM1, green = RRM2, black = RRM1-2 and purple = RRM3-4. *M*, marker and *bp*, base pairs.

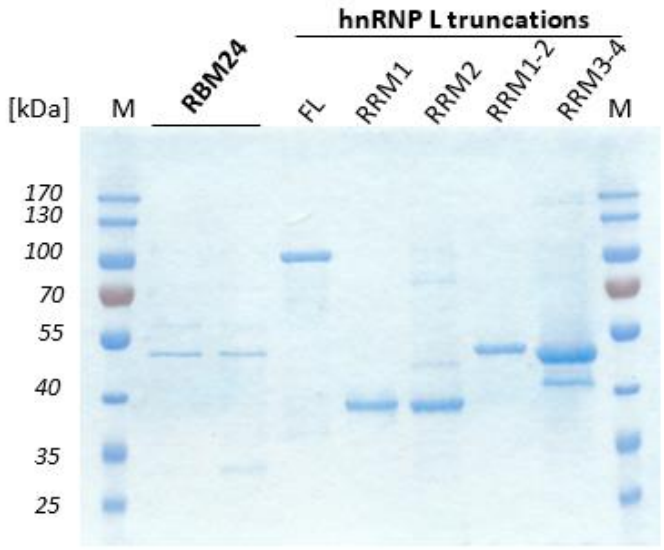

Fig. S2: Expression and purification of recombinant proteins.
SDS-PAGE analysis of purified, GST-tagged RBM24 (~51kDa) and hnRNP L full-length [FL (~93kDa)] and truncation variants - RRM1 (~40kDa), RRM2 (~42kDa), RRM1-2 (~52kDa) and RRM3-4 (53kDa). *M*, marker

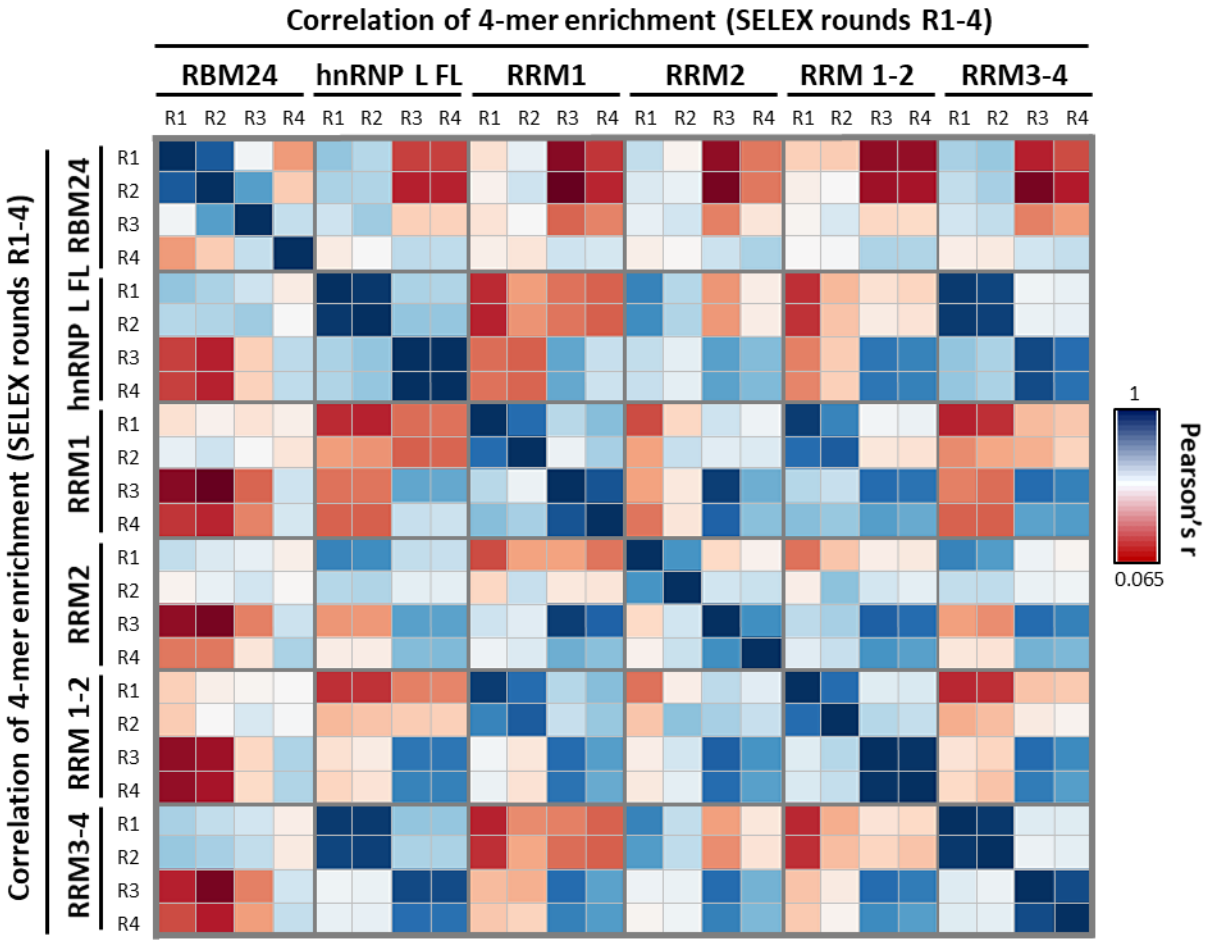

Fig. S3: Correlation of 4-mer motif enrichment
Linear correlation (Pearson's r) of 4-mer motif enrichment is shown as a heat map. Correlation between RBM24, hnRNP L (FL) and its truncation variants is shown for each SELEX round. (blue: high correlation, red: low/no correlation).

L_9/10 (59nts)

GGGCTTATGACACTCCACTTGGACACCTCCACTTGG**ACAT**TCCACTTGG**ACAT**CTTATG

L_9/15 (64nts)

GGGCTTATGACACTCCACTTGGACACGCACCCTCCACTTGG**ACAT**TCCACTTGG**ACAT**CTTATG

L_12/10 (65nts)

GGGCTTATGACACGCACCCTCCACTACACCTCCACTTGG**ACAT**GCACCCTCCACT**ACAT**CTTATG

L_12/15 (70nts)

GGGCTTATGACACGCACCCTCCACTACACGCACCCTCCACTTGG**ACAT**GCACCCTCCACT**ACAT**CTTATG

L_12/10_mut (65nts)

GGGCTTATG**TGTG**GCACCCTCCACT**TGTG**CTCCACTTGG**TGTT**GCACCCTCCACT**TGTT**CTTATG

Fig. S4: SELEX-derived RNA sequences for binding hnRNP L.
DNA sequence of the corresponding RNA used for hnRNP L sponge design is depicted in order of increasing size (in nt). Binding motifs- ACAC and ACAT (in bold font) for binding of the four respective hnRNP L domains, in WT RNA are indicated in red. A representative mut variant of L_12/10 RNA is also shown, with mutated motifs indicated in blue. *WT*, wild type; *mut*, mutant.

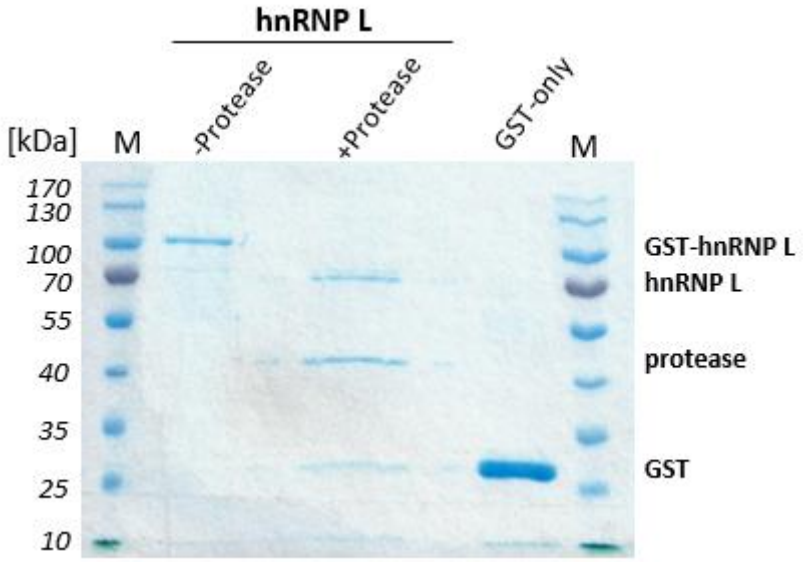

Fig. S5: GST-tag removal by protease cleavage.

SDS-PAGE analysis of GST-tagged hnRNP L and cleavage of GST-tag by prescission protease. Expected sizes of GST-hnRNP L (~93 kDa), hnRNP L-only (~64 kDa) and GST-only (~27 kDa) are indicated on the right. M, *Marker.*

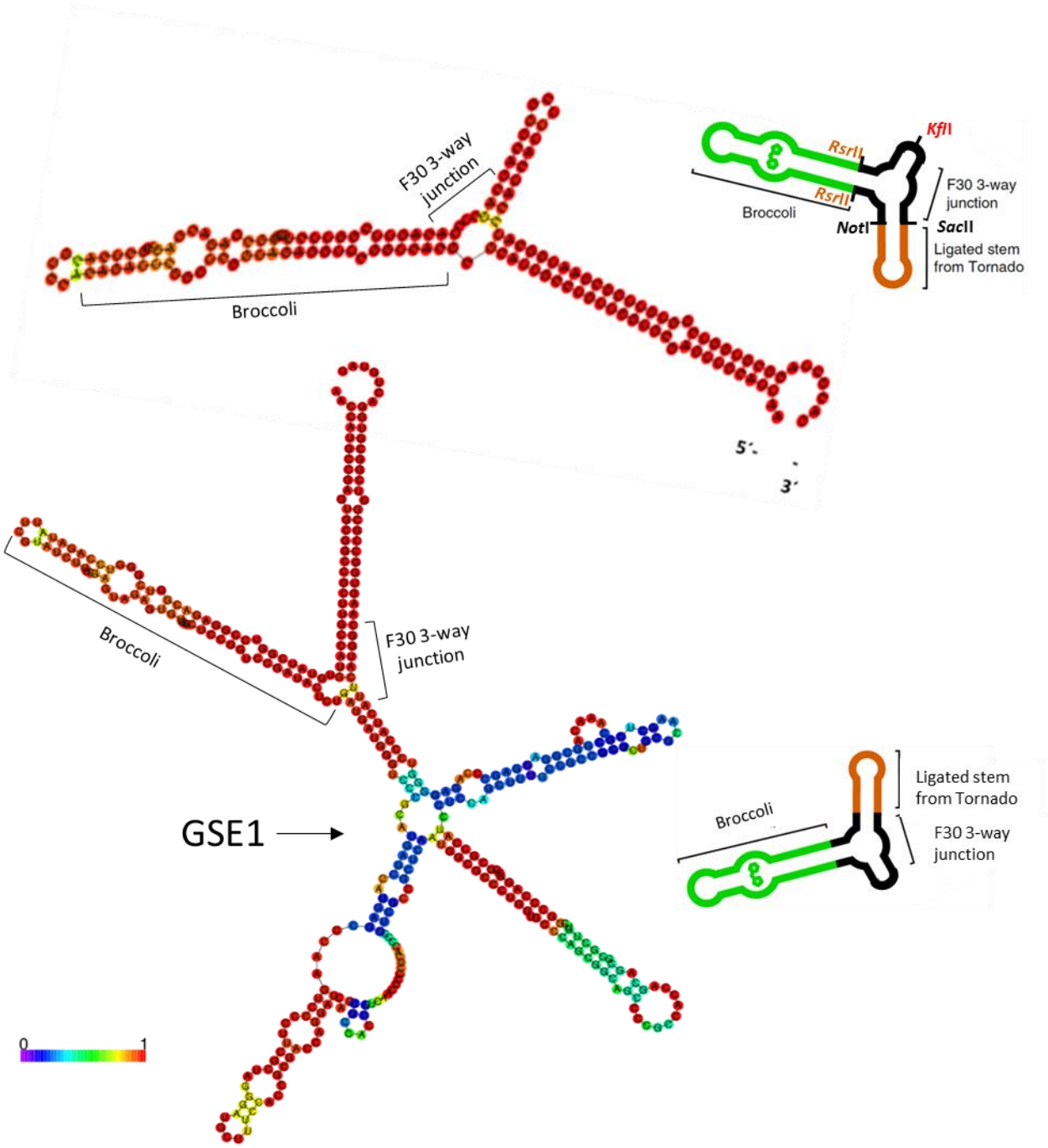

Fig. S6: Secondary structures of Tornado-constructs of endogenous circRNA GSE1.

Predicted Tornado circRNA secondary structures.F30 three-way junction, the Broccoli aptamer and the insert sequence, are marked. The color code denotes the probability of base-pairing from zero to one (from violet to red, respectively). For unpaired regions the color denotes the probability of being unpaired.

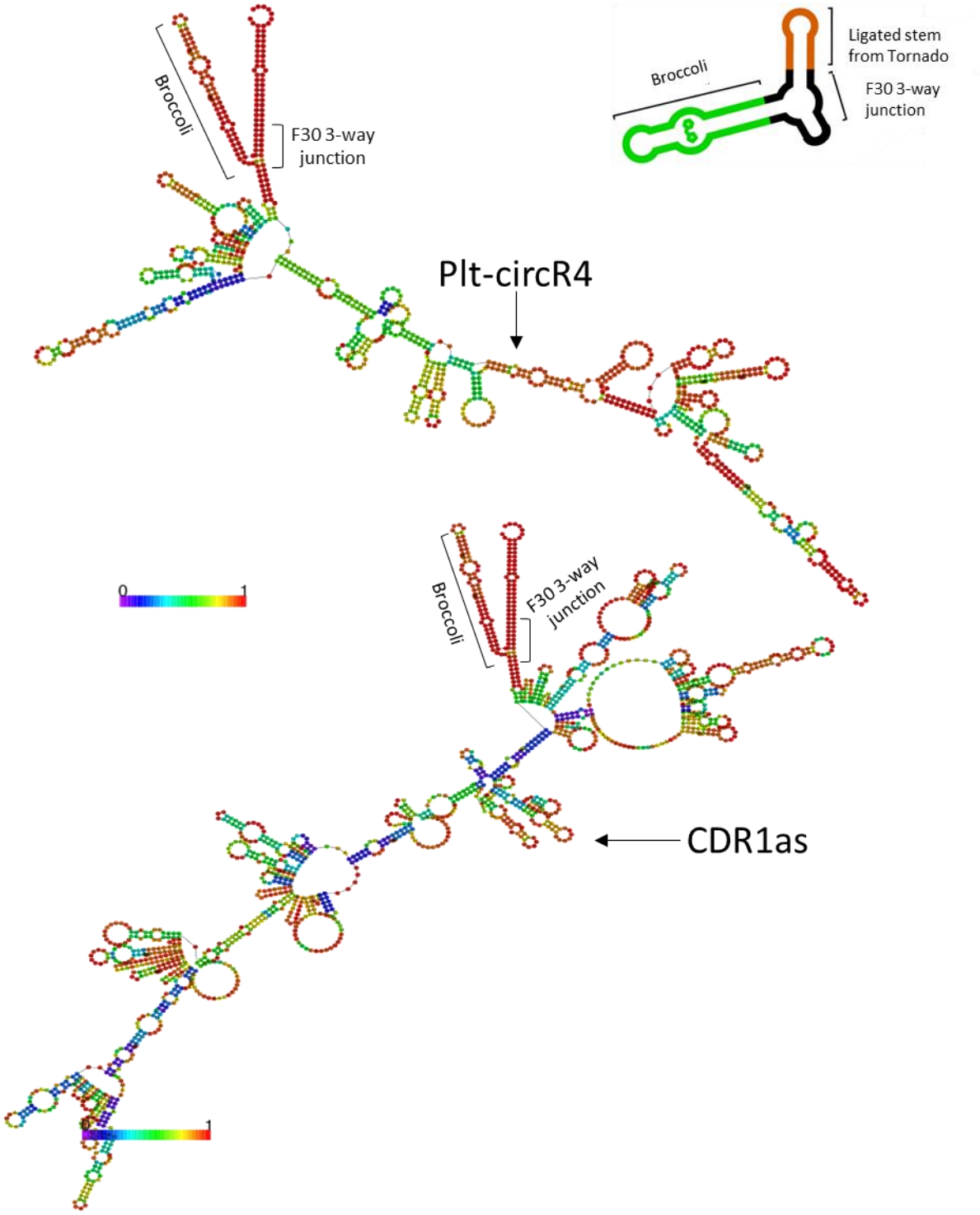

Fig. S7: Secondary structures of Tornado-constructs of endogenous circRNAs.

Predicted Tornado circRNA secondary structures of Plt-circR4 (top) and CDR1as (bottom).F30 three-way junction, the Broccoli aptamer and the insert sequence, are marked. The color code denotes the probability of base-pairing from zero to one (from violet to red, respectively). For unpaired regions the color denotes the probability of being unpaired.

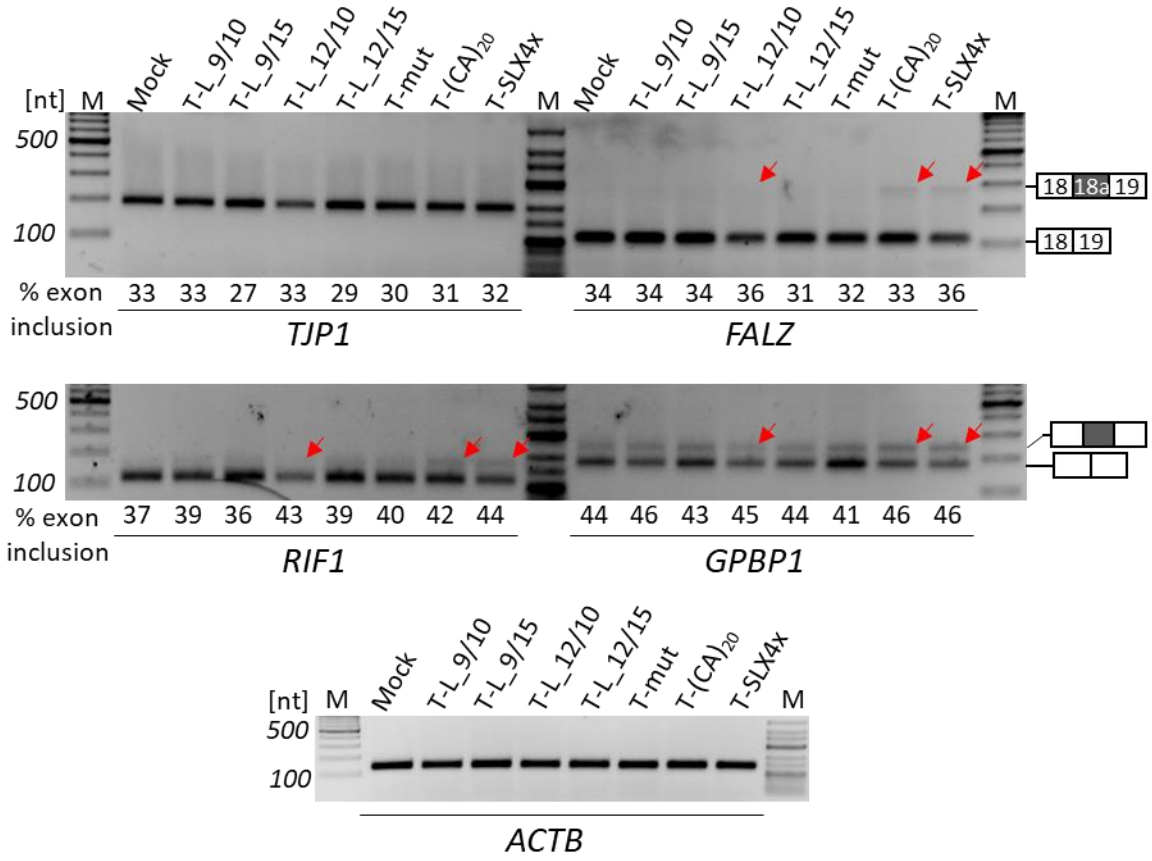

Fig. S8: Tornado-expressed circRNA sponges affect hnRNP L alternative splicing targets in HeLa cells (exon skipping)

Total RNA was extracted from HeLa cells transfected with Tornado-expressed circRNA sponges - T-L_9/10, T-L_9/15, T-L_12/10, T-L_12/15, T-mut, T-$(CA)_{20}$ and T-SLX4x two days post transfection. Isolated RNA was used for RT-PCR with primers against hnRNP L alternative splicing targets – *TJP1, FALZ*, *RIF1* and *GPBP1*. Percentage of exon inclusion is shown below the corresponding lanes. Positions of PCR products corresponding to alternatively spliced mRNAs are indicated with red arrows. *ACTB* served as loading control for RT-PCR. Exon inclusion is schematically represented on the right of each panel. *M*, marker.

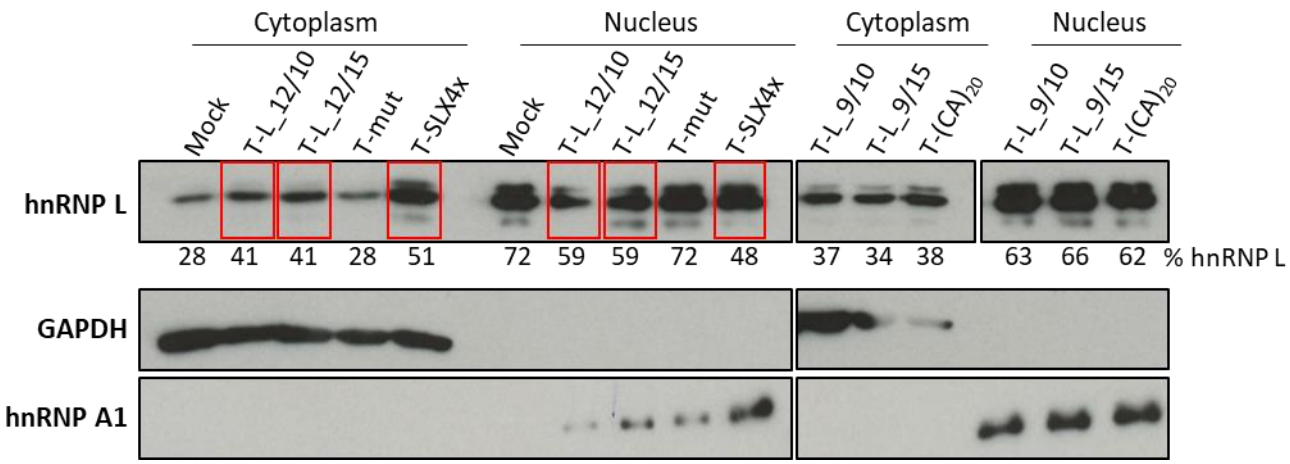

Fig. S9: Transfected Tornado-circRNA shifts nuclear-cytoplasmic distribution of hnRNP L.
Tornado-expressed hnRNP L circRNA-sponge shifts nuclear-cytoplasmic distribution of hnRNP L. T-L_9/10, T-L_9/15, T-L_12/10, T-L_12/15, T-L_mut, T-SLX4x and T-$(CA)_{20}$ were transfected in HeLa cells, followed by cell fractionation after 24h. Equivalent lysate amounts of total cells, cytoplasmic and nuclear fractions were analysed by Western blotting for hnRNP L, GAPDH, and hnRNP A1. The distribution of hnRNP L between nuclear and cytoplasmic fractions was quantitated, based on Western signals (mean values and standard deviations given below the respective lanes; *n* = 3). HnRNP L distribution shifted by circRNA-sponge, is marked with a red box.

Table 1.1: Expression and mechanistic role of hnRNP L in tumor tissues (from Gu *et al*., 2020, modified)

Cancer type	Phenotypes affected	Related carcinogenesis	Role	References
Lung cancer	apoptosis	caspase 9	hnRNP L phosphorylation regulates alternative splicing	Goehe *et al*., 2010
		C9/E3-ESS	RBP	Goehe *et al*., 2010
		p53 tumor suppressor pathway	pathway inhibitor	Siebring-van Olst *et al*., 2017
Hepatocellular carcinoma	proliferation, migration, viability	specific binding to CASC9 affects PI3K/ AKT and DNA damage signalling pathways	hnRNP L upregulated, carcinogenic factor, RBP, tumor associated antigen	Klingenberg *et al*., 2018
Prostate cancer	proliferation, invasion, metastasis, apoptosis, viability	p53/p21/cyclin D1 pathway	hnRNP L upregulated, carcinogenic factor	Gu *et al*., 2020; Cox & Lane *et al*., 1995
		CEACAM1/Caspase-3	alternative splicing regulator	Dery *et al*., 2011
Bladder cancer	proliferation, invasion, metastasis, apoptosis, EMT	p53/bcl-2/caspase signalling pathway	hnRNP L upregulated, carcinogenic factor, RBP	Chen *et al*., 2018
			MAPK pathway	Lv *et al*., 2017
			transcription factor E2F1, LNMATI	Chen *et al*., 2018
Gastric cancer	proliferation, migration	protein-protein interaction	RBP	Dai *et al*., 2016
Pancreatic cancer	proliferation	lncRNA.uc345	RBP	Liu *et al*., 2016
Breast cancer	invasion, EMT	hnRNP L, NSP 5a3a, B23 interaction	carcinogenic factor	D'Agostino *et al*., 2010
	proliferation, invasion	Binding to DSACM-AS1	RBP	Niknafs *et al*., 2016
Ovarian clear cell carcinoma			increased titers of IgA autoantibody against hnRNP L	Yoneyama *et al*., 2015
Acute myeloid leukaemia		combining with THRIL to form a complex that increases TNF transcription	RBP	Sayad *et al*., 2018
Glioblastoma	proliferation, invasion	hnRNP L interacts with SChLAP1 and ACTN4	RBP	Ji *et al*., 2019

B cell lymphoma	apoptosis	altering Bcl2:IGH fusion mRNA and Bcl2 levels	RBP	Kishor *et al*., 2019
Chronic lymphocyte leukaemia		hnRNP L combines with D2S1888 to inhibit miR-155	RBP	Pagotto *et al*., 2019

Table 1.2: Potential functions of RBM24 in development and disease (from Grifone *et al*., 2020, modified)

Tissue or disease	Post-transcriptional regulation	Role in development or disease
Skeletal muscle	muscle-specific pre-mRNA splicing, mRNA stability	myogenic differentiation, somitogenesis and sarcomere organization (Grifone *et al*., 2014; Yang *et al*., 2014)
Cardiac muscle	muscle-specific pre-mRNA splicing, mRNA stability and translation	heart development, sarcomere assembly and cardiac contractility (Liu *et al*., 2019a; Yang *et al*., 2014; Maragh *et al*., 2011)
Lens	cytoplasmic polyadenylation, mRNA stability	lens fiber differentiation, and lens transparency (Shao *et al*., 2020; Dash *et al*., 2020)
Inner ear/ Neuromasts	mRNA stability	hair cell morphogenesis and differentiation (Cheng *et al*., 2020)
Olfactory epithelium	unknown (cytoplasmic localisation in neuronal cells)	may promote neurogenic differentiation by regulating translation of target mRNA (Grifone *et al*., 2020)
Blastula/Gastrula	unknown	germ layer formation (Li *et al*., 2010)
Prostate cancer	interaction with miR-106a-5p	inhibition of tumorigenesis (Wei *et al*., 2020)
Nasopharyngeal carcinoma	interaction with miR-25	inhibition of cell proliferation, migration and invasion (Hua *et al*., 2016)
Hirschsprung disease	interaction with MIR143HG	proliferation or migration of enteric neural crest cells and gangliogenesis (Du *et al*., 2016)
Familial dysautonomia	increased recognition of the mutated 5' splicing site in *IKBKAP* gene by U1 snRNP	possible role in protecting against the aberrant splicing of the mutated gene (Ohe *et al*., 2017)
Hepatitis A and B	pre-genomic RNA packaging and replication	possible host factor for the viruses (Yao *et al*., 2019)

Table 1.3: Biological functions and mechanism of action of key circRNAs (from Kristensen *et al.*, 2019, modified)

circRNA	Location	Mechanism of action	Putative biological function	References
cirS7/ CDR1as	Cytoplasm	miRNA sponge or decoy for miR-7	neuronal development, positive regulator of insulin secretion, oncogenic functions	Memczak *et al.*, 2013; Hansen *et al.*, 2013; Weng *et al.*, 2017; Kleaveland *et al.*, 2018;
		interacts with IGF2BP3	compromises pro-metastatic function	Hanniford *et al.*, 2020
		interacts with p53 and blocks it from MDM2	inhibits gliomagenesis	Lou *et al.*, 2020
circZNF91	Cytoplasm	miRNA sponge or decoy for miR-23b-3p	epidermal stem cell differentiation	Kristensen *et al.*, 2018b
circHIPK3	Cytoplasm	miRNA sponge or decoy for multiple miRNAs	tumor suppressor, positive regulator of insulin secretion	Zheng *et al.*, 2016; Okholm *et al.*, 2017; Stoll *et al.*, 2018
circBIRC6	Cytoplasm	miRNA sponge or decoy for miR-34a and miR-145	maintains pluripotency	Yu *et al.*, 2017
circPVT1	Cytoplasm	miRNA sponge or decoy for miR-497-5p	positive regulator of cell cycle progression	Verduci *et al.*, 2017
circCCDC66	Cytoplasm	miRNA sponge or decoy for multiple tumor suppressor miRNAs	oncogenic functions	Hsiao *et al.*, 2017
circCCAC1	Cytoplasm	miRNA sponge or decoy for miR-514a-5p	CCA progression, induces angiogenesis and disrupts vascular endothelial barriers	Xu *et al.*, 2021
circMbl	Cytoplasm	protein sponge or decoy for mbl and template for translation	regulator of neuronal functions	Pamudurti *et al.*, 2017; Ashwal-Fluss *et al.*, 2014;
circPABPN1	Cytoplasm	protein sponge or decoy for HUR	suppresses PABPN1 translation and decreases cellular proliferation	Abdelmohsen *et al.*, 2017
cia-cGAS	Nucleus	protein sponge or decoy for cGAS	protects long-tern hematopoietic stem cells from exhaustion	Xia *et al.*, 2018
circANRIL	Cytoplasm	protein sponge or decoy for PES1	impairs pre-rRNA processing and ribosome biogenesis to induce nucleolar stress and activate p53	Holdt *et al.*, 2016
circ-Foxo3	Cytoplasm	protein scaffold (facilitates MDM-2 dependent ubiquitylation of p53)	induces apoptosis of cancer cells	Du *et al.*, 2016

		and sponge for MDM2 (to prevent ubiquitylation of FOXO-3)		
circ-Amotl1	Cytoplasm	protein scaffold (facilitates PDK1 dependent phosphorylation of AKT1)	cardioprotective role in doxorubicin- induced cardiomyopathy	Zeng *et al*., 2017
FECR1	Nucleus	protein recruitment (recruits TET1 to the promoter region of its own host gene)	oncogenic functions through upregulation of FLI1	Chen *et al*., 2018
ci-ankrd52	Nucleus	enhances protein function (positive regulator of Pol II transcription)	positively regulates the expression of its parental gene	Zhang *et al*., 2013
circEIF3J	Nucleus	enhances protein function (positive regulator of Pol II transcription)	positively regulates the expression of its parental gene	Li *et al*., 2015
circPAIP2	Nucleus	enhances protein function (positive regulator of Pol II transcription)	positively regulates the expression of its parental gene	Li *et al*., 2015
circ-ZNF609	Cytoplasm	template for translation	regulates myoblast proliferation	Legnini *et al*., 2017
circ-FBXW7	Cytoplasm	template for translation	tumor suppressor	Yang *et al*., 2018
		miRNA sponge or decoy for miR-197-3p	inhibits malignant progression in triple-negative breast cancer	Ye *et al*., 2019
circPINTexon2	Cytoplasm	template for translation	tumor suppressor	Zhang *et al*., 2018a
circ-SHPRH	Cytoplasm	template for translation	tumor suppressor	Zhang *et al*., 2018b
circNfix	Cytoplasm	reinforces interaction of Ybx1 with Nedd4l (an E3 ubiquitin ligase) and induces degradation of Ybx1 by ubuiquitination	cardiac regenerative repair	Huang *et al*., 2019
		miRNA sponge or decoy for miR-214	promotes Gsk3β (glycogen synthase kinase 3 β) expression and repress β-catenin activity.	Huang *et al*., 2019

CDR1as, cerebellar degeneration-related protein 1; CCA, cholangiocarcinoma; cGAS, cyclic GMP–AMP synthase; circRNA, circular RNA; FECR1, FLI1 exonic circular RNA; FLI1, friend leukaemia integration 1 transcription factor; HUR, Hu-antigen R; MDM2, mouse double-minute 2; miRNA, microRNA; PABPN1, poly(A) binding protein 1; PDK1, phosphoinositide-dependent protein kinase 1; PES1, pescadillo homologue 1; Pol II, RNA polymerase II; TET1, methylcytosine dioxygenase TET1; Ybx1, Y-box binding protein 1.

References

Abdelmohsen, K., Panda, A. C., Munk, R., Grammatikakis, I., Dudekula, D. B., De, S., Kim, J., Noh, J. H., Kim, K. M., Martindale, J. L. & Gorospe, M. (2017). Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biology*, *14*(3), 361-369.

Afroz, T., Cienikova, Z., Cléry, A. & Allain, F. H. (2015). One, two, three, four! How multiple RRMs read the genome sequence. *Methods in Enzymology*, *558*, 235-278.

Aktaş, T., Ilık, İ. A., Maticzka, D., Bhardwaj, V., Rodrigues, C. P., Mittler, G., Manke, T., Backofen, R. & Akhtar, A. (2017). DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature*, *544*(7648), 115-119.

Alberti, S., Gladfelter, A. & Mittag, T. (2019). Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell*, *176*(3), 419-434.

Alexander, M. S., Hightower, R. M., Reid, A. L., Bennett, A. H., Iyer, L., Slonim, D. K., Saha, M., Kawahara, G., Kunkel, L. M., Kopin, A. S. & Gupta, V. A. (2021). HnRNP L is essential for myogenic differentiation and modulates myotonic dystrophy pathologies. *Muscle & Nerve*, *63*(6), 928-940.

Alipanahi, B., Delong, A., Weirauch, M. T. & Frey, B. J. (2015). Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature Biotechnology*, *33*(8), 831-838.

Amunts, A., Brown, A., Bai, X. C., Llácer, J. L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. (2014). Structure of the yeast mitochondrial large ribosomal subunit. *Science*, *343*(6178), 1485-1489.

Anderson, P. & Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nature Reviews Molecular Cell Biology*, *10*(6), 430-436.

Anon, 2021. RNA Processing. Available at: https://bio.libretexts.org/@go/page/15198 [Accessed September 02, 2021].

Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N. & Kadener, S. (2014). CircRNA biogenesis competes with pre-mRNA splicing. *Molecular cell*, *56*(1), 55-66.

Bachmayr-Heyda, A., Reiner, A. T., Auer, K., Sukhbaatar, N., Aust, S., Bachleitner-Hofmann, T., Mesteri, I., Grunt, T. W., Zeillinger, R. & Pils, D. (2015). Correlation of circular RNA abundance with proliferation–exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis and normal human tissues. *Scientific Reports*, *5*(1), 1-10.

Banerjee, A. K., Blanco, M. R., Bruce, E. A., Honson, D. D., Chen, L. M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S. A., Loney, C. & Thai, J. (2020). SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell*, *183*(5), 1325-1339.

Baralle, D., & Buratti, E. (2017). RNA splicing in human disease and in the clinic. *Clinical Science*, *131*(5), 355-368.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J. & Frey, B. J. (2010). Deciphering the splicing code. *Nature*, *465*(7294), 53-59.

Barrett, S. P., Wang, P. L. & Salzman, J. (2015). Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *elife*, *4*, e07540.

Baserga, S. J. & Steitz, J. A. (1993). The diverse world of small ribonucleoproteins. *Cold Spring Harbor Monograph Series*, *24*, 359-359.

Begum, S., Yiu, A., Stebbing, J. & Castellano, L. (2018). Novel tumour suppressive protein encoded by circular RNA, circ-SHPRH, in glioblastomas. *Oncogene*, *37*(30), 4055-4057.

Behrmann, E., Loerke, J., Budkevich, T. V., Yamamoto, K., Schmidt, A., Penczek, P. A., Vos, M. R., Bürger, J., Mielke, T., Scheerer, P. & Spahn, C. M. (2015). Structural snapshots of actively translating human ribosomes. *Cell*, *161*(4), 845-857.

Bell, J. L., Wächter, K., Mühleck, B., Pazaitis, N., Köhn, M., Lederer, M. & Hüttelmaier, S. (2013). Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cellular and Molecular Life Sciences*, *70*(15), 2657-2675.

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*, *15*(3), 163-175.

Berget, S. M., Moore, C. & Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences*, *74*(8), 3171-3175.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, *72*(1), 291-336.

Blatter, M., Dunin-Horkawicz, S., Grishina, I., Maris, C., Thore, S., Maier, T., Bindereif, A., Bujnicki, J. M. & Allain, F. H. T. (2015). The signature of the five-stranded vRRM fold defined by functional, structural and computational analysis of the hnRNP L protein. *Journal of Molecular Biology*, *427*(19), 3001-3022.

Boisvert, F. M., Rhie, A., Richard, S. & Doherty, A. J. (2005c). The GAR motif of 53BP1 is arginine methylated by PRMT1 and is necessary for 53BP1 DNA binding activity. *Cell Cycle*, *4*(12), 1834-1841.

Boy, S., Souopgui, J., Amato, M. A., Wegnez, M., Pieler, T. & Perron, M. (2004). XSEB4R, a novel RNA-binding protein involved in retinal cell differentiation downstream of bHLH proneural genes. *Development* 131 (4): 851–862.

Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell*, *152*(6), 1252-1269.

Breuer, J. & Rossbach, O. (2020). Production and purification of artificial circular RNA sponges for application in molecular biology and medicine. *Methods and Protocols*, *3*(2), 42.

Brody, E. & Abelson, J. (1985). The" spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science*, *228*(4702), 963-967.

Brudno, M., Gelfand, M. S., Spengler, S., Zorn, M., Dubchak, I. & Conboy, J. G. (2001). Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic acids research*, *29*(11), 2338-2348.

Buchan, J. R. (2014). mRNP granules: assembly, function, and connections with disease. *RNA Biology*, *11*(8), 1019-1030.

Burd, C. E., Jeck, W. R., Liu, Y., Sanoff, H. K., Wang, Z. & Sharpless, N. E. (2010). Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genetics*, *6*(12), e1001233.

Busslinger, M., Moschonas, N. & Flavell, R. A. (1981). β+ thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell*, *27*(2), 289-298.

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. & Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, *73*(5), 1019-1030.

Caputi, M. & Zahler, A. M. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H′/F/2H9 family. *Journal of Biological Chemistry*, *276*(47), 43850-43859.

Cartegni, L. & Krainer, A. R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature Genetics*, *30*(4), 377-384.

Cartegni L., Chew S. L. & Krainer A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics 3*: 285–298

Castle, J. C., Zhang, C., Shah, J. K., Kulkarni, A. V., Kalsotra, A., Cooper, T. A. & Johnson, J. M. (2008). Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nature Genetics*, *40*(12), 1416-1425.

Cech, T. R. & Steitz, J. A. (2014). The noncoding RNA revolution—trashing old rules to forge new ones. *Cell*, *157*(1), 77-94.

Celotto, A. M. & Graveley, B. R. (2001). Alternative splicing of the Drosophila *Dscam* pre-mRNA is both temporally and spatially regulated. *Genetics*, *159*(2), 599-608.

Cheli, Y. & Kunicki, T. J. (2006). HnRNP L regulates differences in expression of mouse integrin α2β1. *Blood*, *107*(11), 4391-4398.

Chen, C. Y. & Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science*, *268*(5209), 415-417.

Chen, C., He, W., Huang, J., Wang, B., Li, H., Cai, Q., Su, F., Bi, J., Liu, H., Zhang, B. & Jiang, N. (2018). LNMAT1 promotes lymphatic metastasis of bladder cancer via CCL2 dependent macrophage recruitment. *Nature Communications*, *9*(1), 1-18.

Chen, L. L. (2016). The biogenesis and emerging roles of circular RNAs. *Nature Reviews Molecular Cell Biology*, *17*(4), 205-211.

Chen, L. L. (2020). The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nature Reviews Molecular Cell Biology*, *21*(8), 475-490.

Chen, N., Zhao, G., Yan, X., Lv, Z., Yin, H., Zhang, S., Song, W., Li, X., Li, L., Du, Z. & Jia, L. (2018). A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome Biology*, *19*(1), 1-14.

Chen, Y. G., Chen, R., Ahmad, S., Verma, R., Kasturi, S. P., Amaya, L., Broughton, J. P., Kim, J., Cadena, C., Pulendran, B. & Hur, S. (2019). N6-methyladenosine modification controls circular RNA immunity. *Molecular Cell*, *76*(1), 96-109.

Cheng, X. N., Zhang, J. J., & Shi, D. L. (2020). Loss of Rbm24a causes defective hair cell development in the zebrafish inner ear and neuromasts. *Journal of Genetics and Genomics*, ff10.1016/j.jgg.2020.07.002ff.

Chiou, N. T., Shankarling, G. & Lynch, K. W. (2013). HnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly. *Molecular Cell*, *49*(5), 972-982.

Cho, S., Hoang, A., Sinha, R., Zhong, X. Y., Fu, X. D., Krainer, A. R. & Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences*, *108*(20), 8233-8238.

Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977). An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell*, *12*(1), 1-8.

Cléry, A., Blatter, M. & Allain, F. H. (2008). RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, *18*(3), 290-298.

Cocquerelle, C., Mascrez, B., Hétuin, D. & Bailleul, B. (1993). Mis-splicing yields circular RNA molecules. *The FASEB Journal*, *7*(1), 155-160.

Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salmanidis, M., Phillips, C. A., Roslan, S., Schreiber, A. W., Gregory, P. A. & Goodall, G. J. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell*, *160*(6), 1125-1134.

Coulter, L. R., Landree, M. A. & Cooper, T. A. (1997). Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Molecular and Cellular Biology*, *17*(4), 2143-2150.

Cox, L. S. & Lane, D. P. (1995). Tumour suppressors, kinases and clamps: how p53 regulates the cell cycle in response to DNA damage. *Bioessays*, *17*(6), 501-508.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561-563.

Crooke, S. T., Wang, S., Vickers, T. A., Shen, W. & Liang, X. H. (2017). Cellular uptake and trafficking of antisense oligonucleotides. *Nature Biotechnology*, *35*(3), 230-237.

Custódio, N. & Carmo-Fonseca, M. (2016). Co-transcriptional splicing and the CTD code. *Critical Reviews in Biochemistry and Molecular Biology*, *51*(5), 395-411.

D'Agostino, L., Caracciolo, V. & Giordano, A. (2010). NSP 5a3a's link to nuclear-cyto proteins B23 and hnRNP-L between normal and aberrant breast cell lines. *Cell Cycle*, *9*(6), 1131-1142.

Dahl, M., Daugaard, I., Andersen, M. S., Hansen, T. B., Grønbæk, K., Kjems, J. & Kristensen, L. S. (2018). Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Laboratory Investigation*, *98*(12), 1657-1669.

Dai, P., Wang, Q., Wang, W., Jing, R., Wang, W., Wang, F., Azadzoi, K. M., Yang, J. H. & Yan, Z. (2016). Unravelling molecular differences of gastric cancer by label-free quantitative proteomics analysis. *International Journal of Molecular Sciences*, *17*(1), 69.

Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Research*, *40*(7), 3131-3142.

Darnell, J. E. (1982). Variety in the level of gene control in eukaryotic cells. *Nature*, *297*(5865), 365-371.

Dash, S., Brastrom, L. K., Patel, S. D., Scott, C. A., Slusarski, D. C. & Lachke, S. A. (2020). The master transcription factor SOX2, mutated in anophthalmia/microphthalmia, is post-transcriptionally regulated by the conserved RNA-binding protein RBM24 in vertebrate eye development. *Human Molecular Genetics*, *29*(4), 591-604.

Deprey, K., Batistatou, N. & Kritzer, J. A. (2020). A critical analysis of methods used to investigate the cellular uptake and subcellular localization of RNA therapeutics. *Nucleic Acids Research*, *48*(14), 7623-7639.

Dery, K. J., Gaur, S., Gencheva, M., Yen, Y., Shively, J. E. & Gaur, R. K. (2011). Mechanistic control of carcinoembryonic antigen-related cell adhesion molecule-1 (CEACAM1) splice isoforms by the heterogeneous nuclear ribonuclear proteins hnRNP L, hnRNP A1, and hnRNP M. *Journal of Biological Chemistry*, *286*(18), 16039-16051.

Dery, K. J., Silver, C., Yang, L. & Shively, J. E. (2018). Interferon regulatory factor 1 and a variant of heterogeneous nuclear ribonucleoprotein L coordinately silence the gene for adhesion protein CEACAM1. *Journal of Biological Chemistry*, *293*(24), 9277-9291.

Di Timoteo, G., Dattilo, D., Centrón-Broco, A., Colantoni, A., Guarnacci, M., Rossi, F., Incarnato, D., Oliviero, S., Fatica, A., Morlando, M. & Bozzoni, I. (2020). Modulation of circRNA metabolism by m6A modification. *Cell Reports*, *31*(6), 107641.

Domdey, H., Apostol, B., Lin, R. J., Newman, A., Brody, E. & Abelson, J. (1984). Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell*, *39*(3), 611-621.

Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., Van Nostrand, E. L., Pratt, G. A. & Yeo, G. W. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Molecular Cell*, *70*(5), 854-867.

Dong, R., Ma, X. K., Li, G. W. & Yang, L. (2018). CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics, Proteomics & Bioinformatics*, *16*(4), 226-233.

Dreyfuss, G., Kim, V. N. & Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews Molecular Cell Biology*, *3*(3), 195-205.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. & Burd, C. G. (1993). HnRNP proteins and the biogenesis of mRNA. *Annual Review of Biochemistry*, *62*(1), 289-321.

Du, C., Shen, Z., Zang, R., Xie, H., Li, H., Chen, P., Hang, B., Xu, X., Tang, W. & Xia, Y. (2016). Negative feedback circuitry between MIR143HG and RBM24 in Hirschsprung disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1862*(11), 2127-2136.

Du, W. W., Fang, L., Yang, W., Wu, N., Awan, F. M., Yang, Z. & Yang, B. B. (2017). Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell Death & Differentiation*, *24*(2), 357-370.

Du, W. W., Yang, W., Liu, E., Yang, Z., Dhaliwal, P. & Yang, B. B. (2016). Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Research*, *44*(6), 2846-2858.

Dudekula, D. B., Panda, A. C., Grammatikakis, I., De, S., Abdelmohsen, K. & Gorospe, M. (2016). CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biology*, *13*(1), 34-42.

Duque, P. (2011). A role for SR proteins in plant stress responses. *Plant signaling & behavior*, *6*(1), 49-54.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell*, *20*(2), 313-319.

Eger, N., Schoppe, L., Schuster, S., Laufs, U. & Boeckel, J. N. (2018). Circular RNA splicing. *Circular RNAs*, 41-52.

Eisenberg, E. & Levanon, E. Y. (2018). A-to-I RNA editing—immune protector and transcriptome diversifier. *Nature Reviews Genetics*, *19*(8), 473-490.

Ellington, A. D. & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*(6287), 818-822.

Errichelli, L., Modigliani, S.D., Laneve, P., Colantoni, A., Legnini, I., Capauto, D., Rosa, A., De Santis, R., Scarfo, R., Peruzzi, G. & Lu, L. (2017). FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nature Communications*, *8*(1), 1-11.

Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. *Science*, *297*(5583), 1007-1013.

Fei, T., Chen, Y., Xiao, T., Li, W., Cato, L., Zhang, P., Cotter, M. B., Bowden, M., Lis, R. T., Zhao, S. G. & Wu, Q. (2017). Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proceedings of the National Academy of Sciences*, *114*(26), E5207-E5215.

Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P. & Piccirilli, J. A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature*, *503*(7475), 229-234.

Filonov, G. S., Kam, C. W., Song, W. & Jaffrey, S. R. (2015). In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chemistry & Biology*, *22*(5), 649-660.

Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. (2014). Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society*, *136*(46), 16299-16308.

Finan, C., Gaulton, A., Kruger, F. A., Lumbers, R. T., Shah, T., Engmann, J., Galver, L., Kelley, R., Karlsson, A., Santos, R. & Overington, J. P. (2017). The druggable genome and support for target identification and validation in drug development. *Science Translational Medicine*, *9*(383).

Fischer, J. W., Busa, V. F., Shao, Y. & Leung, A. K. (2020). Structure-mediated RNA decay by UPF1 and G3BP1. *Molecular Cell*, *78*(1), 70-84.

Fletcher, S., Meloni, P. L., Johnsen, R. D., Wong, B. L., Muntoni, F. & Wilton, S. D. (2013). Antisense suppression of donor splice site mutations in the dystrophin gene transcript. *Molecular Genetics & Genomic Medicine*, *1*(3), 162-173.

Förch, P. & Valcarcel, J. (2001). Molecular mechanisms of gene expression regulation by the apoptosis-promoting protein TIA-1. *Apoptosis*, *6*(6), 463-468.

Ford, E. & Ares, M. (1994). Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proceedings of the National Academy of Sciences*, *91*(8), 3117-3121.

Fox-Walsh, K. L., Dou, Y., Lam, B. J., Hung, S. P., Baldi, P. F. & Hertel, K. J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proceedings of the National Academy of Sciences*, *102*(45), 16176-16181.

Frugier, T., Nicole, S., Cifuentes-Diaz, C. & Melki, J. (2002). The molecular bases of spinal muscular atrophy. *Current Opinion in Genetics & Development*, *12*(3), 294-298.

Fu, X. D. & Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics*, *15*(10), 689-701.

Gao, Y., Wang, J. & Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biology*, *16*(1), 1-16.

Gaudreau, M.C., Grapton, D., Helness, A., Vadnais, C., Fraszczak, J., Shooshtarizadeh, P., Wilhelm, B., Robert, F., Heyd, F. & Möröy, T. (2016). Heterogeneous Nuclear

Ribonucleoprotein L is required for the survival and functional integrity of murine hematopoietic stem cells. *Scientific reports*, *6*(1), 1-15.

Gebauer, F., Schwarzl, T., Valcárcel, J. & Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nature Reviews Genetics*, *22*(3), 185-198.

Glažar, P., Papavasileiou, P. & Rajewsky, N. (2014). circBase: a database for circular RNAs. *RNA*, *20*(11), 1666-1670.

Goehe, R. W., Shultz, J. C., Murudkar, C., Usanovic, S., Lamour, N. F., Massey, D. H., Zhang, L., Camidge, D. R., Shay, J. W., Minna, J. D. & Chalfant, C. E. (2010). HnRNP L regulates the tumorigenic capacity of lung cancer xenografts in mice via caspase-9 pre-mRNA processing. *The Journal of Clinical Investigation*, *120*(11), 3923-3939.

Good, P. D., Krikos, A. J., Li, S. X. L., Bertrand, E., Lee, N. S., Giver, L., Ellington, A., Zaia, J. A., Rossi, J. J. & Engelke, D. R. (1997). Expression of small, therapeutic RNAs in human cell nuclei. *Gene Therapy*, *4*(1), 45-54.

Grabowski, P. J., Zaug, A. J. & Cech, T. R. (1981). The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of Tetrahymena. *Cell*, *23*(2), 467-476.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *RNA, 6*(9), 1197-1211.

Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. *TRENDS in Genetics*, *17*(2), 100-107.

Grifone, R., Shao, M., Saquet, A., & Shi, D. L. (2020). RNA-binding protein Rbm24 as a multifaceted post-transcriptional regulator of embryonic lineage differentiation and cellular homeostasis. *Cells*, *9*(8), 1891.

Grifone, R., Xie, X., Bourgeois, A., Saquet, A., Duprez, D. & Shi, D. L. (2014). The RNA-binding protein Rbm24 is transiently expressed in myoblasts and is required for myogenic differentiation during vertebrate development. *Mechanisms of Development*, *134*, 1-15.

Gross, H. J., Domdey, H., Lossow, C., Jank, P., Raba, M., Alberty, H. & Sänger, H. L. (1978). Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature*, *273*(5659), 203-208.

Gruenbaum, Y. & Medalia, O. (2015). Lamins: the structure and protein complexes. *Current Opinion in Cell Biology*, *32*, 7-12.

Gu, J., Chen, Z., Chen, X. & Wang, Z. (2020). Heterogeneous nuclear ribonucleoprotein (hnRNPL) in cancer. *Clinica Chimica Acta*, *507*, 286-294.

Guang, S., Felthauser, A. M. & Mertz, J. E. (2005). Binding of hnRNP L to the pre-mRNA processing enhancer of the herpes simplex virus thymidine kinase gene enhances both polyadenylation and nucleocytoplasmic export of intronless mRNAs. *Molecular and Cellular Biology*, *25*(15), 6303-6313.

Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. (2014). Expanded identification and characterisation of mammalian circular RNAs. *Genome biology*, *15*(7), 1-14.

Hahm, B., Cho, O. H., Kim, J. E., Kim, Y. K., Kim, J. H., Oh, Y. L. & Jang, S. K. (1998a). Polypyrimidine tract-binding protein interacts with HnRNP L. *FEBS letters*, *425*(3), 401-406.

Hahm, B., Kim, Y. K., Kim, J. H., Kim, T. Y. & Jang, S. K. (1998b). Heterogeneous nuclear ribonucleoprotein L interacts with the 3′ border of the internal ribosomal entry site of hepatitis C virus. *Journal of Virology*, *72*(11), 8782-8788.

Hamilton, B. J., Nichols, R. C., Tsukamoto, H., Boado, R. J., Pardridge, W. M. & Rigby, W. F. (1999). HnRNP A2 and hnRNP L bind the 3′ UTR of glucose transporter 1 mRNA and exist as a complex *in vivo*. *Biochemical and Biophysical Research Communications*, *261*(3), 646-651.

Hanniford, D., Ulloa-Morales, A., Karz, A., Berzoti-Coelho, M. G., Moubarak, R. S., Sánchez-Sendra, B. & Hernando, E. (2020). Epigenetic silencing of CDR1as drives IGF2BP3-mediated melanoma invasion and metastasis. *Cancer Cell*, *37*(1), 55-70.

Hansen, T. B. (2018). Improved circRNA identification by combining prediction algorithms. *Frontiers in Cell and Developmental Biology*, *6*, 20.

Hansen, T. B. (2021). Signal and noise in circRNA translation. *Methods*.

Hansen, T. B., Venø, M. T., Damgaard, C. K. & Kjems, J. (2016). Comparison of circular RNA prediction tools. *Nucleic Acids Research*, *44*(6), e58-e58.

Hansen, T. B., Wiklund, E. D., Bramsen, J. B., Villadsen, S. B., Statham, A. L., Clark, S. J. & Kjems, J. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *The EMBO Journal*, *30*(21), 4414-4422.

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K. & Kjems, J., (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, *495*(7441), 384-388.

Heiner, M., Hui, J., Schreiner, S., Hung, L. H. & Bindereif, A. (2010). HnRNP L-mediated regulation of mammalian alternative splicing by interference with splice site recognition. *RNA Biology*, *7*(1), 56-64.

Hentze, M. W., & Preiss, T. (2013). Circular RNAs: splicing's enigma variations. *The EMBO Journal*, *32*(7), 923-925.

Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, *19*(5), 327-341.

Holdt, L. M., Stahringer, A., Sass, K., Pichler, G., Kulak, N. A., Wilfert, W. & Teupser, D. (2016). Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nature Communications*, *7*(1), 1-14.

Houseley, J. M., Garcia-Casado, Z., Pascual, M., Paricio, N., O'Dell, K. M., Monckton, D. G. & Artero, R. D. (2006). Noncanonical RNAs from transcripts of the Drosophila muscleblind gene. *Journal of Heredity*, *97*(3), 253-260.

Ho-Xuan, H., Glažar, P., Latini, C., Heizler, K., Haase, J., Hett, R., Anders, M., Weichmann, F., Bruckmann, A., Van den Berg, D., Hüttelmaier, S., Rajewsky, N., Hackl, C. & Meister, G. (2020). Comprehensive analysis of translation from overexpressed circular RNAs reveals pervasive translation from linear transcripts. *Nucleic Acids Research*, *48*(18), 10368-10382.

Hsiao, K. Y., Lin, Y. C., Gupta, S. K., Chang, N., Yen, L., Sun, H. S. & Tsai, S. J. (2017). Noncoding effects of circular RNA CCDC66 promote colon cancer growth and metastasis. *Cancer Research*, *77*(9), 2339-2350.

Hsu, M. T. & Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature*, *280*(5720), 339-340.

Hua, W. F., Zhong, Q., Xia, T. L., Chen, Q., Zhang, M. Y., Zhou, A. J., Tu, Z. W., Qu, C., Li, M. Z., Xia, Y. F. & Wang, H. Y. (2016). RBM24 suppresses cancer progression by upregulating miR-25 to target MALAT1 in nasopharyngeal carcinoma. *Cell Death & Disease*, *7*(9), 2352-2352.

Huang, C., Liang, D., Tatomer, D. C. & Wilusz, J. E. (2018). A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. *Genes & Development*, *32*(9-10), 639-644.

Huang, S., Li, X., Zheng, H., Si, X., Li, B., Wei, G. & Bin, J. (2019). Loss of super-enhancer-regulated circRNA Nfix induces cardiac regeneration after myocardial infarction in adult mice. *Circulation*, *139*(25), 2857-2876.

Hui, J., Hung, L.H., Heiner, M., Schreiner, S., Neumüller, N., Reither, G., Haas, S.A. & Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *The EMBO Journal*, *24*(11), 1988-1998.

Hui, J., Reither, G. & Bindereif, A. (2003a). Novel functional role of CA repeats and hnRNP L in RNA stability. *RNA*, *9*(8), 931-936.

Hui, J., Stangl, K., Lane, W. S. & Bindereif, A. (2003b). HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nature Structural Biology*, *10*(1), 33-37.

Hung, L. H., Heiner, M., Hui, J., Schreiner, S., Benes, V. & Bindereif, A. (2008). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA*, *14*(2), 284-296.

Hwang, B., Lim J. H., Hahm, B., Jang S. K. & Lee, S. (2009). HnRNP L is required for the translation mediated by HCV IRES. *Biochemical and Biophysical Research Communications 378*: 584–588

Iaboni, M., Fontanella, R., Rienzo, A., Capuozzo, M., Nuzzo, S., Santamaria, G., Catuogno, S., Condorelli, G., de Franciscis, V. & Esposito, C. L. (2016). Targeting Insulin Receptor with a Novel Internalizing Aptamer. *Molecular Therapy-Nucleic Acids*, *5*(9), e365.

Inoue, H., Hayase, Y., Imura, A., Iwai, S., Miura, K. & Ohtsuka, E. (1987). Synthesis and hybridisation studies on two complementary nona (2'-O-methyl) ribonucleotides. *Nucleic Acids Research*, *15*(15), 6131-6148.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H. T., Orejuela, M.R., Piechotta, M., Levanon, E. Y., Landthaler, M., Dieterich, C. & Rajewsky, N. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell reports*, *10*(2), 170-177.

Jafarifar, F., Yao, P., Eswarappa, S. M. & Fox, P. L. (2011). Repression of VEGFA by CA-rich element-binding microRNAs is modulated by hnRNP L. *The EMBO Journal*, *30*(7), 1324-1334.

Jeck, W. R. & Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nature Biotechnology*, *32*(5), 453-461.

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., Marzluff, W. F. & Sharpless, N. E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, *19*(2), 141-157.

Jensen, K. B., Dredge, B. K., Stefani, G., Zhong, R., Buckanovich, R. J., Okano, H. J., Yang, Y. Y. & Darnell, R. B. (2000). Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, *25*(2), 359-371.

Ji, J., Xu, R., Ding, K., Bao, G., Zhang, X., Huang, B., Wang, X., Martinez, A., Wang, X., Li, G. & Miletic, H. (2019). Long noncoding RNA SChLAP1 forms a growth-promoting complex with hnRNP L in human glioblastoma through stabilization of ACTN4 and activation of NF-κB signalling. *Clinical Cancer Research*, *25*(22), 6868-6881.

Jost, I., Shalamova, L. A., Gerresheim, G. K., Niepmann, M., Bindereif, A., & Rossbach, O. (2018). Functional sequestration of microRNA-122 from Hepatitis C Virus by circular RNA sponges. *RNA Biology*, *15*(8), 1032-1039.

Kaminski, A., Hunt, S. L., Patton, J. G. & Jackson, R. J. (1995). Direct evidence that polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA. *RNA*, *1*(9), 924-938.

Kastner, B., Will, C. L., Stark, H. & Lührmann, R. (2019). Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *11*(11), a032417.

Ke, S., Anquetil, V., Zamalloa, J. R., Maity, A., Yang, A., Arias, M. A., Kalachikov, S., Russo, J. J., Ju, J. & Chasin, L. A., (2018). Saturation mutagenesis reveals manifold determinants of exon definition. *Genome Research*, *28*(1), 11-24.

Kelaini, S., Chan, C., Cornelius, V. A. & Margariti, A. (2021). RNA-Binding Proteins Hold Key Roles in Function, Dysfunction, and Disease. *Biology*, *10*(5), 366.

Kelly, S., Greenman, C., Cook, P. R. & Papantonis, A. (2015). Exon skipping is correlated with exon circularisation. *Journal of Molecular Biology*, *427*(15), 2414-2417.

Kim, J. H., Hahm, B., Kim, Y. K., Choi, M. & Jang, S. K. (2000). Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *Journal of Molecular Biology*, *298*(3), 395-405.

Kim, J., Hu, C., Moufawad El Achkar, C., Black, L. E., Douville, J., Larson, A., Pendergast, M. K., Goldkind, S. F., Lee, E. A., Kuniholm, A. & Soucy, A. (2019). Patient-customised oligonucleotide therapy for a rare genetic disease. *New England Journal of Medicine*, *381*(17), 1644-1652.

Kishor, A., Ge, Z. & Hogg, J. R. (2019). HnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. *The EMBO Journal*, *38*(3), e99128.

Kjems, J. & Garrett, R. A. (1988). Novel splicing mechanism for the ribosomal RNA intron in the archaebacterium *Desulfurococcus mobilis*. *Cell*, 54, 693–703.

Kleaveland, B., Shi, C. Y., Stefano, J. & Bartel, D. P. (2018). A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell*, *174*(2), 350-362.

Klingenberg, M., Groß, M., Goyal, A., Polycarpou-Schwarz, M., Miersch, T., Ernst, A.S., Leupold, J., Patil, N., Warnken, U., Allgayer, H. & Longerich, T. (2018). The long noncoding RNA cancer susceptibility 9 and RNA binding protein heterogeneous nuclear ribonucleoprotein L form a complex and co-regulate genes linked to AKT signalling. *Hepatology*, *68*(5), 1817-1832.

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annual Review of Biochemistry*, *79*, 213-231.

Kohlberger, M. & Gadermaier, G. (2021). SELEX: Critical factors and optimization strategies for successful aptamers selection. *Biotechnology and Applied Biochemistry*. https://doi.org/10.1002/bab.2244

Komar, A. A. & Hatzoglou, M. (2005). Internal ribosome entry sites in cellular mRNAs: mystery of their existence. *Journal of Biological Chemistry*, *280*(25), 23425-23428.

Kos, A., Dijkema, R., Arnberg, A. C., Van der Meide, P. H. & Schellekens, H. (1986). The hepatitis delta (δ) virus possesses a circular RNA. *Nature*, *323*(6088), 558-560.

Kramer, M. C., Liang, D., Tatomer, D. C., Gold, B., March, Z. M., Cherry, S. & Wilusz, J. E. (2015). Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes & Development*, *29*(20), 2168-2182.

Kristensen, L. S., Andersen, M. S., Stagsted, L. V., Ebbesen, K. K., Hansen, T. B. & Kjems, J. (2019). The biogenesis, biology and characterisation of circular RNAs. *Nature Reviews Genetics*, *20*(11), 675-691.

Kristensen, L. S., Hansen, T. B., Venø, M. T. & Kjems, J. (2018a). Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene*, *37*(5), 555-565.

Kristensen, L. S., Okholm, T. L. H., Venø, M. T. & Kjems, J. (2018b). Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. *RNA Biology*, *15*(2), 280-291.

Lee, Y. & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annual Review of Biochemistry*, *84*, 291-323.

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M. & Laneve, P. (2017). Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Molecular Cell*, *66*(1), 22-37.

Lei, M., Zheng, G., Ning, Q., Zheng, J. & Dong, D. (2020). Translation and functional roles of circular RNAs in human cancer. *Molecular Cancer*, *19*(1), 1-9.

Lerner, M. R. & Steitz, J. A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences*, *76*(11), 5495-5499.

Letunic, I., Khedkar, S. & Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic acids research*, *49*(D1), D458-D460.

Li, H. Y., Bourdelas, A., Carron, C. & Shi, D. L. (2010). The RNA-binding protein Seb4/RBM24 is a direct target of MyoD and is required for myogenesis during Xenopus early development. *Mechanisms of Development*, *127*(5-6), 281-291.

Li, X. F. & Lytton, J. (1999). A circularised sodium-calcium exchanger exon 2 transcript. *Journal of Biological Chemistry*, *274*(12), 8153-8160.

Li, X., Liu, C. X., Xue, W., Zhang, Y., Jiang, S., Yin, Q. F., Wei, J., Yao, R. W., Yang, L. & Chen, L. L. (2017). Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Molecular cell*, *67*(2), 214-227.

Li, X., Liu, S., Zhang, L., Issaian, A., Hill, R. C., Espinosa, S., Shi, S., Cui, Y., Kappel, K., Das, R. & Hansen, K.C. (2019). A unified mechanism for intron and exon definition and back-splicing. *Nature*, *573*(7774), 375-380.

Li, Y., Chen, B., Zhao, J., Li, Q., Chen, S., Guo, T., Li, Y., Lai, H., Chen, Z., Meng, Z. & Guo, W. (2021). HnRNPL Circularises ARHGAP35 to Produce an Oncogenic Protein. *Advanced Science*, 2001701.

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X. & Shan, G. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nature Structural & Molecular Biology*, *22*(3), 256-264.

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L. & Zhu, P. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nature Structural & Molecular Biology*, *22*(3), 256-264.

Liang, D. & Wilusz, J. E. (2014). Short intronic repeat sequences facilitate circular RNA production. *Genes & Development*, *28*(20), 2233-2247.

Liang, D., Tatomer, D. C., Luo, Z., Wu, H., Yang, L., Chen, L. L., Cherry, S. & Wilusz, J.E. (2017). The output of protein-coding genes shifts to circular RNAs when the pre-mRNA processing machinery is limiting. *Molecular cell*, *68*(5), 940-954.

Lim, L. P. & Burge, C. B. (2001). A computational analysis of sequence features involved in recognition of short introns. *Proceedings of the National Academy of Sciences*, *98*(20), 11193-11198.

Lin, Y., Tan, K. T., Liu, J., Kong, X., Huang, Z. & Xu, X. Q. (2018). Global profiling of Rbm24 bound RNAs uncovers a multi-tasking RNA binding protein. *The International Journal of Biochemistry & Cell Biology*, *94*, 10-21.

Lischwe, M. A., Cook, R. G., Ahn, Y. S., Yeoman, L. C. & Busch, H. (1985). Clustering of glycine and NG, NG-dimethylarginine in nucleolar protein C23. *Biochemistry*, *24*(22), 6025-6028.

Litke, J. L. & Jaffrey, S. R. (2019). Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nature Biotechnology*, *37*(6), 667-675.

Liu, C., Wang, J., Yuan, X., Qian, W., Zhang, B., Shi, M., Xie, J., Shen, B., Xu, H., Hou, Z. & Chen, H. (2016). Long noncoding RNA uc. 345 promotes tumorigenesis of pancreatic cancer by upregulation of hnRNPL expression. *Oncotarget*, *7*(44), 71556.

Liu, C. X., Li, X., Nan, F., Jiang, S., Gao, X., Guo, S. K., Xue, W., Cui, Y., Dong, K., Ding, H. & Qu, B. (2019b). Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell*, *177*(4), 865-880.

Liu, J., Kong, X., Zhang, M., Yang, X. & Xu, X. (2019a). RNA binding protein 24 deletion disrupts global alternative splicing and causes dilated cardiomyopathy. *Protein & Cell*, *10*(6), 405-416.

Liu, X. & Mertz, J. E. (1995). HnRNP L binds a cis-acting RNA sequence element that enables intron-dependent gene expression. *Genes & Development*, *9*(14), 1766-1780.

Llorian, M. & Smith, C. W. (2011). Decoding muscle alternative splicing. *Current Opinion in Genetics & Development*, *21*(4), 380-387.

Long, J. C. & Caceres, J. F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, *417*(1), 15-27.

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters*, *579*(9), 1900-1903.

Lou, J., Hao, Y., Lin, K., Lyu, Y., Chen, M., Wang, H. & Jin, B. (2020). Circular RNA CDR1as disrupts the p53/MDM2 complex to inhibit Gliomagenesis. *Molecular Cancer*, *19*(1), 1-19.

Lu, T., Cui, L., Zhou, Y., Zhu, C., Fan, D., Gong, H., Zhao, Q., Zhou, C., Zhao, Y., Lu, D. & Luo, J. (2015). Transcriptome-wide investigation of circular RNAs in rice. *RNA, 21*(12), 2076-2087.

Lucchesi, C. A., Zhang, J., Ma, B., Chen, M. & Chen, X. (2019). Disruption of the Rbm38-eIF4E complex with a synthetic peptide Pep8 increases p53 expression. *Cancer Research*, *79*(4), 807-818.

Luckow, V. A., Lee, S. C., Barry, G. F. & Olins, P. (1993). Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *Journal of Virology*, *67*(8), 4566-4579.

Luo, Y. B., Mastaglia, F. L., & Wilton, S. D. (2014). Normal and aberrant splicing of LMNA. *Journal of Medical Genetics*, *51*(4), 215-223.

Lv, D., Wu, H., Xing, R., Shu, F., Lei, B., Lei, C., Zhou, X., Wan, B., Yang, Y., Zhong, L. & Mao, X. (2017). HnRNP L mediates bladder cancer progression by inhibiting apoptotic signalling and enhancing MAPK signalling pathways. *Oncotarget*, *8*(8), 13586.

Ma, X. K., Wang, M. R., Liu, C. X., Dong, R., Carmichael, G. G., Chen, L. L. & Yang, L. (2019). CIRCexplorer3: a CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics, Proteomics & Bioinformatics*, *17*(5), 511-521.

Majumder, M., Yaman, I., Gaccioli, F., Zeenko, V.V., Wang, C., Caprara, M.G., Venema, R.C., Komar, A.A., Snider, M.D. & Hatzoglou, M. (2009). The hnRNA-binding proteins hnRNP L and PTB are required for efficient translation of the Cat-1 arginine/lysine transporter mRNA during amino acid starvation. *Molecular and Cellular Biology*, *29*(10), 2899-2912.

Maragh, S., Miller, R. A., Bessling, S. L., McGaughey, D. M., Wessels, M. W., De Graaf, B., Stone, E. A., Bertoli-Avella, A. M., Gearhart, J.D., Fisher, S. & McCallion, A. S. (2011). Identification of RNA binding motif proteins essential for cardiovascular development. *BMC Developmental Biology*, *11*(1), 1-13.

Marshall, L. & White, R. J. (2008). Non-coding RNA production by RNA polymerase III is implicated in cancer. *Nature Reviews Cancer*, *8*(12), 911-914.

Matera, A. G. & Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, *15*(2), 108-121.

Matlin, A. J., Clark, F. & Smith, C. W. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology*, *6*(5), 386-398.

Matsukura, M., Shinozuka, K., Zon, G., Mitsuya, H., Reitz, M., Cohen, J. S. & Broder, S. (1987). Phosphorothioate analogs of oligodeoxynucleotides: inhibitors of replication and cytopathic effects of human immunodeficiency virus. *Proceedings of the National Academy of Sciences*, *84*(21), 7706-7710.

Mayrand S.H., Dwen P. & Pederson T. (1993). Serine/threonine phosphorylation regulates binding of C hnRNP proteins to pre-mRNA. *Proceedings of the National Academy of Sciences*. *90*(16):7764-7768.

McGlincy, N. J. & Smith, C. W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in Biochemical Sciences*, *33*(8), 385-393.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M. & Loewer, A. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, *495*(7441), pp.333-338.

Mohanta, A. & Chakrabarti, K. (2020). Dbr1 functions in mRNA processing, intron turnover and human diseases. *Biochimie*, 180, 134-142.

Montes, M., Sanford, B. L., Comiskey, D. F. & Chandler, D. S. (2019). RNA splicing and disease: animal models to therapies. *Trends in Genetics*, *35*(1), 68-87.

Morris, A. R., Mukherjee, N. & Keene, J. D. (2010). Systematic analysis of posttranscriptional gene expression. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *2*(2), 162-180.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621-628.

Mueller, W. F. & Hertel, K. J. (2012). The role of SR and SR-related proteins in pre-mRNA splicing. *In RNA Binding Proteins*, 27-46.

Müller, M., Samel-Pommerencke, A., Legrand, C., Tuorto, F., Lyko, F. & Ehrenhofer-Murray, A. E. (2019). Division of labour: tRNA methylation by the NSun2 tRNA methyltransferases Trm4a and Trm4b in fission yeast. *RNA Biology*, *16*(3), 249-256.

Müller, S. & Appel, B. (2017). In vitro circularisation of RNA. *RNA biology*, *14*(8), 1018-1027.

Niblock, M. & Gallo, J. M. (2012). Tau alternative splicing in familial and sporadic tauopathies. *Biochemical Society Transactions*, *40*(4), 677-680.

Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., Kinzler, K. W. & Vogelstein, B. (1991). Scrambled exons. *Cell*, *64*(3), 607-613.

Niknafs, Y. S., Han, S., Ma, T., Speers, C., Zhang, C., Wilder-Romans, K. & Feng, F. Y. (2016). The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. *Nature Communications*, *7*(1), 1-13.

Nilsen, T. W. & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, *463*(7280), 457-463.

Nimjee, S. M., White, R. R., Becker, R. C. & Sullenger, B. A. (2017). Aptamers as therapeutics. *Annual Review of Pharmacology and Toxicology*, *57*, 61-79.

Noto, J. J., Schmidt, C. A. & Matera, A. G. (2017). Engineering and expressing circular RNAs via tRNA splicing. *RNA Biology*, *14*(8), 978-984.

Ohe, K., Yoshida, M., Nakano-Kobayashi, A., Hosokawa, M., Sako, Y., Sakuma, M. & Hagiwara, M. (2017). RBM24 promotes U1 snRNP recognition of the mutated 5′ splice site in the IKBKAP gene of familial dysautonomia. *RNA*, *23*(9), 1393-1403.

Okholm, T. L. H., Nielsen, M. M., Hamilton, M. P., Christensen, L. L., Vang, S., Hedegaard, J., Hansen, T. B., Kjems, J., Dyrskjøt, L. & Pedersen, J. S. (2017). Circular RNA expression is abundant and correlated to aggressiveness in early-stage bladder cancer. *NPJ Genomic Medicine*, *2*(1), 1-14.

Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science*, *225*(4665), 898-903.

Pagotto, S., Veronese, A., Soranno, A., Balatti, V., Ramassone, A., Guanciali-Franchi, P.E., Palka, G., Innocenti, I., Rassenti, L. Z., Kipps, T. J. & Mariani-Costantini, R. (2019). HnRNP L restrains miR-155 targeting of BUB1 to stabilise aberrant karyotypes of transformed cells in chronic lymphocytic leukaemia. *Cancers*, *11*(4), 575.

Paige, J. S., Wu, K. Y. & Jaffrey, S. R. (2011). RNA mimics of green fluorescent protein. *Science*, *333*(6042), 642-646.

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., Shenzis, S., Samson, M., Dittmar, G., Landthaler, M., Chekulaeva, M., Rajewsky, N. & Kadener, S. (2017). Translation of CircRNAs. *Molecular Cell*, *66*(1), 9–21.e7.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, *40*(12), 1413-1415.

Panda, A.C., De, S., Grammatikakis, I., Munk, R., Yang, X., Piao, Y., Dudekula, D.B., Abdelmohsen, K. & Gorospe, M. (2017). High-purity circular RNA isolation method (RPAD) reveals vast collection of intronic circRNAs. *Nucleic Acids Research*, *45*(12), e116-e116.

Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *The American Journal of Human Genetics*, *102*(1), 11-26.

Peddigari, S., Li, P. W. L., Rabe, J. L. & Martin, S. L. (2013). HnRNP L and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Research*, *41*(1), 575-585.

Peng, Z. Y., Huang, J., Lee, S. C., Shi, Y. L., Chen, X. H. & Xu, P. (2009). The expression pattern of heterogeneous nuclear ribonucleoprotein R in rat retina. *Neurochemical Research*, *34*(6), 1083-1088.

Pérez, I., McAfee, J. G. & Patton, J. G. (1997). Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry*, *36*(39), 11881-11890.

Perriman, R. & Ares, M. (1998). Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA, 4*(9), 1047-1054.

Pfafenrot, C., Schneider, T., Müller, C., Hung, L. H., Schreiner, S., Ziebuhr, J., & Bindereif, A. (2021). Inhibition of SARS-CoV-2 coronavirus proliferation by designer antisense-circRNAs. *Nucleic Acids Research*, *49*(21), 12502-12516.

Pfafenrot, C. & Preußer, C. (2019). Establishing essential quality criteria for the validation of circular RNAs as biomarkers. *Biomolecular Detection and Quantification*, *17*, 100085.

Piñol-Roma, S., Swanson, M. S., Gall, J. G. & Dreyfuss, G. (1989). A novel heterogeneous nuclear RNP protein with a unique distribution on nascent transcripts. *The Journal of Cell Biology*, *109*(6), 2575-2587.

Piwecka M., Glažar P., Hernandez-Miranda L. R., Memczak S., Wolf S. A., Rybak-Wolf A., Filipchyk A., Klironomos F., Cerda Jara C. A., Fenske P., Trimbuch T., Zywitza V., Plass M., Schreyer L., Ayoub S., Kocks C., Kühn R., Rosenmund C., Birchmeier C. & Rajewsky N (2017). Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science*, 357(6357).

Plaschka, C., Lin, P. C. & Nagai, K. (2017). Structure of a pre-catalytic spliceosome. *Nature*, *546*(7660), 617-621.

Plaschka, C., Newman, A. J. & Nagai, K. (2019). Structural basis of nuclear pre-mRNA splicing: lessons from yeast. *Cold Spring Harbour Perspectives in Biology*, *11*(5), a032391.

Popow, J., Schleiffer, A. & Martinez, J. (2012). Diversity and roles of (t) RNA ligases. *Cellular and Molecular Life Sciences*, *69*(16), 2657-2670.

Pozzoli, U., Sironi, M., Cagliani, R., Comi, G. P., Bardoni, A. & Bresolin, N. (2002). Comparative analysis of the human dystrophin and utrophin gene structures. *Genetics*, *160*(2), 793-798.

Prakash, T. P., Lima, W. F., Murray, H. M., Elbashir, S., Cantley, W., Foster, D., Jayaraman, M., Chappell, A. E., Manoharan, M., Swayze, E. E. & Crooke, S. T. (2013). Lipid nanoparticles improve activity of single-stranded siRNA and gapmer antisense oligonucleotides in animals. *ACS Chemical Biology*, *8*(7), 1402-1406.

Preußer, C., Hung, L. H., Schneider, T., Schreiner, S., Hardt, M., Moebus, A., Santoso, S. & Bindereif, A. (2018). Selective release of circRNAs in platelet-derived extracellular vesicles. *Journal of Extracellular Vesicles*, *7*(1), 1424473.

Preußner, M., Schreiner, S., Hung, L. H., Porstner, M., Jäck, H. M., Benes, V., Rätsch, G. & Bindereif, A. (2012). HnRNP L and L-like cooperate in multiple-exon regulation of CD45 alternative splicing. *Nucleic Acids research*, *40*(12), 5666-5678.

Qian, K., Li, M., Wang, J., Zhang, M. & Wang, M. (2020). Structural basis for mRNA recognition by human RBM38. *Biochemical Journal*, *477*(1), 161-172.

Quemener, A. M., Bachelot, L., Forestier, A., Donnou-Fournet, E., Gilot, D. & Galibert, M. D. (2020). The powerful world of antisense oligonucleotides: From bench to bedside. *Wiley Interdisciplinary Reviews: RNA*, *11*(5), e1594.

Raczynska, K. D., Simpson, C. G., Ciesiolka, A., Szewc, L., Lewandowska, D., McNicol, J., Szweykowska-Kulinska, Z., Brown, J.W. & Jarmolowski, A. (2010). Involvement of the nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Research*, *38*(1), 265-278.

Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Research*, *12*(8), 1231-1245.

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecenas, D., Li, W., Laishram, R. S., Qiao, M., Lipshitz, H. D., Piano, F., Corbett, A. H., Carstens, R. P., Frey, B. J., Anderson, R. A., Lynch, K. W., Penalva, L. F., Lei, E.

P., Fraser, A. G., Blencowe, B. J., Morris, Q. D. & Hughes, T. R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature 499* (7457), 172-177.

Ray, D., Kazan, H., Chan, E. T., Peña-Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B. J., Morris, Q. & Hughes, T. R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature Biotechnology, 27,* 667–670.

Reddy, A. S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review Plant Biol*ogy, *58*, 267-294.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Current Opinion in Cell Biology*, *12*(3), 340-345.

Rodriguez, J. R., Pikielny, C. W. & Rosbash, M. (1984). In vivo characterisation of yeast mRNA processing intermediates. *Cell*, *39*(3), 603-610.

Roscigno, R. F. & Garcia-Blanco, M. A. (1995). SR proteins escort the U4/U6. U5 tri-snRNP to the spliceosome. *RNA, 1*(7), 692-706.

Rossbach, O., Hung, L. H., Schreiner, S., Grishina, I., Heiner, M., Hui, J. & Bindereif, A. (2009). Auto-and cross-regulation of the hnRNP L proteins by alternative splicing. *Molecular and Cellular Biology*, *29*(6), 1442-1451.

Roth, A., Weinberg, Z., Chen, A. G., Kim, P. B., Ames, T. D. & Breaker, R. R. (2014). A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nature Chemical Biology*, *10*(1), 56-60.

Ruckman, J., Green, L. S., Beeson, J., Waugh, S., Gillette, W. L., Henninger, D. D., Claesson-Welsh, L. & Janjic, N. (1998). 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165): Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *Journal of Biological Chemistry*, *273*(32), 20556-20567.

Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*, *38*(1), 317-331.

Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S. & Rajewsky, N. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Molecular Cell*, *58*(5), 870-885.

Sakharkar, M. K., Chow, V. T. & Kangueane, P. (2004). Distributions of exons and introns in the human genome. *In silico Biology*, *4*(4), 387-393.

Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. (2013). Cell-type specific features of circular RNA expression. *PLoS Genetics*, *9*(9), e1003777.

Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PloS one*, *7*(2), e30733.

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J. & Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proceedings of the National Academy of Sciences*, *73*(11), 3852-3856.

Sayad, A., Hajifathali, A., Omrani, M. D., Arsang-Jang, S., Hamidieh, A. A., & Taheri, M. (2018). Expression of TNF-and hnRNP L-related immunoregulatory long non-coding RNA (THRIL) in acute myeloid leukaemia: is there any correlation? *Iranian Journal of Allergy, Asthma and Immunology*, 274-280.

Scherly, D., Boelens, W. V., Van Venrooij, W. J., Dathan, N. A., Hamm, J. & Mattaj, I. W. (1989). Identification of the RNA binding segment of human U1A protein and definition of its binding site on U1 snRNA. *The EMBO Journal*, *8*(13), 4163-4170.

Schlundt, A., Tants, J. N. & Sattler, M. (2017). Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. *Methods*, *118*, 119-136.

Schmidt, C. A., Giusto, J. D., Bao, A., Hopper, A. K. & Matera, A. G. (2019). Molecular determinants of metazoan tricRNA biogenesis. *Nucleic Acids Research*, *47*(12), 6452-6465.

Schneider, T. & Bindereif, A. (2017). Circular RNAs: coding or noncoding? *Cell Research*, *27*(6), 724-725.

Schneider, T., Hung, L. H., Aziz, M., Wilmen, A., Thaum, S., Wagner, J., Janowski, R., Müller, S., Schreiner, S., Friedhoff, P., Hüttelmaier, S. & Bindereif, A. (2019). Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nature Communications*, *10*(1), 1-18.

Schneider, T., Hung, L. H., Schreiner, S., Starke, S., Eckhof, H., Rossbach, O., Reich, S., Medenbach, J. & Bindereif, A. (2016). CircRNA-protein complexes: IMP3 protein component defines subfamily of circRNPs. *Scientific Reports*, *6*(1), 1-11.

Schreiner, S., Didio, A., Hung, L. H. & Bindereif, A. (2020). Design and application of circular RNAs with protein-sponge function. *Nucleic Acids Research*, *48*(21), 12326-12335.

Scotti, M. M. & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics*, *17*(1), 19-32.

Setten, R. L., Rossi, J. J. & Han, S. P. (2019). The current state and future directions of RNAi-based therapeutics. *Nature Reviews Drug Discovery*, *18*(6), 421-446.

Shao, M., Lu, T., Zhang, C., Zhang, Y. Z., Kong, S. H. & Shi, D. L. (2020). Rbm24 controls poly (A) tail length and translation efficiency of crystallin mRNAs in the lens via cytoplasmic polyadenylation. *Proceedings of the National Academy of Sciences*, *117*(13), 7245-7254.

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R. & Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research*, *34*(14), 3955-3967.

Shih, S. C. & Claffey, K. P. (1999). Regulation of human vascular endothelial growth factor mRNA stability in hypoxia by heterogeneous nuclear ribonucleoprotein L. *Journal of Biological Chemistry*, *274*(3), 1359-1365.

Siebring-van Olst, E., Blijlevens, M., de Menezes, R. X., van der Meulen-Muileman, I. H., Smit, E. F. & van Beusechem, V. W. (2017). A genome-wide si RNA screen for regulators of tumour suppressor p53 activity in human non-small cell lung cancer cells identifies components of the RNA splicing machinery as targets for anticancer treatment. *Molecular Oncology*, *11*(5), 534-551.

Sikora, D., Greco-Stewart, V. S., Miron, P. & Pelchat, M. (2009). The hepatitis delta virus RNA genome interacts with eEF1A1, p54nrb, hnRNP-L, GAPDH and ASF/SF2. *Virology*, *390*(1), 71-78.

Singh, A. J., Ramsey, S. A., Filtz, T. M. & Kioussi, C. (2018). Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences*, *75*(6), 1013-1025.

Singh, R. K. & Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends in Molecular Medicine*, *18*(8), 472-482.

Song, X., Zhang, N., Han, P., Moon, B. S., Lai, R. K., Wang, K. & Lu, W. (2016). Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic Acids Research*, *44*(9), e87-e87.

Spingola, M., Grate, L., Haussler, D. & Ares, M. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA*, *5*(2), 221-234.

Stagsted, L. V., Nielsen, K. M., Daugaard, I. & Hansen, T. B., (2019). Noncoding AUG circRNAs constitute an abundant and conserved subclass of circles. *Life Science Alliance*, *2*(3).

Staley, J. P. & Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, *92*(3), 315-326.

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L. H. & Bindereif, A. (2015). Exon circularisation requires canonical splice signals. *Cell Reports*, *10*(1), 103-111.

Steitz, T. A. & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences*, *90*(14), 6498-6502.

Stoll, L., Sobel, J., Rodriguez-Trejo, A., Guay, C., Lee, K., Venø, M. T., Kjems, J., Laybutt, D. R. & Regazzi, R. (2018). Circular RNAs as novel regulators of β-cell functions in normal and disease conditions. *Molecular Metabolism*, *9*, 69-83.

Sun, W., Hu, Y., Xu, H., He, H., Han, C., Liu, H., Wang, J. & Li, L. (2016). Characterisation of the duck (*Anas platyrhynchos*) Rbm24 and Rbm38 genes and their expression profiles in myoblast and skeletal muscle tissues. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *198*, 27-36.

Surono, A., Takeshima, Y., Wibawa, T., Ikezawa, M., Nonaka, I. & Matsuo, M. (1999). Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Human molecular genetics*, *8*(3), 493-500.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M. Q., Malhotra, A. & Mayeda, A. (2006). Characterisation of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res*earch, 34, e63–e69.

Swanson, M. S. & Dreyfuss, G. (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Molecular and Cellular Biology*, *8*(5), 2237-2241.

Szabo, L., Morey, R., Palpant, N. J., Wang, P. L., Afari, N., Jiang, C., Parast, M. M., Murry, C. E., Laurent, L. C. & Salzman, J. (2015). Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human foetal development. *Genome Biology*, *16*(1), 1-26.

Tang, C., Xie, Y., Yu, T., Liu, N., Wang, Z., Woolsey, R. J., Tang, Y., Zhang, X., Qin, W., Zhang, Y. & Song, G. (2020). m6A-dependent biogenesis of circular RNAs in male germ cells. *Cell Research*, *30*(3), 211-228.

Tatomer, D. C. & Wilusz, J. E. (2017). An unchartered journey for ribosomes: circumnavigating circular RNAs to produce proteins. *Molecular Cell*, *66*(1), 1-2.

Tazi, J., Bakkour, N. & Stamm, S. (2009). Alternative splicing and disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1792*(1), 14-26.

Tsui, L. C. & Dorfman, R. (2013). The cystic fibrosis gene: a molecular genetic perspective. *Cold Spring Harbor Perspectives in Medicine*, *3*(2), a009472.

Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, *249*(4968), 505-510.

Tuvshinjargal, N., Lee, W., Park, B. & Han, K. (2016). PRIdictor: protein–RNA interaction predictor. *Biosystems*, *139*, 17-22.

Ule, J. & Blencowe, B. J. (2019). Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Molecular Cell*, *76*(2), 329-345.

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A. & Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science*, *302*(5648), 1212-1215.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J. & Darnell, R.B. (2006). An RNA map predicting Nova-dependent splicing regulation. *Nature*, *444*(7119), 580-586.

Ulrich, H., Trujillo, C. A., Nery, A. A., Alves, J. M., Majumder, P., Resende, R. R. & Martins, A. H. (2006). DNA and RNA aptamers: from tools for basic research towards therapeutic applications. *Combinatorial Chemistry & High Throughput Screening*, *9*(8), 619-632.

Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P. & Black, D. L. (2005). Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. *Molecular and Cellular Biology*, *25*(22), 10005-10016.

Valverde, R., Edwards, L. & Regan, L. (2008). Structure and function of KH domains. *The FEBS Journal*, *275*(11), 2712-2726.

Verduci, L., Ferraiuolo, M., Sacconi, A., Ganci, F., Vitale, J., Colombo, T., Paci, P., Strano, S., Macino, G., Rajewsky, N. & Blandino, G. (2017). The oncogenic role of circPVT1 in head and neck squamous cell carcinoma is mediated through the mutant p53/YAP/TEAD transcription-competent complex. *Genome Biology*, *18*(1), 1-24.

Voelker, R. B. & Berglund, J. A. (2007). A comprehensive computational characterisation of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome research*, *17*(7), 1023-1033.

Wachter, A., Rühl, C. & Stauffer, E. (2012). The role of polypyrimidine tract-binding proteins and other hnRNP proteins in plant splicing regulation. *Frontiers in Plant Science*, *3*, 81.

Wan, W. B. & Seth, P. P. (2016). The medicinal chemistry of therapeutic oligonucleotides. *Journal of Medicinal Chemistry*, *59*(21), 9645-9667.

Wan, Y., Anastasakis, D. G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C. C., Hafner, M. & Larson, D. R. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*, *184*(11), 2878-2895.

Wang, B. B. & Brendel, V. (2004b). The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biology*, *5*(12), 1-23.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470-476.

Wang, G. S., & Cooper, T. A. (2007). Splicing in disease: Disruption of the splicing code and the decoding machinery. Nature Reviews Genetics, 8(10), 749–761.

Wang, Z. & Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA 14*(5), 802-813.

Wang, Z., Xiao, X., Van Nostrand, E. & Burge, C. B. (2006). General and specific functions of exonic splicing silencers in splicing control. *Molecular Cell*, *23*(1), 61-70.

Weeland, C. J., van den Hoogenhof, M. M., Beqqali, A. & Creemers, E. E. (2015). Insights into alternative splicing of sarcomeric genes in the heart. *Journal of Molecular and Cellular Cardiology*, *81*, 107-113.

Wei, P., Yang, J., Zhang, D., Cui, M. & Li, L. (2020). LncRNA HAND2-AS1 regulates prostate cancer cell growth through targeting the miR-106a-5p/RBM24 axis. *OncoTargets and Therapy*, *13*, 4523.

Weigand, J. E. & Suess, B. (2009). Aptamers and riboswitches: perspectives in biotechnology. *Applied Microbiology and Biotechnology*, *85*(2), 229-236.

Weiner, A. J., Choo, Q. L., Wang, K. S., Govindarajan, S., Redeker, A. G., Gerin, J. L. & Houghton, M. (1988). A single antigenomic open reading frame of the hepatitis delta virus encodes the epitope (s) of both hepatitis delta antigen polypeptides p24 delta and p27 delta. *Journal of Virology*, *62*(2), 594-599.

Weng, W., Wei, Q., Toden, S., Yoshida, K., Nagasaka, T., Fujiwara, T., Cai, S., Qin, H., Ma, Y. & Goel, A. (2017). Circular RNA ciRS-7—a promising prognostic biomarker and a potential therapeutic target in colorectal cancer. *Clinical Cancer Research*, *23*(14), 3918-3928.

Wesselhoeft, R. A., Kowalski, P. S. & Anderson, D. G. (2018). Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nature Communications*, *9*(1), 1-10.

Wesselhoeft, R. A., Kowalski, P. S., Parker-Hale, F. C., Huang, Y., Bisaria, N. & Anderson, D. G. (2019). RNA circularisation diminishes immunogenicity and can extend translation duration *in vivo*. *Molecular Cell*, *74*(3), 508-520.

Westholm, J. O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S. E., Graveley, B. R. & Lai, E. C. (2014). Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell reports*, *9*(5), 1966-1980.

White, R. J. (2008). RNA polymerases I and III, non-coding RNAs and cancer. *Trends in Genetics*, *24*(12), 622-629.

Wilkinson, M. E., Charenton, C. & Nagai, K. (2020). RNA Splicing by the Spliceosome. *Annual Review of Biochemistry*, *89*, 359-388.

Will, C. L. & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, *3*(7), a003707.

Woerfel, G. & Bindereif, A. (2001). *In vitro* selection of exonic splicing enhancer sequences: identification of novel CD44 enhancers. *Nucleic Acids Research*, *29*(15), 3204-3211.

Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. & Smith, C. W. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Molecular Cell*, *13*(1), 91-100.

Wu, J. Y. & Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, *75*(6), 1061-1070.

Wu, Z., Sun, H., Wang, C., Liu, W., Liu, M., Zhu, Y., Xu, W., Jin, H. & Li, J. (2020). Mitochondrial genome-derived circRNA mc-COX2 functions as an oncogene in chronic lymphocytic leukemia. *Molecular Therapy-Nucleic Acids*, *20*, 801-811.

Xia, P., Wang, S., Ye, B., Du, Y., Li, C., Xiong, Z., Qu, Y. & Fan, Z. (2018). A circular RNA protects dormant hematopoietic stem cells from DNA sensor cGAS-mediated exhaustion. *Immunity*, *48*(4), 688-701.

Xie, W., Zhu, H., Zhao, M., Wang, L., Li, S., Zhao, C., Zhou, Y., Zhu, B., Jiang, X., Liu, W. & Ren, C. (2021). Crucial roles of different RNA-binding hnRNP proteins in Stem Cells. *International Journal of Biological Sciences*, *17*(3), 807.

Xu, Y., Leng, K., Yao, Y., Kang, P., Liao, G., Han, Y. & Cui, Y. (2021). A circular RNA, Cholangiocarcinoma-Associated Circular RNA 1, contributes to cholangiocarcinoma progression, induces angiogenesis, and disrupts vascular endothelial barriers. *Hepatology*, *73*(4), 1419-1435.

Yang, J., Hung, L. H., Licht, T., Kostin, S., Looso, M., Khrameeva, E., Bindereif, A., Schneider, A. & Braun, T. (2014). RBM24 is a major regulator of muscle-specific alternative splicing. *Developmental Cell*, *31*(1), 87-99.

Yang, Y., Gao, X., Zhang, M., Yan, S., Sun, C., Xiao, F. & Zhang, N. (2018). Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis. *JNCI: Journal of the National Cancer Institute*, *110*(3), 304-315.

Yao, Y., Yang, B., Chen, Y., Wang, H., Hu, X., Zhou, Y. & Chen, X. (2019). RNA-binding motif protein 24 (RBM24) is involved in pregenomic RNA packaging by mediating interaction between hepatitis B virus polymerase and the epsilon element. *Journal of Virology*, *93*(6), e02161-18.

Ye, F., Gao, G., Zou, Y., Zheng, S., Zhang, L., Ou, X., Xie, X. & Tang, H. (2019). CircFBXW7 inhibits malignant progression by sponging miR-197-3p and encoding a 185-aa protein in triple-negative breast cancer. *Molecular Therapy-Nucleic Acids*, *18*, 88-98.

Yeo, G. W., Nostrand, E. L. V. & Liang, T. Y. (2007). Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genetics*, *3*(5), e85.

Yin, Y. W., Liu, K. L., Lu, B.S., Li, W., Niu, Y. L., Zhao, C. M., Yang, Z., Guo, P. Y. & Qi, J. C. (2021). RBM24 exacerbates bladder cancer progression by forming a Runx1t1/TCF4/miR-625-5p feedback loop. *Experimental & Molecular Medicine*, *53*(5), 933-946.

Yoneyama, K., Kojima, S., Kodani, Y. Yamaguchi, N., Igarashi, A., Kurose, K., Kawase, R., Takeshita, T., Hattori, S. & Nagata, K. (2015). Proteomic identification of auto-antibodies in sera from patients with ovarian cancer as possible diagnostic bio-markers. *Anticancer Research 35 (2)*, 881–889.

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G. & Chen, W. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nature Neuroscience*, *18*(4), 603-610.

Yu, C. Y., Li, T.C., Wu, Y. Y., Yeh, C.H., Chiang, W., Chuang, C. Y. & Kuo, H. C. (2017). The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. *Nature Communications*, *8*(1), 1-15.

Yu, L., Gong, X., Sun, L., Zhou, Q., Lu, B. & Zhu, L. (2016). The circular RNA Cdr1as act as an oncogene in hepatocellular carcinoma through targeting miR-7 expression. *PloS One*, *11*(7), e0158347.

Yuan, W., Xie, J., Long, C., Erdjument-Bromage, H., Ding, X., Zheng, Y., Tempst, P., Chen, S., Zhu, B. & Reinberg, D. (2009). Heterogeneous nuclear ribonucleoprotein L Is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity *in vivo*. *Journal of Biological Chemistry*, *284*(23), 15701-15707.

Zaphiropoulos, P. G. (1996). Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proceedings of the National Academy of Sciences*, *93*(13), 6536-6541.

Zeng, Y., Du, W. W., Wu, Y., Yang, Z., Awan, F. M., Li, X., Yang, W., Zhang, C., Yang, Q., Yee, A. & Chen, Y. (2017). A circular RNA binds to and activates AKT phosphorylation and nuclear localisation reducing apoptosis and enhancing cardiac repair. *Theranostics*, *7*(16), 3842.

Zhan, Q., Carrier, F. & Fornace Jr, A. J. (1993). Induction of cellular p53 activity by DNA-damaging agents and growth arrest. *Molecular and Cellular Biology*, *13*(7), 4242-4250.

Zhang, D., Ma, Y., Ma, Z., Liu, S., Sun, L., Li, J., Zhao, F., Li, Y., Zhang, J., Li, S. & Jiang, J. (2020). Circular RNA SMARCA5 suppressed non-small cell lung cancer progression by regulating miR-670-5p/RBM24 axis. *Acta Biochimica et Biophysica Sinica*, *52*(10), pp.1071-1080.

Zhang, J., Cho, S. -J., Shu, L., Yan, W., Guerrero, T., Kent, M. S., Skorupski, K., Chen, H. & Chen, X. (2011) Translational repression of p53 by RNPC1, a p53 target overexpressed in lymphomas. *Genes Development*, 25, 1528–1543.

Zhang, L., Hou, C., Chen, C., Guo, Y., Yuan, W., Yin, D., Liu, J. & Sun, Z. (2020). The role of N 6-methyladenosine (m6A) modification in the regulation of circRNAs. *Molecular Cancer*, *19*(1), 1-11.

Zhang, M., Han, Y., Liu, J., Liu, L., Zheng, L., Chen, Y., Xia, R., Yao, D., Cai, X. & Xu, X. (2020). Rbm24 modulates adult skeletal muscle regeneration via regulation of alternative splicing. *Theranostics*, *10*(24), 11159.

Zhang, M., Huang, N., Yang, X., Luo, J., Yan, S., Xiao, F. & Zhang, N. (2018a). A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene*, *37*(13), 1805-1814.

Zhang, M., Zhang, J., Chen, X., Cho, S. J., Chen, X. & Chen, X. (2013). Glycogen synthase kinase 3 promotes p53 mRNA translation via phosphorylation of RNPC1. *Genes Development*, 27, 2246–2258.

Zhang, M., Zhao, K., Xu, X., Yang, Y., Yan, S., Wei, P. & Zhang, N. (2018b). A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nature Communications*, *9*(1), 1-17.

Zhang, X. O., Wang, H. B., Zhang, Y., Lu, X., Chen, L. L. & Yang, L. (2014). Complementary sequence-mediated exon circularisation. *Cell*, *159*(1), 134-147.

Zhang, X. O., Dong, R., Zhang, Y., Zhang, J. L., Luo, Z., Zhang, J., Chen, L. L. & Yang, L. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Research*, *26*(9), 1277-1287.

Zhang, Y., Zhang, X. O., Chen, T., Xiang, J. F., Yin, Q. F., Xing, Y. H. & Chen, L. L. (2013). Circular intronic long noncoding RNAs. *Molecular Cell*, *51*(6), 792-806.

Zhang, Z., Will, C. L., Bertram, K., Dybkov, O., Hartmuth, K., Agafonov, D. E., Hofele, R., Urlaub, H., Kastner, B., Lührmann, R. & Stark, H. (2020). Molecular architecture of the human 17S U2 snRNP. *Nature*, *583*(7815), 310-313.

Zhao, Q., Liu, J., Deng, H., Ma, R., Liao, J. Y., Liang, H., Hu, J., Li, J., Guo, Z., Cai, J. & Xu, X. (2020). Targeting mitochondria-located circRNA SCAR alleviates NASH via reducing mROS output. *Cell*, *183*(1), 76-93.

Zheng, L., Yuan, H., Zhang, M., Wang, C., Cai, X., Liu, J. & Xu, X. Q. (2021). Rbm24 regulates inner-ear-specific alternative splicing and is essential for maintaining auditory and motor coordination. *RNA Biology*, *18*(4), 468-480.

Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., Luo, Y., Lyu, D., Li, Y., Shi, G. & Liang, L. (2016). Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nature Communications*, *7*(1), 1-13.

Zheng, Z. M., Tao, M., Yamanegi, K., Bodaghi, S. & Xiao, W. (2004). Splicing of a cap-proximal human papillomavirus 16 E6E7 intron promotes E7 expression, but can be restrained by distance of the intron from its RNA 5′ cap. *Journal of Molecular Biology*, *337*(5), 1091-1108.

Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C. & Shi, Y. (2014). Crystal structures of the Lsm complex bound to the 3′ end sequence of U6 small nuclear RNA. *Nature*, *506*(7486), 116-120.

Zhou, W. Y., Cai, Z. R., Liu, J., Wang, D. S., Ju, H. Q. & Xu, R. H. (2020). Circular RNA: metabolism, functions and interactions with proteins. *Molecular Cancer*, *19*(1), 1-19.

Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature*, *419*(6903), 182-185.

Zuo, P. & Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes & Development*, *10*(11), 1356-1368.

Abbreviations and symbols

Abbreviations and symbols used in this thesis are listed here. Frequently used gene names are mentioned, for full gene names of other genes see the NCBI Gene Database (http://www.ncbi.nlm.nih.gov/gene/)

~	approximately
×g	times gravity
°C	degree Celsius
µg	microgram
µl	microliter
2’	bonds to the C-2 carbon of ribose in nucleic acids
2’-MOE	2’-O-methoxyethyl
2’Ome	2’-O-methyl
3’	directionality in nucleic acids: in the direction of the C-3 ribose carbon
5’	directionality in nucleic acids: in the direction of the C-5 ribose carbon
aa	amino acid
A	adenosine
A	ampere
ACTB	actin, beta gene
AGO2	argonaute-2 protein
APS	ammonium persulfate
ASO	antisense oligonucleotide
ATP	adenosine triphosphate
bp	base pair(s)
BP-A	branch point adenosine
BSA	bovine serum albumin
C	cytidine
C_T	cycle of threshold
cDNA	complementary DNA
CD45	cluster of differentiation antigen 45
CDR1as	cerebellar degeneration-related protein 1 antisense
CFTR	cystic fibrosis transmembrane conductance regulator
circRNA	circular RNA
CLIP	crosslinking and immune- precipitation
Da	Dalton
dNTP	deoxyribonucleotides
DFHBI	3,5-difluoro-4-hydroxy-benzylidene imidazolinone
DIG	digoxigenin
DMD	Duchenne muscular dystrophy
DMEM	Dulbecco's modified eagle medium
DMPC	dimethyl pyrocarbonate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid

endo	endogenous
eNOS	endothelial nitric oxide synthase
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
ESTs	expressed sequence tags
et al.	*et alia*
ex	exon
ecto	ectopically expressed
FDA	Food and Drug Administration
fwd	forward
g	acceleration due to gravity
g	gram
G	guanosine
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GFP	green fluorescence protein
GMP	guanosine monophosphate
GST	glutathione S-transferase
GTP	guanosine triphosphate
h	hour(s)
HCV	Hepatitis C virus
His	polyhistidine-tag
hnRNP L	heterogenous nuclear ribonucleoprotein L
hnRNP LL	hnRNP L-like
HPLC	high-performance liquid chromatography
IFN-1β	Interferon type I β
intr	intron
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
k	kilo
L	liter
LDL	low-density lipoprotein
M	marker
M	molar
MAPT	microtubule associated protein tau
m6A	N6-methyladenosine
m7G	7-methyl guanosine
min	minute(s)
miRNA	microRNA
ml	milliliter
mRNA	messenger RNA
MS	mass spectrometry
mut	mutant/mutated
MYL6	myosin light polypeptide 6
N-terminal	amino-terminal
ng	nanogram
Ni-NTA	nickel-nitrilotriacetic acid

nM	nanomolar
NMD	nonsense-mediated decay
NP-40	nonyl phenoxypolyethoxylethanol
nt	nucleotide(s)
OH	hydroxyl (-group)
PAGE	polyacrylamide gel electrophoresis
PAPOLA	poly(A) polymerase alpha
PARK7	Parkinson disease protein 7
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PIE	permuted intron-exon
PK	proteinase K
PKR	protein kinase R
PNK	polynnucleotide kinase
Pol II/III	RNA polymerase II/III
pre-mRNA	precursor mRNA
pre-tRNA	precursor tRNA
PS	phosphorothioate
PTC	premature termination codon
qPCR	quantiatative PCR
RBP	RNA-binding protein
rev	reverse
RIP	RNA immunoprecipitation
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
RRM	RNA recognition motif
rRNA	ribosomal RNA
RT	reverse transcription
RT-qPCR	quantiatative reverse transcription-PCR
RtcB	RNA 2’,3’-cyclic phosphate and 5’-OH ligase
SDS	sodium dodecyl sulphate
SELEX	systematic evolution of ligands by exponential enrichment
Seq	sequencing
siRNA	small interfering RNA
SMA	spinal muscular atrophy
SMN	survival of motor-neuron
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SR	serine-arginine-rich
SRSF1	serine/arginine-rich splicing factor 1
SS	splice site
SSO	splice-switching oligonucleotide
TBE	tris-buffered saline EDTA
TJP1	tight junction protein-1
Tornado	twister-optimised RNA for durable overexpression

tricRNA	tRNA intronic circular RNA
Trm4	methyltransferase
tRNA	transfer RNA
U	unit
U	uracil
UV	ultraviolet
U1-U6	small nuclear RNA U1-U6
U2AF	U2-auxillary factor
V	volt
v/v	volume per volume
WT	wild type
w/v	weight per volume
ZKSCAN1	zinc finger with KRAB and SCAN domains 1
α	alpha
β	beta
γ	gamma
Δ	delta, without/lacking

Scientific achievements

Publication

Thankachan, J. M*., **Nuthalapati, S. S***., Tirumala, N. A., & Ananthanarayanan, V. (2017). Fission yeast myosin I facilitates PI (4, 5) P_2-mediated anchoring of cytoplasmic dynein to the cortex. *Proceedings of the National Academy of Sciences*, *114*(13), E2672-E2681.

*co-authors (first)

Poster presentation

International Max-Planck Research School for Heart and Lung Research (IMPRS-HLR) Annual Retreat 2019

9th - 11th October 2019, Hohenroda, Germany

Nuthalapati S. S., Schneider T. and Bindereif A. Human circular RNAs: design, development and functional analysis of protein sponges.

Oral presentations

ITN circRTrain Network meeting

4th - 5th February 2019, Rome, Italy

Nuthalapati S. S. and Bindereif A. Identification and functional characterisation of circRNA-protein complexes (circRNPs) in human cells.

ITN circRTrain Final Meeting (Online)

2nd – 3rd June 2021, Giessen, Germany

Nuthalapati S. S. and Bindereif A. Design and application of circular RNAs for protein sponging and modulation of alternative splicing.

International Max-Planck Research School for Molecular Organ Biology (IMPRS-MOB) Annual Retreat 2021 (Online)

22nd - 24th September 2021, Bad Nauheim, Germany

Nuthalapati S. S., Schneider T. and Bindereif A. Human circular RNAs: design, development and functional analysis of protein sponges.

PhD Graduate program

Successfully completed the structured PhD graduate program from International Max-Planck Research School for Molecular Organ Biology (IMPRS-MOB)

October 2018 – October 2021, Bad Nauheim, Germany.

Acknowledgements

But none of these things move me, neither count I my life dear unto myself, so that **I might finish my course with joy**, and the ministry, which I have received of the Lord Jesus, to testify the gospel of the grace of God.
Acts 20, v24 (Holy Bible)

First and foremost I would like to attribute all praise, glory and honour to GOD Almighty for His Mercy, which is new every morning, Great indeed is God's faithfulness. Without God's amazing Grace my PhD would have been impossible.

I would like to express my utmost gratitude to Prof. Dr. Albrecht Bindereif for this great opportunity to do my research and PhD thesis in his lab. His constant support, encouragement, guidance and critical evaluation of research data have helped shape my scientific career. Thank you for bearing with my mistakes and helping me learn. Thanks also for all scientific discussions, ideas and funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant for my work.
My heartfelt thanks to Prof. Dr. Reinhard Dammann for reviewing my thesis and allowing me to work in his lab as a HiWi. His willingness to help is truly commendable.
A special note of thanks to the Max-Planck Research School (IMPRS, Bad Nauheim) for the structured PhD program and funding; Prof. Dr. Katja Straesser and Dr. Andre Schneider for guidance during TAC meetings.
I would be failing in my duty if I do not recognise the support of my labmates, Anna Didio- my "go-to" person for every need, right from my first days in lab until the completion of my thesis. Thanks to Corinna Ulshoefer & Marie Mossbach for their help, Silke Schreiner for technical help with my experiments; Janina Breuer and Anna Wilmen for being great colleagues. Also, Dr. Oliver Rossbach and Dr. Christian Preußer for valuable and very essential tips for successful experiments. I would like to extend my deepest gratitude to Dr. Tim Schneider and Christina Pfafenrot. Thank you for being like my "Doktorvater und Doktormutter" and supporting me throughout my PhD journey!
I am also grateful to Arokia SusaiRaj for his love and prayers for me. Thanks to my Giessener-friends Ratnal, Francisca and Ajay Adams for feeding me and helping me survive.
I also deeply appreciate the love and support of my German family – "Opa" Alexander und "Oma" Catherine Seibel; Liebe Lisa und ihre Familie und auch die Gemeinde Familie in Laufdorf! Prayer support from brothers and sisters in Christ, is also greatly acknowledged.
Most importantly, I would be ever grateful to my beloved family! Church family and the Berean Teens, who have upheld me with their unceasing prayers. Praise God! for parents and siblings who love me and support me so much, words cannot express my feeling of love and gratitude to my family. THANK YOU one and all!

Now unto the King eternal, immortal, invisible, the only wise God, be honour and glory for ever and ever. Amen.
1 Timothy 1, v17 (Holy Bible)

Eidesstattliche Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Ich stimme einer evtl. Überprüfung meiner Dissertation durch eine Antiplagiat-Software zu. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-LiebigUniversität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

I declare that I have completed this dissertation single-handedly without the unauthorised help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and cited all text passages that are derived verbatim from or are based on the content of published work of others, and all information relating to verbal communications. I consent to the use of an anti-plagiarism software to check the results of my thesis. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University Giessen „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ in carrying out the investigations described in the dissertation.

Gießen, den 22. Dezember 2021

Stephen Sukumar Nuthalapati