



Skolkovo Institute of Science and Technology

LONG-RANGE COMPLEMENTARY INTERACTIONS IN
HUMAN PRE-MRNAS AND THEIR IMPLICATIONS IN SPLICING

Doctoral Thesis

by

MARINA KALININA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Olga Dontsova
Co-supervisors
Assistant Professor Dmitri Pervouchine
Dr. Dmitry Skvortsov

Moscow - 2021

© Marina Kalinina 2021

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Marina Kalinina)

Supervisor (Prof. Olga Dontsova)

Co-supervisor (Assistant Prof. Dmitri Pervouchine)

Co-supervisor (Dr. Dmitry Skvortsov)

Abstract

RNA structure plays an essential role in the maturation of eukaryotic transcripts. While most current studies are focused on locally-occurring RNA structures, long-range base pairings have been increasingly reported as being implicated in the regulation of pre-mRNA splicing. This study aims at the validation of several targets from a recently published catalog of conserved long-range RNA structures. We studied the impact of long-range complementary interactions on splicing in three human genes — *Phf20l1*, *Cask*, and *Ate1*. In *Phf20l1* and *Cask*, in which alternative exons are located in between the complementary regions, we confirmed the looping-out mechanism of splicing regulation using compensatory mutations in minigenes. Additionally, we were able to change the alternative splicing outcome of the endogenous transcripts using steric blocking LNA-based antisense oligonucleotides. In *Ate1*, we demonstrated that two RNA structure modules coexist, one responsible for mutually exclusive exon choice (MXE) and the other regulating the ratio of transcript isoforms. We showed that *Ate1* contains five conserved regulatory intronic elements R1–R5, of which R1 and R4 compete for base pairing with R3, while R2 and R5 form an ultra-long-range RNA structure spanning 30 Kb. In minigenes, single and double mutations that disrupt base pairings in R1R3 and R3R4 lead to the loss of MXE splicing, while compensatory triple mutations that restore RNA structure revert splicing to that of the wild type. In the endogenous *Ate1* pre-mRNA, blocking the competing base pairings by LNA-based antisense oligonucleotides complementary to R3 leads to the loss of MXE splicing, while the disruption of R2R5 interaction changes the ratio of MXE. That is, *Ate1* splicing is controlled by two independent, dynamically interacting, and functionally distinct RNA structure modules. The MXE ratio in *Ate1* changes in response to RNA polymerase II slowdown, however it fails to do so when the ultra-long-range R2R5 interaction is disrupted, indicating that exon ratio depends on co-transcriptional RNA folding. To check whether a similar response also takes places in other genes, we performed an RNA-seq experiment using RNA polymerase II slowdown with α -amanitin. We found that introns

with predicted long-range RNA structures respond to RNA polymerase II slowdown more than introns without such structures do, indicating that co-transcriptional RNA folding affects pre-mRNA splicing in a transcriptome-wide manner. In sum, our results demonstrate that splicing is coordinated both in time and in space over very long distances and that the interaction of these components is mediated by long-range RNA structure.

Publications

1. M. Kalinina, D. Skvortsov, S. Kalmykova, T. Ivanov, O. Dontsova, and D. D. Pervouchine. Multiple competing RNA structures dynamically control alternative splicing in the human ATE1 gene. *Nucleic Acids Res*, 49(1):479–490, 01 2021
2. S. Kalmykova, M. Kalinina, S. Denisov, A. Mironov, D. Skvortsov, R. Guigó, and D. Pervouchine. Conserved long-range base pairings are associated with pre-mRNA processing of human genes. *Nat Commun*, 12(1):2300, 04 2021

Conferences

1. Marina Kalinina, Dmitry Skvortsov, Svetlana Kalmykova, Olga Dontsova, and Dmitri D. Pervouchine “Multiple competing RNA structures dynamically control alternative splicing in human ATE1 gene”, Splicing 2020 (3rd International Caparica Conference in Splicing), online. Short talk
2. Marina Kalinina, Olga Babadei, Dmitry Skvortsov, Svetlana Kalmykova, Olga Dontsova, and Dmitri D. Pervouchine “Long-range RNA-RNA pairings in human pre-mRNAs and their implications in alternative splicing”, RNA 2021, online. Poster presentation
3. Marina Kalinina, Olga Babadei, Dmitry Skvortsov, Svetlana Kalmykova, Olga Dontsova, and Dmitri Pervouchine "Long-range RNA-RNA pairings in human pre-mRNAs and their implications in alternative splicing", MCCMB 2021, Moscow. Poster presentation.

Acknowledgments

I want to express my gratitude to my supervisor, Prof. Olga Dontsova, for the opportunity to work in her laboratory, her support, and valuable advice. I have been working in her lab for ten years, and all this time, she has been promoting my scientific career and helping in challenging situations.

I would like to express my great appreciation to my co-supervisor, Dr. Dmitri Pervouchine, for his immense contribution to the development of this project. I'll be forever grateful for the time he devoted to helping me to be a better scientist.

I am immensely grateful to my co-supervisor, Dr. Dmitry Skvortsov, for mentoring me during the last ten years. He taught me all the experimental techniques I know. I would like to thank him for his help with troubleshooting experiments, fruitful discussions, and his patience and understanding.

I would like to thank all the thesis defense committee members Prof. Éric Westhof, Prof. Alain Laederach, Prof. Yuri Kotelevtsev, Dr. Timofei Zatsepin, and Prof. Konstantin Lukyanov for their time and critical reviews of this thesis.

I would like to thank all members of Prof. Dontsova's lab. It was a valuable experience working in such a big and diverse laboratory. Many thanks to my former student Olga Babadei, whose master thesis project I co-supervised. She analyzed many targets during the screening of long-range RNA structures. I also want to thank my colleagues and friends Ivan Laptev, Alexander Malyavko, and Radik Shafikov for their invaluable support during my work in the lab and meaningful discussions. Special thanks to Radik Shafikov for proofreading this thesis.

I would also like to thank all the present and former members of Dr. Dmitri Pervouchine's lab. Without their bioinformatic help, this project would not have progressed. I would like to extend special thanks to Svetlana Kalmykova for providing bioinformatics support throughout this project and for analyzing RNA-seq data in experiments with RNAPII slowdown. I also want to thank Timofey Ivanov for his help with the Ate1 mutually exclusive exons and Alexey Mironov for his help and discussions about Phf2011 and Cask alternative splicing regulation.

I would like to thank my parents for giving me this amazing opportunity to study all this time, and for their support and love.

I would like to acknowledge all teachers I have ever met. Your role in my life is invaluable.

And finally, I am very thankful to my Skoltech classmates and friends Ksenia Safina, Alexandra Bezmenova, Anastasia Stolyarova, and my close friends Dmitrii Travin and Anna Olina. Thank you all for believing in me and supporting me all the time. You are the best.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 14 |
| 1.1 | Thesis Structure | 16 |
| 2 | Background | 17 |
| 2.1 | RNA structure and its role | 17 |
| 2.1.1 | Basic examples of functional RNA structures | 19 |
| 2.2 | Splicing | 21 |
| 2.2.1 | Alternative splicing | 23 |
| 2.3 | Mechanisms of AS regulation through RNA structures | 26 |
| 2.3.1 | Local RNA structures | 26 |
| 2.3.2 | Long-range complementary interactions | 30 |
| 2.4 | Mutually exclusive alternative splicing | 33 |
| 2.5 | RNA structure analysis using physical and chemical methods | 38 |
| 2.6 | Computational methods of RNA structure prediction | 39 |
| 2.6.1 | Computationally predicted long-range RNA structures | 40 |
| 3 | Thesis Objectives | 42 |
| 4 | Materials and methods | 43 |
| 4.1 | Minigene constructs and mutagenesis | 43 |
| 4.2 | Antisense oligonucleotides | 44 |
| 4.3 | Cell culture | 44 |
| 4.4 | Cell culture transfections | 44 |
| 4.5 | Cell culture treatments | 45 |
| 4.6 | RT-PCR experiments | 46 |
| 4.7 | Western blotting | 47 |
| 4.8 | Library preparation and RNA-seq experiment | 47 |
| 4.9 | Statistical analysis | 48 |
| 5 | Long-range RNA structures | 49 |
| 5.1 | Screening of targets | 49 |
| 5.2 | Phf2011 | 53 |
| 5.2.1 | Biological function | 53 |
| 5.2.2 | RNA secondary structure controls exon 6 inclusion | 54 |
| 5.3 | Cask | 58 |
| 5.3.1 | Biological function | 58 |
| 5.3.2 | RNA secondary structure controls exon 19 inclusion | 58 |

| | | |
|----------|--|------------|
| 5.4 | <i>Ate1</i> | 61 |
| 5.4.1 | Biological function | 61 |
| 5.4.2 | <i>Ate1</i> alternative isoforms | 61 |
| 5.4.3 | Conserved complementary regions in exon 7 cluster | 63 |
| 5.4.4 | Competition between R1R3 and R3R4 controls mutually exclusive splicing | 65 |
| 5.4.5 | The ultra-long-range R2R5 base pairing controls isoform bias | 69 |
| 5.4.6 | Crosstalk between competing and ultra-long-range RNA structures | 74 |
| 6 | AS regulation under RNAPII slowdown | 77 |
| 6.1 | <i>Ate1</i> splicing regulation | 78 |
| 6.2 | Genome-wide analysis | 83 |
| 7 | Discussion | 85 |
| 7.1 | Screening of the predicted RNA structures | 85 |
| 7.2 | Two structural modules in <i>Ate1</i> | 87 |
| 7.3 | RNAPII slowdown and long-range RNA structures | 89 |
| 8 | Conclusion | 92 |
| | Bibliography | 93 |
| A | Supplementary information | 114 |

List of Figures

| | | |
|------|--|-----|
| 2-1 | The elements of RNA secondary structure | 18 |
| 2-2 | Scheme of pre-mRNA splicing | 22 |
| 2-3 | Main types of alternative splicing events | 24 |
| 2-4 | Positive and negative control of pre-mRNA splicing | 25 |
| 2-5 | <i>Dscam1</i> exon 6 cluster | 35 |
| 5-1 | Scheme of <i>Phf20l1</i> exon 6 region | 55 |
| 5-2 | Mutagenesis of the <i>Phf20l1</i> minigene | 56 |
| 5-3 | Treatment of cells with AONs complementary to <i>Phf20l1</i> regulatory elements | 57 |
| 5-4 | Scheme of <i>Cask</i> cassette exon 19 region. | 59 |
| 5-5 | Mutagenesis of the <i>Cask</i> minigene | 60 |
| 5-6 | Treatment of cells with AONs complementary to <i>Cask</i> regulatory elements | 62 |
| 5-7 | Scheme of <i>Ate1</i> mutually exclusive cluster 7 | 64 |
| 5-8 | Complementary interactions in <i>Ate1</i> pre-mRNA | 64 |
| 5-9 | The scheme of the <i>Ate1</i> minigene | 65 |
| 5-10 | R1R3 and R3R4 mutagenesis | 67 |
| 5-11 | Pairwise comparisons of mutants disrupting the same helix. | 68 |
| 5-12 | The predicted hybridization energies of R1R3 and R3R4 in <i>Ate1</i> mutants | 69 |
| 5-13 | R1-R5 RNA editing | 70 |
| 5-14 | AON1 against R3 sequence | 71 |
| 5-15 | R2R5 mutagenesis | 72 |
| 5-16 | AON2-1 and AON2-2 disrupting R2 and R5 helix | 73 |
| 5-17 | Crosstalk between structures in the <i>Ate1</i> | 75 |
| 6-1 | Relative expression level in samples with RNAPII elongation slowdown | 78 |
| 6-2 | Effects of RNAPII elongation inhibitor and NELFE overexpression on AS in several genes | 79 |
| 6-3 | Effects of α -amanitin treatment on <i>Ate1</i> alternative splicing | 80 |
| 6-4 | <i>Ate1</i> alternative splicing changes under RNAPII slowdown | 82 |
| 6-5 | The change of inclusion rate of exons following introns with PCCRs in response to RNAPII slowdown. | 84 |
| A-1 | RIC-seq signal for <i>Phf20l1</i> | 119 |
| A-2 | RIC-seq signal for <i>Ate1</i> | 120 |

List of Tables

| | | |
|-----|--|-----|
| 5.1 | PCCRs selected for AONs and/or minigene testing | 50 |
| A.1 | Primers for cloning <i>Ate1</i> , <i>Phf20l1</i> , and <i>Cask</i> minigenes | 114 |
| A.2 | Primers for mutagenesis | 115 |
| A.3 | Sequences of AONs | 116 |
| A.4 | Primers for RT-PCR analysis | 117 |
| A.5 | qRT-PCR primers | 118 |

List of symbols, abbreviations

AON – antisense oligonucleotide
AS – alternative splicing
BPS - branch point sequence
CDS – coding sequence
Cryo-EM – cryogenic electron microscopy
DNA - deoxyribonucleic acid
ESE – exonic splicing enhancer
ESS – exonic splicing silencer
GQ - G-quadruplex
hnRNP – heterogeneous nuclear ribonucleoprotein
icSHAPE - in vivo click selective 2-hydroxyl acylation and profiling experiment
IRE – iron-responsive element
ISE – intronic splicing enhancer
ISS – intronic splicing silencer
LNA - locked nucleic acid
lncRNA – long non-coding RNA
MFE – minimum free energy
miRNA – microRNA
mRNA – messenger RNA
MXE – mutually exclusive exons
NMD – nonsense-mediated mRNA decay
NMR - nuclear magnetic resonance

PCR – polymerase chain reaction
PCCR - pair of conserved complementary regions
PPT - polypyrimidine tract
qRT-PCR – quantitative reverse-transcription polymerase chain reaction
RBP – RNA binding protein
RIC-seq – RNA in situ conformation sequencing
RNA - ribonucleic acid
RNAPII – RNA polymerase II
rRNA – ribosomal RNA
RT– room temperature
snRNA – small nuclear RNA
snRNP – small nuclear ribonucleoprotein
SR protein – serine-arginine-rich protein
SRE - splicing regulatory element
TPP - thiamine pyrophosphate
tRNA – transport RNA
UTR – untranslated region
WC – Watson-Crick pairing
WT – wild-type

Chapter 1

Introduction

Ribonucleic acid (RNA) molecules play a key role in the functioning of all living cells on Earth. The monomeric units of RNAs are ribonucleotides, which are organized hierarchically in several structural levels. The primary RNA structure is the sequence of its nucleotides. The secondary structure, which determines pairwise interactions of nucleotides, is organized in characteristic elements such as double-stranded regions (helices), bulges, and loops. The tertiary structure determines the interaction of the elements of the secondary structure in the three-dimensional space, e.g., the interaction between loops. The structural organization of RNA is essential for the maturation and functioning of many non-coding RNA classes including ribosomal RNAs, transfer RNAs, small nuclear RNAs, pre-microRNAs, as well as for the processing and regulation of many messenger RNAs (mRNAs).

Experimental approaches to study the RNA structure include physical methods such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, chemical structure probing methods, which are based on different reactivity of structured and unstructured regions to various probing agents, and biological methods, which examine the function of the RNA structure by disruptive and restoring compensatory mutations. Lately, a number of high-throughput genome-wide RNA structure probing methods have become available to investigate the base pairs involved in inter- and intramolecular structure (for example, icSHAPE, PARIS, and RIC-seq

[3]). Computational methods of RNA structure prediction focus on RNA secondary structure and include thermodynamic models as well as methods based on comparative genomics. Most of the methods described above are accurate for local RNA secondary structure, i.e., interactions of nucleotides on short distances within the primary sequence, but not for long-range interactions.

A recent study presented a list of predicted pairs of conserved complementary regions in introns of human protein-coding genes [2]. These regions tend to occur within introns, suppress intervening exons, and obstruct cryptic and inactive splice sites. Their double-stranded structure is supported by decreased icSHAPE nucleotide accessibility, high abundance of RNA editing sites, and frequent occurrence of forked probing peaks. In this thesis, we experimentally confirm some long-range RNA structures from this list and propose a new potential mechanism of alternative splicing (AS) regulation through long-range RNA structures and RNA polymerase II (RNAPII) elongation speed. The experimentally confirmed structures include the first example of competing RNA structures in the human transcript (*Ate1*) and the longest RNA structure known to date, in which the interacting base pairs span over 30 kb. To investigate the role of long-range RNA structure in co-transcriptional splicing, we performed RNA-seq experiments, in which we used α -amanitin to slow down the RNAPII elongation speed, and additionally analyzed publicly available RNA-seq data obtained for slow RNAPII mutants. We showed that structured introns exhibit a distinct splicing pattern in response to RNAPII slowdown suggesting that splicing is widely affected by co-transcriptional RNA folding. To screen long-range RNA structures in human transcripts, we widely used a protocol with steric blocking by antisense oligonucleotides (AONs). However, AONs' treatments led to the predicted AS outcomes not for all selected targets, indicating that the problem of long-distance RNA structure prediction is still challenging, and a more comprehensive approach is required. Perhaps, the use of genome-wide experimental data on long-range RNA interactions such as RIC-seq data may help to make such predictions more accurate.

1.1 Thesis Structure

Chapter 2 contains the literature review of the RNA structures, splicing, and the role of RNA structures in splicing regulation.

Chapter 3 states the thesis objectives.

Chapter 4 describes all methods used in this thesis.

Chapter 5 lists the criteria used for targets selection, presents two examples of RNA structure-mediated exon skipping in *Phf20l1* and *Cask* genes, and describes a complex system of competing RNA structures and ultra-long-range RNA structure in the *Ate1* gene.

Chapter 6 is dedicated to the investigation of the role of RNAPII elongation slowdown on alternative splicing outcome, firstly for the *Ate1* transcript, and then transcriptome-wide using RNA-seq experiments.

Chapter 7 discusses the results on competing and long-range RNA structures in the *Ate1* transcript and the genome-wide association of predicted long-range RNA structures in introns with the RNAPII slowdown.

Chapter 8 lists main conclusions of the Thesis.

Chapter 2

Background

2.1 RNA structure and its role

Ribonucleic acid (RNA) is a biopolymer that consists of ribonucleotide monomers. Each ribonucleotide consists of a phosphate, a ribose residue, and one of the four different bases: adenine (A), cytosine (C), uracil (U), and guanine (G). Ribose residues are joined together by phosphodiester bonds between the third and the fifth carbon atoms of the adjacent sugar rings. This asymmetry specifies the direction of the RNA strand. The sequence of ribonucleotides, given from 5' end to 3' end, is called the primary structure of the RNA.

In contrast to DNA (deoxyribonucleic acid), which normally forms a double helix, RNA is present in living cells in both double-stranded and single-stranded forms. The most frequent and thermodynamically stable are the canonical Watson-Crick base pairings (A and U, C and G). Additionally, wobble base pairings (G and U) and other non-canonical base pairings may be formed. According to the Leontis and Westhof classification [4], each nucleotide has three edges for interaction: Watson-Crick (WC), Hoogsteen (H), and Sugar (S) edges. Interactions between edges might be in *cis* or in *trans* depending on the backbone orientation. There are twelve possible different base pairings between any two RNA bases [4].

RNA secondary structure, i.e., the set of base-paired residues, may be viewed as consisting of several elementary structural units. When base pairs stack directly on top of each other, they may form an RNA helix. The stacking of base pairs is stabilized by the overlap of the π -orbitals of the nucleotides' aromatic ring systems, which gives the major stabilizing contribution to the structure. Besides helices, the RNA structure contains unpaired regions usually referred to as loops. Depending on the structural context, they may be classified as hairpin loops, bulges, internal loops, and multiloops (Figure 2-1). The RNA molecule tends to maximize the stabilizing effect from stacking energies and minimize the destabilizing effect from loop regions. It is believed that the actual RNA structure corresponds to the the minimum free energy (MFE) of the molecule, however many RNAs also adapt several biologically important suboptimal structures.

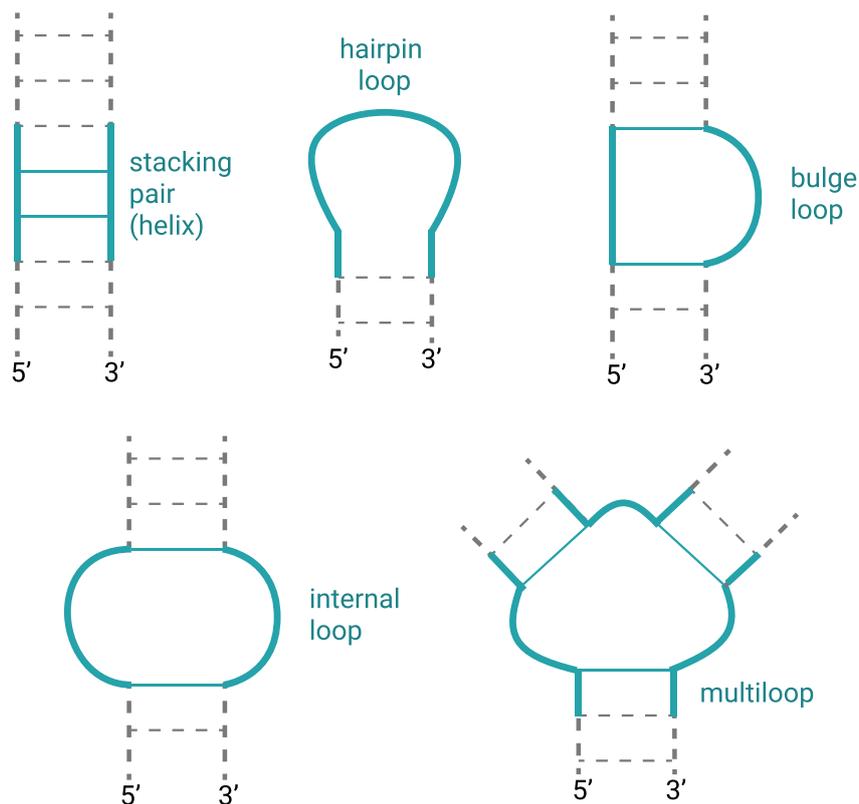


Figure 2-1: The elements of RNA secondary structure.

The tertiary structure of RNA is formed due to three-dimensional interactions of its secondary structure elements. These interactions occur mainly between dis-

tant structural elements. Tertiary structural motifs include G-bulges, pseudoknots, loop-loop interactions (or kissing loops), and many others [5]. Moreover, mono- and divalent metal cations can interact with negatively charged phosphate groups and contribute to RNA folding. All tertiary interactions are difficult to predict computationally. They may be confirmed experimentally using physical methods, e.g., X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Nevertheless, tertiary motifs play a crucial role in many biological processes, many of which were first discovered in viral genomes such as turnip yellow mosaic virus [6] and HIV-1 [7] and later identified in other RNA classes.

2.1.1 Basic examples of functional RNA structures

Prokaryotic mRNAs use RNA structures for transcriptional and post-transcriptional regulation. For instance, there are two mechanisms of transcription termination in prokaryotes: the rho-independent (or intrinsic) and rho-dependent mechanisms. The former doesn't utilize auxiliary proteins and depends solely on pre-mRNA structure [8]. Rho-independent terminators consist of a GC-rich sequence followed by a U-rich tract in the RNA. The termination is achieved by RNA polymerase dissociation due to the formation of a GC-rich stem that is facilitated by decreased binding of the U-rich tract to DNA [9]. Intrinsic terminators were found in 80% of *E. coli* transcripts, including *rpoC* and *crp* genes [10, 11, 12]. Rho-dependent termination is mediated by an RNA helicase Rho. Rho-binding sites (called *rut*) are C-rich sequences with little or no secondary structure [13]. Rho-dependent terminators are usually located at the end of transcription units, for example, in *E. coli* *tyrT* locus [14] or *trpEDCBA* operon [15].

Besides transcription termination, mRNA secondary structure may regulate translation initiation in prokaryotes. Thermodynamically stable mRNA structures may overlap with the initiation site thus preventing the recruitment of the 30S ribosomal subunit [16].

Riboswitches are (mostly prokaryotic) RNA structures that regulate gene expression by occluding or exposing the transcription or translational signals in response to the presence of a ligand. Riboswitches are mainly found in the 5'-leader sequence of the mRNAs and typically consist of two domains: an aptamer that binds a ligand, and a regulatory domain that undergoes a conformational change due to ligand binding [17]. The ligand binding induces a structural conformation, in which a rho-independent terminator stem is formed to abort transcription or to prevent the occlusion of the translation initiation site. Riboswitches may bind metal ions, amino acids, nucleotide precursors, or enzyme cofactors as ligands, thus allowing the cells to react rapidly to changes in metabolite concentrations without protein synthesis. Eukaryotic pre-mRNAs also utilize riboswitches, but they mainly regulate the accessibility of splice sites (see below).

The mature eukaryotic mRNAs may employ RNA secondary structure to regulate translation. In general, 5'UTRs in eukaryotic mRNAs are important for translation initiation due to scanning mechanism and thermodynamically stable secondary structures in the 5'UTRs (including Kozak sequence and start codon itself) prevent efficient interaction with pre-initiation complex and thus can inhibit translation [18]. Secondary structures within a 3'UTR, on the other hand, correlate with high protein expression levels [19]. For example, structured iron-responsive elements (IREs), which were reported in both 5'UTR and 3'UTR of twelve human mRNAs, interact with iron regulatory proteins that stabilize the mRNAs and affect translation in iron starvation conditions [20]. Another example is the RNA switch in the 3'UTR of the VEGF mRNA, which undergoes a conformational change in response to the binding of a regulatory protein affecting *VEGF* translation [21]. For polyadenylated RNAs, the distance between poly(A)-signal and cleavage site is important for efficient polyadenylation. Transcripts with a longer distance between these sites tend to form structured elements to bring them closer [22]. Structural motifs within the coding sequence (CDS) may also cause frame-shifting events [23]. However, the majority of functional eukaryotic RNA structures known to date regulate pre-mRNA

processing and, in particular, pre-mRNA splicing. They are discussed in detail in section 2.3.

2.2 Splicing

Eukaryotic genes are organized in the genome in a discontinuous manner and contain exons and introns. Exons are defined by the sequences that are present in mRNA, while introns could be found only in pre-mRNA. Additionally, the 5' untranslated region (5' UTR) is located before the first exon, and the 3' untranslated region (3' UTR) is located after the last exon. They both are needed for the regulation of proper mRNA functioning. Human exons are usually shorter than 200 nucleotides, while introns can reach 1 Mb [24, 25]. During a process called pre-mRNA splicing, introns are removed and exons are joined together. On average, the length of a pre-mRNA shortens 10-fold after splicing [26].

It is important for the splicing process to distinguish between exons and introns, as well as to identify exon-intron borders. The latter are defined by *cis*-regulatory sequences called the donor (5') and the acceptor (3') splice sites. The 5' splice site (5'ss) identifies the beginning of an intron, while the 3' splice site (3'ss) corresponds to its end. The consensus nucleotide sequences of the 5'ss and 3'ss are usually described as GU and AG, respectively, although important deviations from these standard dinucleotides exist. Another splicing *cis*-regulatory element is the branch point sequence (BPS), which contains an adenosine residue and is located 15-50 nucleotides upstream from the 3'ss. This adenosine initiates a nucleophilic attack during the first transesterification reaction of splicing. The polypyrimidine tract (PPT) is a pyrimidines-rich sequence that is located between the BPS and 3'ss. The 5'ss, 3'ss, the branch point, and the polypyrimidine tract are the major splicing *cis*-regulatory signals (Figure 2-2A) [27].

For the chemical standpoint, the splicing reaction consists of two transesterification steps. First, the 2'-hydroxyl group of the adenosine in BPS makes a nucleophilic

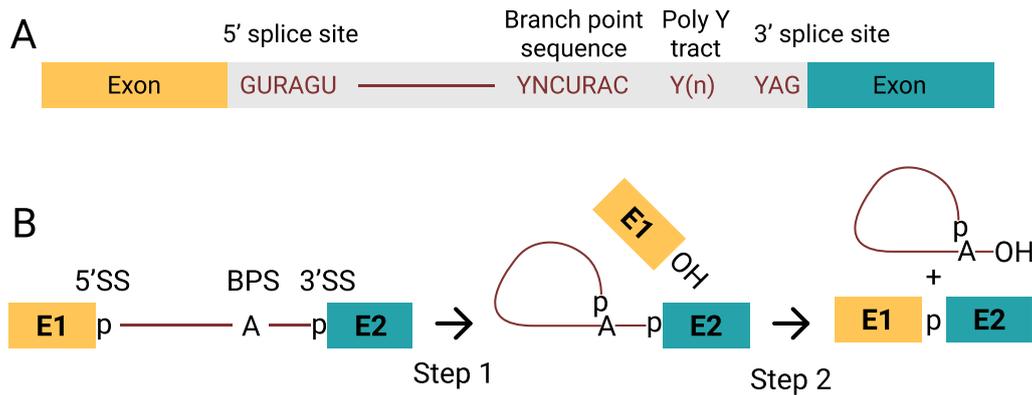


Figure 2-2: A. Conserved sequences in U2-type pre-mRNA introns in metazoans. "Y" is for pyrimidine, "R" - for purine. Y(n) indicates the polypyrimidine tract. B. Two-step mechanism of pre-mRNA splicing. "A" indicates the adenosine in the branch point sequence, "p" is for phosphate group. The figure is adapted from [27].

attack to the phosphate group of the 5'ss. This reaction cleaves the phosphodiester bond in the exon/intron junction, releases exon 1, and forms a new phosphodiester bond between BPS and the 5' end of the intron. The intron remains bound to exon 2 and forms the so-called lariat structure. Next, the 3' hydroxyl group of exon 1 makes a nucleophilic attack to the phosphate group of the 3'ss. As a result, the exons are ligated by a new phosphodiester bond, and the intron is released in the form of a lariat. The two steps of pre-mRNA splicing are shown in [Figure 2-2B](#).

The spliceosome, a macromolecular complex that catalyzes pre-mRNA splicing, is composed of five small nuclear ribonucleoproteins (snRNPs) – U1, U2, U4, U5, and U6 and several dozens of auxiliary factors [27]. These snRNPs contain small nuclear RNAs (snRNAs) that can base pairs with the pre-mRNA. The process of splicing starts with U1 snRNP recognition of 5'ss, in which the U1 snRNA base pairs with the nucleotides at the exon/intron border. Next, or simultaneously with the previous step, splicing factor 1 (SF1) recognizes the branch point adenosine residue and activates it through a bulge formation. The PPT and 3'ss are recognized by the U2 snRNP auxiliary factor (U2AF). Then, U2 snRNP displaces SF1 and U2AF and binds to the BPS. Next, with the recruitment of pre-assembled U4/U6 and U5 snRNPs, the first transesterification reaction takes place. Then the spliceosome complex undergoes remodeling, in which U1 and U4 snRNPs are removed. The

second transesterification reaction occurs, and the intron is released with U2, U5, and U6 snRNPs bound to it. After intron removal, all snRNPs and proteins are recycled to catalyze other reactions.

Besides the major spliceosome, which recognizes the GT/AG consensus sequences, there is the U12-dependent, or the minor spliceosome. It consists of the specific U11, U12, U4atac, U6atac snRNPs, and the common U5 snRNP. Minor spliceosome catalyzes splicing of less than 1% of introns, which contain non-consensus /AT..AC/ splice sites, distinct BPS and PPT sequences [28]. However, minor splice site sequences are evolutionarily highly conserved [29], and deficiencies in its activity may lead to developmental defects or lethality [30, 31].

2.2.1 Alternative splicing

While introns are removed from the pre-mRNA, not all exons always remain included in it. More than 95% of multi-exon human genes are alternatively spliced [32] and provide on average three or more different mRNA isoforms per gene [33]. Alternative splicing produces an additional layer of gene regulation and greatly expands the variety of proteins produced from the same genome [34]. Different mRNA isoforms may result in proteins with different properties because of the presence of functional domains or unstructured polypeptide regions that are important for protein catalytic properties, interaction capabilities, and localization [35, 36]. In some cases, alternative splicing affects the stability of transcript rather than protein properties. Alternative splicing may be coupled with nonsense-mediated decay (NMD), different RNA-binding proteins might be bound to different transcript isoforms, thus controlling its stability and translation efficiency [37]. Alternative splicing serves the needs of cells in different tissues and developmental stages, maintaining a specific balance of mRNA and protein isoforms. [38, 39].

According to the current classification, there are several major types of alternative splicing events, including cassette exons, alternative 5'SS, alternative 3'SS, intron retention, and mutually exclusive exons (Figure 2-3), and also a variety of

minor subtypes [40]. In comparison to constitutive exons, alternative exons often have weaker splice site sequences, i.e., they have lower affinity to the spliceosome due to deviations from the splice site consensus sequences [41].

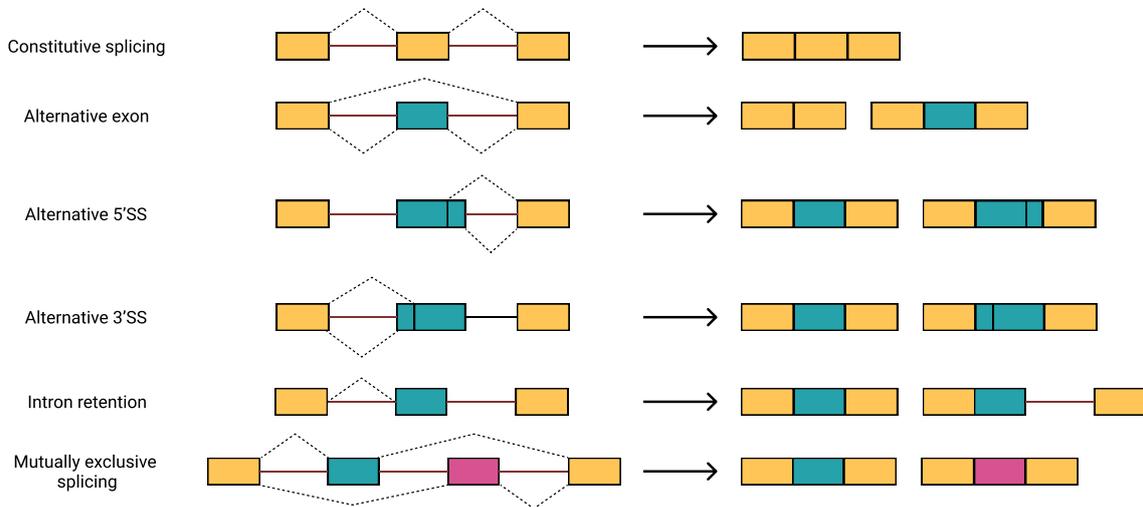


Figure 2-3: Main types of alternative splicing events. Figure is adapted from [40].

Other *cis*-acting elements within pre-mRNA are represented by splicing regulatory elements (SRE), and they have an affinity for *trans*-acting factors. Depending on their positioning and way of action, SREs are divided into intronic splicing silencers (ISS), intronic splicing enhancers (ISE), exonic splicing silencers (ESS), and exonic splicing enhancers (ESE). The *trans*-acting RNA-binding proteins (also called splicing factors) may interact with these sequences and either enhance or repress the binding of spliceosome components (Figure 2-4)[33]. SREs that are located close to splice sites tend to have the strongest effect on splicing [42].

The two main classes of splicing factors are heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine-arginine-rich proteins (SR). Usually, SR proteins bind to enhancer sequences and promote the inclusion of alternative exons, while hnRNPs tend to bind to silencer sequences and promote the skipping of exons (Figure 2-4 [33, 43]). However, both SR proteins and hnRNPs may play antagonistic roles depending on the context of the binding site [41]. Mainly, splicing factors influence splice site recognition either by facilitating or inhibiting the binding of U1 and U2 snRNP to splice sites. Apart from these two families, there are many other RNA-

binding proteins influencing splicing, and some of them are tissue-specific splicing factors [32]. Moreover, the abundance of the core spliceosomal proteins also affects alternative splicing and may be associated with disease states [44].

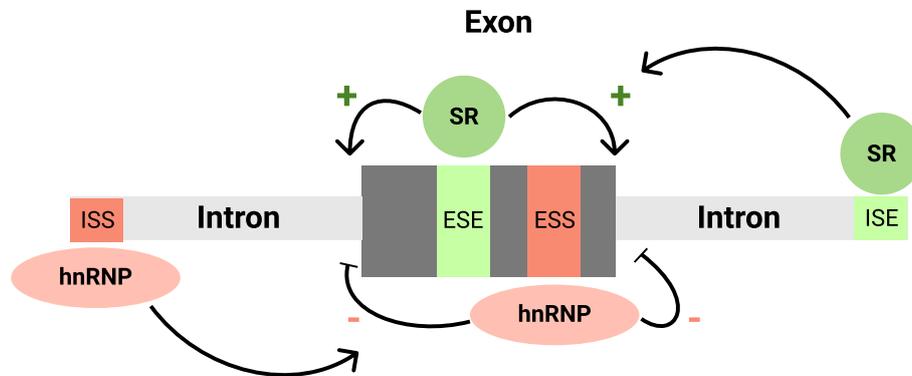


Figure 2-4: Positive and negative regulation of pre-mRNA splicing by *cis*-acting and *trans*-acting elements. SR-proteins are generally promote splicing by interacting with splicing enhancers, while hnRNP proteins generally inhibit splicing by interacting with splicing silencers. Figure is adapted from [33].

The interplay between pre-mRNA splicing and transcription has long been a focus of intense research [43]. Firstly, transcription factors may influence alternative splicing outcome in both direct and indirect ways [45], which demonstrates the ability of DNA-binding proteins to bind RNA. Secondly, histone modifications in a position-dependent manner may play a role in splicing outcome [46]. There are two possible models linking chromatin modifications to pre-mRNA splicing. First, the interaction may be mediated by slowing down the RNA polymerase II (RNAPII) and thus creating the “window of opportunity” for splicing factors to bind pre-mRNA (the so-called kinetic model); or chromatin modifications may recruit splicing factors to the pre-mRNA (so called recruitment model) [47]. Nucleosome positioning can shape the splicing outcome due to specific marking of intron/exon junctions [48, 49], and chromatin remodeling factor PARP1 may modify activity of hnRNPs and SR-proteins through their polyADP-ribosylation [50].

The modifications in the pre-mRNA may affect splicing as well. The most studied RNA modification, m⁶A methylation is recognized by m⁶A readers, which may interact either with either splicing factors [51] or with the phosphorylated C-terminal

domain of RNAPII [52], thereby affecting splicing outcome. Besides splicing, m⁶A modifications may affect transcript stability [53] and facilitate translation initiation [54]. Another RNA modification that may affect splicing is adenosine-to-inosine (A-to-I) editing by ADAR proteins. This modification may alter many downstream interactions of the RNA since inosine is mostly recognized as guanosine. The RNA editing sites near 3'ss may alter the transcription of the adjacent exons [55], and ADAR2 regulates its own pre-mRNA splicing via converting the intronic AA to AI (mimics conserved AG) 3'ss [56].

2.3 Mechanisms of AS regulation through RNA structures

The secondary structure of a pre-mRNA molecule, the presence or absence of double-stranded regions in it, may strongly affect the splicing outcome. RNA secondary structure may expose or mask splicing *cis*-regulatory elements or serve as a substrate for modifications or interaction with other factors. In this Thesis, we will distinguish between short-range and long-range RNA structure based in the distance between interacting bases within primary RNA structure, i.e., within its nucleotide sequence. While there is no natural threshold that separates short-range and long-range RNA structures, we will broadly ascribe base pairing interactions to long-range scale if they cover from hundreds to several thousands nucleotides.

2.3.1 Local RNA structures

Hairpins

One of the main types of RNA secondary structure influencing splicing is hairpin. Hairpin consists of a double-stranded stem and a single-stranded loop, usually not more than dozens of nucleotides, and the position of the splicing regulatory element inside the hairpin defines its function. Mainly, SREs that are involved in double-

stranded structures are non-functional since *trans*-acting splicing factors are unable to bind stem-forming sequences. Some examples of splicing regulation by hairpins are reviewed below.

The first example is the fibronectin gene. One of the alternatively spliced exons, the EDA exon, is highly structured [57]. The proper RNA conformation displays the ESE sequence in the loop of hairpin V, which enables the regulation of this exon through the interaction of ESE with SR-proteins [58]. ESS sequence, which was previously mapped in this exon, determines changes in the RNA secondary structure. A loop-to-stem change in ESE decreases its regulatory ability indicating that ESE exposure in a loop is critical for its function.

A hairpin structure in intron 4 of the cardiac troponin T (*cTnT*) gene is recognized by MBNL1 protein [59]. At the same time, the binding site of U2AF65, which is important for the recognition of the 3'-end of the intron and U2 recruiting, is located in the hairpin loop [60]. The inhibitory effect of MBNL1 results from the competitive binding of U2AF645 and MBNL1, and the hairpin structure plays an important role in these interactions.

The *MAPT* gene encodes the tau protein, alternatively spliced isoforms of which result in proteins with different function. Exon 10 encodes a microtubule-binding domain, and the balance between proteins with and without this domain (isoforms 4R and 3R, respectively) is important for healthy cells. Changes in the 3R/4R ratio leads to tau aggregation in neuronal cells [61]. The hairpin structure at the border of exon 10 and intron 11 is responsible for alternative splicing of exon 10. Studies showed that this RNA fragment contains six hairpins and one pseudoknot [62]. The regulatory hairpin should be unwound by the helicase p68 in order to include exon 10 into the transcript [63]. Alternatively, the same structure is stabilized by PSF protein, which results in exon 10 skipping [64]. Together, these proteins and the hairpin regulate the alternative splicing of exon 10 in the *MAPT* transcript and provide the equal level of 3R and 4R isoforms of tau protein.

Another gene, in which a hairpin structure plays a major role in splicing, is the survival of motor neuron (*SMN*). There are two copies of *SMN* genes, *SMN1* and *SMN2*. These genes are nearly identical, but *SMN2* generates a shorter transcript because of exon 7 skipping. The resulting SMN protein is unstable, and together with the loss of *SMN1*, it leads to spinal muscular atrophy (SMA) [65]. *SMN2* contains mutations in exon 7 and intron 7, which lead to the loss of an exonic enhancer and create exonic and intronic silencers [66]. Also, there is a hairpin structure at the border of exon 7 and intron 7, named stem-loop structure 2 (TSL-2) [67]. The 5' splice site is involved in base pairing within this hairpin, which inhibits U1 snRNA binding. Small molecules that are able to disrupt this interaction (for example, homocarbonyltopsentin) may be used as a therapy for SMA [68]. Additionally, long-distance interactions in intron 7 were shown to play a role in the control of exon 7 splicing [69]. These interactions could be disrupted by antisense oligonucleotides, which also leads to the correction of splicing in SMA. There are already two drugs for *SMN2* splicing correction approved by the U.S. FDA (antisense oligonucleotide nusinersen in December 2016 and small molecule risdiplam in August 2020), but other therapeutic solutions may also soon be available in clinical practice [70], such as Zolgensma, a gene therapy approved by the U.S. FDA in May 2019.

G-quadruplexes

This type of RNA structure is formed by sequences containing tracts of guanine nucleotides. In a G-quadruplex, four guanines form a cyclic complex, in which all the residues interact via Hoogsteen bonds. The stack of these G-quartets forms a four-stranded helical structure, which is called a G-quadruplex (GQ). This structure might be intramolecular, as well as intermolecular depending on whether all guanines are located in one strand. GQ motifs are enriched in telomeric regions, gene promoters, and untranslated regions of mRNA [71].

In vitro studies showed that GQs in intronic regions may act as *cis*-regulatory elements in alternative splicing. Depending on the position of GQs, they could be

both silencers and enhancers [72, 73, 74]. Some splicing factors (hnRNPH, hnRNPF, SRSF1, SRSF9, hnRNPU, and U2AF65) are suspected to bind GQ motifs [75, 76]. Several examples of the regulatory role of GQs in alternative splicing are described below.

GQs structures were found within intron 6 of the *hTERT* gene [72]. Alternative splicing of *hTERT* transcript results in 22 isoforms, of which only the full-length mRNA is translated into the active protein with reverse transcriptase activity [77]. The use of a GQ-stabilizing agent, named 12459, results in reduced levels of active telomerase. It was shown that the stabilization of GQs modulates alternative splicing by skipping exons 7 and 8, which leads to inactive the hTERT-b protein. Probably, some *cis*-acting sites in intron 6 are blocked by GQ formation. Another GQ-stabilizing agent, CX-5461, also showed the same effect on hTERT alternative splicing [78]. Both agents do not change the total level of hTERT protein but regulate the ratio between active/inactive isoforms.

Another example of the regulatory role of GQ is described for the *TP53* gene. The alternative splicing of one of its isoforms, *p53I2*, that results from intron 2 retention and the use of the second start codon ATG₄₀, leads to the production of inactive Δ 40p53 protein [79]. In intron 3, two six-guanosine tracts crucial for GQ formation have been found [80]. The usage of the GQ-stabilizing agent, 360A, increases the level of full-length *p53* mRNA. It demonstrates that the formation of GQ in the intron 3 regulates the splicing of intron 2, which leads to the change in the the ratio between active and inactive isoforms.

Riboswitches

RNA structures can be altered by the binding of small molecules, and these changes are important for alternative splicing outcomes. One of such small molecules is thiamine pyrophosphate (TPP), and it can regulate eukaryotic riboswitches. In contrast to prokaryotes, which riboswitches affect transcription and translation, the TPP riboswitches regulate alternative splicing. They are usually located in introns

and can hide or expose splice sites depending on the structure they possess. This mechanism of alternative splicing regulation has been confirmed in algae, plants, and fungi [81].

The TPP riboswitch in *Neurospora Crassa NMT1* gene regulates alternative splicing [82]. When TPP concentration is low, the structural conformation of the riboswitch allows the usage of the upstream 5'ss, and this leads to the translation of *NMT1*. When TPP concentration is high, the conformation of the riboswitch changes and a downstream 5'ss is used. As a result, the upstream open reading frame is included in the transcript, which competes with the translation of the main open reading frame and decreases the level of the NMT1 protein. Other examples include the alternative usage of 3'ss in *Chlamydomonas reinhardtii THIC* gene [83], and the usage of 5'ss through long-range RNA complementary interaction in *Neurospora Crassa NCU01977* transcript [84].

2.3.2 Long-range complementary interactions

Similar to local structural motifs, long-range complementary interactions may change splicing patterns dramatically. Experimentally confirmed long-range complementary interactions may be divided into functional sub-groups: structures that regulate alternative 5' or 3' ss, structures that regulate exon skipping, and structures that approximate RBP sites to the place of action.

RNA base pairings in alternative 5'and 3' splice site choice

Long-range RNA complementary interactions may regulate alternative donor and acceptor site usage in *Drosophila* transcripts. The *CG33298* gene encodes an ATPase, and contains a base pair that overlaps with the proximal donor site of exon 13. In this case, the structure loops out 185 nt and is needed for the suppression of the proximal donor splice site. The atrophin gene encodes a transcriptional co-repressor and contains a base pair that overlaps with the proximal acceptor splice site of exon 10. In this case, the structure is needed for the equal usage of the two

acceptor splice sites, which have different strengths. The *Nmnat* gene encodes a nicotinamide mononucleotide adenylyltransferase. Long-range RNA helix loops out about 350 nt regions and leads to the exclusion of exon 5 and poly(A) signal from pre-mRNA. In this case, the RNA structure brings the distal acceptor splice site close to the donor splice site and thereby promotes splicing [85].

The human splicing factor 1 (*SF1*) gene contains 14 exons, and some of them are alternatively spliced. A long-range RNA complementary interaction in the intron between exons 9 and 10, which was confirmed experimentally, loops out 262 nts [86]. Mutations in the minigene construct that disrupted this base pairing changed the choice of 3'ss (a suboptimal acceptor splice site 21 nts downstream of the major one), and the compensatory mutations restored the wild-type splicing. The usage of the distal 3'ss was also noticed in different cancer cell lines and brain tissues, which may indicate the corresponding change of function on the protein level.

The human proteolipid protein 1 (*PLP1*) gene encodes a transmembrane proteolipid present in central nervous system. This gene can produce two alternatively splices isoforms, *PLP1* and *DM20*. These isoforms differ by alternative donor splice sites of exons 3a and 3b, while the acceptor splice site of exon 4 is the same. A decrease in *PLP1/DM20* ratio may lead to X-linked leukodystrophy Pelizaeus-Merzbacher disease (PMD). *PLP1/DM20* ratio is regulated by long-range RNA complementary interactions between conserved sequences separated by 581 nts. The formation of the regulatory loop leads to the production of the *PLP1* isoform. Mutations that disrupt long-range RNA helix result in the formation of the *DM20* isoform, and compensatory mutations return the *PLP1/DM20* ratio to the normal level [87].

RNA bridges

RNA binding proteins may bind to splicing enhancer sequences and promote the inclusion of nearby exons. However, in some genes such as the myosin heavy chain B gene [88], *FGFR2* [89], *ENAH*, and *KIF21A* [90], the splicing outcome depends

on an enhancer site located more than 1 kb downstream of the regulated exon. It has been demonstrated that in the *ENAH* gene, a long-range RNA complementary interaction in the intron downstream of exon 11a brings the distal *RBFOX* motif close to the alternative exon. This mechanism may be computationally extrapolated to many other transcripts as soon as many RBP binding sites are predicted to be deeply intronic, and conserved RNA bridges are more thermodynamically stable near alternative exons than constitutive ones [90].

RNA base pairings in exon skipping regulation

The human *TERT* gene encodes the catalytic subunit of the human telomerase, and only the full-length alternative isoform is active. There are a variable number of tandem repeats in intron 6, and nine of them are required to promote exons 7 and 8 skipping [91]. Mutations introduced within the repeats lead to the inclusion of exons, while compensatory mutations promote exon skipping again. The GQ motif described previously is also located in the same regulatory area, but the impact of GQ formation on long-range RNA complementary interactions is still unknown.

Local hairpin structure TSL-2 in the human *SMN2* gene was already described in the [section 2.3.1](#). In addition, long-range RNA complementary interactions in intron 7 are also involved in exon 7 skipping [92]. Three relatively short helices (8-nt, 7-nt, and 8-nt-long) are located near each other and loop out approximately 40-nt, 120-nt, and 80-nt-long regions. This long-range RNA structure inhibits the inclusion of exon 7, but the addition of antisense oligonucleotides complementary to stem-forming sequences promotes exon 7 inclusion into the final transcript.

Long-range RNA base pairings may also create the system of competing RNA structures, and thus regulate the splicing of mutually exclusive exons. In the next section, we describe mutually exclusive splicing in detail.

2.4 Mutually exclusive alternative splicing

During mutually exclusive alternative splicing, one and only one exon from a cluster of exons is included into the final transcript. Mutually exclusive exons (MXEs) are often tandemly arranged and highly homologous, which indicates that they originate from tandem exon duplications [93]. Nevertheless, the use of different MXEs may provide sufficient variability of protein properties, such as for ion channels and receptors [94, 95]. MXE variants are often specific to tissue and developmental stages in *Drosophila* [96, 95]. The same is true for the regulation of the majority of human MXEs (65% of the total 1,399 MXEs) [97].

Several mechanisms may regulate mutually exclusive splicing. First, steric hindrance prevents the inclusion of several MXEs together. The intron length should be no less than 50-60 nt for the pre-mRNA splicing of *Drosophila* and mammalian transcripts [98, 99]; otherwise, the spliceosome cannot assemble efficiently [100]. This mechanism takes place, for instance, in the mammalian α -tropomyosin transcript [101] and in the insect *Dscam1* exon 17 regulation [102].

The second possible mechanism to regulate MXE splicing is to use different spliceosomes. A specific arrangement of U2- and U12-types splice sites on either of the exons does not allow introns with mixed splice sites to be spliced [103, 104]. This mechanism of spliceosome incompatibility is employed in *MAPK8/JNK1* [103] and other *MAPK* genes [97].

The third mechanism is linked with nonsense-mediated mRNA decay (NMD). If the lengths of both MXEs are not multiples of 3 nt and the rest of the transcript complements them to maintain the reading frame, then the inclusion of both MXEs, or skipping of both of them would lead to a frame shift and consequent introduction of a premature stop codon. The latter is recognized by the NMD pathway, which leads to the degradation of aberrantly spliced mRNA [105]. This mechanism has been demonstrated for the *FGFR2* [106] and *CACNA1C* transcripts [107].

According to current estimates, only a quarter of human MXE clusters are regulated by the mechanisms described above [97]. Other mechanisms include the

regulation by RBPs, interactions between snRNPs with splicing factors [33], and competing RNA secondary structures. The most striking example of long-range competing RNA structures that regulate alternative splicing is in the *Drosophila Dscam1* gene. Additionally, competing RNA structures are responsible for the choice of MXEs in several other *Drosophila* genes, which are described in detail below.

Dscam1

The *Dscam1* gene in *D.melanogaster* encodes an axon guidance receptor and theoretically can produce 38,016 isoforms [96]. This gene contains 95 alternative exons organized into four mutually exclusive clusters: 12 alternative exons in cluster 4, 48 alternative exons in cluster 6, 33 alternative exons in cluster 9, and 2 exons in cluster 17. The ectodomain of *Dscam1* consists of 10 Ig domains and 6 fibronectin type III repeats, in which N-terminal parts of Ig2 and Ig3 and the whole Ig7 domain are encoded by exon clusters 4, 6, and 9 [108]. Alternative transmembrane domains are encoded by exon cluster 17. *Dscam1* with exon 17.1 regulates dendritic development, while *Dscam1* with exon 17.2 governs axonal arborization [109].

Three mutually exclusive exon clusters (exon 4, 6, and 9 clusters) display different patterns of expression bias. Variable exon 6s has the least expression bias, and its splicing pattern is conserved in different cells and tissue types [110]. In contrast, the expression of exon 4s varies not only in the same cells in different individuals but also the splicing profile of a given neuron can be changed over time [111]. The splicing pattern of exon 9 is characterized by the preferred inclusion of exons 9.6, 9.9, 9.13, 9.30, and 9.31 [110]. It is also noteworthy that alternative splicing is independent in different clusters [112]. Moreover, the expression of *Dscam1* isoforms in individual cells is not determined but is rather stochastic [110, 111]. It is almost impossible to see identical subsets of *Dscam1* splice isoforms in two different neurons. Thus, the diversity of combinations of isoforms could provide a mechanism for neuronal self-avoidance [113].

Remarkably, mutually exclusive splicing in all these clusters is regulated by competing RNA base pairings. A seminal paper by Graveley identified two classes of conserved intronic elements in the exon 6 cluster and hypothesized that they are able to form competing RNA structures [114]. According to the hypothesis, there is a single docking site located downstream of constitutive exon 5 and multiple selector sequences located upstream of each exon 6 variant. Each selector sequence is partially complementary to the docking site. Competing RNA structures can form only one at a time, and thus the spliceosome can include only one exon 6 variant into the *Dscam1* mRNA (Figure 2-5). Similar base-pairing interactions were later found in exon 4 and 9 clusters [115] [116]. However, besides the interactions between upstream docking site and downstream selector sequences, the inclusion of some 4 (or 9) variants depends on interactions between upstream selector sequences and the downstream docking site. These sites in both clusters are highly divergent in different clades, but their secondary structures remain conserved.

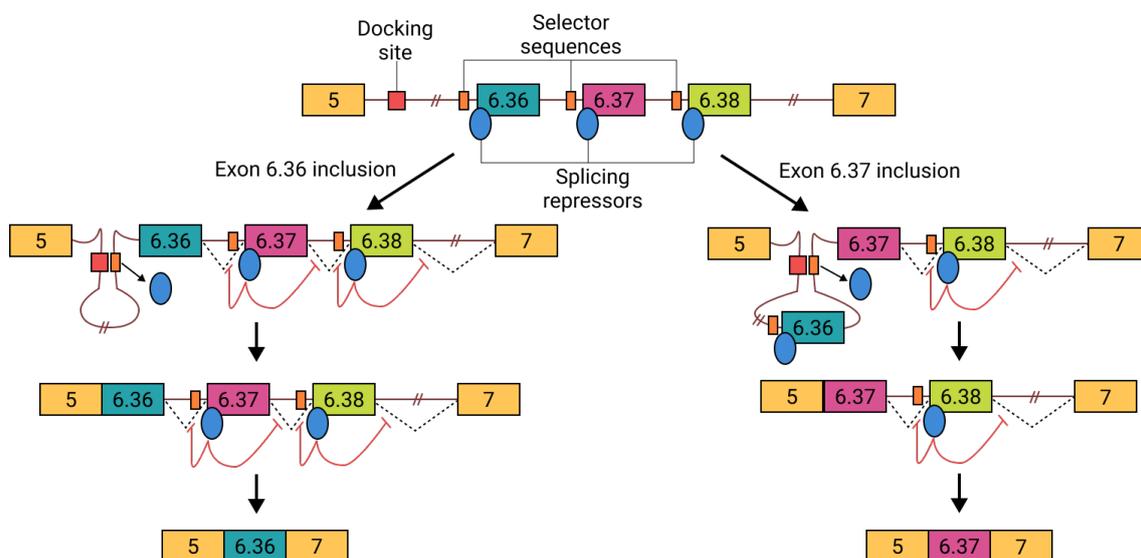


Figure 2-5: The model of *Dscam1* exon 6 cluster regulation. Adapted from [114].

Additionally, intra-intron RNA structures are formed in the exon 4 and 6 clusters. The inclusion stem (iStem) is identified in the intron between exon 3 and the first exon 4 variant [117]. This structural element is required for the effective inclusion of the entire exon 4 cluster but does not determine which variant of exon 4 would be

selected. Another structure that is located immediately upstream of the docking site in the exon 6 cluster is also involved in the inclusion of the entire exon 6 cluster [118]. The proposed mechanism of action involves structure-binding protein factors that promote the recognition of weak splice sites and inactivate the splicing repressors (i.e., hrp36 in exon 6 cluster) [119].

RNA secondary structures are also involved in AS regulation in the exon 17 cluster [115] [102]. The inclusion of both MXEs is inhibited by steric hindrance (the distance between 5'ss and branch point is not enough for effective spliceosome assembly), but the choice between exons is regulated by RNA structures. In *D.melanogaster*, alternative intronic secondary structures could either activate exon 17.1 inclusion by approximation mechanism or inhibit its inclusion by the hindrance of PPT accessibility. In non-Drosophila insect species, an intron-exon RNA structure may be formed, which leads to masking the 3'ss upstream of exon 17.1 and the inhibition of its inclusion. Therefore, the mutually exclusive splicing in the exon 17 cluster is regulated by both steric hindrance and RNA secondary structure.

Competing RNA base pairings

14-3-3 ξ gene in *D.melanogaster* encodes a protein responsible for memory and learning [120]. This gene contains an exon 5 mutually exclusive cluster with 3 alternative exons: 5a, 5b, and 5c. The choice between alternative isoforms depends on competing RNA structures [116]. There are two selector sequences, IE1 and IE2, located downstream of the 5'ss, and a docking sequence IEa in an intron upstream of constitutive exon 6. Interaction between selector sequence (IE1, IE2) and the docking site (IEa) forms a splicing-activating complex which leads to the inclusion of exon 5a and 5b, respectively. Without base pairings, exon 5c is included, whereas exons 5a and 5b are spliced out. In the proposed model, the choice of alternative exons is regulated by the strength of base pairings and the distance between them, as well as by splicing repressors and activators.

The *Drosophila* GATA factor *Serpent* (*Srp*) plays important roles in different processes such as fat body development and the humoral immune response [121, 122]. This gene has an exon cluster with two mutually exclusive exon forms, exon 4.1 and exon 4.2. The splicing choice is directed by two highly conserved RNA pairings with two different docking sites and two different selector sequences [123]. These structures may be mutually exclusive, and the exon 4.2 inclusion depends on the upstream base pairings (stem I), while the exon 4.1 inclusion depends on the downstream base pairings (stem II). These structures may form long-range pseudoknot as well, and in this case, the exon choice may depend on the kinetics of RNA folding and splicing.

The same dual secondary structures were predicted for another *Drosophila* gene, *RIC-3* [123]. *RIC-3* encodes a transmembrane protein responsible for the maturation of neuronal nicotinic acetylcholine receptors [124]. Stems I and II are located around alternative exons 6.1 and 6.2, and these sequences are highly conserved among *Drosophila* species. The same bidirectional competing structures were also experimentally demonstrated for the *Lepidopteran MRP1* exon 8 cluster [125]. *MRP1* encodes an ABC-transporter associated with drug resistance.

PGRP-LC gene encodes a receptor involved in the innate immune response [126]. There are three major isoforms of this protein, *PGRP-LCx*, *PGRP-LCy*, and *PGRP-LCa* [127]. These isoforms differ by 3'-regions, and competing RNA structures control the mutually exclusive choice of 3'-variable regions of *Drosophila PGRP-LC* pre-mRNA [128]. These two mutually exclusive base pairings regulate splicing by activating the proximal 3'ss and masking the intron-proximal 5'ss, and the selection of 3' region variants is correlated with RNA pairing strength. A similar mechanism is used in the regulation of 3' regions of other genes in *Drosophila*, *CG4235*, and *Pip* [128].

2.5 RNA structure analysis using physical and chemical methods

Experimental approaches for RNA secondary structure analysis can be divided into physical and enzymatic/chemical.

Physical methods of determination of RNA structure include X-ray crystallography, cryogenic electron microscopy (Cryo-EM), nuclear magnetic resonance (NMR) spectroscopy, and others. These methods are the most accurate for RNA structure determination, but they are also expensive and time-consuming, and can be effectively used only for short RNAs. Moreover, it is hard to implement crystallization techniques for long and highly flexible RNA molecules.

Chemical approaches provide information about RNA structure by measuring the accessibility of nucleotides to chemical modifications or RNase cleavage [129, 130]. In these experiments, several types of reagents may be used: base-specific reagents, backbone-cleaving reagents, and reagents that modify the 2'-OH of the RNA backbone. After the reverse transcription, each modification results in either termination or mutation at the modified position. The pool of resulting cDNA molecules is sequenced, which enables determination of modification frequency and, consequently, structural information. These approaches are often used together with computational RNA folding algorithms to improve the accuracy of RNA structure prediction [131]. However, RNA structure probing method based on the difference in reactivities of single-stranded and double-stranded residues can tell whether a nucleotide is paired, but they cannot tell to which other nucleotide.

However, it is possible to determine the interacting RNA nucleotides by a group of approaches that are based on proximity ligation. They were first developed to assess the chromatin structure [132], but similar techniques exist for RNA. The main steps in these protocols include crosslinking, proximity ligation, and consequent sequencing of RNA-RNA interactions. These methods may be used to identify intermolecular RNA-RNA interactions (for example, miRNA-mRNA pairs with

CLASH-seq [133]), as well as for both inter- and intramolecular RNA-RNA interactions (for example, LIGR-seq [134] and RIC-seq[3]). These methods can be used *in vivo* and don't require prior knowledge about RNAs or RNA-binding proteins. Moreover, like all chemical approaches, they can be used at high throughput to obtain a global snapshot of RNA structure and interactome.

2.6 Computational methods of RNA structure prediction

Computational methods may be broadly divided into thermodynamic and comparative approaches. Thermodynamic methods assume additive contributions to RNA equilibrium free energy of elementary structure units. They use dynamic programming, in which the most thermodynamically stable RNA secondary structure (the minimum free energy structure) or a number of suboptimal structures are determined. The dynamic programming works by recursively computing the equilibrium free energy from shorter to longer fragments of the molecule, thus assuming that RNA structure is nested, i.e., its diagram doesn't contain crossing base pairings. Therefore, most thermodynamic methods are unable to predict RNA structure with pseudoknots, and if they do so, the computational complexity increases dramatically [135]. Thus, the thermodynamic methods may work well for small RNAs, but not for long-range interactions and pseudoknots in large RNAs.

The comparative approaches to RNA structure prediction make use of compensatory nucleotide substitutions, which can be detected in multiple sequence alignments of homologous sequences [136]. Several variants of this approach exist, including methods based on stochastic context-free grammars [137]. However, the alignment of homologous RNA sequences is not always available and, moreover, stochastic context-free grammars are based on hidden Markov models, thus being computationally extensive and not applicable to long-range RNA interactions. However, if a candidate RNA structure is found, its evolutionary support can be

estimated by measuring pairwise covariations on a phylogenetic tree [138]. Comparative approaches could be combined with thermodynamic approaches for predicting a consensus RNA secondary structure [139], but the applicability of such methods to long-range RNA interactions until recently has been limited.

2.6.1 Computationally predicted long-range RNA structures

In this section, we discuss long-range RNA base pairings that were predicted computationally, but not yet confirmed experimentally.

Myosin heavy chain (*Mhc*) gene in *D.melanogaster* may have 480 different isoforms because of five exon clusters: exon 3 cluster (2 MXEs), exon 7 cluster (4 MXEs), exon 9 cluster (3 MXEs), exon 11 cluster (5 MXEs), and exon 15 cluster (2 MXEs). Remarkably, competing RNA structures including a docking site and selector sequences have been identified in exon clusters 7, 9, and 11 [116]. Due to difficulties with minigene constructs, structures in these clusters were not experimentally demonstrated yet; however, selector sequences and docking sites are highly conserved among *Drosophila* species, which strongly suggests their importance in alternative splicing regulation.

The human *CD55* gene contains a predicted exon 10 MXE cluster with 5 exons [97]. The alternative splicing of this cluster may be regulated by competing RNA structures. The docking site was found between exon 10e and exon 11, and selector sequences were identified downstream of each exon 10 variant, however the experimental validation of this gene is still missing.

The human *Dynamin 1* contains a mutually exclusive exon 10 cluster. The regulation of alternative splicing in this gene is still unclear. Using computational approach, two conserved complementary regions have been found in this cluster [140]. The predicted selector sequence is located downstream of exon 10a and the predicted docking site is located upstream of exon 11. The proposed model of regulation of alternative splicing involves RNA base pairings together with a hypothetical repressor (like *hrp36* in *Dscam1* exon 6 cluster).

The human dystonin (*DST*) gene encodes a protein from the plakin family. The group of exons, 47-52, code the plectin type of repeats. Computational research revealed conserved complementary RNA sequences around this group of exons [141]. Different *DST* isoforms have tissue- and stage of development-dependent expression [142], and the inclusion of this group of exons may be regulated by long-range RNA complementary interactions.

Chapter 3

Thesis Objectives

The main objectives of this research project are:

- Screening of the predicted long-range RNA structures that are associated with alternative splicing in human transcripts.
- Characterization of long-range RNA structures and their impact on alternative splicing in *Phf20l1*, *Cask*, and *Ate1* human genes.
- Study of the impact of long-range RNA structure on alternative splicing in the *Ate1* human gene under RNAPII slowdown.
- Genome-wide characterization of the impact of RNAPII slowdown on alternative splicing using RNA-sequencing.

Chapter 4

Materials and methods

4.1 Minigene constructs and mutagenesis

Whole genomic DNA was isolated from the A549 cell line using the Quick-gDNA MiniPrep kit (Zymo Research). *Ate1* minigene was assembled from three fragments: the first one was inserted into pRK5 vector (kindly gifted by Prof. Petr M. Rubtsov, the Addgene accession number 3944) using restriction-free cloning (protocol from [143]), and the next two fragments were inserted into the resulting plasmid using NEBuilder®HiFi DNA Assembly Cloning Kit (New England Biolabs) according to manufacturer's protocol. *Phf20l1* minigene was created on the basis of the pRK5 vector using restriction-free cloning (*Phf20l1* fragment was inserted after the CMV promoter). *Cask* fragment of DNA was amplified using phosphorylated primers and then inserted into pRK5 vector with Rapid DNA Ligation Kit (Thermo Scientific) according to manufacturer's protocol. Phosphorylated primers were obtained using T4 PNK (Thermo Scientific). Fragments for minigenes assembly were amplified with primers listed in the [Table A.1](#) and Q5 High-Fidelity DNA polymerase (New England Biolab). All primers were synthesized by Evrogen (Moscow, Russia) and Lumiprobe (Moscow, Russia). All minigenes were verified by sequencing in the Shared Resource Centre "Genome" (Moscow, Russia).

Minigene mutagenesis was performed using either protocol from [144] or using phosphorylated primers that introduced the desired changes with subsequent ligation using a Rapid DNA ligation kit (Thermo Scientific). All primers for mutagenesis are listed in Table A.2; all PCR reactions were performed using Phusion High-Fidelity PCR Master Mix (Thermo Scientific). All mutants were verified by sequencing.

4.2 Antisense oligonucleotides

LNA/DNA mixmers were designed based on [145]. All oligonucleotides were locked nucleic acid (LNA)-based with a DNA substitution at every other nucleotide. Antisense oligonucleotides (AONs) that target *Ate1* and *Cask* were 13-mers. AONs that target *Phf20l1* were 18-mers. All LNA/DNA mixmers had phosphonothioated backbones for protection from cellular nucleases. Synthesis of LNA/DNA mixmers was carried out by Syntol JSC (Moscow, Russia). The sequences of used AONs are listed in Table A.3.

4.3 Cell culture

Human A549 lung adenocarcinoma cell line (kindly gifted by Dr. Ilya Terenin, Moscow State University) was maintained in DMEM/F-12 medium containing 10% fetal bovine serum, 50 U/ml penicillin, and 0.05 mg/ml streptomycin (all products from Thermo Fisher Scientific) at 37°C in 5% CO₂. Trypsin-EDTA (Thermo Fisher Scientific) was used for subculturing of cells according to ATCC recommendations. Cell line authentication was confirmed using short tandem repeat analysis. The cell line was tested for the absence of mycoplasma using MycoReport kit (EuroGene).

4.4 Cell culture transfections

Minigene plasmids were transfected using Lipofectamine 3000 (Invitrogen) with reverse transfection protocol. 200,000 cells were seeded in wells of 24-well plate prior

to the transfection. For one well, 1 μ l of Lipofectamine, 1 μ l of P3000 reagent, and 500 ng of plasmid were used in 50 μ l of RPMI medium (Thermo Fisher Scientific). Mixes were incubated for 15 minutes at room temperature and then added on top of seeded cells. 24-hour post-transfection cells were harvested by lysis buffer from PureLink RNA minikit (Invitrogen).

AONs treatment was performed using Lipofectamine RNAiMAX (Invitrogen) on 50–70% confluent cells in a 12-well plate at 5, 25, and 100 nM concentrations of AONs. The transfection procedure was performed according to manufacturer's protocol. Cells were harvested by lysis buffer from PureLink RNA minikit (Invitrogen) after 48 h of treatment. α -amanitin (Sigma) was added at concentrations 1 or 2 μ g/ml to 50–70% confluent cells. After 24 h of treatment, cells were harvested.

In the experiments, when cells were transfected with minigenes and AONs together, the conditions were as follows. Plasmids and AONs were mixed together prior to the transfection, and then these mixtures were transfected using Lipofectamine 3000 (Invitrogen) to 50–70% confluent cells. After 24 h of treatment, cells were harvested.

The experiments with overexpression of NELFE were performed using pCMV3-NELFE plasmid (Cat: HG15217- UT, Sino Biological). Cells were transfected with a plasmid using Lipofectamine 3000 (Invitrogen) and harvested after 24 hours. Cells without transfected plasmid were used as a control.

4.5 Cell culture treatments

For α -amanitin (Sigma) treatments, 1 and 2 μ g/mL of α -amanitin was added to cells at 50–70% confluency. After 24 h of treatment, cells were harvested.

The experiments with α -amanitin and AONs/minigenes treatments were performed as follows. Cells were transfected with AONs/minigenes using reverse transfection. After 12–14 h of transfection, the media was changed, and α -amanitin was added. After 24 h of α -amanitin treatment, cells were harvested.

4.6 RT-PCR experiments

Total RNA was extracted using PureLink RNA minikit (Invitrogen) and treated with RNase-Free DNase I (Thermo Scientific) at 37°C for 60 min, followed by inactivation at 75°C for 10 min in the presence of 5 mM of EDTA (Thermo Scientific). First-strand cDNA was synthesized using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific) according to the manufacturer's instructions. All primers for PCR analysis are listed in [Table A.4](#) (for RT-PCR analysis) and in [Table A.5](#) (for qRT-PCR analysis).

We used competitive RT-PCR analysis for the assessment of the ratio between different splice isoforms in one PCR tube. We used 20–30 ng of cDNA obtained from the total RNA for each PCR reaction. Reactions were carried out using PCR Master Mix (2×) (Thermo Scientific). RT-PCR was carried out under the following conditions: denaturing at 95°C for 3 min, 35 cycles of denaturing at 95°C for 30 s, annealing at 54°C for 30 s, and extension at 72°C for 1 min, followed by extension at 72°C for 5 min. The resulting products were analyzed on a 3% agarose gel stained with ethidium bromide and visualized using ChemiDoc XRS+ (Bio-Rad).

For RT-qPCR experiments, we used 20–30 ng of cDNA obtained from the total RNA and Maxima SYBR Green qPCR Master Mix (2X) (Thermo Scientific) on a CFX96 or CFX384 Real-Time system (Bio-Rad). Minus-RT controls without RT enzyme in the cDNA synthesis reaction were performed in every RT-qPCR analysis. For each mutant and/or treatment experiment, at least three biological replicates in PCR triplicates were analyzed. The PCR cycle parameters were as follows: 95°C for 10 min and 35 cycles with denaturation at 95°C for 10 s, annealing at 54°C for 20 s and extension at 72°C for 30 s. For each pair of primers in RT-qPCR analysis, the primer efficiencies were estimated using a calibration curve. The expression of isoforms was calculated by the efficiency method [146], with primers efficiency of more than 90% in all cases. The expression levels of all isoforms were normalized to the expression of the constitutive exon in the corresponding sample. Then, the sum of all isoform expression levels was taken as 100% to enable comparative analysis of

different biological replicates. We additionally checked the sum of expression levels of all isoforms, and it was not statistically different from the expression level of the constitutive exon in all cases.

4.7 Western blotting

The plasmid pCMV3-NELFE (Cat: HG15217-UT, Sino Biological) was transfected into A549 cells using Lipofectamine 3000 (Invitrogen). Cells were lysed with RIPA buffer after 24 h. Cells without transfected plasmid were used as a negative control. Cell lysates (3 μ g of total protein) were separated by gel electrophoresis on a 10% sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) under denaturing conditions and transferred onto a nitrocellulose membrane. Immunoblotting was done first at 4°C overnight using antibodies against NELFE from [147] (1:500) and GAPDH (Thermo Fisher Scientific, #39-8600, 1:3000). Next, a horseradish peroxidase-conjugated goat anti-rabbit IgG (Invitrogen, G21234, 1:2500) was added for 1 h at the room temperature, followed by detection using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences) and Bio-Rad ChemiDoc XRS + imaging system.

4.8 Library preparation and RNA-seq experiment

A549 cell line was treated with 1 and 2 μ g/mL of α -amanitin (Sigma) and non-treated cells were used for a control. After 24 h of treatment, cells were harvested, total RNA was isolated using PureLink RNA Mini Kit (Thermo Fisher Scientific). Poly(A)⁺ mRNA was purified using Dynabeads Oligo(dT) 25 (Thermo Fisher Scientific) following the manufacturer's instructions.

Illumina cDNA libraries were constructed using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England BioLabs) following the manufacturer's protocol with the only modification: the change of fragmentation time from 15 to 10 min. Complementary DNA libraries were sequenced using the NextSeq500

(Illumina, San Diego, CA, USA) instrument; 33–41 million raw reads were obtained for each sample with a 75 bp read length. The results of RNA-seq have been deposited at Gene Expression Omnibus under the accession number GSE153303.

4.9 Statistical analysis

The data were analyzed and visualized using R statistics software version 3.4.1 and ggplot2 package [148]. The difference between exon inclusion rates in experiments with *Phf20l1* and *Cask* minigenes was assessed and visualized by the ggsignif package, PSI values were compared by t-test. Ternary diagrams were constructed using the ggtern package [149]. Confidence intervals (regions) in ternary diagrams were constructed using the Mahalanobis distance [150]. Error bars in the plots and numbers after the \pm sign represent 95% confidence intervals. When appropriate, the comparison of ψ values was done by a *post hoc* t-test with pooled variances after Mahalanobis distance comparison based on confidence intervals [150]. In all figures, we used standard notation for boxplots including the median, upper and lower quartiles and upper and lower fences without outliers. In RNA-seq experiments under RNAPII slowdown conditions two-tailed Mann–Whitney and Wilcoxon’s tests were used.

Chapter 5

Long-range RNA structures in alternative splicing regulation

5.1 Screening of predicted long-range RNA structures

To prioritize the candidates for experimental validation, we used a list of 916,360 pairs of conserved complementary regions (PCCRs) located in introns of human protein-coding genes that was obtained using a computational pipeline for prediction of long-range RNA structures [2]. These regions were at least 10-nt-long and had the hybridization free energy $\Delta G \leq -15$ kcal/mol. Since the false discovery rate among these predictions drops with increasing PCCR energy, we used the hybridization free energy as the main covariate for prioritizing the candidates. Another useful covariate was the degree of evolutionary conservation, measured as the average phastCons score [151] within base paired nucleotides [2]. PCCRs that occur in more constrained conserved regions were viewed as more probable, however only few of them had compensatory nucleotide substitutions and, consequently, we decided not to use R-scape E-value as a filter [138]. Finally, we used RIC-seq data from [3] as an additional experimental support for RNA structure.

Additionally, we imposed other criteria for PCCRs that were related to their location and complexity. In order to have association with splicing, we selected PCCR that loop out alternative exons and characterized these exons using splice junctions counts calculated from A549 RNA-seq data [152]. PCCRs that overlap with many other PCCRs were excluded from the analysis for the simplicity of tested systems. We additionally considered the functional annotation of selected target genes. We were interested in onco-associated targets, splicing factors, and transcription factors. Also, we aimed at selecting genes with functionally distinct splice isoforms. The lists of isoforms with different functions, tissue- or development-specific expression and localization were taken from [153, 154, 155, 156].

The list of all tested targets is shown in Table 5.1.

Table 5.1: **A short summary of all experimentally tested targets by AONs and/or minigenes.** Columns are (left to right): NCBI gene name; cell line; id of PCCR from [2] or location of the alternative splicing event; AON sequences; AON effect classified by size (no effect, small, large), predictability (the opposite effect identifies effect different from that predicted for the secondary structure regulation) and reproducibility (non-reproducible effects); mutagenesis effect classified by size (no effect, large) and problems (aberrant splicing denotes problems with minigene splicing pattern, contradicting effect describes problems with the explanation of mutagenesis results). N/A – there were no such experiments with the particular target.

| Gene name | Exp | PCCR id | AONs sequences | AON effect | Mutations effect |
|-----------|------|---------|--------------------|------------|------------------|
| SRSF7 | A549 | 466058 | N/A | N/A | Aberrant |
| DDR1 | A549 | 727966 | N/A | N/A | No effect |
| ATE1 | A549 | 148881 | TGCTTCTGAAGGT | Large | Large |
| | | 148879 | N/A | N/A | Large |
| | | exon 7b | TGCACTTTCAGAA | Large | Large |
| | | | TTCTGAAAGTGCA | Large | Large |
| MAP3K4 | A549 | 758066 | N/A | N/A | No effect |
| PHF20L1 | A549 | 838701 | TTGCTGCTATTTGGGGCT | Large | Large |
| | | | AATCCCAAATAACAGCAG | Large | Large |

Table 5.1: Continued on next page

Table 5.1: continued from previous page

| Gene name | Exp | PCCR id | AONs sequences | AON effect | Mutations effect |
|-----------|--------|------------|----------------|------------------|------------------|
| CAMK2G | A549 | exon 13/21 | N/A | N/A | Aberrant |
| FUBP1 | A549 | exon 3/21 | N/A | N/A | No effect |
| CREB1 | A549 | 533237 | N/A | N/A | Aberrant |
| OPA1 | A549 | exon 5/31 | N/A | N/A | No effect |
| MAZ | A549 | 343724 | GACGGCTGTGTCC | Small | No effect |
| | | | TGGGACACAGCCG | Small | No effect |
| MBNL1 | A549 | 623895 | TGTGTAGGCCAAC | No effect | No effect |
| | | | GTTGGCCTACACA | No effect | No effect |
| MAPK14 | A549 | 730865 | ATACGGTGGCAAT | Small | N/A |
| | | 730848 | GATGCCAGCCATA | Small | N/A |
| DNM1 | A549 | 883325 | CCCCACCCCGCTG | Small | N/A |
| | | 883328 | GAGGTCTGTACAT | Small | N/A |
| | | | ATGAGCAGTACCC | Small | N/A |
| DST | A549 | 739752 | TGTTTACAGCATA | Non-reproducible | N/A |
| P4HA1 | A549 | exons 9a/b | GGGTTTCAGATTGG | Large | N/A |
| | | | CCAGTTTCAATCC | No effect | N/A |
| | | | CCAGTCAGAATTC | No effect | N/A |
| | | | GGTGATTGGAAGG | Small | N/A |
| H2AFY | A549 | exons 6a/b | GTAGTTACCTGAG | No effect | N/A |
| | | | CCTGAGACTTCAG | No effect | N/A |
| | | | GGACAGCTGGAAG | No effect | N/A |
| SRSF6 | MCF10A | 554605 | AACATGCAACAAT | Opposite | N/A |
| | | | AACTGATTACATG | Opposite | N/A |
| SRSF3 | A549 | 731025 | CACTCAGCACAAC | Large | No effect |
| | | | TGTGCCGAAGAAC | Large | No effect |
| ANKRD13A | A549 | exon 7/12 | CTGGATGTGACAT | No effect | N/A |
| | | | ACATCACATCCCA | No effect | N/A |
| SFPQ | A549 | 21347 | TGAGTCGATCCAC | Opposite | N/A |
| | | | TGGATCATGTCAC | Opposite | N/A |

Table 5.1: Continued on next page

Table 5.1: continued from previous page

| Gene name | Exp | PCCR id | AONs sequences | AON effect | Mutations effect |
|-----------|-------|---------|----------------|------------------|------------------|
| | | | GAAATGTGATGCC | No effect | N/A |
| TCF3 | A549 | 439607 | CAAGTGCGAGGTG | Large | Contradicting |
| HNRNPD | A549 | 651231 | TGTGCAGCAGGCC | Small | Large |
| | | | ACTTTAATTAACC | No effect | Contradicting |
| CACNA1A | U87MG | 443839 | CTAGGTGGGGAGC | Large | N/A |
| CACNA1D | U87MG | 587278 | CCATGGCACCCAG | No effect | N/A |
| GRK6 | A549 | 720432 | GAGGCAGAGCAGG | Opposite | N/A |
| KIF1B | A549 | 2420 | GAGCACAGATGCC | No effect | N/A |
| OCRL | A549 | 930417 | GGCTCTGGGGAAA | Opposite | N/A |
| KIAA1468 | A549 | 436534 | GCATCCCTTCTTG | Non-reproducible | N/A |
| | | | ACCAAGAGAAAGA | No effect | N/A |
| | | 436542 | CCAAGAAGAATGC | No effect | N/A |
| | | 436539 | TGATGGTTATTTC | No effect | N/A |
| | | | GAAACAGTCCATC | No effect | N/A |
| CASP2 | A549 | 804350 | AGCCACCACCCAG | No effect | N/A |
| KRAS | A549 | 214619 | CAATGCAGAGAGA | Large | N/A |
| | | | TGGAAGCCAATAA | No effect | N/A |
| CASK | A549 | 902118 | GCACACCAATTTCG | Large | Large |
| | | | AAATTGGTGTGCA | Large | Large |
| RBM15 | A549 | 62473 | TTATTGGTGCTAT | No effect | N/A |
| WSB1 | A549 | 375359 | TGACACCGAGCAC | No effect | N/A |
| AGAP3 | A549 | 805022 | GGGCTCCTGGGAG | No effect | N/A |
| DDX46 | A549 | 695689 | N/A | N/A | Contradicting |
| | | 695692 | N/A | N/A | No effect |
| | | 695693 | N/A | N/A | No effect |
| CAMTA1 | HAP1 | 1595 | GTCAGCGTTTGCA | No effect | N/A |
| ABCC5 | A549 | 633851 | ACAGAAGCTCTTC | Small | N/A |
| | | | CCTGTAGTGAAAC | Small | N/A |
| SORBS1 | A549 | 130983 | TTGGAAAGGATTC | Opposite | N/A |

Table 5.1: Continued on next page

Table 5.1: continued from previous page

| Gene name | Exp | PCCR id | AONs sequences | AON effect | Mutations effect |
|-----------|------|-----------|----------------|------------|------------------|
| | | | CATTGCAAACCGC | Small | N/A |
| PRC1 | A549 | 334064 | GGATTCTCAGCCT | Small | No effect |
| | | 334059 | GCACACCAATTTCG | No effect | Contradicting |
| UBP1 | A549 | exon 9/17 | GCAAAAACCCCAGC | No effect | N/A |
| | | | TTCTTGCACAATG | Small | N/A |
| | | | GTGTTAGATTTAG | Small | N/A |
| | | | GTGAAATCCCAGG | Small | N/A |
| WHSC1 | A549 | 639563 | CTCTTCCTGCCTC | No effect | N/A |
| CCNL1 | A549 | 624768 | N/A | N/A | No effect |
| RBMX | A549 | 937662 | CAGCCACTTTCCA | No effect | N/A |
| HNRNPDL | A549 | 651333 | GATCTTTAAATTA | Opposite | N/A |
| | | | TCAATAATTAAAG | Opposite | N/A |

Table 5.1: It ends from the previous page.

Among all tested targets, the greatest effect from AONs treatment and minigene mutagenesis was in *Ate1*, *Phf2011*, and *Cask*. In what follows, we will experimentally demonstrate long-range RNA structures involved in alternative splicing regulation in these three human genes.

5.2 Phf2011

5.2.1 Biological function

The plant homeodomain finger protein 20-like 1 (*Phf2011*) interacts with mono- and dimethylated lysines of histones and non-histone proteins. Its Tudor and PHD domains interact with H3K4me1, H4K20me1, H3K27me2, and DNMT1 [157, 158, 159]. *Phf2011* is also involved in the stability of methylated SOX2 and pRb proteins [160, 161]. The regulatory role of *Phf2011* is essential for epigenetic inheritance in

mammals, cell pluripotency and differentiation, and the maintenance of the G1-S phase checkpoint. Aberrations in *Phf2011* gene expression are common for breast, colorectal, and ovarian cancers [162, 163, 164].

The *Phf2011* gene produces three protein isoforms, and all of them contain two Tudor domains [159]. The two longest isoforms, *Phf2011*-a and -b, differ by the alternative cassette exon 6, when the short *Phf2011*-c isoform doesn't contain exon 6 as well as has a poly(A)-signal after exon 8. All three isoforms were overexpressed in a panel of breast cancer cell lines, but the level of *Phf2011*-c was dramatically higher in several cell lines [165].

5.2.2 RNA secondary structure controls exon 6 inclusion

One of the *Phf2011* gene exons, exon 6, is alternatively spliced. Flanking introns of exon 6 contain conserved complementary regions, which could contribute to alternative splicing regulation (Figure 5-1). The chosen sequences can create a stable RNA duplex, and the published RIC-seq data confirm the presence of this duplex in the *PHF20L1* transcript expressed in the HeLa cell line Figure A-1. Two stem-forming sequences are separated by more than 500 nt and functionally may loop out the alternative exon. In order to investigate the functional role of these conserved complementary regions, we used minigene mutagenesis and anti-sense oligonucleotides.

We constructed a minigene that contains a part of *Phf2011* spanning exons 5 and 7 (Figure 5-2A). We introduced mutations that changed the stem forming sequences to the complementary ones (Figure 5-2B). All effects of mutations on alternative splicing of Phf12011 minigene were assessed by RT-PCR. Mutated minigenes with disrupted base-pairing (which are called m1 and m2) generate only transcripts with exon 6 included. The compensatory double mutant m1m2 restores the base-pairing and reverts the splicing pattern to that of the WT (Figure 5-2C,D). The m1 and m2 mutants had splicing pattern statistically different from the WT, while the double mutant m1m2 was not statistically different from the WT.

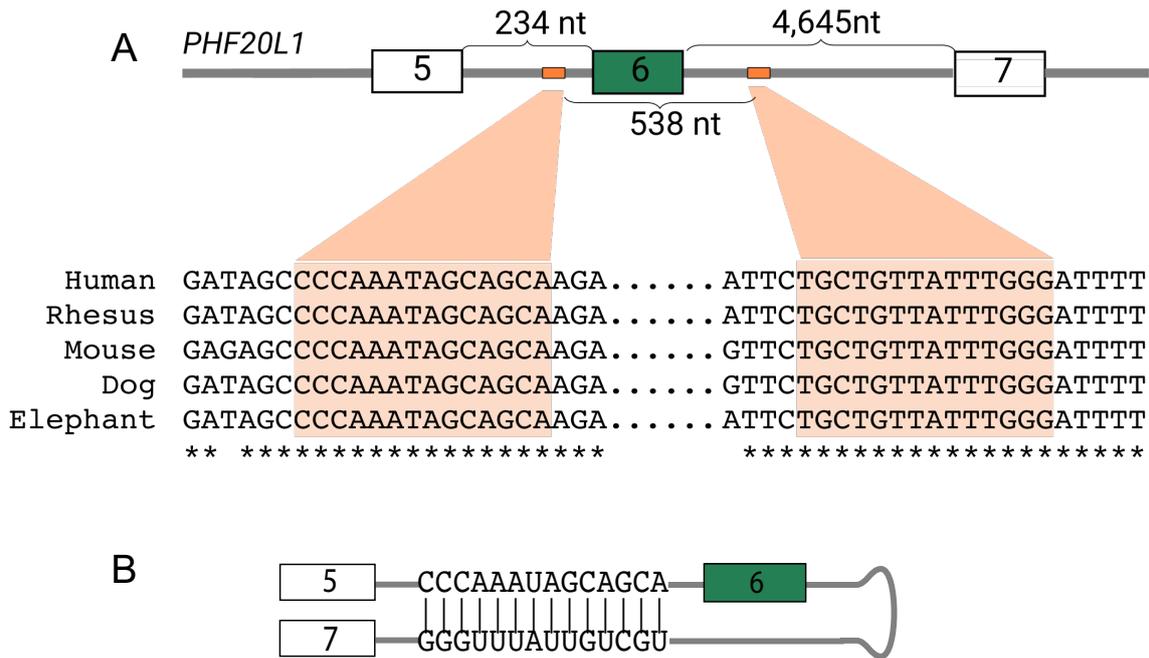


Figure 5-1: A. *Phf20l1* contains two evolutionarily-conserved intronic elements. The conserved positions are indicated by asterisks. Stem-forming sequences are highlighted in orange. B. The scheme of long-range RNA complementary interaction in the exon 6 region.

To evaluate the role of this looping-out stem in the endogenous transcript, we used antisense oligonucleotides (AONs) complementary to both stem-forming sequences (called here AON1 and AON2, Figure 5-3). RT-PCR analysis revealed that 5 nM or higher concentration of either of two AONs is efficient to increase exon 6 usage in the endogenous transcript. For the control, non-treated cells and cells treated with the control AON were used. Control AON has no targets in the *Phf20l1* transcript. Two AONs have the same effect on *Phf20l1* alternative splicing, and the efficacy of AON2 is comparable to that of AON1.

In sum, the effects of AON1 and AON2 on splicing of *Phf20l1* are concordant with each other and consistent with the results of the mutagenesis. Taken together, they indicate that the RNA structure formed around exon 6 in the *Phf20l1* transcript controls the inclusion of this exon.

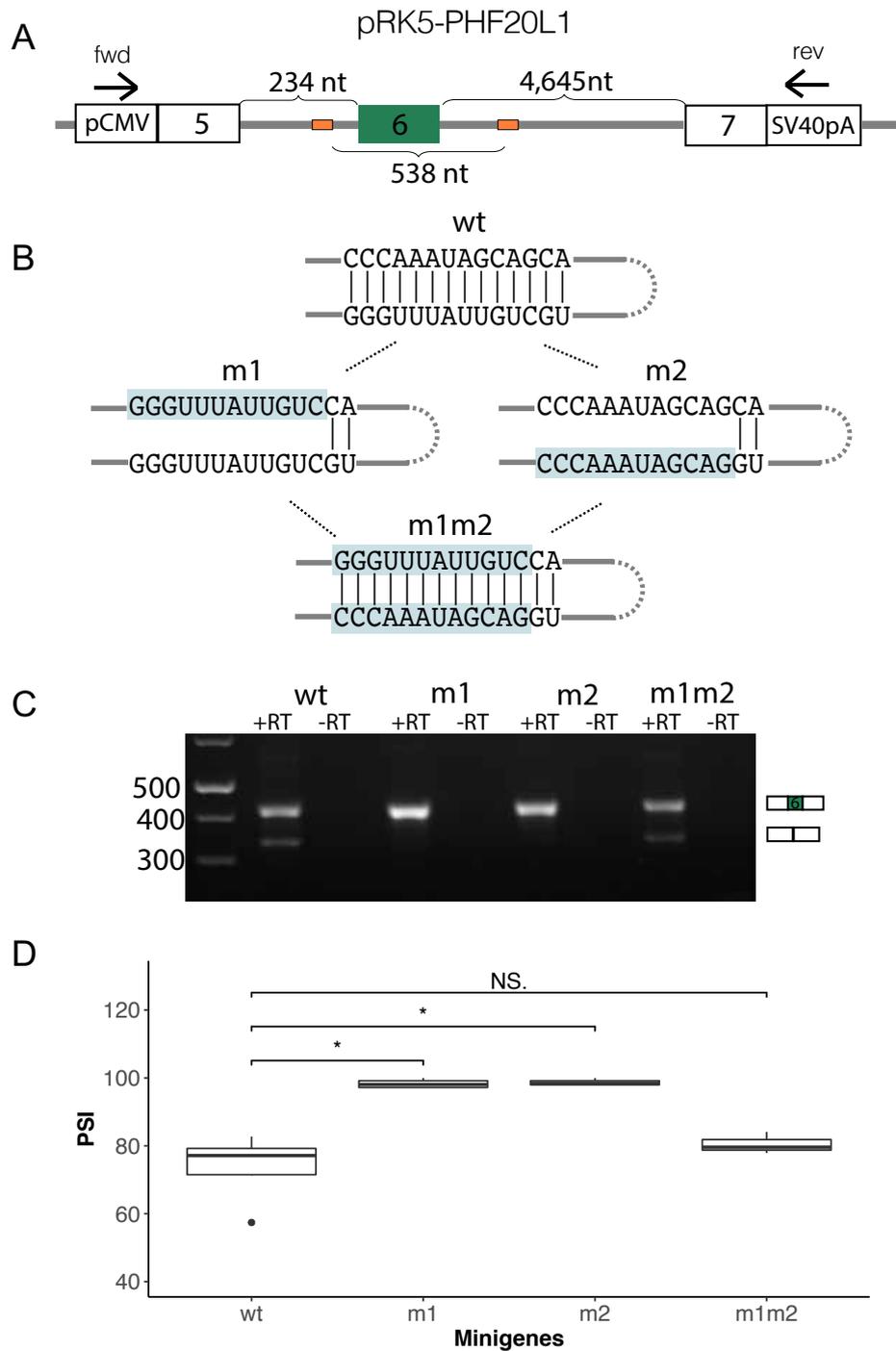


Figure 5-2: A. The scheme of *Phf20l1* minigene. Arrows indicate primers for RT-PCR analysis. B. Disruptive and compensatory mutations. The mutated nucleotides are highlighted in blue. C. The rate of exon 6 inclusion in single and double mutants. D. Exon 6 inclusion rate in three bioreplicates. PSI levels were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***) ; not significant (NS).

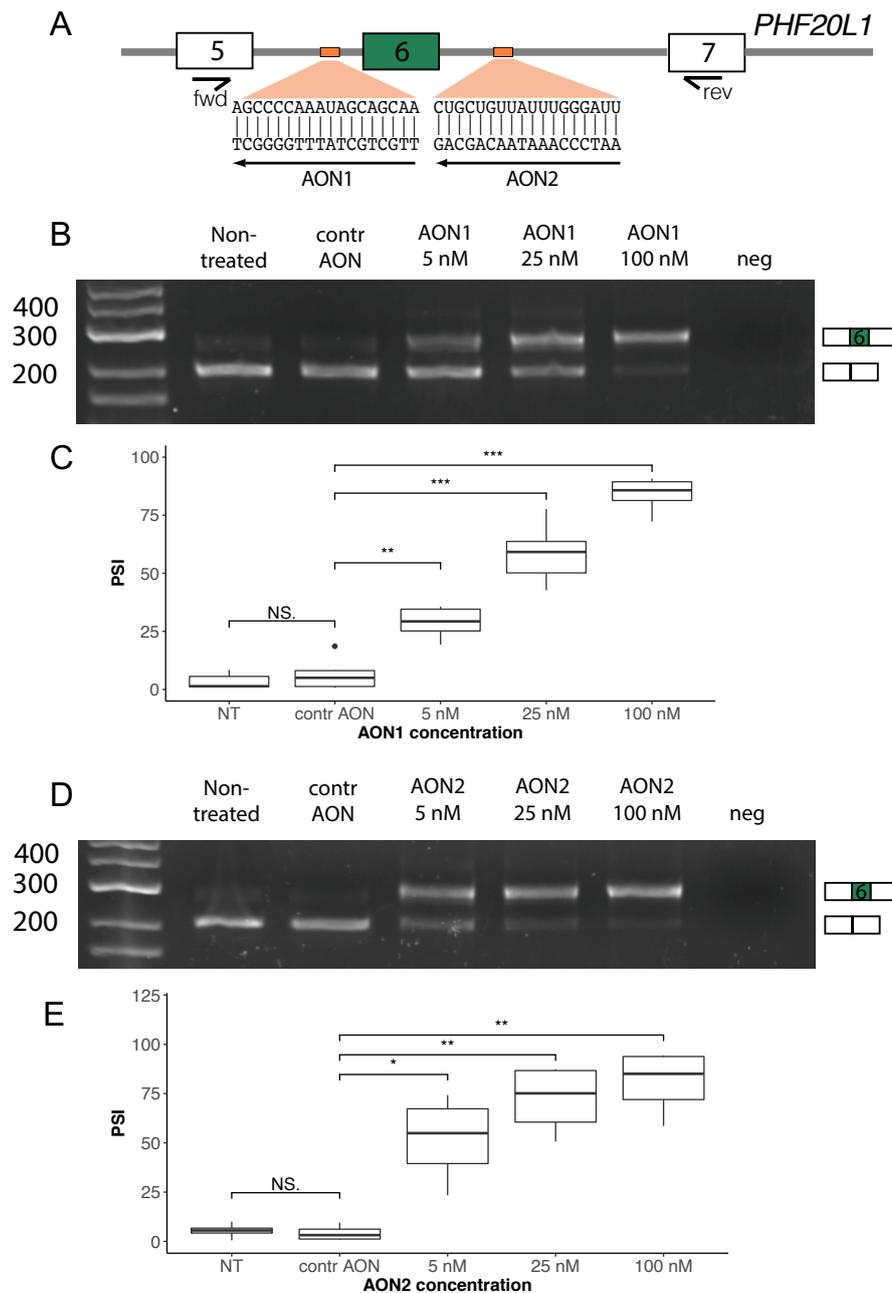


Figure 5-3: A. The sequences of AONs are shown on the scheme of *Phf20l1* gene fragment. Primers used for RT-PCR analysis are also schematically represented on the scheme. B and D. Gel electrophoresis image shows the result of RT-PCR analysis after AON1 and AON2 treatments. C and E. Quantification of electrophoresis gels images with results of AON1 and AON2 treatments. NT corresponds to non-treated cells. PSI levels were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***) ; not significant (NS). All experiments were performed in three bioreplicates.

5.3 *Cask*

5.3.1 Biological function

Cask represents calcium/calmodulin-dependent serine protein kinase, but residues in CaMK-like are mutated, and the protein does not possess kinase activity [166]. In contrast, *Cask* functions as a scaffolding protein involved in presynaptic and postsynaptic processes [167, 168]. Also, *Cask* can enter the nuclei of neurons and interact with Tbr-1, a T-box transcription factor involved in forebrain development [169]. Knockout of the *Cask* gene is lethal for mice within a few hours after birth [170], which suggests its important function in development. Though the expression level of *Cask* is the highest in the brain, its expression is not restricted to neurons [166, 166]. Based on multiple studies, *Cask* is also suspected of playing a role in epithelial cell polarity establishment in mammals [171].

Eight variants of the *Cask* mRNA were identified in the human fetal brain [172]. All variants differ by three cassette exons (11, 19, and 20), as well as the usage of alternative donor sites of exon 23. Exons 19 and 20 encode a linker between PDZ and SH3 domains. In murine neurons, the inclusion of exon 19 (69 bp) or 20 (36 bp) is induced by KCl treatment which mimics neuronal excitation [173]. Probably, different isoforms have altered binding properties to *Cask* partners during development stages, as well as in different cell populations with distinct neuronal activity [172].

5.3.2 RNA secondary structure controls exon 19 inclusion

The *Cask* gene consists of 27 exons, four of which can be alternatively spliced. Exon 19 is a cassette exon, and its flanking introns contain conserved complementary regions, which could regulate alternative splicing (Figure 5-4A). These sequences are highly conserved among almost all vertebrates and can create a stable RNA duplex. Stem-forming sequences are separated by more than 3,000 nt and potentially could loop out the alternative exon (Figure 5-4B). We used minigene mutagenesis

and antisense oligonucleotides for the investigation of the impact of this stem in alternative splicing regulation.

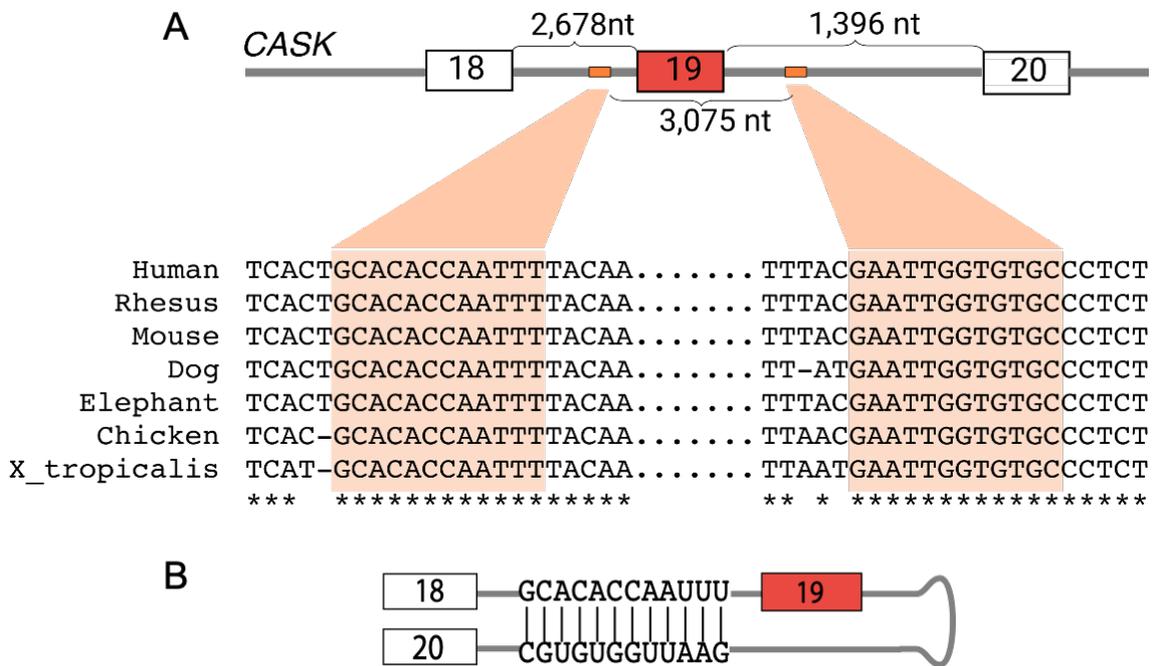


Figure 5-4: A. *Cask* contains two evolutionarily-conserved intronic elements. The conserved positions are indicated by asterisks. Stem-forming sequences are highlighted in orange. B. The scheme of long-range RNA complementary interaction in the exon 19 region.

In the minigene that contains a part of *Cask* spanning exons 18 and 20 (Figure 5-5A), we introduced mutations that changed the stem forming sequences to the complementary ones (Figure 5-5B). All effects of mutations on alternative splicing of *Cask* minigene were assessed by RT-PCR. Mutated minigenes with disrupted base-pairing (which are called m1 and m2) generate more transcripts with exon 19 included compared to the WT. The compensatory double mutant m1m2 restores the base-pairing and reverts the splicing pattern to that of the WT (Figure 5-5C,D). The m1 and m2 mutants have a splicing pattern statistically different from the WT, while double mutant m1m2 is not statistically different from the WT.

To evaluate the role of this looping-out stem in the endogenous transcript, we used antisense oligonucleotides (AONs) complementary to both stem-forming sequences (called here AON1 and AON2, Figure 5-6). RT-PCR analysis revealed that

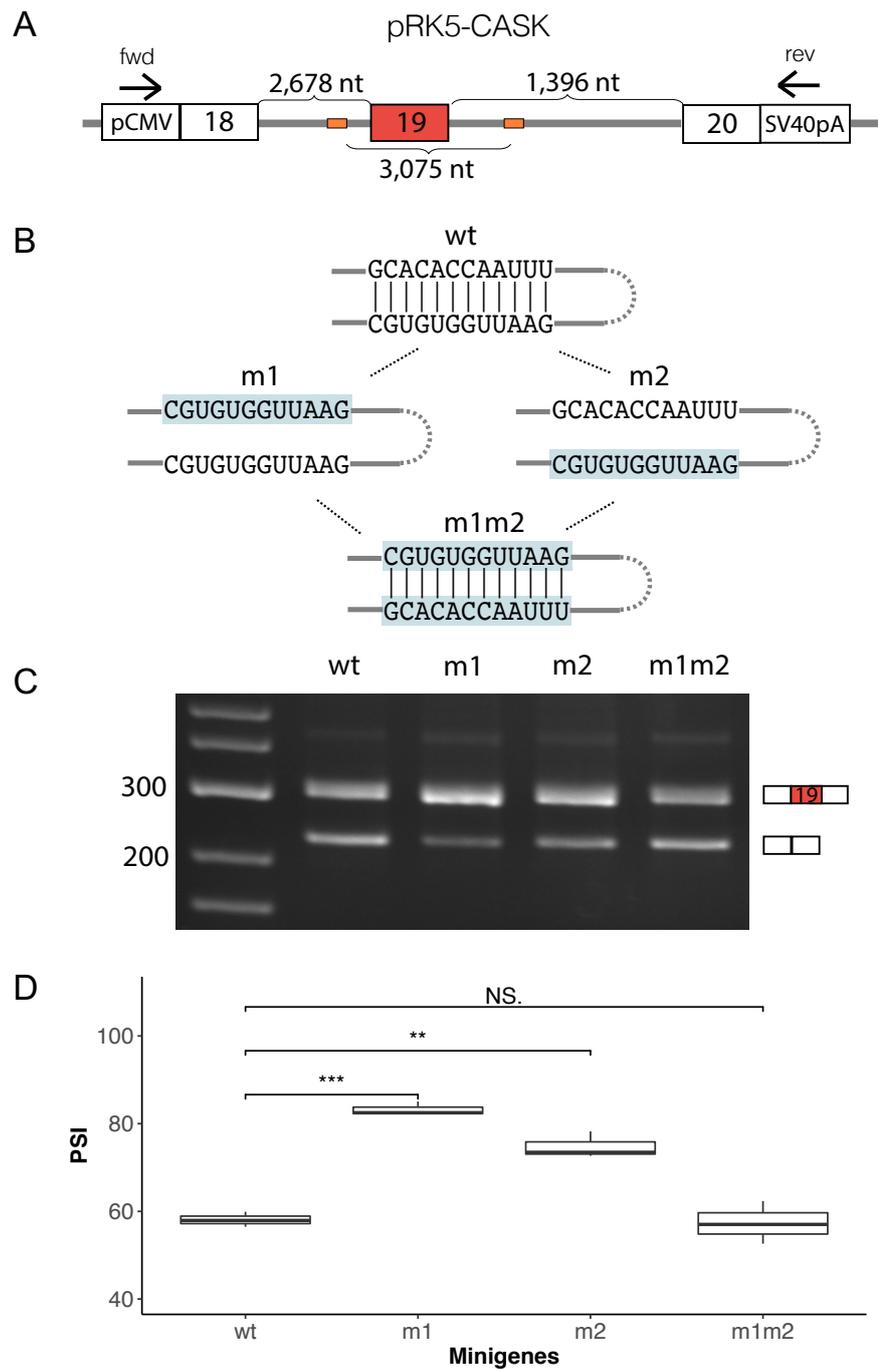


Figure 5-5: A. The scheme of *Cask* minigene. Arrows indicate primers for RT-PCR analysis. B. Disruptive and compensatory mutations. The mutated nucleotides are highlighted in blue. C. The rate of exon 19 inclusion in single and double mutants. D. Exon 19 inclusion rate in three bioreplicates. PSI levels were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***); not significant (NS).

5 nM or higher concentration of either of two AONs is efficient to increase exon 19 usage in the endogenous transcript. For the control, non-treated cells and cells treated with the control AON were used. Two AONs have the same effect on *Cask* alternative splicing, and the efficacy of AON2 is comparable to that of AON1.

In sum, the effects of AON1 and AON2 on the splicing of *Cask* are concordant with each other and consistent with the results of the mutagenesis. Taken together, they indicate that the RNA structure formed around exon 19 in the *Cask* transcript controls the inclusion of this exon.

5.4 *Ate1*

5.4.1 Biological function

Arginylation is a widespread post-translational protein modification that transfers an L-arginyl residue from the Arg-tRNA onto the polypeptide chain [174]. It is mediated by the arginyl transferase encoded within the *Ate1* gene [175]. *Ate1* is essential in most eukaryotic systems and is implicated in the regulation of physiological pathways including proteolysis [176, 177], response to stress and heat shock [178, 179, 180], embryogenesis [181, 182, 183], regenerative processes [184, 185, 186], and aging [187, 188]. *Ate1* has recently been identified as a master regulator affecting disease-associated pathways [189, 190, 191], and its knockout results in embryonic lethality and severe developmental defects in mice [182, 183, 192, 193].

5.4.2 *Ate1* alternative isoforms

Like many other eukaryotic genes, *Ate1* generates several mRNA isoforms through alternative splicing [194]. In mammals, they differ by mutually exclusive choice of two adjacent, homologous 129-bp exons (7a or 7b) and by alternative choice of the initial exon (1a or 1b) [195]. The two major mRNA isoforms of *Ate1* are *Ate1-1* (1b7a) and *Ate1-2* (1b7b), while the isoforms that contain both exon 7a and 7b are

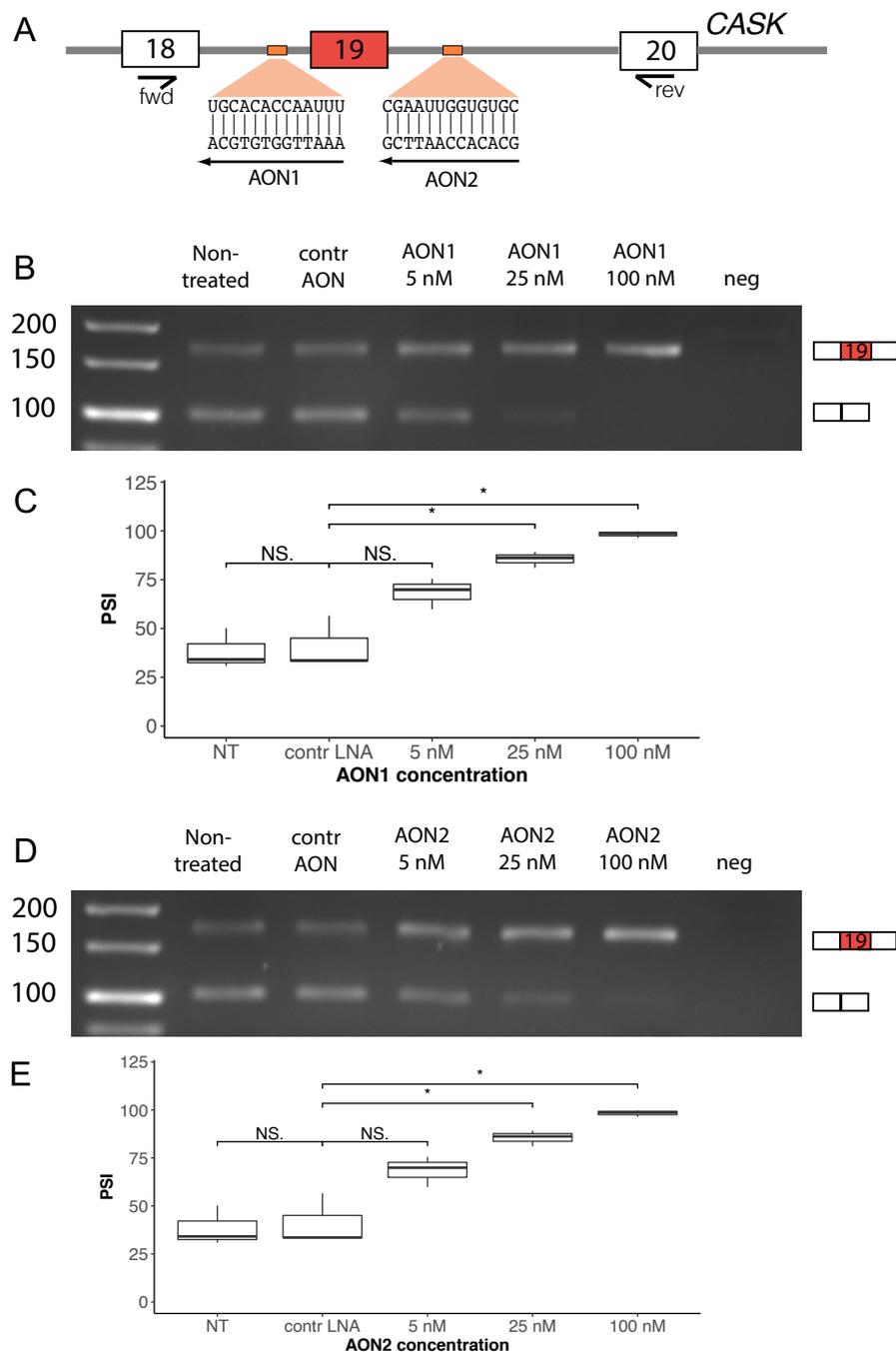


Figure 5-6: A. The sequences of AONs are shown on the scheme of *Cask* gene fragment. Primers used for RT-PCR analysis are also schematically represented on the scheme. B and D. Gel electrophoresis image shows the result of RT-PCR analysis after AON1 and AON2 treatments. C and E. Quantification of electrophoresis gels images with results of AON1 and AON2 treatments. NT corresponds to non-treated cells. PSI levels were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***); not significant (NS). All experiments were performed in three bioreplicates.

suppressed by mutually exclusive splicing [191]. In mice, *Ate1-1* and *Ate1-2* are expressed stably in all tissues, but their ratio varies from 0.1 in the skeletal muscle to 10 in the testis [194, 196, 197]. While *Ate1-2* is almost completely cytosolic, *Ate1-1* localizes in both cytosol and nucleus [194] and can specifically interact with Liat1, a testis-specific molecule [198]. Furthermore, *Ate1*-knockout cells can form tumors in subcutaneous murine xenograft assays, in which the tumor growth can be partially rescued by the reintroduction of stably expressed *Ate1-1*, but not *Ate1-2* [189]. The isoforms *Ate1-3* (1a7a) and *Ate1-4* (1a7b) encode a variant of arginyl transferase that is specific for N-terminal cysteine with tissue-specific expression, cellular localization, and carcinogenic potential similar to those of *Ate1-1* and *Ate1-2*, respectively [195]. The ratio of *Ate1* isoforms containing exons 7a and 7b switches substantially during male meiosis in mice suggesting a role in the mitotic-to-meiotic transition of the germ cell cycle [197]. All these observations suggest that the sequences of amino acids encoded by exons 7a and 7b result in functionally distinct arginyl transferases.

5.4.3 Conserved complementary regions in exon 7 cluster

In order to identify the mechanism responsible for mutual exclusive splicing of exons 7a and 7b, we used comparative sequence analysis to search for potential regulatory sequences in their intervening and flanking introns (Figure 5-7). The regions immediately upstream of exons 7a and 7b, termed here as R1 and R4, are highly similar to each other and show remarkable sequence conservation across vertebrate species. The intron between exons 7a and 7b contains two conserved regions, termed here as R2 and R3, where R3 is complementary to both R1 and R4, while R2 is complementary to another highly conserved region R5 located ~30 kb downstream in the intron between exons 7b and 8. The base pairing between R1 and R3 was predicted in earlier works [141]. The pattern of complementarity between these regions suggests that R1 and R4 could compete with each other for base pairing with R3 and, together with base pairing of R2 with R5, they form a pseudoknot (Figure 5-8).

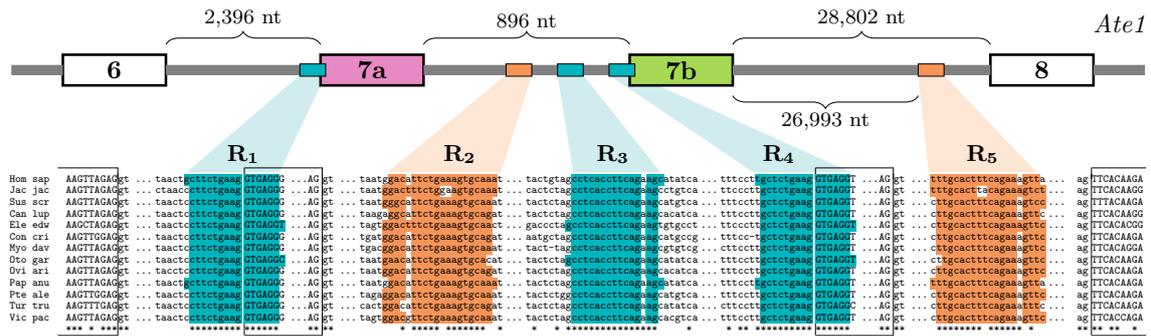


Figure 5-7: Scheme of *Ate1* mutually exclusive cluster 7. *Ate1* contains five evolutionarily-conserved intronic elements (R1-R5). The conserved positions are indicated by asterisks.

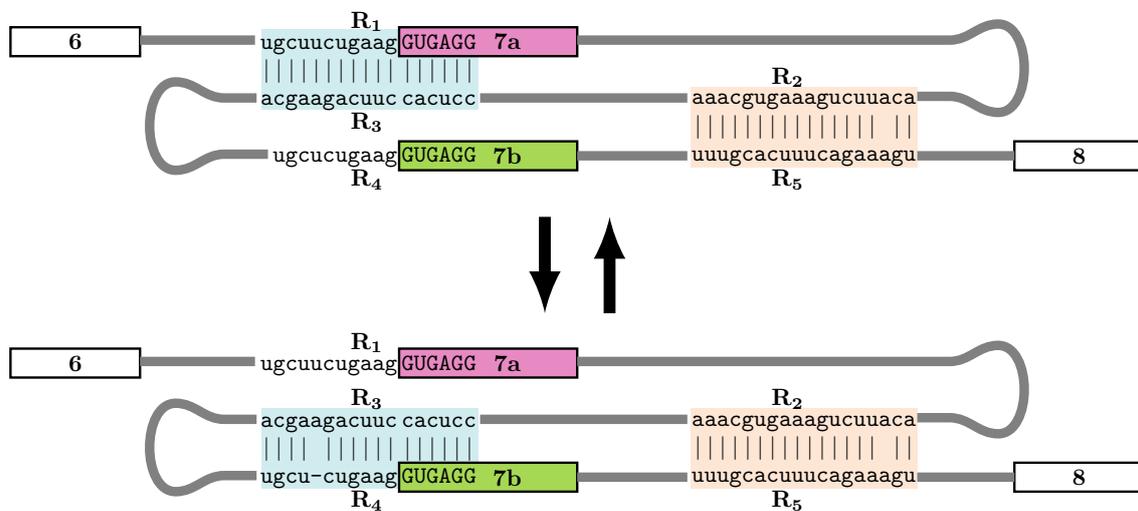


Figure 5-8: Proposed complementary interactions in the *Ate1* pre-mRNA. R1 and R4 are highly similar to each other and are both complementary to R3; R2 is complementary to R5, which is located 40 Kb downstream. R1 and R4 compete for base pairing with R3.

The occurrence of competing RNA structures in mutually exclusive exons in *Ate1* is reminiscent of splicing control mechanisms in other genes [199]. We therefore analyzed in detail the function of these complementary regions using site-directed mutagenesis and antisense oligonucleotides.

5.4.4 Competition between R1R3 and R3R4 controls mutually exclusive splicing

To check whether RNA structure is implicated in the regulation of splicing in *Ate1*, we created a minigene construct containing a fragment spanning between exons 6 and 8 under the constitutive CMV promoter and quantitatively assessed splice isoforms in transfection experiments using human adenocarcinoma A549 cells (Figure 5-9). In the minigene, the endogenous intron downstream of exon 7b was reduced in size to ~2 kb due to obvious limitations of cloning large fragments.

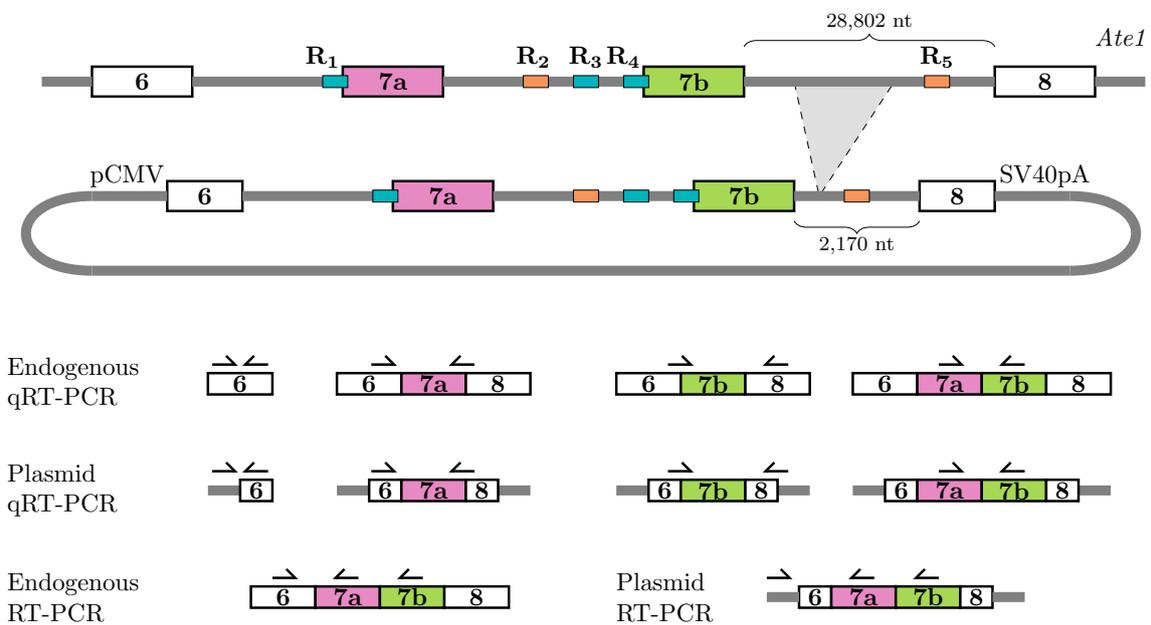


Figure 5-9: The scheme of the *Ate1* minigene. The intron between exon 7b and exon 8 was reduced from 29 kb in the genome to 2 kb in the minigene (top); primers for quantitative and qualitative assessment of alternative splicing (bottom).

Our strategy was to make point mutations that disrupt RNA structure when introduced alone, but restore it when introduced in different combinations. First, we tested the effect of disruptive and compensatory mutations on double-stranded structures of R1R3 and R3R4 (Figure 5-10A) and assessed the corresponding splicing changes in *Ate1* minigene by RT-PCR (Figure 5-10B). The mutation disrupting R1R3 base pairing (m2) increased the usage of exon 7a, whereas the mutation disrupting R3R4 (m11) increased the usage of exon 7b (Figure 5-12A and Fig-

ure 5-10B). The mutated R3 (m1), which is unable to base-pair with R1 or R4, increased the proportion of transcripts containing double exons. The compensatory double mutation m1m2, which restored R1R3 but disrupted R3R4, increased the efficiency of exon 7b inclusion while exon 7a inclusion became almost fully suppressed. Conversely, the compensatory double mutation m1m11, which restored R3R4 while disrupting R1R3, led to a more efficient inclusion of exon 7a compared to the WT. Finally, the proportion of splice isoforms in the triple mutant (m1m2m11), in which both R1R3 and R3R4 were restored, was similar to that in the WT.

To evaluate the effect of disruptive and compensatory mutations quantitatively, we repeated the same measurements using qRT-PCR and constructed a ternary plot for the inclusion rates of exon 7a, exon 7b, and double exons (Figure 5-10C). The mutants formed four separate clusters depending on which double-stranded structure (R1R3, R3R4, none, or both) was disrupted. The 95% confidence regions around these clusters confirmed that the exon inclusion rates were significantly different when different helix sets were disrupted. Despite small, but statistically discernible differences between individual mutants within each cluster (Figure 5-11), the pattern of exon inclusion in the triple mutant (m1m2m11) with both R1R3 and R3R4 restored was more similar to that in the WT than it was in all other mutants (Figure 5-10B). Notably, the ratio of exon 7a/7b inclusion changed proportionally to the difference of thermodynamic stabilities of R1R3 and R3R4 in all mutants except m2 and m11 (Figure 5-12B), and the proportion of isoforms containing both exon 7a and 7b increased whenever R1R3 base pairing was disrupted.

It could be questioned whether the double-stranded regions formed by R1–R5 are substrates of ADAR adenosine deaminases [200]. To address this question, we sequenced nucleotide sequences surrounding the regions R1–R5 in the minigene and in the corresponding transcript, but no pattern specific to adenosine-to-inosine RNA editing was detected (Figure 5-13). RADAR, currently the largest database of RNA editing sites, contains no evidence of RNA editing in R1–R5. Additionally, we examined the responses of exons 7a and 7b to shRNA depletion of ADAR1

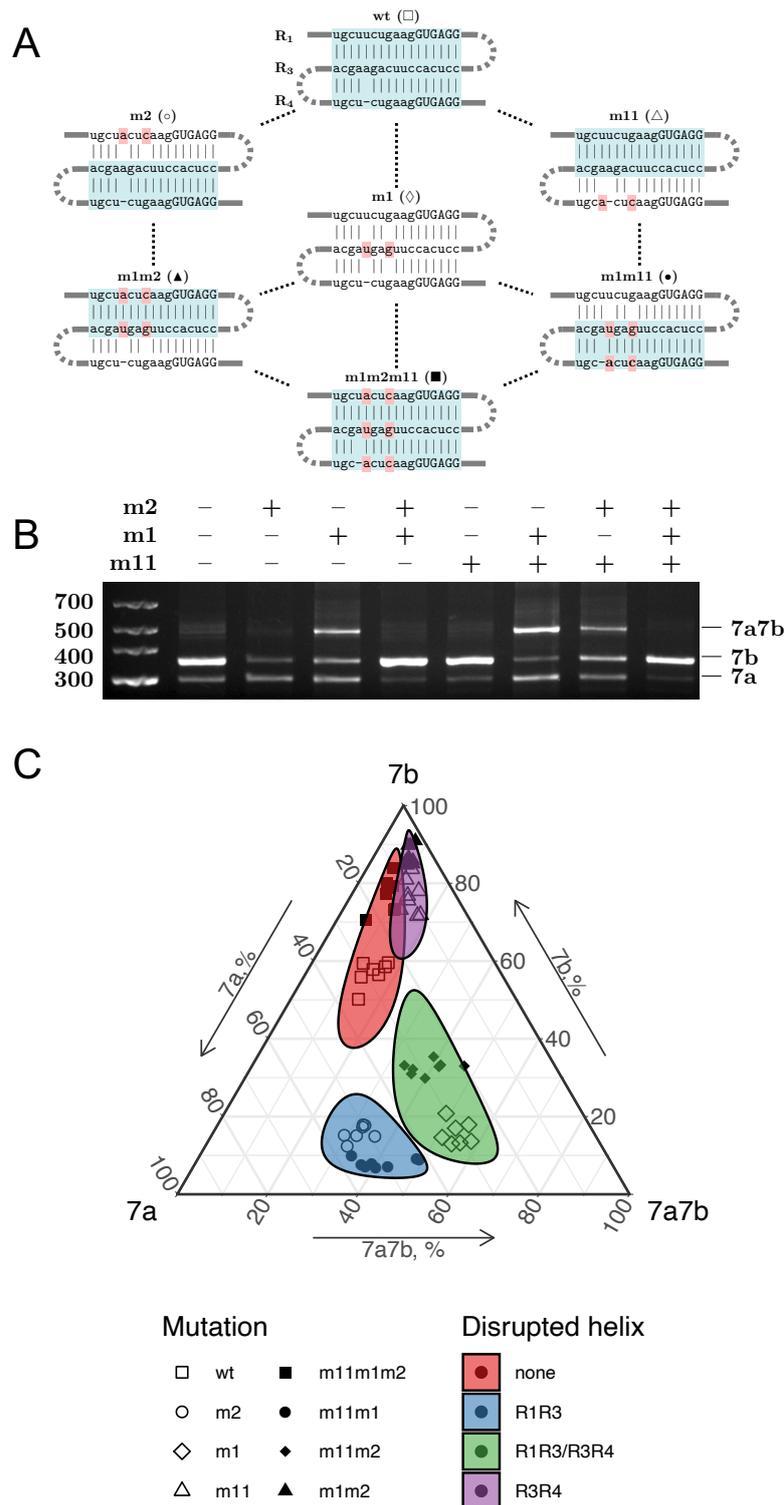


Figure 5-10: R1R3 and R3R4 mutagenesis. A. Disruptive and compensatory mutations in R1, R3 and R4. B. The rate of inclusion of exons 7a, 7b, and double exons (7a7b) changes in single and double mutants; WT splicing is qualitatively restored in the m1m2m11 triple mutant. C. Ternary plot of exons 7a, 7b and 7a7b inclusion rate measured by qRT-PCR. Colored areas represent the 95% confidence intervals.

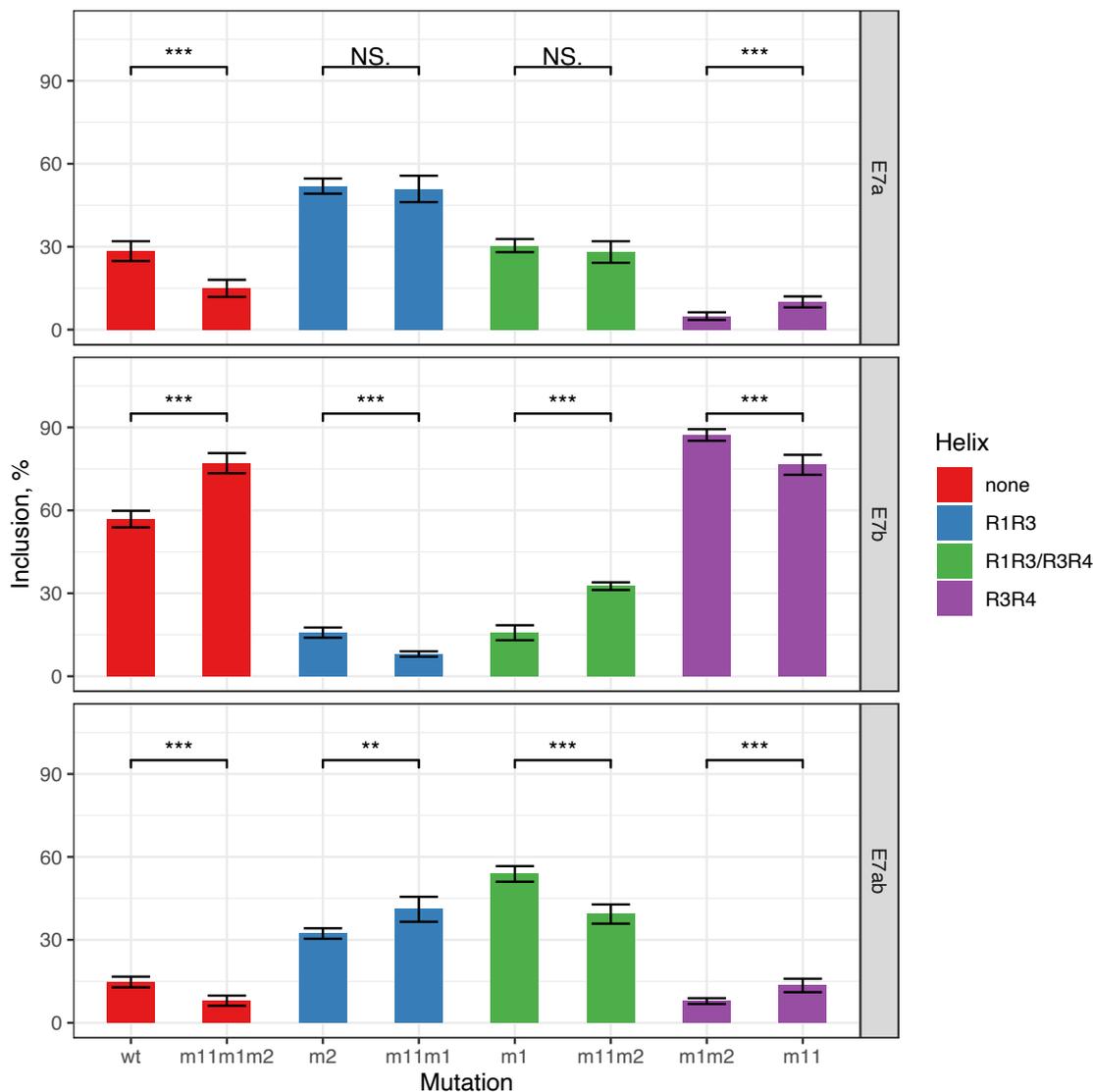


Figure 5-11: Pairwise comparisons of mutants disrupting the same helix. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***); not significant (NS).

profiled by RNA-seq within ENCODE consortium [201], but no significant changes of exon inclusion were detected. Together, these results indicate that R1–R5 are not substrates of ADAR enzymes.

The cloned fragment of *Ate1* lacks a substantial part of intron 7, which may affect splicing. We therefore independently examined the role of R1, R3, and R4 in the endogenous *Ate1* transcript using locked nucleic acid (LNA)/DNA mixmer as antisense oligonucleotides (AONs) that interfere with RNA secondary structure

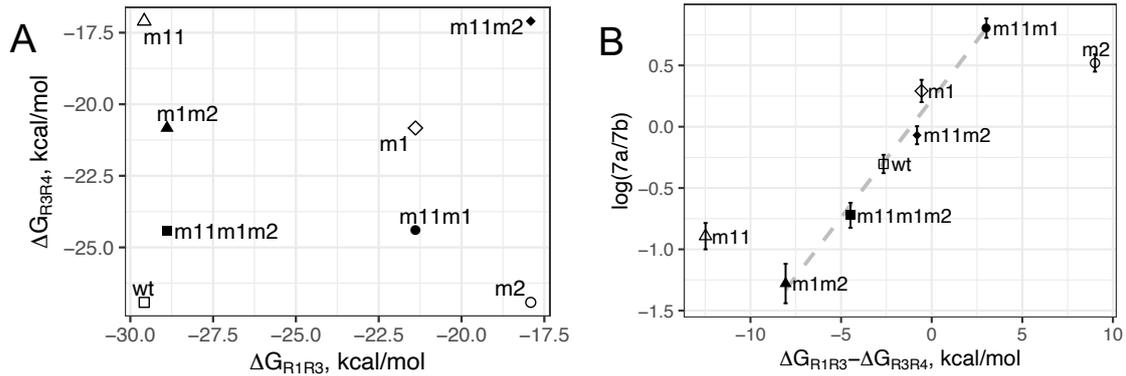


Figure 5-12: A. The predicted hybridization energies of R1R3 and R3R4 in *Ate1* mutants. B. Log-ratio of exon 7a/7b isoforms depends on the difference of thermodynamic stabilities of R1R3 and R3R4.

[145]. Since R1 and R4 overlap with splice sites, we chose to use AON with specific sequence complementary to R3 (AON1) (Figure 5-14A). RT-PCR analysis revealed that the transfection of 5 nM or a higher concentration of AON1 efficiently induced the inclusion of double exons and suppressed the inclusion of individual exons, while the transfection of the control AON didn't show any difference from the non-treated cells (Figure 5-14B). Quantitative RT-PCR (qRT-PCR) with isoform-specific primers confirmed consistent dose dependence of AON1 treatment (Figure 5-14C). Taken together, the mutagenesis and AON1 treatment indicate that the function of competitive base pairings between R1, R3, and R4 is to control mutually exclusive splicing of exons 7a and 7b.

5.4.5 The ultra-long-range R2R5 base pairing controls isoform bias

To elucidate the function of the other two conserved regions, R2 and R5, we applied a similar mutation strategy to the minigene carrying a reduced *Ate1* fragment (Figure 5-15A). When introduced separately, the disruptive mutations m3 and m4 almost completely abrogated the inclusion of exon 7a and strongly enhanced the inclusion of exon 7b, while the compensatory double mutant m3m4 reverted the splicing pattern to that of the WT (Figure 5-15B). Remarkably, the disruptive mu-

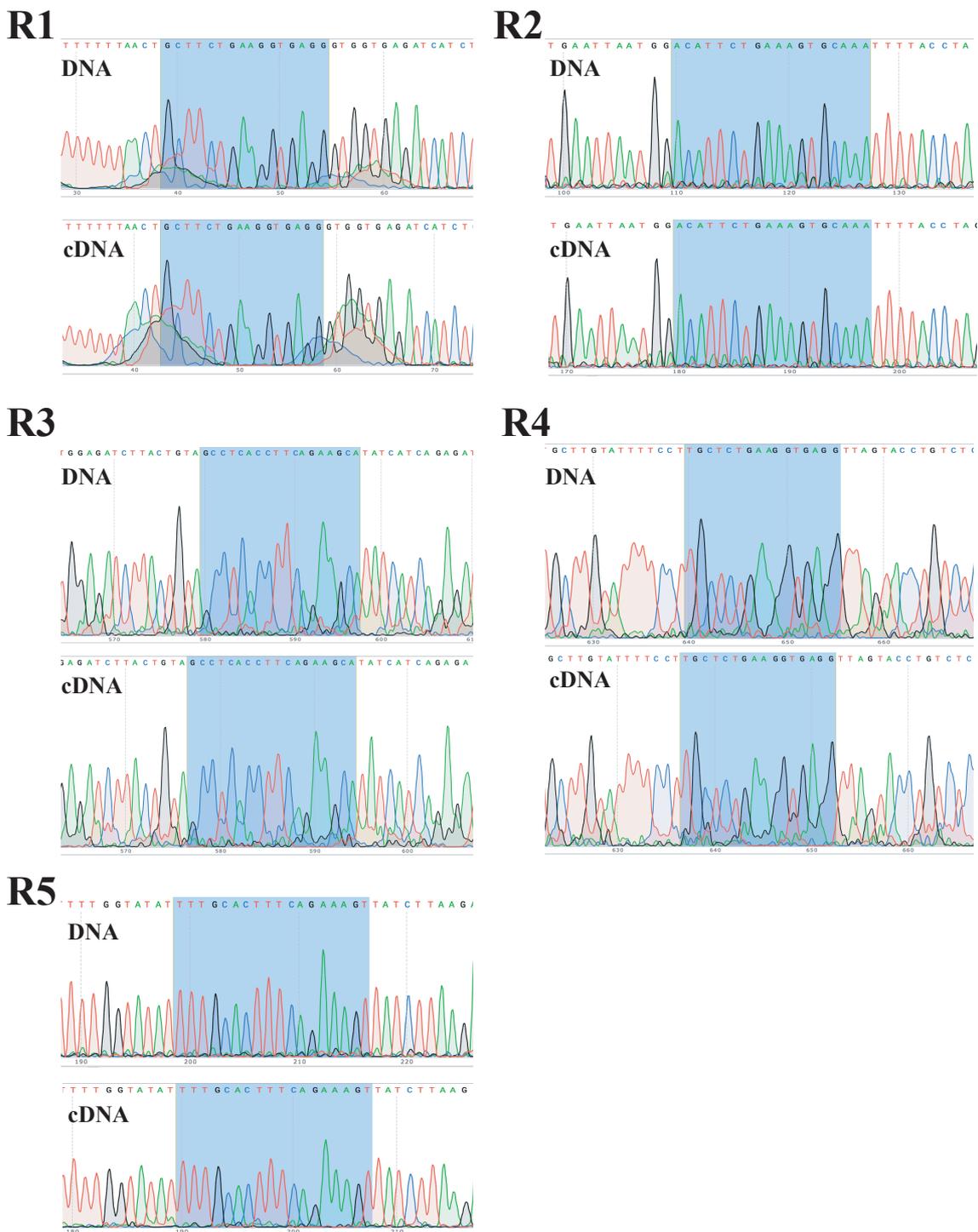


Figure 5-13: The profiles of Sanger sequencing of regions R1–R5 in the minigene and in the mRNA show no traces of RNA editing.

tations affected only the ratio of isoforms carrying mutually exclusive exons, but not the proportion of transcripts with both exon 7a and 7b (Figure 5-15C).

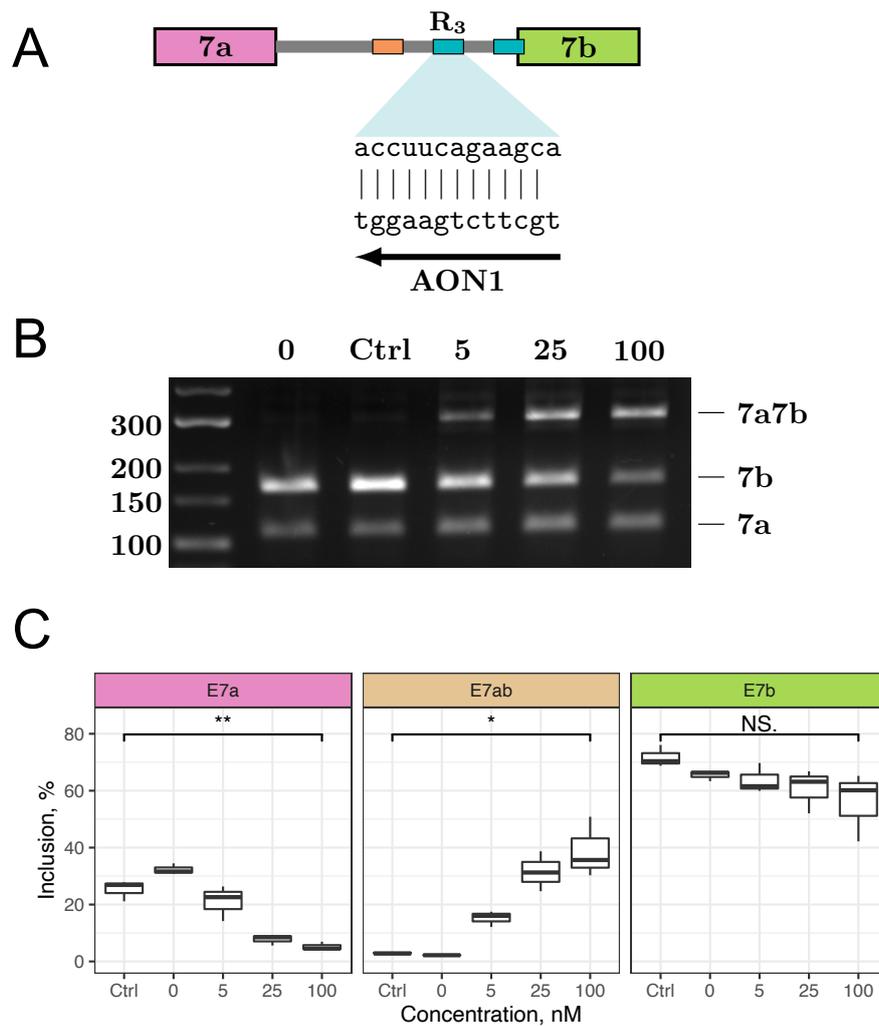


Figure 5-14: A. AON1 disrupts R1R3 and R3R4 via base pairing to R3. B,C. The inclusion of double exons increases with increasing AON1 concentration in the endogenous *Ate1* transcript. Inclusion rates were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***) ; not significant (NS). All experiments were performed in three bioreplicates.

In the endogenous gene, however, R2 and R5 are located ~ 30 Kb apart from each other, while in the minigene construct the distance between them is reduced to ~ 2 Kb. Hence, we used AONs complementary to R2 and R5 (AON2-1 and AON2-2, respectively) to confirm that the ultra-long-range base pairing between R2 and R5 also modulates alternative splicing in the endogenous *Ate1* transcript (Figure 5-16A). RT-PCR and qRT-PCR analyses revealed that 5 nM or a higher concentration of AON2-1 was sufficient to decrease exon 7a usage and increase exon 7b usage in comparison with non-treated cells and cells treated with the control AON, while the

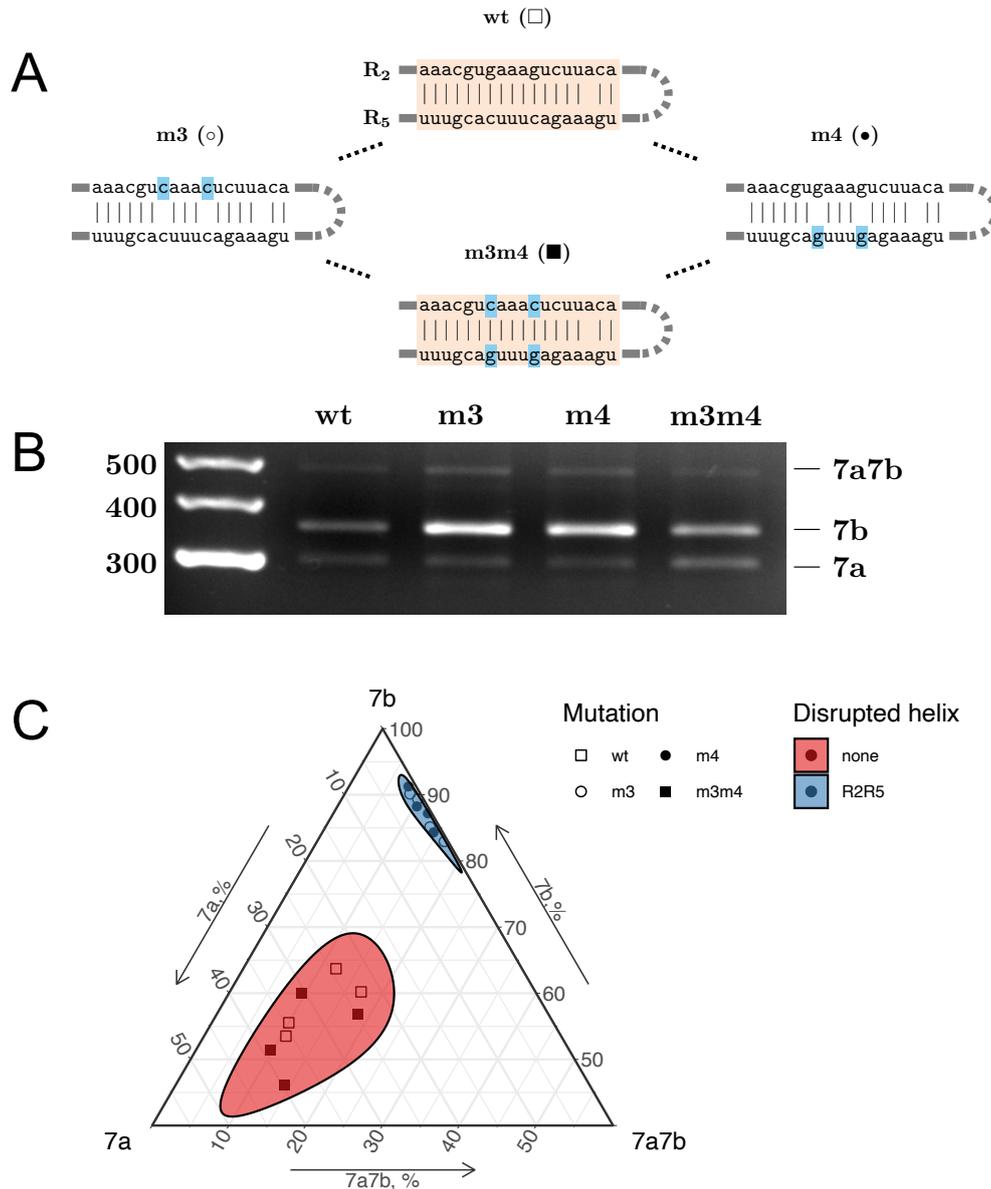


Figure 5-15: A. Disruptive and compensatory mutations in R2 and R5. B,C. The mutations disrupting R2R5 base pairing (m3 and m4) promote exon 7b inclusion, while the compensatory mutation (m3m4) restores WT splicing.

proportion of transcripts with double exons remained low (Figure 5-16B and Figure 5-16C). The efficacy of AON2-2 in reducing the inclusion of exon 7a decreased to a comparable extent only at higher concentrations of AON2-2 (100 nM). Since the hybridization strengths was similar to that of AON2-1.

In sum, the effects of AON2-1 and AON2-2 on splicing are concordant with each other and consistent with the results of the mutagenesis. Taken together, they

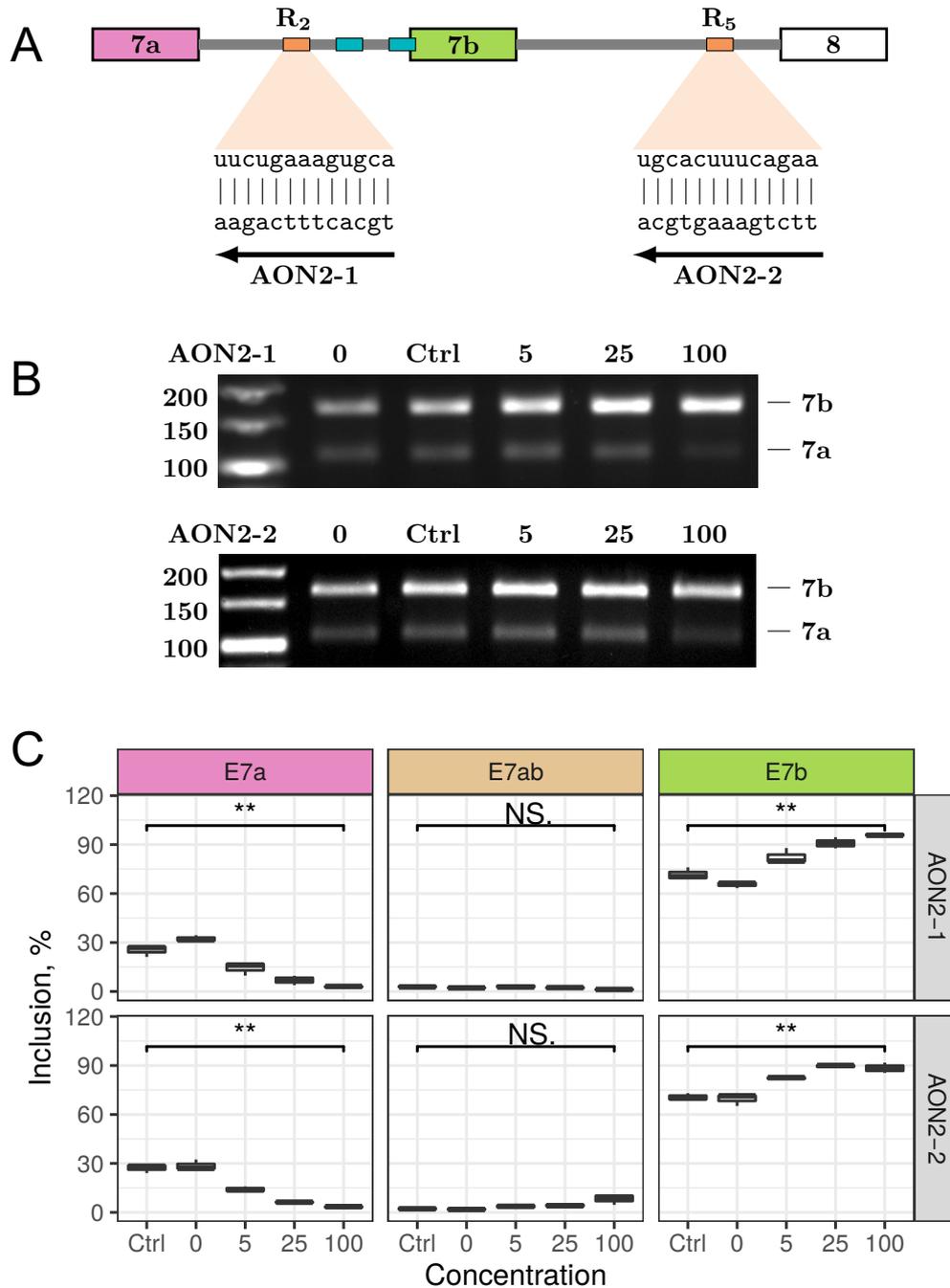


Figure 5-16: A. AON2-1 and AON2-2 disrupt R2R5 interaction. B,C. Increasing concentration of AON2-1 and AON2-2 suppresses exon 7a and promotes exon 7b usage without inducing double (7a7b) exons in the endogenous *Ate1* transcript. Inclusion rates were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***) ; not significant (NS). All experiments were performed in three bioreplicates.

indicate that the RNA structure formed by R2 and R5 is functionally distinct from

that of R1R3 and R3R4, and it serves to control the isoform ratio rather than the mutually exclusive choice of exons 7a and 7b.

5.4.6 Crosstalk between competing and ultra-long-range RNA structures

We have demonstrated that the secondary structure of *Ate1* pre-mRNA contains two distinct modules, R1R3/R3R4 and R2R5, where the former ensures mutually exclusive exon inclusion, and the latter regulates the respective isoform ratio. To investigate how these modules interact with each other, we examined the response of alternative splicing in mutants with disruptive and compensatory mutations within R1, R3, and R4 to the treatment with AON2-1, which blocks the interaction between R2 and R5.

Towards this end, we treated the minigenes carrying mutations in R1, R3, and R4 with AON2-1 and measured the changes in alternative splicing with respect to cells treated with control AON (Figure 5-17A). The effect of AON2-1 was equivalent to that of the point mutations that disrupted the interaction between R2 and R5 regardless of mutations changing base pairings within R1R3 and R3R4, i.e., it suppressed the inclusion of exon 7a and promoted the inclusion of exon 7b without affecting the proportion of double exons. This experiment demonstrated that the ultra-long-range RNA structure formed by R2 and R5 plays a dominant role in choosing between exons 7a and 7b, while not being directly responsible for their mutually exclusive choice.

To discern the interplay between R1R3/R3R4 and R2R5 in the endogenous *Ate1* transcript, we examined the effect of simultaneous disruption of base pairings with AON1 and AON2-1 on exon selection. We treated A549 cells with the combination of 25 nM AON1 and 25 nM AON2-1 and compared the splicing pattern with the effect from single AON treatment and with that of the control AON (Figure 5-17B). Remarkably, the treatment with AON1 alone increased the proportion of double exons by 28% without changing the rate of exon 7b inclusion, while the treatment

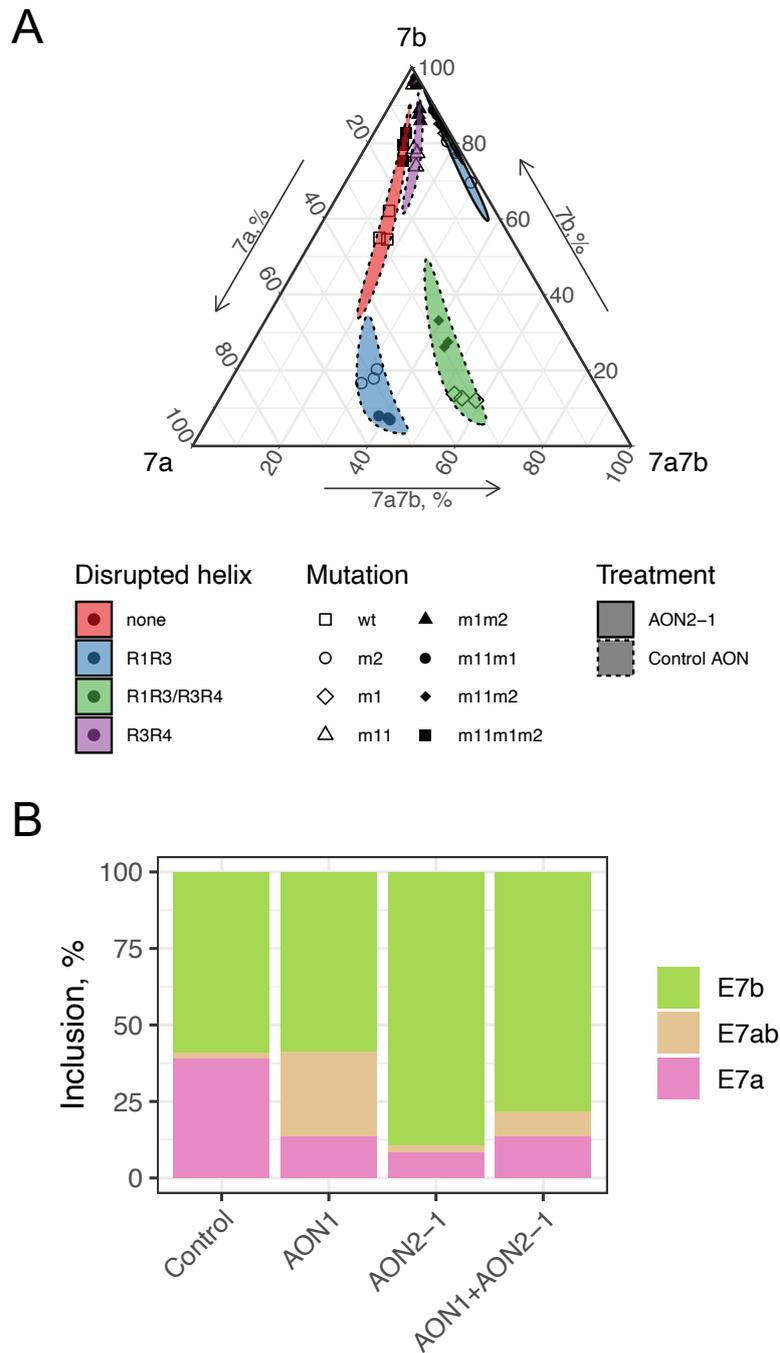


Figure 5-17: A. Treatment with AON2-1 suppresses the inclusion of exon 7a and promotes the inclusion of exon 7b without affecting the proportion of double exons regardless of mutations in R1, R3, and R4. B. The disruption of R1R3 and R3R4 base pairings in the endogenous *Ate1* with AON1 increases the proportion of double exons without affecting exon 7b. The disruption of R2R5 with AON2-1 changes exon 7a/7b ratio without inducing double exons. Simultaneous disruption of R1R3, R3R4 and R2R5 leads to an intermediate result.

with AON2-1 alone, conversely, increased the inclusion of exon 7b by 31% without introducing double exons. Simultaneous addition of AON1 and AON2-1 led to an intermediate result, in which exon 7b inclusion increased by 16% and the inclusion of double exons increased by 7%. This effect was similar to the response of m1 mutant to the treatment with AON2-1, in which the interaction between R1, R3, and R4 was disrupted by point mutations (Figure 5-15C).

These results support our hypotheses about functional distinction between two RNA structure modules in *Ate1* pre-mRNA. The module of competing base pairings (R1R3/R3R4) is responsible for mutual exclusivity of exons 7a and 7b, whereas the module of ultra-long-range base pairings (R2R5) controls the isoform balance.

Chapter 6

The impact of RNAPII slowdown on alternative splicing regulation

The transcription elongation speed strongly impacts alternative splicing [202]. It is currently accepted that slow transcription elongation opens a window of opportunity for the upstream splice sites to be recognized, which promotes the inclusion of exons that are otherwise skipped, although in some cases the effect can be the opposite [203, 204, 205]. We therefore evaluated how the transcription elongation speed influences *Ate1* splicing, as well as splicing of all introns with predicted pairs of long-range conserved complementary regions, using α -amanitin, a selective inhibitor that interacts with the core subunit of RNAPII and switches transcription to the “slow mode” [206, 207].

To assess the efficiency of elongation inhibition by α -amanitin, we used 28S rRNA/GAPDH ratio measured with qRT-PCR and additionally assessed splicing changes in the genes that are known to respond to RNAPII slowdown [208]. Indeed, upon α -amanitin treatment, the 28S rRNA/GAPDH ratio increased almost twofold (Figure 6-1A), and splicing products that were previously reported for slow RNAPII were detected (Figure 6-2).

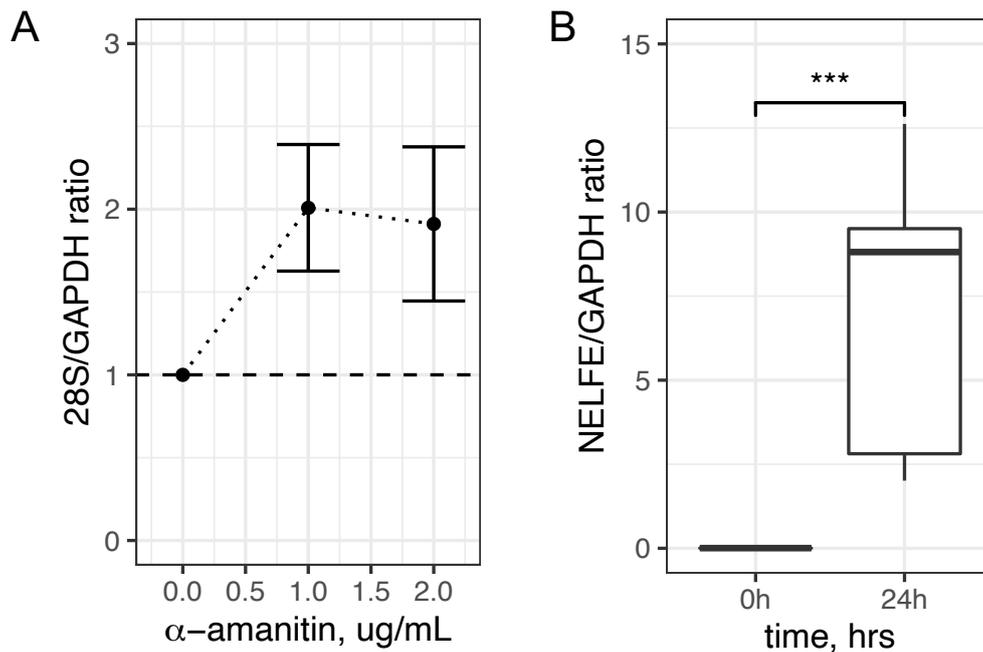
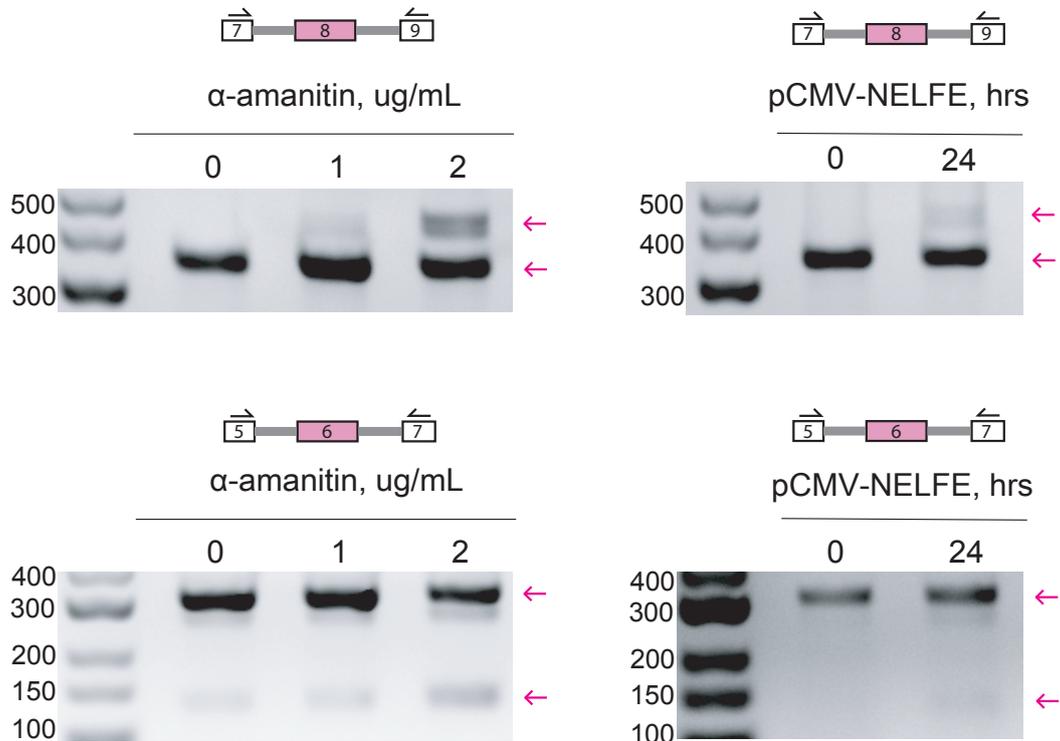


Figure 6-1: A. 28S rRNA/GAPDH ratio in samples treated with α -amanitin. B. NELFE overexpression relative to GAPDH in samples with overexpressed NELFE subunit. NELFE/GAPDH ratios were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***) ; not significant (NS). All experiments were performed in three bioreplicates.

6.1 Splicing pattern of Ate1 depends on RNAPII elongation rate

In the endogenous Ate1 transcript, the exposure to 2 $\mu\text{g/mL}$ α -amanitin led to a notable decrease of exon 7b usage, increased exon 7a usage, and slight elevation of inclusion of double exons (Figure 6-3A). This pattern was opposite to the effect of AON2-1 treatment, in which the exon 7a/7b ratio has decreased (Figure 5-16C). Consistent with this, the reanalysis of RNA-seq data for the R749H mutant from [209] confirmed that the usage of exon 7a increases, while the usage of exon 7b decreases when RNAPII is slowed down (Figure 6-4C). In the minigene construct, however, the ratio of exon 7a/7b isoforms did not change significantly (Figure 6-3B). This suggests that transcription elongation slowdown could promote the interaction of R2 with R5 by allowing sufficient time for RNA to fold, and that the absence of

HNRNPDL



HNRNPDL

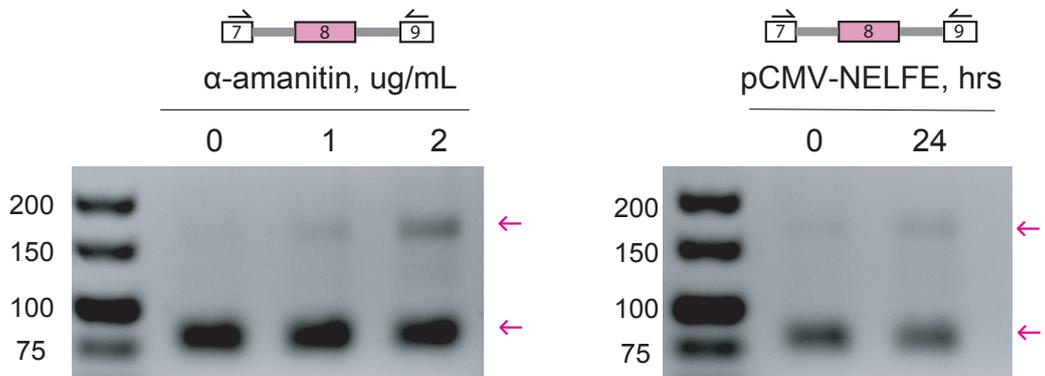


Figure 6-2: RNAPII elongation inhibition by α -amanitin (left) and NELFE over-expression (right) leads to splicing changes in *hnRNPDL* and *hnRNPDL* genes. RNAPII slowdown promotes exon 8 inclusion in *hnRNPDL* and *hnRNPDL*. The inclusion rate of exon 6 in *hnRNPDL* decreases in RNAPII slowdown.

the effect in the minigene construct could be related to the shortening of the loop between R2 and R5.

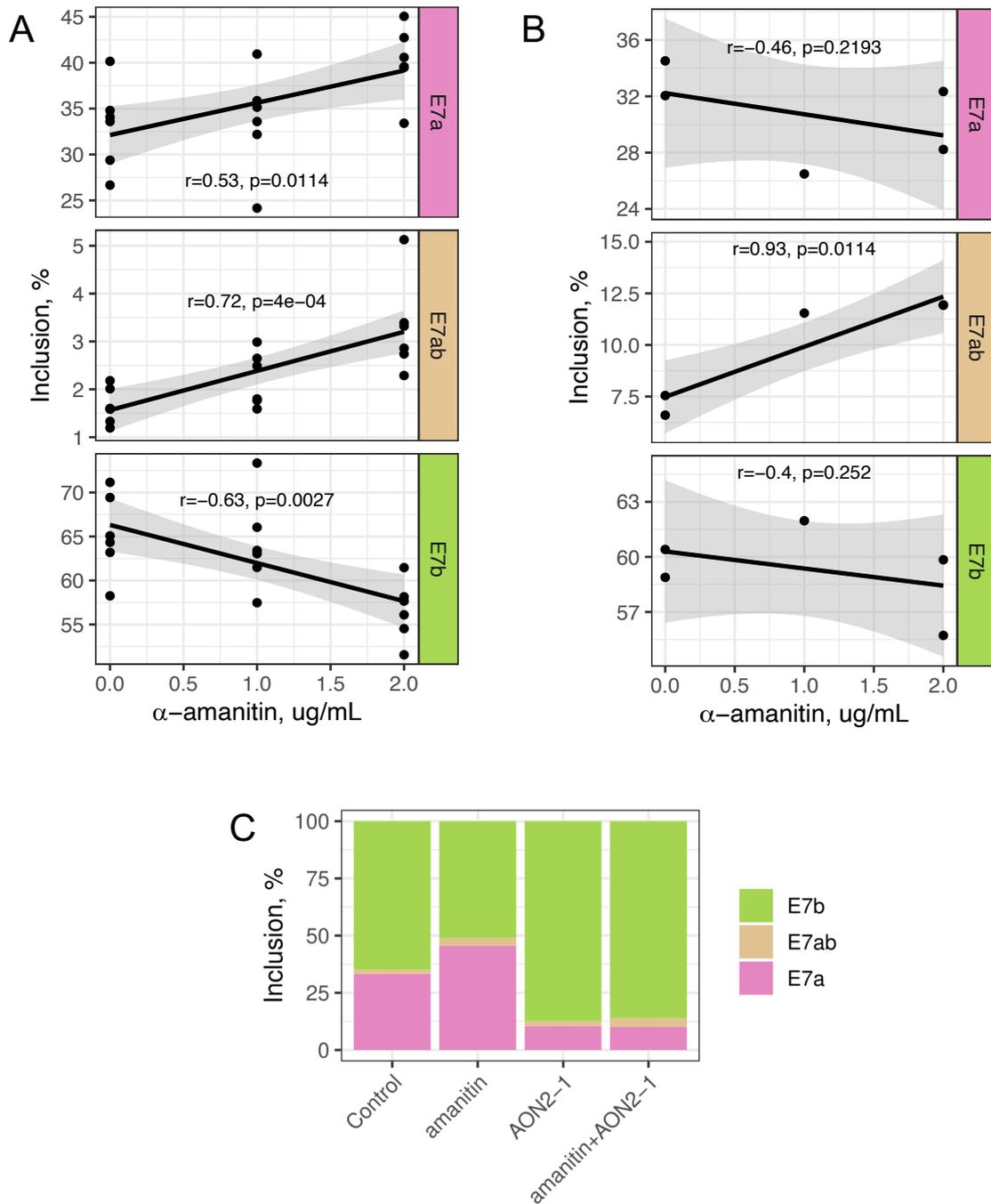


Figure 6-3: A. Endogenous *Ate1* exon inclusion in response to α -amanitin treatment. r denotes the Pearson correlation coefficient. P -values indicate the significance of the regression slope ($H_0: \beta = 0$). Note that the panels have different scales of y-axes. B. Minigene *Ate1* exon inclusion in response to α -amanitin treatment. C. Exon inclusion changes in response to the treatment with AON2-1 and α -amanitin. The addition of α -amanitin affects exon 7a/7b ratio only when R2R5 base pairing is present. The proportions are the averages of three bioreplicates.

To dissect the interplay between transcription elongation speed and RNA structure, we tested the effect of α -amanitin on the endogenous *Ate1* transcript in cells treated with AON2-1 at the concentration disrupting R2R5 base pairing. The addition of AON2-1 abolished the effect of α -amanitin, i.e., the inclusion rate of exon 7a decreased, the inclusion rate of exon 7b increased, and double exons were not affected by α -amanitin (Figure 6-3C). This indicates that the increase of exon 7a/7b ratio after α -amanitin treatment in the endogenous *Ate1* with intact R2R5 structure was not due to a longer opportunity window for the spliceosome to recognize exon 7a, but rather due to a longer time for the RNA structure to fold. We therefore conclude that the formation of ultra-long-range RNA pairing between R2 and R5 depends on transcript elongation speed, and that the impact from its slowdown on *Ate1* splicing is mediated by the ultra-long-range RNA base pairing R2R5.

One of the mechanisms of RNAPII pausing in mammalian cells involves NELF complex, which has been shown to operate not only in the promoter-proximal regions, but also to interact with the integrator complex that specifically controls NELF-mediated RNAPII pause and release in coding genes [210, 211]. The NELFE subunit, the binding of which to RNA has been linked to RNAPII pausing [212, 213, 214], is highly expressed in testis, where the inclusion level of exon 7a of *Ate1* is also the largest. Furthermore, we found that the inactivation of NELFE by shRNA in HepG2 and K562 cell lines [201] results in a significant decrease of exon 7a inclusion from 94% to 48%, and that CUGAGG, the canonical motif of NELFE in *Drosophila* [215], occurs 14 times in the intron downstream of exon 7b, while only 6.5 such occurrences would be expected by chance alone (Poisson test, $P = 0.007$). This observation suggests that NELFE-mediated RNAPII pausing could influence the ratio of exon 7a/7b isoforms by affecting cotranscriptional folding of R2R5.

To further interrogate the role of NELFE in the regulation of *Ate1* splicing, we expressed it under the CMV promoter in A549 cells and confirmed the overexpression using NELFE/GAPDH ratio measured by qRT-PCR (Figure 6-1B) and by Western blot analysis (Figure 6-4B). Then, we measured the inclusion levels of exons 7a and

7b 24 hours past CMV-NELFE transfection and found that NELFE promotes the inclusion of exon 7a and suppresses the inclusion of exon 7b (Figure 6-4A). Notably, the overexpression of NELFE also induced splicing products that were previously reported for slow RNAPII (Figure 6-2). The same pattern was observed upon α -amanitin treatment, from which we conclude that testis-specific inclusion of exon 7a could be due to slow RNAPII elongation that is caused by testis-specific expression of NELFE.

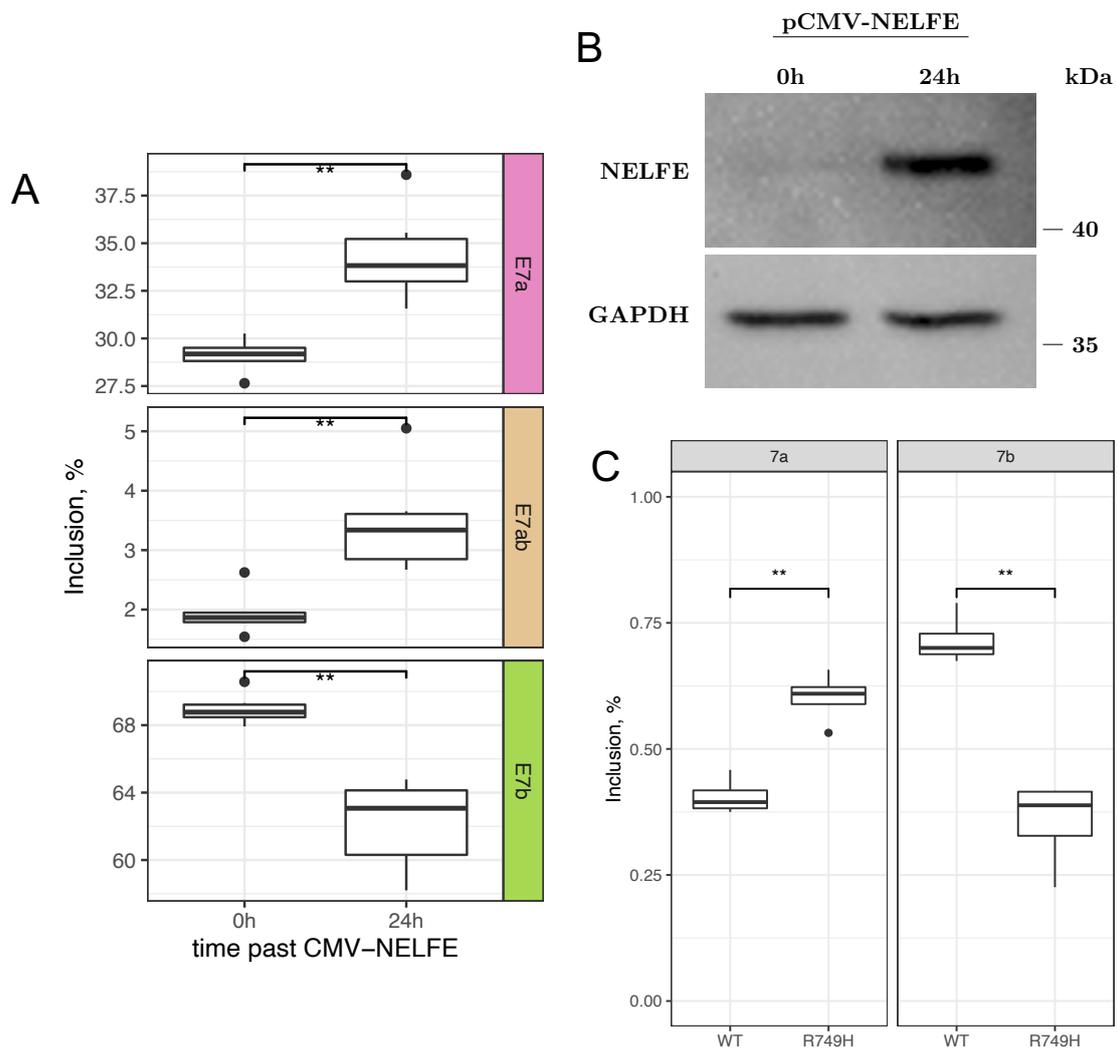


Figure 6-4: A. Exon inclusion changes 24 h after NELFE overexpression. B. Western blot confirms the overexpression of NELFE protein. C. The inclusion rate of exons 7a and 7b in slow RNAPII mutant R749H from the experiments of Fong et al [209]. Inclusion rates were compared by t-test. Asterisks indicate the range of p-values: $0.05 \leq p < 0.1$ (*); $0.01 \leq p < 0.05$ (**); $p < 0.01$ (***) ; not significant (NS). All experiments were performed in three bioreplicates.

6.2 RNAPII slowdown affects splicing of introns with long-range RNA structures

The kinetic model of co-transcriptional splicing suggests that RNAPII elongation slowdown expands the "window of opportunity" for the recognition of weak splice sites, thereby increasing the rate of inclusion of upstream exons [209, 216]. Besides this direct impact on splice site recognition, slow RNAPII elongation may also affect the way the transcript folds, which is another important determinant of how the transcript will be processed by the splicing machinery [203]. To investigate the role of long-range RNA structure in co-transcriptional splicing, we performed RNA-seq experiments, in which we used α -amanitin to slow down the RNAPII elongation speed [217], and additionally analyzed publicly available data on the impact of the RNAPII elongation speed on splicing [209].

The expected consequence of RNAPII slowdown is that the inclusion rate of exons that follow short introns will increase, and the inclusion rate of exons that follow long introns will decrease. All introns shorter than a median value (925 nt) were considered "short", all introns longer than the median were considered "long". Indeed, this trend was observed both when the RNAPII elongation speed was decreased by α -amanitin and in the slow RNAPII mutant R749H (Figure 6-5A) [209]. To check whether RNAPII slowdown differently affects introns with and without PCCRs, we matched each exon that follows an intron containing a PCCR with a randomly chosen exon that follows an intron of the same length, but without PCCRs. The difference in inclusion rates of these matched exons showed that exons that follow an intron with a PCCR tend to be more included than exons following an intron without PCCRs at both concentrations of the inhibitor and in R749H RNAPII mutant (Figure 6-5B). This can be considered as evidence for RNAPII slowdown to affect exon inclusion through pre-mRNA folding, in addition to modulation of splice site recognition. Namely, slower RNAPII elongation speed may not only facilitate the processing of upstream splice sites by the spliceosome but also allow sufficient time

for the intronic RNA structure to fold, thus promoting exon inclusion. A particular example of such a kinetic mechanism linked to RNA structure takes place in the *Ate1* gene, in which a long-range base pairing dynamically regulates the ratio of mutually exclusive exons.

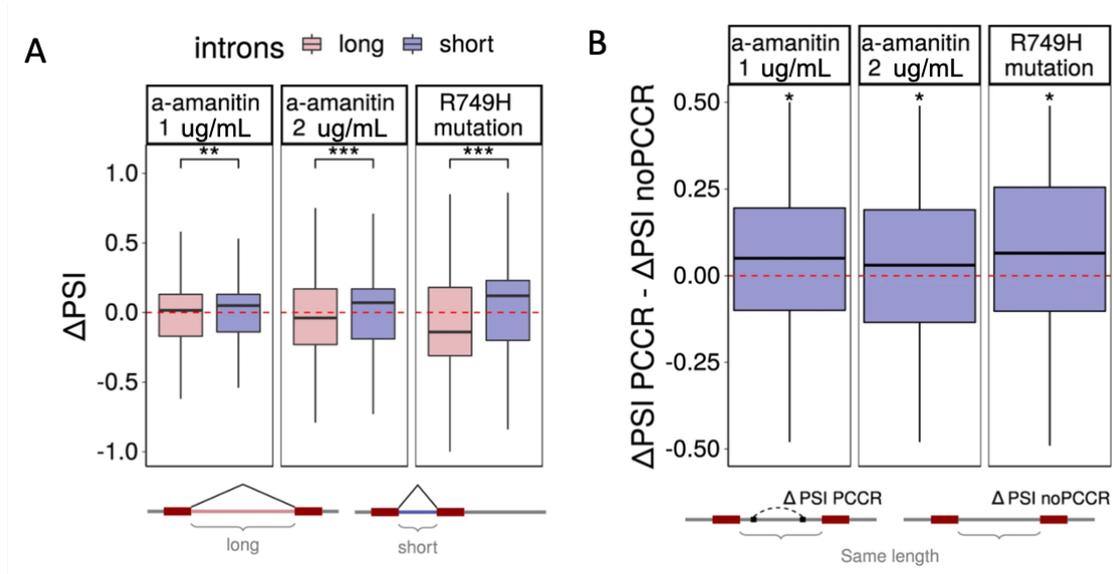


Figure 6-5: A. The change of inclusion rate of exons following short introns ($n = 2844, 2650,$ and 4032 for $1 \mu\text{g/mL}, 2 \mu\text{g/mL},$ and R749H mutant, respectively) vs. exons following long introns ($n = 2931, 2762,$ and 3807 for $1 \mu\text{g/mL}, 2 \mu\text{g/mL},$ and R749H mutant, respectively) in response to RNAPII slowdown with α -amanitin and in the slow RNAPII mutant R749H. B. The difference between the inclusion rate change of exons following introns with a PCCR and the inclusion rate change of exons following introns of the same length, but without PCCRs in response to RNAPII slowdown ($n = 191, 184,$ and 156 for $1 \mu\text{g/mL}, 2 \mu\text{g/mL},$ and R749H mutant, respectively). In all panels, boxplots are represented by the median, upper and lower quartiles, upper and lower fences without outliers; *, **, and *** denote a statistically discernible difference at the 5%, 1%, and 0.1% significance level, respectively (two-tailed Mann–Whitney and Wilcoxon’s tests, in panel B with respect to $H_0: \Delta\psi_{\text{PCCR}} - \Delta\psi_{\text{noPCCR}} = 0$).

Chapter 7

Discussion

The regulation of splicing by RNA structures has always been considered as an exceptional mechanism that works in only a few rare cases [218]. The structures in *Ate1*, *Phf20l1*, and *Cask* identified in this work add more examples to this bulk of anecdotal evidence.

7.1 Screening of the predicted RNA structures

In this study, we used a list of 916,360 pairs of conserved complementary regions (PCCRs) in introns of human protein-coding genes that were predicted by a computational pipeline for long-range RNA structures [2]. We selected several dozen structures for the experimental validation, however we could not provide experimental confirmation for all tested long-range RNA structures. This raises a question of whether some of the predicted long-range RNA structures are not functional or our experimental validation procedure fails to confirm their function. Obviously, if the target gene is not expressed in the cell line that was used for testing, or if the splicing pattern of the cloned fragment in the minigene is different from that of the endogenous gene, this does not invalidate the function of the predicted RNA structure.

The discrepancy between the predicted long-range RNA structures and their functionality in the experiment could be due to several caveats of the screening protocol. First, we used steric blocking LNA-based antisense oligonucleotides (AONs) to interfere with the predicted long-range structures and registered the changes in the splicing outcome. In some cases, we had no effect from the AON treatment, which could be interpreted as the absence of the RNA structure or lack of regulatory connection between RNA structure and splicing outcome. In principle, it is possible that the specificity of the designed AON is not sufficient to induce changes in RNA structure due to discrepancies between the genome of the cell line and the reference genome, thus resulting in a false negative result. In other cases, we observed an opposite effect from the AON treatment, which could be due to complex non-direct interactions in gene regulatory networks, or due to interference with important *cis*-regulatory elements other than RNA structure that also affect alternative splicing. In other cases, we observed a small effect, which could be due to a short dynamic range of splicing outcomes in cell lines or due to the action of a stronger regulator. Another source of discrepancies between the predicted RNA structures and the splicing outcome stems from the point mutagenesis, which in some cases inevitably affects important *cis*-regulatory elements such as the polypyrimidine tract. The point mutations that disrupt or restore RNA structure may also affect the affinity of the spliceosome binding, which confounds the splicing outcome. Finally, RT-PCR and qRT-PCR analyses can detect only splicing changes of a relatively big magnitude, i.e., the changes of exon inclusion rate below 20% changes were hard to detect using these methods. In sum, the absence of the expected response in the experiment does not ultimately invalidate the predicted long-range RNA structures, notwithstanding the fact that the estimated false discovery rate for long-range RNA structures predictions in [2] can be as large as 50%.

However, it must be noted explicitly that in selecting long-range RNA structures that are potentially associated with splicing, we had no evidence for their impact on splicing other than exon loop-out. In two cases, *Phf20l1* and *Cask*, we clearly

demonstrated that RNA structure looping-out an exon indeed promotes exon skipping, and the disruption of RNA structure leads to the increased exon inclusion rate. The functional relevance of exon skipping in *Phf20l1* and *Cask* and their possible regulators are not known and deserve further investigation. Interestingly, the exon looping-out mechanism was observed in the absence of RNA structure, where it was mediated by protein factors. For instance, hnRNP A1 [219] and PTB [220] splicing factors bound to the pre-mRNA tend to form dimers, thus placing an exon to the loop and promoting its skipping. These factors may also have other functions, e.g., hnRNP A1 also reduces U1 recruitment to 5' ss downstream of alternative exon [221], and PTB blocks an interaction between U1 and U2 snRNPs [222].

7.2 Two structural modules in *Ate1*

Mutually exclusive splicing is among the top five most abundant alternative splicing classes after exon skipping, alternative 5'- and 3'-splice site usage, and retained introns [223]. Instances of MXEs have been described in diverse phyla including *C. elegans*, *D. melanogaster*, and plants [125, 97, 199, 224]. Pre-mRNAs of many essential human genes such as glutamate receptor subunits 1-4 (GluR1-4) and voltage-gated sodium channels (*SCN* genes) undergo mutually exclusive splicing [94, 225]. MXE clusters often have tissue- and developmental stage-specific expression [97], and mutations in them have been linked to hereditary diseases [226, 227, 228, 229] and cancer [230, 231]. MXEs within a cluster often share high similarity at the sequence level indicating that some of them have emerged through tandem genomic duplications [93]. Furthermore, this mechanism could also generate competing RNA structures by duplicating one of the two arms of an ancestral stem-loop resulting in two selector sequences that compete for the same docking site [232].

Mutually exclusive splicing can be regulated by several distinct mechanisms including spliceosome incompatibility [104], steric hindrance of splice sites [224], or frame shifting coupled with degradation by nonsense-mediated decay [106], but com-

peting RNA secondary structure represents the major mechanism reported in most documented cases [199]. Here, we demonstrate for the first time an example of MXE in a human gene with two independent structural modules that have distinct functions: a competing RNA structure module (R1R3/R3R4) that controls mutually exclusive splicing, and an ultra-long-range base pairing module (R2/R5) that spans over 30 kb and regulates the ratio of transcript isoforms. The latter makes *Ate1* a gene with the longest such RNA structure known to date.

The competing base pairings in R1R3 and R3R4 regulate MXE choice by the hindrance of one of the two 3'-splice sites that mediates exposure of only one of them to the spliceosome. On one hand, the relative abundances of these two structures, and consequently the ratio of exon 7a and 7b transcript isoforms, must depend on the difference of their folding energies. On the other hand, mutations in R1 and R4 affect the polypyrimidine tracts and decrease the recognition of exon 7a and 7b by the spliceosome. In all mutants except m2 and m11, the sequences of R1 and R4 either become obstructed by RNA structure or concordantly decrease their affinity to the spliceosome. Consequently, the log-ratio of exon 7a and 7b isoforms is proportionate to the difference of R1R3 and R3R4 hybridization energies in all mutants except m2 and m11 (Figure 5-12B). This is consistent with the previous findings, in which the rate of exon inclusion inversely correlated with the thermodynamic stability of the surrounding RNA structure in *CFTR* minigene system [233], and explains incomplete reversal of the exon 7a/ 7b ratio in the triple mutant (m1m2m11) since the balance between the free energies of R1R3 and R3R4 was restored to a different level compared to the WT.

The fact that the ratio of exon 7a/7b is proportional to the difference of hybridization energies of R1R3 and R3R4 despite the former has a longer loop than the latter indicates that *Ate1* pre-mRNA is a highly folded molecule with many more complementary interactions than we have identified here. Chemical RNA structure probing methods are insufficient to determine long-range base pairings since they can only reveal which bases are single-stranded, but cannot identify the interact-

ing partners [234, 235, 236, 237, 238]. Other rapidly emerging technologies such as RNA *in situ* conformation sequencing (RIC-seq) can be used instead to profile long-range RNA structures [3, 239, 240, 241]. The published RIC-seq data confirm the presence of R1R3 and R3R4 base pairings in *Ate1* pre-mRNA expressed in HeLa cells, but they also reveal over 50 other base paired regions in exon 7a/7b cluster with potential impact on splicing [3] (Figure A-2). Therefore, the regulation of *Ate1* alternative splicing could, in fact, be much more complex than we have described here.

7.3 The effect of RNAPII slowdown on transcripts with long-range RNA structures

Extensive evidence indicates that the speed of transcription elongation may affect the choice of splice sites by changing timing in which they are presented to the spliceosome [208, 216, 205]. Slow RNA Pol II elongation generally opens a window of opportunity for weak splice sites to be recognized, leading to higher inclusion of alternative exons, although in some cases, the effect can be quite opposite [204, 205].

For instance, a DNA-binding factor CTCF can promote inclusion of weak upstream exons by mediating local RNA polymerase II pausing [242]. For the *Ate1* gene, we report for the first time that beyond co-transcriptional recognition of splice sites, slow transcription elongation also can affect splicing through RNA structure by opening a longer opportunity window for it to fold because exon 7a/7b ratio responds to the slowdown of RNA Pol II in the presence of R2R5 long-range base pairing, but fails to do so when the base pairing is disrupted. In bacterial organisms, kinetic mechanisms that involve RNA structure represent a commonly used strategy of regulating premature transcription termination, in which the structure typically senses specific molecules [243]. In eukaryotes, the kinetics of RNA folding could be influenced by many other factors including changes in the composition of RNA-binding proteins and environmental signals, thus adding to the already complex and

dynamic regulatory landscape of alternative splicing. As we show here, one of such factors is NELFE, a testis-specific subunit of NELF complex, which could promote exon 7a inclusion by inhibiting RNA Pol II elongation.

To investigate the global effect of RNAPII slowdown on transcripts with long-range RNA structures, we performed an RNA-seq experiment with RNAPII slowdown induced by α -amanitin and reanalyzed publicly available RNA-seq experiments with slow RNAPII mutant. It is known that the kinetic profile of RNA Pol II elongation has a significant impact on alternative splicing [202]. Consistent with this, we observed a global increase of inclusion of exons preceded by shorter introns under the conditions of RNAPII slowdown. For longer introns, the results were quite opposite, i.e., we observed a decreased inclusion of exons following long introns in samples with lower elongation speed. A possible explanation for these observations is that in normal RNAPII speed conditions, the spliceosome may assemble on the pre-mRNA in several ways, thus leading to alternative intron definitions, some of which result in exon skipping. In contrast, under RNAPII slowdown, the spliceosome assembly is limited to shorter intron definitions, thus resulting in decreased exon skipping rate. Simultaneously, the exons that follow longer introns will likely be skipped because the spliceosome will prefer to use an intermediate, more proximal acceptor splice site instead.

Another important conclusion that follows from the RNA-seq analysis is that structured and unstructured introns respond differently to RNAPII slowdown. In comparing introns with RNA structure to a matched sample of introns without RNA structure, we observed that exons following structured introns tend to be included more in RNAPII slowdown conditions than exons following unstructured introns. This indicates that RNAPII slowdown provides more time for RNA structure to be folded, and RNA structure makes the intron effectively shorter for the splicing machinery. The example of long-range RNA structure in the *Ate1* gene supports this observation and demonstrates that mechanisms similar to bacterial attenuation may also take place in eukaryotic cells [244, 243]. In sum, long-range RNA structure could

coordinate the interaction between spatial and temporal components of splicing regulation.

Chapter 8

Conclusion

In this study, we performed screening of long-range RNA structures with potential impact on pre-mRNA splicing and investigated the role of such structures in alternative splicing of three human genes, *Phf20l1*, *Cask*, and *Ate1*. In *Phf20l1* and *Cask*, RNA secondary structure loops out the respective alternative exons and promotes their skipping. The third gene, *Ate1*, contains two functionally distinct RNA structure modules, one of which consists of competing RNA structures and controls the mutually exclusive choice of exons 7a and 7b. The other consists of an ultra-long-range RNA structure spanning over 30,000 nts and controls the ratio of exons 7a and 7b in transcript isoforms. Additionally, we showed that transcription elongation slowdown by α -amanitin treatment or transient overexpression of NELFE increases the inclusion rate of exon 7a, but it fails to do so when the ultra-long-range RNA structure is disrupted. Using an RNA-seq experiment, we demonstrated that a similar RNA-structure-dependent response to transcription elongation slowdown takes place transcriptome-wide, namely that exons that follow structured introns tend to respond more to α -amanitin treatment than exons that follow unstructured introns. These results demonstrate that splicing is coordinated both in time and in space over very long distances, and that the interaction of these components is mediated by RNA structure.

Bibliography

- [1] M. Kalinina, D. Skvortsov, S. Kalmykova, T. Ivanov, O. Dontsova, and D. D. Pervouchine. Multiple competing RNA structures dynamically control alternative splicing in the human ATE1 gene. *Nucleic Acids Res*, 49(1):479–490, 01 2021.
- [2] S. Kalmykova, M. Kalinina, S. Denisov, A. Mironov, D. Skvortsov, R. Guigó, and D. Pervouchine. Conserved long-range base pairings are associated with pre-mRNA processing of human genes. *Nat Commun*, 12(1):2300, 04 2021.
- [3] Z. Cai, C. Cao, L. Ji, R. Ye, D. Wang, C. Xia, S. Wang, Z. Du, N. Hu, X. Yu, J. Chen, L. Wang, X. Yang, S. He, and Y. Xue. RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature*, May 2020.
- [4] N. B. Leontis and E. Westhof. Geometric nomenclature and classification of RNA base pairs. *RNA*, 7(4):499–512, Apr 2001.
- [5] Z. Miao and E. Westhof. RNA Structure: Advances and Assessment of 3D Structure Prediction. *Annu Rev Biophys*, 46:483–503, 05 2017.
- [6] K. Rietveld, R. Van Poelgeest, C. W. Pleij, J. H. Van Boom, and L. Bosch. The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res*, 10(6):1929–1946, Mar 1982.
- [7] M. Laughrea and L. Jetté. A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry*, 33(45):13464–13474, Nov 1994.
- [8] Y. d'Aubenton Carafa, E. Brody, and C. Thermes. Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. *J Mol Biol*, 216(4):835–858, Dec 1990.
- [9] A. Ray-Soni, M. J. Bellecourt, and R. Landick. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem*, 85:319–347, Jun 2016.
- [10] R. Reynolds and M. J. Chamberlin. Parameters affecting transcription termination by Escherichia coli RNA. II. Construction and analysis of hybrid terminators. *J Mol Biol*, 224(1):53–63, Mar 1992.

- [11] J. C. McDowell, J. W. Roberts, D. J. Jin, and C. Gross. Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. *Science*, 266(5186):822–825, Nov 1994.
- [12] H. Abe and H. Aiba. Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabilization. *Biochimie*, 78(11-12):1035–1042, 1996.
- [13] C. Y. Chen and J. P. Richardson. Sequence elements essential for rho-dependent transcription termination at lambda tR1. *J Biol Chem*, 262(23):11292–11299, Aug 1987.
- [14] J. Rossi, J. Egan, L. Hudson, and A. Landy. The tyrT locus: termination and processing of a complex transcript. *Cell*, 26(3 Pt 1):305–314, Nov 1981.
- [15] A. M. Wu, G. E. Christie, and T. Platt. Tandem termination sites in the tryptophan operon of Escherichia coli. *Proc Natl Acad Sci U S A*, 78(5):2913–2917, May 1981.
- [16] M. H. de Smit and J. van Duin. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci U S A*, 87(19):7668–7672, Oct 1990.
- [17] R. R. Breaker. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol*, 4(2), Feb 2012.
- [18] M. Kozak. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361:13–37, Nov 2005.
- [19] D. M. Mauger, B. J. Cabral, V. Presnyak, S. V. Su, D. W. Reid, B. Goodman, K. Link, N. Khatwani, J. Reynders, M. J. Moore, and I. J. McFadyen. mRNA structure regulates protein expression through changes in functional half-life. *Proc Natl Acad Sci U S A*, 116(48):24075–24083, 11 2019.
- [20] M. U. Muckenthaler, B. Galy, and M. W. Hentze. Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network. *Annu Rev Nutr*, 28:197–213, 2008.
- [21] P. S. Ray, J. Jia, P. Yao, M. Majumder, M. Hatzoglou, and P. L. Fox. A stress-responsive RNA switch regulates VEGFA expression. *Nature*, 457(7231):915–919, Feb 2009.
- [22] X. Wu and D. P. Bartel. Widespread Influence of 3'-End Structures on Mammalian mRNA Processing and Stability. *Cell*, 169(5):905–917, May 2017.
- [23] J. D. Dinman. Programmed Ribosomal Frameshifting Goes Beyond Viruses: Organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift. *Microbe Wash DC*, 1(11):521–527, 11 2006.

- [24] M. K. Sakharkar, V. T. Chow, and P. Kanguane. Distributions of exons and introns in the human genome. *In Silico Biol*, 4(4):387–393, 2004.
- [25] E. A. Grzybowska. Human intronless genes: functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. *Biochem Biophys Res Commun*, 424(1):1–6, Jul 2012.
- [26] R. Sperling. The nuts and bolts of the endogenous spliceosome. *Wiley Interdiscip Rev RNA*, 8(1), 01 2017.
- [27] A. G. Matera and Z. Wang. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*, 15(2):108–121, Feb 2014.
- [28] J. J. Turunen, E. H. Niemelä, B. Verma, and M. J. Frilander. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA*, 4(1):61–76, 2013.
- [29] M. K. Basu, W. Makalowski, I. B. Rogozin, and E. V. Koonin. U12 intron positions are more strongly conserved between animals and plants than U2 intron positions. *Biol Direct*, 3:19, May 2008.
- [30] K. Doggett, B. B. Williams, S. Markmiller, F. S. Geng, J. Coates, S. Mieruszynski, M. Ernst, T. Thomas, and J. K. Heath. -deficient mice. *RNA*, 24(12):1856–1870, 12 2018.
- [31] B. Verma, M. V. Akinyi, A. J. Norppa, and M. J. Frilander. Minor spliceosome and disease. *Semin Cell Dev Biol*, 79:103–112, 07 2018.
- [32] M. Chen and J. L. Manley. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*, 10(11):741–754, Nov 2009.
- [33] Y. Lee and D. C. Rio. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*, 84:291–323, 2015.
- [34] T. W. Nilsen and B. R. Graveley. Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280):457–463, Jan 2010.
- [35] J. D. Ellis, M. Barrios-Rodiles, R. Colak, M. Irimia, T. Kim, J. A. Calarco, X. Wang, Q. Pan, D. O’Hanlon, P. M. Kim, J. L. Wrana, and B. J. Blencowe. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell*, 46(6):884–892, Jun 2012.
- [36] X. Yang, J. Coulombe-Huntington, S. Kang, G. M. Sheynkman, T. Hao, A. Richardson, S. Sun, F. Yang, Y. A. Shen, R. R. Murray, K. Spirohn, B. E. Begg, M. Duran-Frigola, A. MacWilliams, S. J. Pevzner, Q. Zhong, S. A. Trigg, S. Tam, L. Ghamsari, N. Sahni, S. Yi, M. D. Rodriguez, D. Balcha, G. Tan, M. Costanzo, B. Andrews, C. Boone, X. J. Zhou, K. Salehi-Ashtiani, B. Charloteaux, A. A. Chen, M. A. Calderwood, P. Aloy, F. P. Roth, D. E.

- Hill, L. M. Iakoucheva, Y. Xia, and M. Vidal. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell*, 164(4):805–817, Feb 2016.
- [37] U. Braunschweig, S. Gueroussov, A. M. Plocik, B. R. Graveley, and B. J. Blencowe. Dynamic integration of splicing within gene regulatory pathways. *Cell*, 152(6):1252–1269, Mar 2013.
- [38] Q. Xu, B. Modrek, and C. Lee. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res*, 30(17):3754–3766, Sep 2002.
- [39] A. Kalsotra and T. A. Cooper. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet*, 12(10):715–729, Sep 2011.
- [40] Y. Wang, J. Liu, B. O. Huang, Y. M. Xu, J. Li, L. F. Huang, J. Lin, J. Zhang, Q. H. Min, W. M. Yang, and X. Z. Wang. Mechanism of alternative splicing and its regulation. *Biomed Rep*, 3(2):152–158, Mar 2015.
- [41] X. D. Fu and M. Ares. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet*, 15(10):689–701, Oct 2014.
- [42] H. Dvinge, E. Kim, O. Abdel-Wahab, and R. K. Bradley. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer*, 16(7):413–430, 07 2016.
- [43] A. R. Kornblihtt, I. E. Schor, M. Alló, G. Dujardin, E. Petrillo, and M. J. Muñoz. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol*, 14(3):153–165, Mar 2013.
- [44] P. Papasaikas, J. R. Tejedor, L. Vigevani, and J. Valcárcel. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol Cell*, 57(1):7–22, Jan 2015.
- [45] H. Han, U. Braunschweig, T. Gonatopoulos-Pournatzis, R. J. Weatheritt, C. L. Hirsch, K. C. H. Ha, E. Radovani, S. Nabeel-Shah, T. Sterne-Weiler, J. Wang, D. O’Hanlon, Q. Pan, D. Ray, H. Zheng, F. Vizeacoumar, A. Datti, L. Magomedova, C. L. Cummins, T. R. Hughes, J. F. Greenblatt, J. L. Wrana, J. Moffat, and B. J. Blencowe. Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Mol Cell*, 65(3):539–553, Feb 2017.
- [46] E. Agirre, A. J. Oldfield, N. Bellora, A. Segelle, and R. F. Luco. Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. *Nat Commun*, 12(1):682, 01 2021.
- [47] R. F. Luco, M. Allo, I. E. Schor, A. R. Kornblihtt, and T. Misteli. Epigenetics in alternative pre-mRNA splicing. *Cell*, 144(1):16–26, Jan 2011.

- [48] H. Tilgner, C. Nikolaou, S. Althammer, M. Sammeth, M. Beato, J. Valcárcel, and R. Guigó. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, 16(9):996–1001, Sep 2009.
- [49] R. Andersson, S. Enroth, A. Rada-Iglesias, C. Wadelius, and J. Komorowski. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res*, 19(10):1732–1741, Oct 2009.
- [50] T. V. Ramanouskaya and V. V. Grinev. The determinants of alternative RNA splicing in human cells. *Mol Genet Genomics*, 292(6):1175–1195, Dec 2017.
- [51] S. Adhikari, W. Xiao, Y. L. Zhao, and Y. G. Yang. m(6)A: Signaling for mRNA splicing. *RNA Biol*, 13(9):756–759, 09 2016.
- [52] K. I. Zhou, H. Shi, R. Lyu, A. C. Wylder, Ż. Matuszek, J. N. Pan, C. He, M. Parisien, and T. Pan. A through the Low-Complexity Protein hnRNPG. *Mol Cell*, 76(1):70–81, 10 2019.
- [53] Y. Wang, Y. Li, J. I. Toth, M. D. Petroski, Z. Zhang, and J. C. Zhao. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol*, 16(2):191–198, Feb 2014.
- [54] F. P. Lemaigre, S. J. Courtois, D. A. Lafontaine, and G. G. Rousseau. Evidence that the upstream stimulatory factor and the Sp1 transcription factor bind in vitro to the promoter of the human-growth-hormone gene. *Eur J Biochem*, 181(3):555–561, May 1989.
- [55] Y. E. Hsiao, J. H. Bahn, Y. Yang, X. Lin, S. Tran, E. W. Yang, G. Quinones-Valdez, and X. Xiao. RNA editing in nascent RNA affects pre-mRNA splicing. *Genome Res*, 28(6):812–823, 06 2018.
- [56] S. M. Rueter, T. R. Dawson, and R. B. Emeson. Regulation of alternative splicing by RNA editing. *Nature*, 399(6731):75–80, May 1999.
- [57] A. F. Muro, M. Caputi, R. Pariyarath, F. Pagani, E. Buratti, and F. E. Baralle. Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol Cell Biol*, 19(4):2657–2671, Apr 1999.
- [58] E. Buratti, A. F. Muro, M. Giombi, D. Gherbassi, A. Iaconcig, and F. E. Baralle. RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol Cell Biol*, 24(3):1387–1400, Feb 2004.
- [59] M. B. Warf and J. A. Berglund. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA*, 13(12):2238–2251, Dec 2007.
- [60] M. B. Warf, J. V. Diegel, P. H. von Hippel, and J. A. Berglund. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc Natl Acad Sci U S A*, 106(23):9203–9208, Jun 2009.

- [61] W. Qian and F. Liu. Regulation of alternative splicing of tau exon 10. *Neurosci Bull*, 30(2):367–377, Apr 2014.
- [62] J. Lisowiec, D. Magner, E. Kierzek, E. Lenartowicz, and R. Kierzek. Structural determinants for alternative splicing regulation of the MAPT pre-mRNA. *RNA Biol*, 12(3):330–342, 2015.
- [63] A. Kar, K. Fushimi, X. Zhou, P. Ray, C. Shi, X. Chen, Z. Liu, S. Chen, and J. Y. Wu. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol*, 31(9):1812–1821, May 2011.
- [64] P. Ray, A. Kar, K. Fushimi, N. Havlioglu, X. Chen, and J. Y. Wu. PSF suppresses tau exon 10 inclusion by interacting with a stem-loop structure downstream of exon 10. *J Mol Neurosci*, 45(3):453–466, Nov 2011.
- [65] S. Tisdale and L. Pellizzoni. Disease mechanisms and therapeutic approaches in spinal muscular atrophy. *J Neurosci*, 35(23):8691–8700, Jun 2015.
- [66] N. N. Singh, E. J. Androphy, and R. N. Singh. In vivo selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes. *RNA*, 10(8):1291–1305, Aug 2004.
- [67] N. N. Singh, R. N. Singh, and E. J. Androphy. Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res*, 35(2):371–389, 2007.
- [68] A. Garcia-Lopez, F. Tessaro, H. R. A. Jonker, A. Wacker, C. Richter, A. Comte, N. Berntenis, R. Schmucki, K. Hatje, O. Petermann, G. Chiriano, R. Perozzo, D. Sciarra, P. Konieczny, I. Faustino, G. Fournet, M. Orozco, R. Artero, F. Metzger, M. Ebeling, P. Goekjian, B. Joseph, H. Schwalbe, and L. Scapozza. Targeting RNA structure in SMN2 reverses spinal muscular atrophy molecular phenotypes. *Nat Commun*, 9(1):2032, 05 2018.
- [69] N. N. Singh, B. M. Lee, and R. N. Singh. Splicing regulation in spinal muscular atrophy by an RNA structure formed by long-distance interactions. *Ann N Y Acad Sci*, 1341:176–187, Apr 2015.
- [70] S. Messina and M. Sframeli. New Treatments in Spinal Muscular Atrophy: Positive Results and New Challenges. *J Clin Med*, 9(7), Jul 2020.
- [71] D. Rhodes and H. J. Lipps. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res*, 43(18):8627–8637, Oct 2015.
- [72] D. Gomez, T. Lemarteleur, L. Lacroix, P. Mailliet, J. L. Mergny, and J. F. Riou. Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res*, 32(1):371–379, 2004.

- [73] H. Huang, J. Zhang, S. E. Harvey, X. Hu, and C. Cheng. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev*, 31(22):2296–2309, 11 2017.
- [74] C. Weldon, J. G. Dacanay, V. Gokhale, P. V. L. Boddupally, I. Behm-Ansmant, G. A. Burley, C. Branlant, L. H. Hurley, C. Dominguez, and I. C. Eperon. Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Res*, 46(2):886–896, 01 2018.
- [75] E. G. Conlon, L. Lu, A. Sharma, T. Yamazaki, T. Tang, N. A. Shneider, and J. L. Manley. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife*, 5, 09 2016.
- [76] A. von Hacht, O. Seifert, M. Menger, T. Schütze, A. Arora, Z. Konthur, P. Neubauer, A. Wagner, C. Weise, and J. Kurreck. Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res*, 42(10):6630–6644, Jun 2014.
- [77] M. S. Wong, W. E. Wright, and J. W. Shay. Alternative splicing regulation of telomerase: a new paradigm? *Trends Genet*, 30(10):430–438, Oct 2014.
- [78] G. Li, J. Shen, J. Cao, G. Zhou, T. Lei, Y. Sun, H. Gao, Y. Ding, W. Xu, Z. Zhan, Y. Chen, and H. Huang. Alternative splicing of human telomerase reverse transcriptase in gliomas and its modulation mediated by CX-5461. *J Exp Clin Cancer Res*, 37(1):78, Apr 2018.
- [79] S. Courtois, G. Verhaegh, S. North, M. G. Luciani, P. Lassus, U. Hibner, M. Oren, and P. Hainaut. DeltaN-p53, a natural isoform of p53 lacking the first transactivation domain, counteracts growth suppression by wild-type p53. *Oncogene*, 21(44):6722–6728, Oct 2002.
- [80] V. Marcel, P. L. Tran, C. Sagne, G. Martel-Planche, L. Vaslin, M. P. Teulade-Fichou, J. Hall, J. L. Mergny, P. Hainaut, and E. Van Dyck. G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis*, 32(3):271–278, Mar 2011.
- [81] A. Serganov and E. Nudler. A decade of riboswitches. *Cell*, 152(1-2):17–24, Jan 2013.
- [82] M. T. Cheah, A. Wachter, N. Sudarsan, and R. R. Breaker. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447(7143):497–500, May 2007.
- [83] M. T. Croft, M. Moulin, M. E. Webb, and A. G. Smith. Thiamine biosynthesis in algae is regulated by riboswitches. *Proc Natl Acad Sci U S A*, 104(52):20770–20775, Dec 2007.
- [84] S. Li and R. R. Breaker. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Res*, 41(5):3022–3031, Mar 2013.

- [85] V. A. Raker, A. A. Mironov, M. S. Gelfand, and D. D. Pervouchine. Modulation of alternative splicing by long-range RNA structures in *Drosophila*. *Nucleic Acids Res*, 37(14):4533–4544, Aug 2009.
- [86] D. D. Pervouchine, E. E. Khrameeva, M. Y. Pichugina, O. V. Nikolaienko, M. S. Gelfand, P. M. Rubtsov, and A. A. Mironov. Evidence for widespread association of mammalian splicing and conserved long-range RNA structures. *RNA*, 18(1):1–15, Jan 2012.
- [87] J. R. Taube, K. Sperle, L. Banser, P. Seeman, B. C. Cavan, J. Y. Garbern, and G. M. Hobson. PMD patient mutations reveal a long-distance intronic interaction that regulates PLP1/DM20 alternative splicing. *Hum Mol Genet*, 23(20):5464–5478, Oct 2014.
- [88] N. Guo and S. Kawamoto. An intronic downstream enhancer promotes 3' splice site usage of a neural cell-specific exon. *J Biol Chem*, 275(43):33641–33649, Oct 2000.
- [89] A. P. Baraniak, J. R. Chen, and M. A. Garcia-Blanco. Fox-2 mediates epithelial cell-specific fibroblast growth factor receptor 2 exon choice. *Mol Cell Biol*, 26(4):1209–1222, Feb 2006.
- [90] M. T. Lovci, D. Ghanem, H. Marr, J. Arnold, S. Gee, M. Parra, T. Y. Liang, T. J. Stark, L. T. Gehman, S. Hoon, K. B. Massirer, G. A. Pratt, D. L. Black, J. W. Gray, J. G. Conboy, and G. W. Yeo. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol*, 20(12):1434–1442, Dec 2013.
- [91] M. S. Wong, J. W. Shay, and W. E. Wright. Regulation of human telomerase splicing by RNA:RNA pairing. *Nat Commun*, 5:3306, Feb 2014.
- [92] N. N. Singh, M. N. Lawler, E. W. Ottesen, D. Upreti, J. R. Kaczynski, and R. N. Singh. An intronic structure enabled by a long-distance interaction serves as a novel target for splicing correction in spinal muscular atrophy. *Nucleic Acids Res*, 41(17):8144–8165, Sep 2013.
- [93] F. A. Kondrashov and E. V. Koonin. Origin of alternative splicing by tandem exon duplication. *Hum Mol Genet*, 10(23):2661–2669, Nov 2001.
- [94] R. R. Copley. Evolutionary convergence of alternative splicing in ion channels. *Trends Genet*, 20(4):171–176, Apr 2004.
- [95] K. Hatje and M. Kollmar. Expansion of the mutually exclusive spliced exome in *Drosophila*. *Nat Commun*, 4:2460, 2013.
- [96] B. R. Graveley, A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J. van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas, C. A. Davis, A. Dobin, R. Li, W. Lin, J. H. Malone, N. R. Mattiuzzo, D. Miller, D. Sturgill, B. B. Tuch, C. Zaleski, D. Zhang, M. Blanchette, S. Dudoit, B. Eads, R. E. Green, A. Hammonds, L. Jiang,

- P. Kapranov, L. Langton, N. Perrimon, J. E. Sandler, K. H. Wan, A. Willingham, Y. Zhang, Y. Zou, J. Andrews, P. J. Bickel, S. E. Brenner, M. R. Brent, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, B. Oliver, and S. E. Celniker. The developmental transcriptome of *Drosophila melanogaster*. *Nature*, 471(7339):473–479, Mar 2011.
- [97] K. Hatje, R. U. Rahman, R. O. Vidal, D. Simm, B. Hammesfahr, V. Bansal, A. Rajput, M. E. Mickael, T. Sun, S. Bonn, and M. Kollmar. The landscape of human mutually exclusive splicing. *Mol Syst Biol*, 13(12):959, 12 2017.
- [98] C. F. Kennedy and S. M. Berget. Pyrimidine tracts between the 5' splice site and branch point facilitate splicing and recognition of a small *Drosophila* intron. *Mol Cell Biol*, 17(5):2774–2780, May 1997.
- [99] B. Ruskin, J. M. Greene, and M. R. Green. Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell*, 41(3):833–844, Jul 1985.
- [100] C. W. Smith. Alternative splicing—when two's a crowd. *Cell*, 123(1):1–3, Oct 2005.
- [101] C. W. Smith and B. Nadal-Ginard. Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell*, 56(5):749–758, Mar 1989.
- [102] Y. Yue, G. Li, Y. Yang, W. Zhang, H. Pan, R. Chen, F. Shi, and Y. Jin. Regulation of Dscam exon 17 alternative splicing by steric hindrance in combination with RNA secondary structures. *RNA Biol*, 10(12):1822–1833, Dec 2013.
- [103] I. Letunic, R. R. Copley, and P. Bork. Common exon duplication in animals and its role in alternative splicing. *Hum Mol Genet*, 11(13):1561–1567, Jun 2002.
- [104] P. A. Sharp and C. B. Burge. Classification of introns: U2-type or U12-type. *Cell*, 91(7):875–879, Dec 1997.
- [105] N. Hug, D. Longman, and J. F. Cáceres. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res*, 44(4):1483–1495, Feb 2016.
- [106] R. B. Jones, F. Wang, Y. Luo, C. Yu, C. Jin, T. Suzuki, M. Kan, and W. L. McKeehan. The nonsense-mediated decay pathway and mutually exclusive expression of alternatively spliced FGFR2IIIb and -IIIc mRNAs. *J Biol Chem*, 276(6):4158–4167, Feb 2001.
- [107] Z. Z. Tang, S. Sharma, S. Zheng, G. Chawla, J. Nikolic, and D. L. Black. Regulation of the mutually exclusive exons 8a and 8 in the CaV1.2 calcium channel transcript by polypyrimidine tract-binding protein. *J Biol Chem*, 286(12):10007–10016, Mar 2011.

- [108] D. Schmucker, J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon, and S. L. Zipursky. *Drosophila Dscam* is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, 101(6):671–684, Jun 2000.
- [109] L. Shi, H. H. Yu, J. S. Yang, and T. Lee. Specific *Drosophila Dscam* juxtamembrane variants control dendritic elaboration and axonal arborization. *J Neurosci*, 27(25):6723–6728, Jun 2007.
- [110] G. Neves, J. Zucker, M. Daly, and A. Chess. Stochastic yet biased expression of multiple *Dscam* splice variants by individual cells. *Nat Genet*, 36(3):240–246, Mar 2004.
- [111] S. K. Miura, A. Martins, K. X. Zhang, B. R. Graveley, and S. L. Zipursky. Probabilistic splicing of *Dscam1* establishes identity at the level of single neurons. *Cell*, 155(5):1166–1177, Nov 2013.
- [112] W. Sun, X. You, A. Gogol-Döring, H. He, Y. Kise, M. Sohn, T. Chen, A. Klebes, D. Schmucker, and W. Chen. Ultra-deep profiling of alternatively spliced *Drosophila Dscam* isoforms by circularization-assisted multi-segment sequencing. *EMBO J*, 32(14):2029–2038, Jul 2013.
- [113] B. J. Matthews, M. E. Kim, J. J. Flanagan, D. Hattori, J. C. Clemens, S. L. Zipursky, and W. B. Grueber. Dendrite self-avoidance is controlled by *Dscam*. *Cell*, 129(3):593–604, May 2007.
- [114] B. R. Graveley. Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures. *Cell*, 123(1):65–73, Oct 2005.
- [115] B. Xu, Y. Shi, Y. Wu, Y. Meng, and Y. Jin. Role of RNA secondary structures in regulating *Dscam* alternative splicing. *Biochim Biophys Acta Gene Regul Mech*, 1862(11-12):194381, 2019.
- [116] Y. Yang, L. Zhan, W. Zhang, F. Sun, W. Wang, N. Tian, J. Bi, H. Wang, D. Shi, Y. Jiang, Y. Zhang, and Y. Jin. RNA secondary structure in mutually exclusive splicing. *Nat Struct Mol Biol*, 18(2):159–168, Feb 2011.
- [117] J. M. Kreamling and B. R. Graveley. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila Dscam* pre-mRNA. *Mol Cell Biol*, 25(23):10251–10260, Dec 2005.
- [118] X. Wang, G. Li, Y. Yang, W. Wang, W. Zhang, H. Pan, P. Zhang, Y. Yue, H. Lin, B. Liu, J. Bi, F. Shi, J. Mao, Y. Meng, L. Zhan, and Y. Jin. An RNA architectural locus control region involved in *Dscam* mutually exclusive splicing. *Nat Commun*, 3:1255, 2012.
- [119] S. Olson, M. Blanchette, J. Park, Y. Savva, G. W. Yeo, J. M. Yeakley, D. C. Rio, and B. R. Graveley. A regulator of *Dscam* mutually exclusive splicing fidelity. *Nat Struct Mol Biol*, 14(12):1134–1140, Dec 2007.

- [120] N. Philip, S. F. Acevedo, and E. M. Skoulakis. Conditional rescue of olfactory learning and memory defects in mutants of the 14-3-3zeta gene leonardo. *J Neurosci*, 21(21):8417–8425, Nov 2001.
- [121] U. M. Petersen, L. Kadalayil, K. P. Rehorn, D. K. Hoshizaki, R. Reuter, and Y. Engström. Serpent regulates *Drosophila* immunity genes in the larval fat body through an essential GATA motif. *EMBO J*, 18(14):4013–4022, Jul 1999.
- [122] K. Senger, K. Harris, and M. Levine. GATA factors participate in tissue-specific immune responses in *Drosophila* larvae. *Proc Natl Acad Sci U S A*, 103(43):15957–15962, Oct 2006.
- [123] Y. Yue, Y. Yang, L. Dai, G. Cao, R. Chen, W. Hong, B. Liu, Y. Shi, Y. Meng, F. Shi, M. Xiao, and Y. Jin. Long-range RNA pairings contribute to mutually exclusive splicing. *RNA*, 22(1):96–110, Jan 2016.
- [124] S. J. Lansdell, T. Collins, A. Yabe, V. J. Gee, A. J. Gibb, and N. S. Millar. Host-cell specific effects of the nicotinic acetylcholine receptor chaperone RIC-3 revealed by a comparison of human and *Drosophila* RIC-3 homologues. *J Neurochem*, 105(5):1573–1581, Jun 2008.
- [125] Y. Yue, S. Hou, X. Wang, L. Zhan, G. Cao, G. Li, Y. Shi, P. Zhang, W. Hong, H. Lin, B. Liu, F. Shi, Y. Yang, and Y. Jin. Role and convergent evolution of competing RNA secondary structures in mutually exclusive splicing. *RNA Biol*, 14(10):1399–1410, 10 2017.
- [126] J. Royet and R. Dziarski. Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. *Nat Rev Microbiol*, 5(4):264–277, Apr 2007.
- [127] K. M. Choe, H. Lee, and K. V. Anderson. *Drosophila* peptidoglycan recognition protein LC (PGRP-LC) acts as a signal-transducing innate immune receptor. *Proc Natl Acad Sci U S A*, 102(4):1122–1126, Jan 2005.
- [128] H. Pan, Y. Shi, S. Chen, Y. Yang, Y. Yue, L. Zhan, L. Dai, H. Dong, W. Hong, F. Shi, and Y. Jin. Competing RNA pairings in complex alternative splicing of a 3' variable region. *RNA*, 24(11):1466–1480, 11 2018.
- [129] M. Kubota, C. Tran, and R. C. Spitale. Progress and challenges for chemical probing of RNA structure inside living cells. *Nat Chem Biol*, 11(12):933–941, Dec 2015.
- [130] P. C. Bevilacqua, L. E. Ritchey, Z. Su, and S. M. Assmann. Genome-Wide Analysis of RNA Secondary Structure. *Annu Rev Genet*, 50:235–266, Nov 2016.
- [131] R. Lorenz, M. T. Wolfinger, A. Tanzer, and I. L. Hofacker. Predicting RNA secondary structures from sequence and probing data. *Methods*, 103:86–98, 07 2016.

- [132] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, Feb 2002.
- [133] A. Helwak, G. Kudla, T. Dudnakova, and D. Tollervey. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153(3):654–665, Apr 2013.
- [134] E. Sharma, T. Sterne-Weiler, D. O’Hanlon, and B. J. Blencowe. Global Mapping of Human RNA-RNA Interactions. *Mol Cell*, 62(4):618–626, 05 2016.
- [135] E. Rivas and S. R. Eddy. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol*, 285(5):2053–2068, Feb 1999.
- [136] D. D. Pervouchine. Towards Long-Range RNA Structure Prediction in Eukaryotic Genes. *Genes (Basel)*, 9(6), Jun 2018.
- [137] R. D. Dowell and S. R. Eddy. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics*, 5:71, Jun 2004.
- [138] E. Rivas, J. Clements, and S. R. Eddy. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods*, 14(1):45–48, 01 2017.
- [139] A. Tuplin, J. Wood, D. J. Evans, A. H. Patel, and P. Simmonds. Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA*, 8(6):824–841, Jun 2002.
- [140] M. Suyama. Mechanistic insights into mutually exclusive splicing in dynamin 1. *Bioinformatics*, 29(17):2084–2087, Sep 2013.
- [141] D. D. Pervouchine. IRBIS: a systematic search for conserved complementarity. *RNA*, 20(10):1519–1531, Oct 2014.
- [142] D. Karolchik, R. Baertsch, M. Diekhans, T. S. Furey, A. Hinrichs, Y. T. Lu, K. M. Roskin, M. Schwartz, C. W. Sugnet, D. J. Thomas, R. J. Weber, D. Haussler, and W. J. Kent. The UCSC Genome Browser Database. *Nucleic Acids Res*, 31(1):51–54, Jan 2003.
- [143] F. van den Ent and J. Löwe. RF cloning: a restriction-free method for inserting target genes into plasmids. *J Biochem Biophys Methods*, 67(1):67–74, Apr 2006.
- [144] O. Edelheit, A. Hanukoglu, and I. Hanukoglu. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol*, 9:61, Jun 2009.
- [145] A. Touznic, R. Maruyama, K. Hosoki, Y. Echigoya, and T. Yokota. LNA/DNA mixmer-based antisense oligonucleotides correct alternative splicing of the SMN2 gene and restore SMN protein expression in type 1 SMA fibroblasts. *Sci Rep*, 7(1):3672, 06 2017.

- [146] M. W. Pfaffl. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, 29(9):e45, May 2001.
- [147] M. Y. Mazina, E. V. Kovalenko, P. K. Derevyanko, J. V. Nikolenko, A. N. Krasnov, and N. E. Vorobyeva. One signal stimulates different transcriptional activation mechanisms. *Biochim Biophys Acta Gene Regul Mech*, 1861(2):178–189, Feb 2018.
- [148] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [149] Nicholas E. Hamilton and Michael Ferry. ggtern: Ternary diagrams using ggplot2. *Journal of Statistical Software, Code Snippets*, 87(3):1–17, 2018.
- [150] Benjamin Reiser. Confidence intervals for the mahalanobis distance. *Communications in Statistics. Simulation and Computation*, 30, 03 2001.
- [151] A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, 15(8):1034–1050, Aug 2005.
- [152] C. A. Davis, B. C. Hitz, C. A. Sloan, E. T. Chan, J. M. Davidson, I. Gabdank, J. A. Hilton, K. Jain, U. K. Baymuradov, A. K. Narayanan, K. C. Onate, K. Graham, S. R. Miyasato, T. R. Dreszer, J. S. Strattan, O. Jolanki, F. Y. Tanaka, and J. M. Cherry. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*, 46(D1):D794–D801, 01 2018.
- [153] O. Kelemen, P. Convertini, Z. Zhang, Y. Wen, M. Shen, M. Falaleeva, and S. Stamm. Function of alternative splicing. *Gene*, 514(1):1–30, Feb 2013.
- [154] F. E. Baralle and J. Giudice. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol*, 18(7):437–451, 07 2017.
- [155] H. K. Kim, M. H. C. Pham, K. S. Ko, B. D. Rhee, and J. Han. Alternative splicing isoforms in health and disease. *Pflugers Arch*, 470(7):995–1016, 07 2018.
- [156] S. A. Bhuiyan, S. Ly, M. Phan, B. Huntington, E. Hogan, C. C. Liu, J. Liu, and P. Pavlidis. Systematic evaluation of isoform function in literature reports of alternative splicing. *BMC Genomics*, 19(1):637, Aug 2018.
- [157] J. Kim, J. Daniel, A. Espejo, A. Lake, M. Krishna, L. Xia, Y. Zhang, and M. T. Bedford. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep*, 7(4):397–403, Apr 2006.
- [158] Y. Hou, W. Liu, X. Yi, Y. Yang, D. Su, W. Huang, H. Yu, X. Teng, Y. Yang, W. Feng, T. Zhang, J. Gao, K. Zhang, R. Qiu, and Y. Wang. PHF20L1 as a H3K27me2 reader coordinates with transcriptional repressors to promote breast tumorigenesis. *Sci Adv*, 6(16):eaaz0356, Apr 2020.

- [159] P. O. Estève, J. Terragni, K. Deepti, H. G. Chin, N. Dai, A. Espejo, I. R. Corrêa, M. T. Bedford, and S. Pradhan. Methyllysine reader plant homeodomain (PHD) finger protein 20-like 1 (PHF20L1) antagonizes DNA (cytosine-5) methyltransferase 1 (DNMT1) proteasomal degradation. *J Biol Chem*, 289(12):8277–8287, Mar 2014.
- [160] C. Zhang, N. Hoang, F. Leng, L. Saxena, L. Lee, S. Alejo, D. Qi, A. Khal, H. Sun, F. Lu, and H. Zhang. LSD1 demethylase and the methyl-binding protein PHF20L1 prevent SET7 methyltransferase-dependent proteolysis of the stem-cell protein SOX2. *J Biol Chem*, 293(10):3663–3674, 03 2018.
- [161] S. M. Carr, S. Munro, C. A. Sagum, O. Fedorov, M. T. Bedford, and N. B. La Thangue. Tudor-domain protein PHF20L1 reads lysine methylated retinoblastoma tumour suppressor protein. *Cell Death Differ*, 24(12):2139–2149, 12 2017.
- [162] H. Yu, Y. Jiang, L. Liu, W. Shan, X. Chu, Z. Yang, and Z. Q. Yang. Integrative genomic and transcriptomic analysis for pinpointing recurrent alterations of plant homeodomain genes and their clinical significance in breast cancer. *Oncotarget*, 8(8):13099–13115, Feb 2017.
- [163] Z. Zhang, T. He, L. Huang, Y. Ouyang, J. Li, Y. Huang, P. Wang, and J. Ding. Two precision medicine predictive tools for six malignant solid tumors: from gene-based research to clinical application. *J Transl Med*, 17(1):405, 12 2019.
- [164] K. O. Wrzeszczynski, V. Varadan, J. Byrnes, E. Lum, S. Kamalakaran, D. A. Levine, N. Dimitrova, M. Q. Zhang, and R. Lucito. Identification of tumor suppressors and oncogenes from genomic and epigenetic features in ovarian cancer. *PLoS One*, 6(12):e28503, 2011.
- [165] Y. Jiang, L. Liu, W. Shan, and Z. Q. Yang. An integrated genomic analysis of Tudor domain-containing proteins identifies PHD finger protein 20-like 1 (PHF20L1) as a candidate oncogene in breast cancer. *Mol Oncol*, 10(2):292–302, Feb 2016.
- [166] A. R. Cohen, D. F. Woods, S. M. Marfatia, Z. Walther, A. H. Chishti, J. M. Anderson, and D. F. Wood. Human CASK/LIN-2 binds syndecan-2 and protein 4.1 and localizes to the basolateral membrane of epithelial cells. *J Cell Biol*, 142(1):129–138, Jul 1998.
- [167] Y. Li, O. Spangenberg, I. Paarmann, M. Konrad, and A. Lavie. Structural basis for nucleotide-dependent regulation of membrane-associated guanylate kinase-like domains. *J Biol Chem*, 277(6):4159–4165, Feb 2002.
- [168] K. Tabuchi, T. Biederer, S. Butz, and T. C. Sudhof. CASK participates in alternative tripartite complexes in which Mint 1 competes for binding with caskin 1, a novel CASK-binding protein. *J Neurosci*, 22(11):4264–4273, Jun 2002.

- [169] T. F. Wang, C. N. Ding, G. S. Wang, S. C. Luo, Y. L. Lin, Y. Ruan, R. Hevner, J. L. Rubenstein, and Y. P. Hsueh. Identification of Tbr-1/CASK complex target genes in neurons. *J Neurochem*, 91(6):1483–1492, Dec 2004.
- [170] D. Atasoy, S. Schoch, A. Ho, K. A. Nadasy, X. Liu, W. Zhang, K. Mukherjee, E. D. Nosyreva, R. Fernandez-Chacon, M. Missler, E. T. Kavalali, and T. C. Südhof. Deletion of CASK in mice is lethal and impairs synaptic function. *Proc Natl Acad Sci U S A*, 104(7):2525–2530, Feb 2007.
- [171] G. Caruana. Genetic studies define MAGUK proteins as regulators of epithelial cell polarity. *Int J Dev Biol*, 46(4):511–518, 2002.
- [172] D. Tibbe, Y. E. Pan, C. Reißner, F. L. Harms, and H. J. Kreienkamp. Functional analysis of CASK transcript variants expressed in human brain. *PLoS One*, 16(6):e0253223, 2021.
- [173] J. A. Dembowski, P. An, M. Scoulos-Hanson, G. Yeo, J. Han, X. D. Fu, and P. J. Grabowski. Alternative Splicing of a Novel Inducible Exon Diversifies the CASK Guanylate Kinase Domain. *J Nucleic Acids*, 2012:816237, 2012.
- [174] A. S. Kashina. Protein Arginylation: Over 50 Years of Discovery. *Methods Mol. Biol.*, 1337:1–11, 2015.
- [175] E. Balzi, M. Choder, W. N. Chen, A. Varshavsky, and A. Goffeau. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, 265(13):7464–7471, May 1990.
- [176] Y. D. Yoo, S. R. Mun, C. H. Ji, K. W. Sung, K. Y. Kang, A. J. Heo, S. H. Lee, J. Y. An, J. Hwang, X. Q. Xie, A. Ciechanover, B. Y. Kim, and Y. T. Kwon. N-terminal arginylation generates a bimodal degron that modulates autophagic proteolysis. *Proc. Natl. Acad. Sci. U.S.A.*, 115(12):E2716–E2724, 03 2018.
- [177] V. Solomon, S. H. Lecker, A. L. Goldberg, and A. L. Goldberg. The N-end rule pathway catalyzes a major fraction of the protein degradation in skeletal muscle. *J. Biol. Chem.*, 273(39):25216–25222, Sep 1998.
- [178] K. Deka, A. Singh, S. Chakraborty, R. Mukhopadhyay, and S. Saha. Protein arginylation regulates cellular stress response by stabilizing HSP70 and HSP40 transcripts. *Cell Death Discov*, 2:16074, 2016.
- [179] K. D. Lamon, W. H. Vogel, and H. Kaji. Stress-induced increases in rat brain arginyl-tRNA transferase activity. *Brain Res.*, 190(1):285–287, May 1980.
- [180] G. Bongiovanni, S. Fissolo, H. S. Barra, and M. E. Hallak. Posttranslational arginylation of soluble rat brain proteins after whole body hyperthermia. *J. Neurosci. Res.*, 56(1):85–92, Apr 1999.
- [181] Y. T. Kwon, A. S. Kashina, I. V. Davydov, R. G. Hu, J. Y. An, J. W. Seo, F. Du, and A. Varshavsky. An essential role of N-terminal arginylation in cardiovascular development. *Science*, 297(5578):96–99, Jul 2002.

- [182] S. Kurosaka, N. A. Leu, F. Zhang, R. Bunte, S. Saha, J. Wang, C. Guo, W. He, and A. Kashina. Arginylation-dependent neural crest cell migration is essential for mouse development. *PLoS Genet.*, 6(3):e1000878, Mar 2010.
- [183] R. Rai, C. C. Wong, T. Xu, N. A. Leu, D. W. Dong, C. Guo, K. J. McLaughlin, J. R. Yates, and A. Kashina. Arginyltransferase regulates alpha cardiac actin function, myofibril formation and contractility during heart development. *Development*, 135(23):3881–3889, Dec 2008.
- [184] Y. Tanaka and H. Kaji. Incorporation of arginine by soluble extracts of ascites tumor cells and regenerating rat liver. *Cancer Res.*, 34(9):2204–2208, Sep 1974.
- [185] G. Chakraborty and N. A. Ingoglia. N-terminal arginylation and ubiquitin-mediated proteolysis in nerve regeneration. *Brain Res. Bull.*, 30(3-4):439–445, 1993.
- [186] Y. M. Wang and N. A. Ingoglia. N-terminal arginylation of sciatic nerve and brain proteins following injury. *Neurochem. Res.*, 22(12):1453–1459, Dec 1997.
- [187] H. Kaji and A. Kaji. Correlated Measurement of Endogenous ATE1 Activity on Native Acceptor Proteins in Tissues and Cultured Cells to Detect Cellular Aging. *Methods Mol. Biol.*, 1337:39–48, 2015.
- [188] K. D. Lamon and H. Kaji. Arginyl-tRNA transferase activity as a marker of cellular aging in peripheral rat tissues. *Exp. Gerontol.*, 15(1):53–64, 1980.
- [189] R. Rai, F. Zhang, K. Colavita, N. A. Leu, S. Kurosaka, A. Kumar, M. D. Birnbaum, B. Győrffy, D. W. Dong, M. Shtutman, and A. Kashina. Arginyltransferase suppresses cell tumorigenic potential and inversely correlates with metastases in human cancers. *Oncogene*, 35(31):4058–4068, 08 2016.
- [190] M. D. Birnbaum, N. Zhao, B. T. Moorthy, D. M. Patel, O. N. Kryvenko, L. Heidman, A. Kumar, W. M. Morgan, Y. Ban, I. M. Reis, X. Chen, M. L. Gonzalgo, M. Jorda, K. L. Burnstein, and F. Zhang. Reduced Arginyltransferase 1 is a driver and a potential prognostic indicator of prostate cancer metastasis. *Oncogene*, 38(6):838–851, 02 2019.
- [191] M. R. Galiano, V. E. Goitea, and M. E. Hallak. Post-translational protein arginylation in the normal nervous system and in neurodegeneration. *J. Neurochem.*, 138(4):506–517, 08 2016.
- [192] N. A. Leu, S. Kurosaka, and A. Kashina. Conditional Tek promoter-driven deletion of arginyltransferase in the germ line causes defects in gametogenesis and early embryonic lethality in mice. *PLoS ONE*, 4(11):e7734, Nov 2009.
- [193] C. S. Brower and A. Varshavsky. Ablation of arginylation in the mouse N-end rule pathway: loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. *PLoS ONE*, 4(11):e7757, Nov 2009.

- [194] Y. T. Kwon, A. S. Kashina, and A. Varshavsky. Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. *Mol. Cell Biol.*, 19(1):182–193, Jan 1999.
- [195] R. Rai and A. Kashina. Identification of mammalian arginyltransferases that modify a specific subset of protein substrates. *Proc. Natl. Acad. Sci. U.S.A.*, 102(29):10123–10128, Jul 2005.
- [196] R. G. Hu, C. S. Brower, H. Wang, I. V. Davydov, J. Sheng, J. Zhou, Y. T. Kwon, and A. Varshavsky. Arginyltransferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. *J. Biol. Chem.*, 281(43):32559–32573, Oct 2006.
- [197] R. Schmid, S. N. Grellscheid, I. Ehrmann, C. Dalgliesh, M. Danilenko, M. P. Paronetto, S. Pedrotti, D. Grellscheid, R. J. Dixon, C. Sette, I. C. Eperon, and D. J. Elliott. The splicing landscape is globally reprogrammed during male meiosis. *Nucleic Acids Res.*, 41(22):10170–10184, Dec 2013.
- [198] C. S. Brower, C. E. Rosen, R. H. Jones, B. C. Wadas, K. I. Piatkov, and A. Varshavsky. Liat1, an arginyltransferase-binding protein whose evolution among primates involved changes in the numbers of its 10-residue repeats. *Proc. Natl. Acad. Sci. U.S.A.*, 111(46):E4936–4945, Nov 2014.
- [199] Y. Jin, H. Dong, Y. Shi, and L. Bian. Mutually exclusive alternative splicing of pre-mRNAs. *Wiley Interdiscip Rev RNA*, 9(3):e1468, 05 2018.
- [200] B. L. Bass and H. Weintraub. A developmentally regulated activity that unwinds RNA duplexes. *Cell*, 48(4):607–613, Feb 1987.
- [201] Eric L Van Nostrand, Peter Freese, Gabriel A Pratt, Xiaofeng Wang, Xintao Wei, Rui Xiao, Steven M Blue, Jia-Yu Chen, Neal A.L. Cody, Daniel Dominguez, Sara Olson, Balaji Sundararaman, Lijun Zhan, Cassandra Bazile, Louis Philip Benoit Bouvrette, Julie Bergalet, Michael O Duff, Keri E. Garcia, Chelsea Gelboin-Burkhart, Myles Hochman, Nicole J Lambert, Hairi Li, Thai B Nguyen, Tsultrim Palden, Ines Rabano, Shashank Sathe, Rebecca Stanton, Amanda Su, Ruth Wang, Brian A. Yee, Bing Zhou, Ashley L Louie, Stefan Aigner, Xiang-dong Fu, Eric Lécuyer, Christopher B. Burge, Brenton R. Graveley, and Gene W. Yeo. A large-scale binding and functional map of human rna binding proteins. *bioRxiv*, 2018.
- [202] M. de la Mata, M. J. Muñoz, M. All’o, J. P. Fededa, I. E. Schor, and A. R. Kornblihtt. RNA Polymerase II Elongation at the Crossroads of Transcription and Alternative Splicing. *Genet Res Int*, 2011:309865, 2011.
- [203] T. Saldi, M. A. Cortazar, R. M. Sheridan, and D. L. Bentley. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J. Mol. Biol.*, 428(12):2623–2635, 06 2016.

- [204] M. de la Mata, C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley, and A. R. Kornblihtt. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell*, 12(2):525–532, Aug 2003.
- [205] G. Dujardin, C. Lafaille, M. de la Mata, L. E. Marasco, M. J. Muñoz, C. Le Jossic-Corcós, L. Corcos, and A. R. Kornblihtt. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell*, 54(4):683–690, May 2014.
- [206] X. Q. Gong, Y. A. Nedialkov, and Z. F. Burton. Alpha-amanitin blocks translocation by human RNA polymerase II. *J. Biol. Chem.*, 279(26):27422–27427, Jun 2004.
- [207] C. D. Kaplan, K. M. Larsson, and R. D. Kornberg. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol. Cell*, 30(5):547–556, Jun 2008.
- [208] J. Y. Ip, D. Schmidt, Q. Pan, A. K. Ramani, A. G. Fraser, D. T. Odom, and B. J. Blencowe. Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.*, 21(3):390–401, Mar 2011.
- [209] N. Fong, H. Kim, Y. Zhou, X. Ji, J. Qiu, T. Saldi, K. Diener, K. Jones, X. D. Fu, and D. L. Bentley. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev*, 28(23):2663–2676, Dec 2014.
- [210] CH Wu, Y Yamaguchi, LR Benjamin, M Horvat-Gordon, J Washinsky, E Enerly, J Larsson, A Lambertsson, H Handa, and Gilmour D. Nelf and dsif cause promoter proximal pausing on the hsp70 promoter in drosophila. *Genes Dev*, 17(11):1402–14, Jun 2003.
- [211] B Stadelmayer, G Micas, A Gamot, P Martin, N Malirat, S Koval, R Raffel, B Sobhian, D Severac, S Rialle, H Parrinello, O Cuvier, and M Benkirane. Integrator complex regulates nelf-mediated rna polymerase ii pause/release and processivity at coding genes. *Nat Commun*, 20(5):5531, Nov 2014.
- [212] SM Vos, Pöllmann D, L Caizzi, KB Hofmann, P Rombaut, T Zimniak, F Herzog, and P Cramer. Architecture and rna binding of the human negative elongation factor. *Elife*, 10(5):e14981, Jun 2016.
- [213] Y Yamaguchi, N Inukai, T Narita, T Wada, and H Handa. Evidence that negative elongation factor represses transcription elongation through binding to a drb sensitivity-inducing factor/rna polymerase ii complex and rna. *Mol Cell Biol*, 22(9):2918–27, May 2002.
- [214] K Fujinaga, D Irwin, Y Huang, R Taube, T Kurosu, and BM Peterlin. Dynamics of human immunodeficiency virus transcription: P-tefb phosphorylates rd and dissociates negative effectors from the transactivation response element. *Mol Cell Biol*, 24(2):787–95, Jan 2004.

- [215] JM Pagano, H Kwak, CT Waters, RO Sprouse, BS White, A Ozer, K Szeto, D Shalloway, HG Craighead, and JT Lis. Defining nelf-e rna binding in hiv-1 and promoter-proximal pause regions. *PLoS Genet*, 10(1):e1004090, Jan 2014.
- [216] I. E. Schor, L. I. Gómez Acuña, and A. R. Kornblihtt. Coupling between transcription and alternative splicing. *Cancer Treat. Res.*, 158:1–24, 2013.
- [217] M. D. Rudd and D. S. Luse. Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *J Biol Chem*, 271(35):21549–21558, Aug 1996.
- [218] M. B. Warf and J. A. Berglund. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.*, 35(3):169–178, Mar 2010.
- [219] M. Blanchette and B. Chabot. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J*, 18(7):1939–1952, Apr 1999.
- [220] F. C. Oberstrass, S. D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D. L. Black, and F. H. Allain. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, 309(5743):2054–2057, Sep 2005.
- [221] F. U. Nasim, S. Hutchison, M. Cordeau, and B. Chabot. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA*, 8(8):1078–1089, Aug 2002.
- [222] S. Sharma, L. A. Kohlstaedt, A. Damianov, D. C. Rio, and D. L. Black. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol*, 15(2):183–191, Feb 2008.
- [223] E. T. Wang, R. Sandberg, S. Luo, I. Khrebtkova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, and C. B. Burge. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221):470–476, Nov 2008.
- [224] H. Kuroyanagi, S. Takei, and Y. Suzuki. Comprehensive analysis of mutually exclusive alternative splicing in *C. elegans*. *Worm*, 3:e28459, 2014.
- [225] B. Sommer, K. Keinänen, T. A. Verdoorn, W. Wisden, N. Burnashev, A. Herb, M. Köhler, T. Takagi, B. Sakmann, and P. H. Seeburg. Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science*, 249(4976):1580–1585, Sep 1990.
- [226] I. Splawski, K. W. Timothy, L. M. Sharpe, N. Decher, P. Kumar, R. Bloise, C. Napolitano, P. J. Schwartz, R. M. Joseph, K. Condouris, H. Tager-Flusberg, S. G. Priori, M. C. Sanguinetti, and M. T. Keating. Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell*, 119(1):19–31, Oct 2004.

- [227] I. Splawski, K. W. Timothy, N. Decher, P. Kumar, F. B. Sachse, A. H. Beggs, M. C. Sanguinetti, and M. T. Keating. Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proc. Natl. Acad. Sci. U.S.A.*, 102(23):8089–8096, Jun 2005.
- [228] E. J. Bhoj, M. Li, R. Ahrens-Nicklas, L. C. Pyle, J. Wang, V. W. Zhang, C. Clarke, L. J. Wong, N. Sondheimer, C. Ficicioglu, and M. Yudkoff. Pathologic Variants of the Mitochondrial Phosphate Carrier SLC25A3: Two New Patients and Expansion of the Cardiomyopathy/Skeletal Myopathy Phenotype With and Without Lactic Acidosis. *JIMD Rep*, 19:59–66, 2015.
- [229] J. A. Mayr, F. A. Zimmermann, R. Horváth, H. C. Schneider, B. Schoser, E. Holinski-Feder, B. Czermin, P. Freisinger, and W. Sperl. Deficiency of the mitochondrial phosphate carrier presenting as myopathy and cardiomyopathy in a family with three affected children. *Neuromuscul. Disord.*, 21(11):803–808, Nov 2011.
- [230] C. J. David, M. Chen, M. Assanah, P. Canoll, and J. L. Manley. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature*, 463(7279):364–368, Jan 2010.
- [231] A. Valletti, A. Anselmo, M. Mangiulli, I. Boria, F. Mignone, G. Merla, V. D’Angelo, A. Tullo, E. Sbisà, A. M. D’Erchia, and G. Pesole. Identification of tumor-associated cassette exons in human cancer through EST-based computational prediction and experimental validation. *Mol. Cancer*, 9:230, Sep 2010.
- [232] T. M. Ivanov and D. D. Pervouchine. An Evolutionary Mechanism for the Generation of Competing RNA Structures Associated with Mutually Exclusive Exons. *Genes (Basel)*, 9(7), Jul 2018.
- [233] T. W. Hefferon, J. D. Groman, C. E. Yurk, and G. R. Cutting. A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. *Proc. Natl. Acad. Sci. U.S.A.*, 101(10):3504–3509, Mar 2004.
- [234] Y. Ding, Y. Tang, C. K. Kwok, Y. Zhang, P. C. Bevilacqua, and S. M. Assmann. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, 505(7485):696–700, Jan 2014.
- [235] S. Rouskin, M. Zubradt, S. Washietl, M. Kellis, and J. S. Weissman. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, 505(7485):701–705, Jan 2014.
- [236] R. C. Spitale, R. A. Flynn, Q. C. Zhang, P. Crisalli, B. Lee, J. W. Jung, H. Y. Kuchelmeister, P. J. Batista, E. A. Torre, E. T. Kool, and H. Y. Chang. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, 519(7544):486–490, Mar 2015.

- [237] R. A. Flynn, Q. C. Zhang, R. C. Spitale, B. Lee, M. R. Mumbach, and H. Y. Chang. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat Protoc*, 11(2):273–290, Feb 2016.
- [238] Y. Wan, K. Qu, Q. C. Zhang, R. A. Flynn, O. Manor, Z. Ouyang, J. Zhang, R. C. Spitale, M. P. Snyder, E. Segal, and H. Y. Chang. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, 505(7485):706–709, Jan 2014.
- [239] V. Ramani, R. Qiu, and J. Shendure. High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.*, 33(9):980–984, Sep 2015.
- [240] S. A. Mortimer, M. A. Kidwell, and J. A. Doudna. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.*, 15(7):469–479, Jul 2014.
- [241] B. R. Graveley. RNA Matchmaking: Finding Cellular Pairing Partners. *Mol. Cell*, 63(2):186–189, 07 2016.
- [242] S. Shukla, E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, and S. Oberdoerffer. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, 479(7371):74–79, Nov 2011.
- [243] M. Naville and D. Gautheret. Transcription attenuation in bacteria: theme and variations. *Brief Funct Genomic Proteomic*, 8(6):482–492, Nov 2009.
- [244] S. Wright. Regulation of eukaryotic gene expression by transcriptional attenuation. *Mol Biol Cell*, 4(7):661–668, Jul 1993.

Appendix A

Supplementary information

Table A.1: Primers for cloning *Ate1*, *Phf20l1*, and *Cask* minigenes

| ATE1 cloning | |
|------------------------|--|
| pRK5_Ate1_rf_fwd | GCACCTCGGTTCTATCGATTGAATTCCCCTGAGTAAGCCTCCA TGTCGAAAAG |
| pRK5_Ate1_rf_rev | CTGCAGGTCGACTCTAGAGGATCCCCGTATGGCCACTTGATAC TTGACATACAA |
| pRK5_ate1_fwd | ATTCCTTTGCGGGGATCCTCTAGAGTCG |
| pRK5_ate1_rev | GGATCCTGGTGTATGGCCACTTGATACTTG |
| ate1_add1_fwd | GTGGCCATACACCAGGATCCACCCGATG |
| ate1_add1_rev | TTCATTGCAGCTGTTGACCTAGAGTGGATGAC |
| ate1_add2_fwd | AGGTCAACAGCTGCAATGAAAAGAATCCAAG |
| ate1_add2_rev | GAGGATCCCCGCAAAGGAATCTTGTGAAC |
| PHF20L1 cloning | |
| Phf20l1_rf_fwd | GTGGTTGATAAAGGTTTTGCTGCCGCTCCAAAGGAAAAACAC AAAACATG |
| Phf20l1_rf_rev | ACATCCTGGCTACCAGCTTCATGTAGTAAGAGTGTTTCAGTATT CCTAGCT |
| CASK cloning | |
| pRK5_fwd | GGGGATCCTCTAGAGTCGACCTGC |
| pRK5_rev | GGGGAATTCAATCGATAGAACCGAGG |
| cask_fwd | GGGAAATGCGGGGGAGTATT |
| cask_rev | TGTCGTCCTTTTGGTTGGGT |

Table A.2: Primers for mutagenesis

ATE1 mutations

| | |
|------------------|---|
| ate1_m1_fwd | CTCTGATGATATGCTACTCAAGGTGAGGCTACAG |
| ate1_m1_rev | CTGTAGCCTCACCTTGAGTAGCATATCATCAGAG |
| ate1_m2_fwd | ACCACCCTCACCTTGAGTAGCAGTTAAAAACAAGT |
| ate1_m2_rev | ACTTGTTTTTTTAACTGCTACTCAAGGTGAGGGTGGT |
| ate1_m11_fwd | ACTCAAGGTGAGGTTAGTACCTGTC |
| ate1_m11_rev | GCAAGGAAAATACAAGCAAATGC |
| ate1_m1_rear_fwd | GGAAGTCTTCGTTATCATCAGAGATGTTTCAATAGCATTTC |
| ate1_m1_rear_rev | TGAGGCTACAGTAAGATCTCCACAAATC |
| ate1_m4-3_fwd | GACAAAGTTATCTTAAGAGTTAATGTAATGGAAC |
| ate1_m4-3_rev | AAAGTGCAAATATACCAAATTTTAAGT |
| ate1_m3-3_fwd | TTGTCAAAGTGCAAATTTTACCTAGTGCC |
| ate1_m3_rev | TGTCCATTAATTCAGTCTTAGG |

PHF20L1 mutations

| | |
|-----------------|--|
| phf_m1_rear_f | GGGTTTATCGTCCAAGAAAAGGAAGCTTAGACTGTAAA |
| phf_m1_rear_rev | GCTATCAGCACATAACCCTACAATTA |
| phf_m2_rear_fwd | GACAATAAACCCATTTTCAGCACCTGGACCTACAAT |
| phf_m2_rear_rev | CAGAATAGAAAAGAAAATATTTTAAATTTCTAAT |

Cask mutations

| | |
|-------------|--------------------------------------|
| cask_m1_fwd | CGTGTGGTTAAATACAACAAAGATAACGGAACAG |
| cask_m1_rev | AGTGATCTAACAGCACAGG |
| cask_m2_fwd | CTTAACCACACGCCTCTGTTATTGTGGTACGTGCAC |
| cask_m2_rev | GTAAACTTAGATAGCGGAAACTAAA |

Table A.3: Sequences of AONs. DNA nucleotides: G,A,T,C. LNA nucleotides (red): +G, +A, +T, +C. Phosphorothioated DNA nucleotides: G*, A*, T*, C*. Phosphorothioated LNA nucleotides (red): +G*, +A*, +T*, +C*.

| | |
|----------------------------|--|
| control AON | +T*G*+G*A*+A*G*+T*C*+T*T*+C*G*+T |
| AONs against Ate1 | |
| AON2-1 | +T*G*+C*A*+C*T*+T*T*+C*A*+G*A*+A |
| AON2-2 | +T*T*+C*T*+G*A*+A*A*+G*T*+G*C*+A |
| AON1 | +T*G*+C*T*+T*C*+T*G*+A*A*+G*G*+T |
| AONs against Phf201 | |
| AON1 | T*+T*G*+C*T*+G*C*+T*A*+T*T*+T*G*+G*G*+G*C*+T |
| AON2 | A*+A*T*+C*C*+C*A*+A*A*+T*A*+A*C*+A*G*+C*A*+G |
| AONs against Cask | |
| AON1 | +A*A*+A*T*+T*G*+G*T*+G*T*+G*C*+A |
| AON2 | +G*C*+A*C*+A*C*+C*A*+A*T*+T*C*+G |

Table A.4: Primers for RT-PCR analysis

Ate1

| | |
|--------------|---------------------------|
| ate1_rt_fwd | CCAACCAGCCAAAATCACTCG |
| ate1_ex1_rev | GAACTGCGAACTTGGTGGAGATG |
| ate1_ex2_rev | GTATGGCCACTTGATACTTGACATA |

Phf2011

| | |
|-------------|--------------------------|
| phf2011_fwd | GCCATGCCCCGAGGATGCTAA |
| phf2011_rev | TCTCTACGTCTGGTGTGGCTATAC |

Cask

| | |
|----------|----------------------|
| cask_fwd | GGGAAATGCGGGGGAGTATT |
| cask_rev | TGTCGTCCTTTTGGTTGGGT |

HNRNPDL

| | |
|--------------|-----------------------|
| HNRNPDL_fwd | GGCAGCAACAGCAACAACAA |
| HNRNPDL_rev | GTTTTGGTGATTGCCACCCC |
| HNRNPDL2_fwd | GGGGTGGCAATCACCAAAAC |
| HNRNPDL2_rev | AGTACCTGACGCAGAAAAGCA |

HNRPLL

| | |
|------------|-------------------------|
| HNRPLL_fwd | AGGAATGACAATGACAGTTGGGA |
| HNRPLL_rev | TTGTCTCTGGCGACCCTTTC |

Table A.5: Primers for qRT-PCR analysis

| Minigene | | |
|-----------------------------|--------------|---------------------------|
| constitutive fragment | pRK5_seq_f | CACTCCCAGGTCCAAC TG |
| | ate1_rt_rev | AGCTGGGTTCTGCTGCATTAG |
| isoform with exon 7a | ate1_rt_fwd1 | TACCAGAGAATGCATCACACA |
| | ate1_rt_rev1 | AGGAATCTTGTGAACTGGCTTT |
| isoform with exon 7b | ate1_rt_f2_2 | TCACACAAGTTAGAGGTGAGGTTAG |
| | ate1_rt_rev2 | AGGATCCCCGCAAAGGAATCT |
| isoform with both exons | ate1_ex1_fwd | ACGTTACCAGATGGTTATTCACAAG |
| | ate1_ex2_rev | GTATGGCCACTTGATACTTGACATA |
| Endogenous ATE1 mRNA | | |
| constitutive exon | ate1_const_f | GCTCTGGTGAACCGTCACATTCAG |
| | ate1_const_r | GACATGGAGGCTTACTCAAATCAGC |
| isoform with exon 7a | ate1_rt_fwd1 | TACCAGAGAATGCATCACACA |
| | ate1_rt_rev1 | AGGAATCTTGTGAACTGGCTTT |
| isoform with exon 7b | ate1_rt_f2 | TCACACAAGTTAGAGGTGAGGTTAG |
| | ate1_rt_r2 | CCAAGGGTGAAGTCAAAGGA |
| isoform with both exons | ate1_ex1_fwd | ACGTTACCAGATGGTTATTCACAAG |
| | ate1_ex2_rev | GTATGGCCACTTGATACTTGACATA |
| 28S rRNA | | |
| 28S_fwd | | GGAATGCAGCCCAAAGCGG |
| 28S_rev | | GGACCCACCCGTTTACCTCTT |
| GAPDH | | |
| gapdh_fwd | | GTCTCCTCTGACTTCAACAGCG |
| gapdh_rev | | ACCACCCTGTTGCTGTAGCCAA |
| NELFE | | |
| nelfe_fwd | | AGCGTTCTCGAACCCCTTGAG |
| nelfe_rev | | TATGCTCCTCTGGAACGGCT |

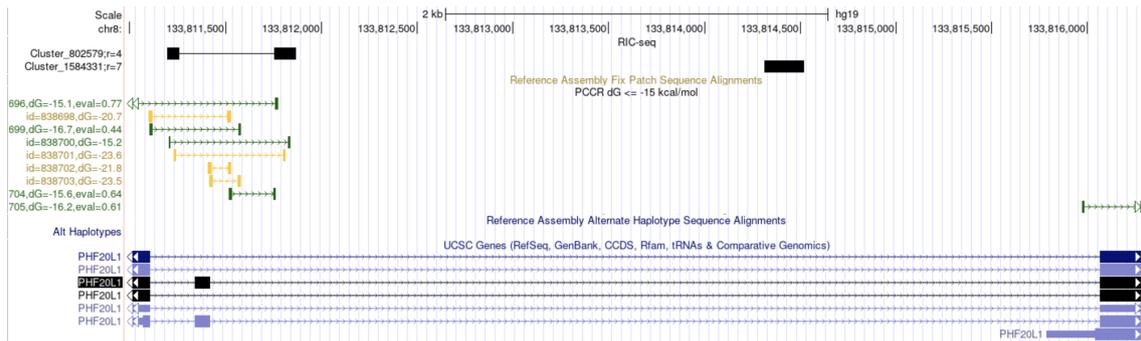


Figure A-1: The published RIC-seq data for HeLa cells confirms id838701 base pairings in *Phf20l1* [3]. The data for mRNAs obtained as a courtesy of Dr. Yuanchao Xue was converted and displayed as a UCSC Genome browser custom track.

