Regulation of NUMB alternative splicing by RBM10 and control of cancer cell proliferation

Jordi Hernández Ribera

TESI DOCTORAL UPF 2016

Thesis Director

Dr. Juan Valcárcel Júarez

Department of Gene Regulation, Stem Cells and Cancer

upf. Universitat Pompeu Fabra Barcelona

CRG Centre for Genomic Regulation

”It is well known that a vital ingredient of success is not knowing that what you're attempting can't be done."

Equal Rites
Sir Terry Pratchett

A la meva familia. Als que vénen i als que se'n van.

Agraïments

“Un d’entre tants”

Voldria començar demanant disculpes, coneixent-me, segur que no vos sorpreneu si m’he oblidat de mencionar-vos…Ja sabeu com sóc.

“Història d’un amic”

Quería dar las gracias a Juan. Por muchas cosas. Por darme la oportunidad de trabajar en el laboratorio. Por creer en mí desde el principio. Por haberme enseñado a hacer y a discutir ciencia. Por ver siempre el lado bueno. Ojalá algún día disfrute esto tanto como tú. Muchas gracias, de verdad.
Gràcies de tot cor. Moitas grazas.

“*Als Companys”*

Esto no habría acabado así si no hubiera sido por vosotros Guio y Lore.
Pff que dir? Puede que no fuesen mis mejores momentos científicos…pero compañeros como vosotros, no hay. Ni en el lab ni en la vida. Muchas gracias por la ayuda, muchas gracias por estar siempre ahí. Més gràcies encara, per la vostra amistat. Grazie.

Y hablando de amistad, gracias a Franky y a Carlos. Porque no todo es el laboratorio, por las birras, las barbacoas, los barrancos, los puentings, los paintballs, en fin, por los buenos ratos.

Gràcies Joan. Tants anys i tantes hores. Moltes gràcies tío per tot, per ser un amic. Dels de veritat.

I also wanted to thank Jofre, Pan, Elena and Ramón. I guess sharing food tightens bonds, sort of, thank you guys so much, for both the scientific discussion and the fun.

I'll keep a warm memory of the lab. Camilla and her shoes, Elias (behave), Villy, Luisa (which I am sure is writing her thesis at this very moment) Cate, Keiko sensei, Claudia (the princess of the OCD). Thanks also to Gwendal, Eli and Gosia. Moltes gràcies Anna per la infinita paciència que tens amb mi! També a gent de fora del lab, con l'Alex, Ilda, Sara,Francesco, Jack, Agata, Simone, Rafa, Miquel i Francisco. Gracias por hacer esta época de mi vida mejor.

"*Va com va"*

I also wanted to thank Sophie for her help when I needed it. Specially for important things, not only for science.

"L'escola de Ribera"

Volia també donar les gràcies a la gent que, m'acompanyen de fa temps: gràcies a Mar, per donar-me sempre forces per a superar-me. Gràcies a Anna, Maria, Clara i Amaia. No sé com haguéreu aprovat la carrera sense mi….Eh, igual haguéreu aprovat però no ens hauríem rigut tant.

También gracias a Pablo, por los buenos ratos que compartimos, con las neuronas o los ordenadores.

"El meu poble Alcoi"

Als meus amics, Nando, Luis, Pablo, Emili, Victor i Jessy. Perquè algunes coses bones si que duren.

Per descomptat, voldria agrair a la meva família tot el que han fet per mi. Per ensenyar-me que hi ha coses per les que si que mereix la pena donar-ho tot. Per estar sempre amb mi Jeanine, Rafael, Tere, Juanito, Luisa, Lucia, Juan, Adrián, Sandra, Berto i Leire. Gràcies.

Als meus pares, perquè sent qui més em coneix encara m'estimen.

A Pau, "*perquè vull*".

“Els Amants”

A Jessi, per tot.

“Bon vent i barca nova”

Summary

Alternative pre-mRNA splicing of the Notch regulator NUMB exon 9 switches between isoforms that promote (exon inclusion) or prevent (exon skipping) cell proliferation. We have studied NUMB alternative splicing regulation by RBM10, an RNA binding protein frequently mutated in lung adenocarcinomas. The study of cancer-associated RBM10 variants revealed domains and residues important for RBM10 function. Structure-function analysis identified residues in RRM2 that, without compromising RNA binding, fail to regulate NUMB exon 9. Immunoprecipitation and mass spectrometry analyses revealed interactors that include U2 snRNP and PRP19 spliceosomal components, allowing us to propose a mechanism for 3' splice site repression by RBM10. Antisense oligonucleotide (AON) experiments delineated regions important for NUMB exon 9 inclusion. Using these reagents, we provide initial evidence that modulation of NUMB alternative splicing by AONs can reduce cancer cell proliferation and tumor growth.

Resum

L'empalmament alternatiu del pre-mARN de NUMB, un regulador de Notch, varia entre dos isoformes que promouen (inclusió de l'exó 9) o prevenen (salt de l'exó 9) la proliferació cel·lular. Hem estudiat la regulació de l'empalmament alternatiu de NUMB controlada per RBM10, una proteïna d'unió l'ARN freqüentment mutada en càncer de pulmó. L'estudi de variants d'RBM10 associades a càncer revela dominis i residus importants per a la funció d'RBM10. Mitjançant anàlisis estructurals i funcionals hem identificat residus l'RRM2 que, quan mutats, no comprometen la unió l'ARN però són incapaços d'alterar la regulació de l'exó 9 de NUMB. Anàlisis de co-immunoprecipitació i espectrometria de masses assenyalen interactors que inclouen U2 snRNP i PRP19 que ens permeten proposar un mecanisme de repressió del lloc d'empalmament al 3' reprimit per RBM10. Experiments amb AONs delimiten regions importants per a la inclusió de l'exó 9 de NUMB. Mitjançant l'ús d'aquests reactius tenim evidències inicials sobre com la modulació de l'empalmament alternatiu de NUMB amb AONs produeix la proliferació de cèl·lules canceroses I el creixement tumoral.

Preface

The relationship between alternative splicing alterations and disease has been known for long time. Several maladies have been related to misregulation of splicing events. One of the splicing events that appears frequently altered in several types of cancer is NUMB exon 9 alternative splicing. The alternative splicing of NUMB controls cell proliferation, with the inclusion of exon 9 generating an mRNA that encodes a protein isoform that promotes cell proliferation, while skipping of exon 9 prevents cell growth. This adds clinical relevance to the study of NUMB alternative splicing.

Previous work form our laboratory characterized RBM10 as a strong regulator of NUMB exon 9 alternative splicing. Overexpression of RBM10 was shown to promote NUMB exon 9 skipping. Our laboratory also identified a new mutant of RBM10 in a human lung cancer cell line, which was unable to modulate NUMB splicing, although the reason for its loss-of-function was unknown.

We want to understand the modulation of NUMB exon 9 alternative splicing by RBM10 and to identify key regions in NUMB exon 9 important for its alternative splicing.

Objectives

1) To better understand how RBM10 modulates NUMB exon 9 alternative splicing, characterizing important domains, residues and protein partners for RBM10 function.

2)To identify sequences in NUMB exon 9 which are important for its alternative splicing regulation. The final goal was the modulation of NUMB exon 9 alternative splicing in tumors as a possible therapeutic tool for preventing tumor growth.

Table of Contents

General Introduction

Cell Theory

Biology, from the greek βίος (life) and –λόγος (reason), is the science that studies life and the living organisms. In modern biology the cell is considered the elementary unit of life. The term cell was coined by Robert Hooke in his *Micrographia*, a publication describing his observations with the microscope, where he saw “cell” structures in a piece of cork (Hooke).However it was not until 1838 that “the cell theory” was formulated by Matthias Schleiden and Theodor Schawnn. The theory states that the cell is the fundamental unit of life, the smallest unit capable of self-replication, and that all the cells come from a previous cell and organisms are composed of one or more cells (Reviewed in (Mazzarello, 1999)). Schleiden and Schawnn showed that complex multicellular organisms are composed of organs, which are formed by tissues, which in turn are built of cells. In complex multicellular organisms, tissues are groups of cells that, together, perform a specific function (reviewed in (Mayr, 1982)), and organs acquire a high level of specialization from the tissues that compose them. An apparent paradox arises when one considers the common lineage of all somatic cells and the vast diversity in morphology and function of the cells in the mature tissue; how the variety of cell and tissue types arise from a single fertilized egg has occupied the minds of biologists until today.

Genes

Genes were postulated as the units of inheritance by Gregor Mendel, capable of transmitting traits from the progenitors to the off-spring (Mendel, 1865). The link between the idea of gene and the molecule responsible for inheritance was established in the 1940s in a famous experiment where Avery, MacLeod and McCarty proved that the Deoxyribonucleic Acid (DNA) carried the genetic traits for the infectious potential of pneumococcus (Avery et al., 1944). At around the same time, Beadle and Tatum showed how lesions on the DNA can cause alterations in the enzymes of specific metabolic routes, proposing the idea of one-gene-one-enzyme (Beadle et al., 1941). These two milestone discoveries settled the idea that the genes where defined entities in the DNA molecule and that each of them coded for a single protein. The differences in morphology and functionality within somatic cells of an organism are not due to differences in their genetic material, which is almost identical among them, but rather to the distinct assortments of expressed genes and to the different levels of their expression.

Although DNA stores the genetic information the final effector of most cellular functions are proteins, chains of amino acids with specific enzymatic or structural functions. The discovery of how the information was transmitted from the DNA to the proteins came in the 1950s and 1960s. The transmission of

information from DNA to protein is mediated by Ribonucleic Acid (RNA). In the 1950s, several experiments from Brachet suggested a link between RNA and protein synthesis; he observed a reduction in protein biosynthesis after treatment with RNase, in onion-root cells and amoebas (Brachet, 1954, 1955). It was also in the 50s when Goldstein and Plaut proved that RNA was synthetized in the nucleus and then exported to the cytoplasm (Goldstein et al., 1955). Goldstein's experiments suggested directionality in the flux of information, whereby from its storage site in the nucleus, the DNA (Levene, 1910, 1912) is transcribed to RNA that will be exported to the cytoplasm to guide protein synthesis.

This flow of information is the central dogma of molecular biology as stated by Crick on "*On protein synthesis"* (F. Crick, 1958). Although several exceptions have since been found to the unidirectional flow of information posited by the central dogma (Baltimore, 1970; Weiss, 1998), its formulation was a milestone in molecular biology and pointed towards the next breakthrough in the field: the cracking of the genetic code. The genetic code establishes how the genetic information is translated from nucleotides into amino acids. Matthaei, Khorana, Nirenberg, Ochoa and colleagues deciphered the relationship between triplets of nucleotides in the RNA, codons, and the corresponding amino acid in the protein, establishing the relationship between the sequence of a gene and the sequence of its coded protein (Lengyel et al., 1961;

M. Nirenberg et al., 1965; M. W. Nirenberg et al., 1961; Nishimura et al., 1964). The genetic code is universal, the same triplets in the RNA sequence code for the same amino acids in all organisms and all cell types. This universality is a reference to and perhaps the strongest evidence for the common origin of life on earth. However only a minor fraction of all nucleotides found on a cell's DNA or even of the portion that is transcribed to RNA are going to have protein coding function.

Non-coding DNA

Non-coding DNA sequences correspond to non-transcribed regulatory regions (promoters, transcriptional silencers or enhancers) whose main function is to modulate the expression of protein coding genes by controlling gene transcription. The existence of such regulatory elements was first documented by Jacob and Monod in the Lac operon in *Escherichia coli,* which included the concept of regulatory genes whose function is to control the expression of other genes (Jacob et al., 1961).

Other non-coding pieces of DNA include specialized regions in the chromosome ends (telomeres) and their anchorage sites to microtubules in spindles (centrosomes).

As it happens with the DNA, not all the transcribed RNA sequences have a protein coding function. Different types of

RNAs exist inside a cell. The most abundant type are ribosomal RNAs (rRNA), major components of the ribosomes, the machinery responsible of protein synthesis (Goldstein et al., 1955); Ribosomes are ribonucleoprotein complexes that consist of two different subunits, composed by proteins and rRNAs. During the process of protein synthesis, the small subunit recognizes the initiation codon (ATG) and subsequent recognition of codons by amino acid-charged transfer RNA molecules, while the large subunit harbors the catalytic site where the peptide bond between consecutive amino acids of the nascent peptide is made (reviewed in (Zaher et al., 2009)).

As obvious from the above, transfer RNAs (tRNAs) are another type of RNAs that do not code for proteins, but however play a key role during protein synthesis by acting as adaptor molecules between the codons in the template messenger RNA and the amino acid of the protein (F. H. Crick et al., 1961).

Messenger RNA (mRNA) is the RNA used as template for protein biogenesis by the ribosomes. Non-coding sequences flank the open reading frame at the 5' and 3' ends of the mRNA (5' and 3' UTRs), playing regulatory roles to control mRNA decay and translation (Wilkie et al., 2003). Also, a polyadenosine (polyA) tail can be found at the 3’ end of mature mRNAs, which is important for nuclear export, RNA

stability and translational regulation (Jackson et al., 1990). In eukaryotes mRNAs are synthetized as a precursor messenger RNA (pre-mRNA) that require several processing steps before being exported to the cytoplasm and translated. A key-processing step is the removal of internal non-coding sequences, known as introns, intercalated within coding sequences, known as exons. During the process of pre-mRNA splicing, introns are removed from the pre-mRNA and exons joined together (Wahl et al., 2009).

There are other RNAs that do not play a direct role in transcription but that can modulate protein synthesis at several levels, such as micro RNAs (miRNAs) and long-non-coding RNAs (lncRNAs). miRNAs are non-coding RNAs, of 19 to 25 nucleotides, that play important roles for the regulation of mRNA abundance and translation efficiency (reviewed in detail in (Bartel, 2004)). In mammals, several families of miRNAs have been implicated cancer progression and prognosis (a brief snapshot in (Spizzo et al., 2009)).

lncRNAs are defined as RNAs longer than 200nt with no identifiable protein coding potential. The GENECODE project has identified more than 15.000 lncRNAs in the human genome (GENECODEv22). The first lncRNAs were characterized as nuclear-restricted epigenetic modifiers, such as *XIST* (Brown et al., 1991) and *AIR* (Lyle et al., 2000; Wutz et al., 1997). However, reports point towards different

regulatory activities, including cytoplasmic functions in mRNA translation (Carrieri et al., 2012), sequestering proteins (Kino et al., 2010) or controlling microRNA activity (Cesana et al., 2011). There is an ongoing debate about the capacity of these transcripts to encode small peptides. Indeed some studies have found them associated with the translation machinery (Ingolia et al., 2011; van Heesch et al., 2014) and functions for some of these peptides have been proposed (Anderson et al., 2015).

Pre-mRNA Splicing

First observations of splicing were made in adenovirus, where Roberts, Sharp and co-workers demonstrated that mRNA sequences were not collinear with the viral genomic DNA and that internal pieces in the primary copy of the DNA molecule (pre-mRNA introns) are removed and the flanking sequences (exons) spliced together (Berget et al., 1977; Berk et al.; Chow et al., 1977). These observations were shortly followed by the identification of introns in the genes of mammalian cells (Garapin et al., 1978; Leder et al., 1978; Tilghman et al., 1978). These observations established that in eukaryotic cells the information for protein synthesis is not collinear with the genetic information in the DNA and reshaped the concept of the architecture of genes. As Walter Gilbert stated in “Why Genes in Pieces?”, genes are stretches of protein-coding sequence, exons, interrupted by “translationally silent” introns (Gilbert, 1978).

Chemically speaking, splicing consists of two consecutive transesterification reactions that take place in the same RNA molecule. Initially, the phosphate bond between the last nucleotide of the exon and the first of the intron (known as the 5' splice site (5'SS)) is subject of a nucleophilic attack by the 2'-hydroxyl group of an intronic adenosine located in the last 20-50 nucleotides of the intron at the branch point sequence (BPS). The result of this first transesterification reaction is a 3' end-cleaved upstream exon and an intron lariat harboring a 2'-5' phosphodiester bond between the BPS adenosine and the 5' end of the intron. After the first reaction, the free 3' hydroxyl group of upstream exon carries out a nucleophilic attack on the phosphodiester bond of the 3' splice site (3'SS).

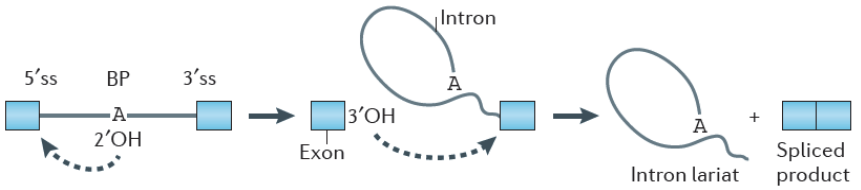


Figure 1 The Splicing Reaction. Splicing consists of two consecutive transesterification reactions. In the first one, the 2'OH of the branch point adenosine performs a nucleophilic attack on the phosphate linking the last nucleotide of the upstream exon and the first nucleotide of the intron. During the second reaction, the free 3'OH generated by the first reaction, attacks the phosphodiester bound between the last nucleotide of the intron and the first of the downstream exon, generating an intron lariat and the spliced product. Adapted from (Scotti et al., 2016).

This second reaction results in the ligation of the exons and the release of the intron in a lariat configuration. This phosphoric-transfer reactions are facilitated by two Mg^{2+} ions

that help activating the attacking groups and stabilizing the leaving groups at each of the two catalytic steps (S. M. Fica et al., 2013; Papasaikas et al., 2016; Steitz et al., 1993; Wahl et al., 2009).

Although there is still some ongoing debate about the evolution of splicing (Koonin, 2006), the characteristic partition of the eukaryotic genes in exons and introns is probably due to an invasion of mobile elements similar to the extant autocatalytic type II introns, which likely occurred in the nucleus of an early eukaryotic ancestor. These introns are still present in contemporary bacteria and eukaryotic plastids of endosymbiotic origin and are capable of self-splicing and inserting in a genome as mobile elements (Koonin et al., 2005). According to the exon shuffling theory of W. Gilbert, the compartmentalization of the genetic information in exons, induced by the invading introns, allowed the possibility of scrambling of the exons between genes, thus efficiently increasing domain shuffling and protein diversity during evolution (Gilbert, 1978).

Despite deriving from the autonomous type II introns, eukaryotic introns have lost their capacity to self-splice and thus require a huge ribonucloprotein enzyme, the spliceosome, to splice them. The main function of the spliceosome is the proper recognition and positioning of the splicing groups, ensuring the efficiency and fidelity of the

reaction. The recognition process involves the recruitment of the spliceosomal components to the specific pre-mRNA sequence elements. These sequences are typically categorized as the 5'SS, the BP and the 3'SS. The degree of conservation of these sequences varies between organisms, they are very well conserved in yeast but their conservation is much lower in Metazoa; the closer their sequence to the consensus the better they will be recognized by trans-acting factors (reviewed in (Lee et al., 2015)). The spliceosomal components are a cohort of protein-RNA complexes, including 5 small nuclear ribonucleoproteins (snRNPs) that play major roles in the recognition of the cis-acting elements. snRNPs are complexes of one snRNA molecule (U1, U2, U4/U6, and U5 (Lerner et al., 1979)) with several associated proteins. Seven Sm proteins are common to four of the five spliceosomal snRNPs, while LSm proteins associated with U6 snRNA. In addition, each snRNP contains a set of specific proteins that, together with specific sequences within the snRNAs, confer their properties in splice site recognition (Papasaikas et al., 2016; Wahl et al., 2009).

The canonical spliceosome assembly pathway consists in the reversible (Tseng et al., 2008) step-wise binding, reorganization and dynamic change in composition of the different components of the splicing machinery (Papasaikas et al.; Wahl et al., 2009). The binding of most of the complexes to the RNA is mediated by individual weak

interactions, which are further stabilized through cooperative and synergistic interactions. In fact, due to the multiple and cooperative interactions, mutations in different components of the spliceosome cause different effects on splicing and are associated with a wide variety of diseases (Cooper et al., 2009).

Spliceosome Assembly

E-complex

The first step is the formation of the E-complex. The 5’ SS is recognized in this initial stage by the binding of the U1 snRNP in an ATP independent manner. The 5' end of U1 snRNA base pairs with six nucleotides of the intron and up to three nucleotides of the upstream exon. The interaction is also stabilized by interactions between the RNA and the U1 associated protein U1C (Kondo et al., 2015). U1 recruitment can be helped by SR proteins (Crispino et al., 1994) and TIA1/TIAR (Del Gatto-Konczak et al., 2000; Forch et al., 2000; Z. Wang et al., 2010)but it can also be blocked by hnRNP A1 and hnRNP H (Buratti et al., 2004; De Conti et al., 2012). At the other side of the intron the 3’SS is recognized by the cooperative binding of the U2AF complex (U2AF65-U2AF35) and SF1/BBP. U2AF35 binds to the conserved AG sequence at the end of the intron and to the first nucleotides of the downstream exon (Merendino et al., 1999; Wu et al., 1999; Zorio et al., 1999); The heterodimer U2AF35/65 binds to the poly-pyrimidine tract and interacts with SF1/BBP, bound to the BPS (Kramer et al., 1991 Berglund, 1998 #300 ; Merendino et al., 1999; Ruskin et al., 1988 ; Wu et al., 1999; Zhang et al., 1992; Zorio et al., 1999). Mutations in elements that take part in E-complex assembly, such as U2AF35 (encoded by the gene U2AF1) are associated with lung cancer (Collisson et al., 2014).

A-complex

After proper assembly of E-complex, U2 snRNP will displace SF1/BBP. The DExD/H box helicases Prp5 and Sub2/UAP56 will help in the substitution of SF1/BBP by U2, using ATP (reviewed in (Pyle, 2008)). The branch point adenosine is bulged out of the key base pairing interaction between nucleotides flanking the branched adenosine and the branch point recognition sequence in U2 snRNA. Bulging out of the branch adenosine is a requirement for splicing catalysis, (Berglund et al., 2001; Smith et al., 2009) and the nucleotide is bound by SF3b14a, a component of U2 snRNP (Will et al., 2001). Other members of U2 snRNP help stabilizing its binding. For example SF3b155 interacts with U2AF65 (Gozani et al., 1998) that remains bound to the PPT through complex A formation. As it happens with the E-complex, the stabilization of the binding requires the cooperative effects of the different members of the complex. Point mutations in components of A complex, including SF3B1, have been associated with different types of Leukemia and other tumors (Quesada, Conde, et al., 2012; Quesada, Ramsay, et al., 2012; L. Wang et al., 2011).

B-complex

The next step in spliceosome assembly is the formation of the pre-catalytic B-complex; this complex is characterized by the recruitment of the tri-snRNP that consists of a pre-assembled U4/U6-U5 and other protein complexes, including the

Prp19/CDC5L complex (known as NTC complex in yeast) (reviewed in (Wahl et al., 2009)).

The heteromeric PRPF19 complex is formed by seven to nine different proteins (of which PRPF19, CDC5L, PRLG1 and SPF27 are strongly associated (Grote et al., 2010; Makarova et al., 2004), and it is required for the second catalytic step (Ajuh et al., 2000). PRPF19 complex does not contain any protein known to directly bind to the RNA, suggesting that is recruitment is based on protein-protein interactions (David et al., 2011)). And, although PRPF19 complex has classically been associated with the catalytically active spliceosome, several reports suggest an earlier recruitment (Chung et al., 1999; Tardiff et al., 2006; Q. Wang et al., 2003)

In yeast, this pre-catalytic B-complex can be isolated (Deckert et al., 2006) but it requires additional factors, specially the NTC complex in order to become B-active-complex (Stevens et al., 2002).

B-active-complex

Prior to the activation of the spliceosome (B-active-complex), in B complex U4/U6 are extensively base paired together, hiding the U6 AGC triad (positions 47-49) and other nucleotides critical for catalysis.

The DExD/H box helicase Brr2 plays a critical role in the formation of the B-active-complex, unwinding the U4-U6 duplex and allowing U2 to base pairing with U6. The proper control of Brr2 activation is key to prevent the premature assembly of a catalytic core. Brr2 function is controlled by Prp8, the most conserved protein of the spliceosome, which contains an RNase H-like domain that resembles proteins coded by group II introns implicated in ribozyme folding and catalysis (Dlakic et al., 2011; Galej et al., 2013; Lambowitz et al., 2011). The RNase H-like and the Jab-1 domains of Prp8 prevent Brr2 from functioning before the catalytic activation. Prp8 is also a core spliceosomal component and a major player in orchestrating the correct tempo of the catalysis. Mutations in Prp8 Jab-1 domain have been associated with retinitis pigmentosa, showing how mutations in core spliceosomal components can have a tissue-specific defect, in this case retinal degeneration (McKie et al., 2001).

During the transition towards the B-active complex, Prp8 releases its double lock on Brr2, allowing Brr2 to unwind U4/U6. SRPK2, a kinase associated with the tri-snRNP will activate Prp28, a DExD/H box helicase that will also replace the U1 interaction with the 5'SS by U6, that will anneal to the 5'SS though its ACAGAG box (Kandels-Lewis et al., 1993; Lesser et al., 1993; Staley et al., 1999). After unwinding of U4/U6, U6 extensively interacts with U2 to form the catalytic

site and U1 and U4 are de-stabilized from the complex or released, together with other proteins.

The U6-U2 interaction generates a triple helix involving nucleotides of U2 Helix Ib (U22-C21-G20) interacting with U6 AGC triad, that also interacts with distant nucleotides of U6 (G40-A41); the formation of this triple helix positions the pair of coordinated Mg^{2+} ions at the catalytic center. At this stage the 5' end of the intron is base paired with U6 and the BPS is base paired with U2 (but with the BP A residue bulged out, a feature important for catalytic activation).

B-complex*

This particular arrangement of the spliceosome, where there is extensive base pairing between U2 and U6, just ready for the first transesterification reaction is known as B*-complex. Prp2 will (using ATP) remodel the B*-complex, so it can carry out the step I reaction, in a Prp2 and ATP dependent manner (Kim et al., 1996; Yan et al., 2016). The first transesterification reaction will generate a free 3' end of the upstream exon and an intron lariat bond to the downstream exon.

The NTC complex, important for both transesterification reactions, will regulate the U5 and U6 interaction with the RNA before and after step I and it is also important for the

formation of the catalytic center (Chan et al., 2005; Chan et al., 2003).

C-complex

Spliceosome C-complex is an intermediate state en route between completion of step I and catalytic step II. Prp18 and U5 snRNA will place the 3'OH of the upstream exon to attack the phosphate at the intron-exon boundary. The 3'OH of the upstream exon is bound by NTC/PRP19 complex subunits Slu7 and the helicase Prp22 (Smith et al., 2008). As a requirement for the second transesterification reaction, the 5'SS is released from U6 (Konarska et al., 2006), involving Prp8 (bound to the upstream exon and lariat intron-3′ exon intermediates (Grainger et al., 2005)), Prp16 (involved in both catalytic steps, it supervises the recognition of the branch point and the 5'SS (Koodathingal et al., 2010; Tseng et al., 2011) and Isy1 (a member of NTC). During this second reaction, the role of the Mg^{2+} ions will be inverted, the one that was previously activating the donor group will now be stabilizing the leaving group and vice versa (Sebastian M. Fica et al., 2013).

After this second reaction the spliceosome will disassemble, the spliced mRNA and an intron lariat will be released from the spliceosome, induced by Prp22, which disrupts the interaction of Prp8 and U5 from the exonic sequences. U2, U5 and U6 are liberated from the complex by Prp43, Brr2 and

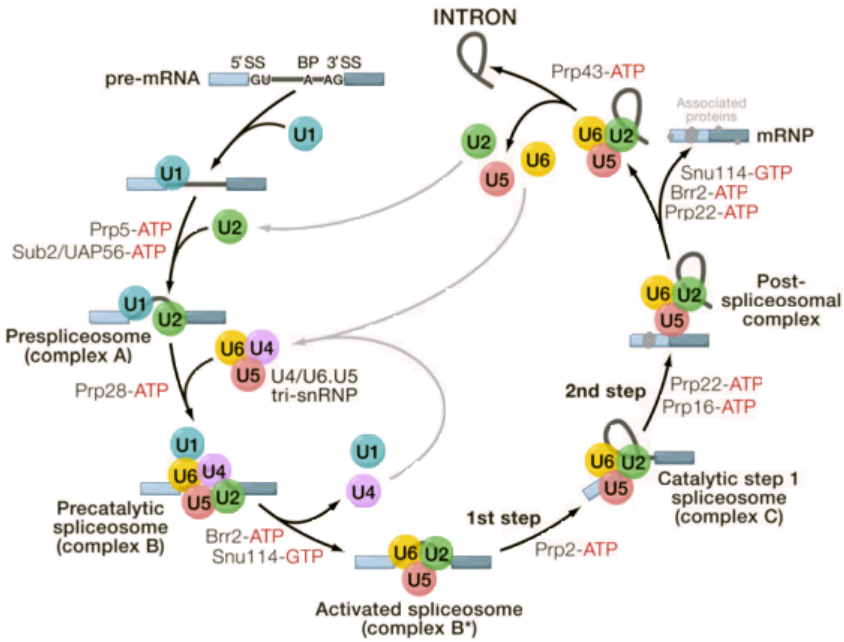


Figure 2 Canonical Spliceosomal assembly pathway. Spliceosome assembly consists in the reversible step-wise association of the different members of the spliceosome with the pre-messenger RNA. Several members will join and leave the complex at different stages and important compositional and structural rearrangements will take place for facilitating the splicing reaction. Adapted from (Wahl et al., 2009).

Snu114 (Valadkhan et al., 2010). The snRNPs and protein complexes will be recycled for subsequent splicing reactions.

Minor spliceosome

In most eukaryotes a “minor spliceosome” can also be found, the role of which is to remove U12 introns. U12 introns differ from standard introns (U2 introns) in that they have a different set of characteristic sequences in their 5’, 3’ SS and branch sites, which are highly conserved (Hall et al., 1994). There are around 700-800 U12 introns in the human genome (Alioto, 2007; Sheth et al., 2006) and U12 introns can be

found in orthologous genes in evolutionary distant species, suggesting an early evolutionary origin (probably present in the last eukaryotic common ancestor) but they have been independently lost in different phyla (Basu et al., 2008; Davila Lopez et al., 2008). Trans-acting factors of the minor spliceosome include U11, U12, U4atac and U6atac snRNPs, which replace U1, U2, U4 and U6 snRNPs respectively (both spliceosomes share U5 snRNP) (Schneider et al., 2002). The minor spliceosome is also characterized by less efficient and slower splicing and are rarely alternatively splicing (Lin et al., 2010; Turunen et al., 2013)

Alternative Splicing

For its proper function, the spliceosome needs to be recruited in the right place and at the right time. In non-yeast eukaryotes, the low degree of conservation of cis-acting sequence elements, combined with the diversity of intron lengths, makes the recognition of the proper splicing sequences more difficult. Cells have capitalized on this weaker recognition of the splicing sequences to regulate splicing, generating different mature transcripts from the same original pre-mRNA. While some introns in the pre-mRNA will always be spliced (constitutive splicing), the splicing of some other introns or exons will depend on the context (alternative splicing).

There are four major types of alternative splicing events:

-Exon Skipping: the most common type of alternative splicing event, up to 40% of events in eukaryotes, where a cassette exon is spliced out with their flanking introns (Alekseyenko et al., 2007 ; Sugnet et al., 2004). An example of this type of event relevant for the work in this thesis occurs in NUMB exon 9, whose inclusion or skipping directly impacts on cell proliferation (Bechara et al., 2013).

-Alternative 5'/3' Splice site: some transcripts contain more than one 5' or 3' splice site within the exon and, depending on the context, the spliceosome will utilize one or another. In some cases, the difference between one SS and another represents changes as limited as one amino acid in the protein, as it is the case for the gene RBM10 that will be discussed later on in the thesis (Hernandez et al., 2016).

-Intron retention: in this case an intron is not spliced out of the transcript. Although not very common in metazoan, it is the most frequent type of alternative splicing event in plants, fungi and protozoa (Alekseyenko et al., 2007 ; Sakabe et al., 2007 ; Sugnet et al., 2004).

-Mutually exclusive exons: these events occur when a choice between two alternative exons is made, such that only one of them is found in the mature mRNA. An example of this type of events is found on the pyruvate kinase gene, that contains two mutually exclusive exons: exon 9, that codes for the adult isoform of the protein, and exon 10, that codes for the embryonic version; there is a switch in this splicing event associated with altered energy metabolism of tumor cells (see below).

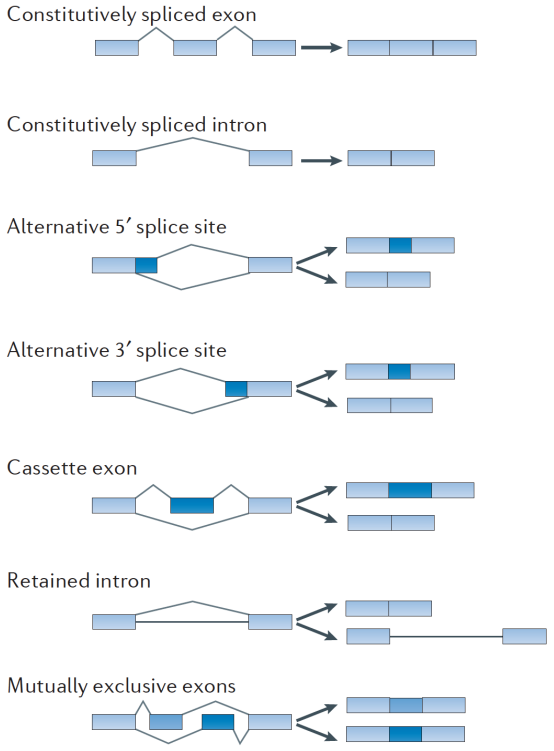


Figure 3 Types of Alternative Splicing Events. Graphical representation of the most common types of alternative splicing events. Adapted from (Dvinge et al., 2016).

- Complex events: several splicing events can be combined together in the same gene to generate tremendously complex combinatorial patterns. The oncogene MDM2, involved in the regulation of p53, is a good example. MDM2 combines 8 alternative exons in tandem, and its alternative splicing has been related with its capacity to regulate p53 levels (Jeyaraj et al., 2009).

The modulation of alternative splicing requires additional cis- and trans-acting elements and the splicing outcome is determined by the combinatorial effect of multiple elements. Aside from the constitutive cis-acting elements (5'SS, BP and 3'SS) other sequences in the pre-mRNA will help or hamper the recruitment of the spliceosome. These cis-acting elements can be intronic (Intronic Splicing Silencers, ISS or Intronic Splicing Enhancers, ISE) or exonic (Exonic Splicing Silencers, ESS or Exonic Splicing Enhancers, ESE) (Wahl et al., 2009). Splicing regulatory factors bind to these sequence elements and their regulatory roles depend on the binding affinity, the position of their binding sites relative to the alternative splice sites (RNA Maps) and their functional interactions with other regulatory sequences and factors (Lee et al., 2015). As a rule of thumb SR proteins will positively promote splicing (Zahler et al., 1993) while hnRNPs will hinder the recruitment of the spliceosome (Buratti et al., 2004; De Conti et al., 2012). As mentioned above, accessory

factors play different roles depending of where they bind, RNA maps that help to predict the regulation of the alternative splicing by several factors, for instance, by RBM proteins (RBM5, RBM6 and RBM10 (Bechara et al., 2013)).

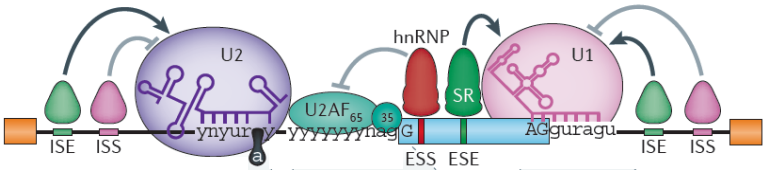


Figure 4 Regulation of Alternative Splicing through cis-acting elements. The combinatorial effect of the splicing factors is a key determinant in the splicing outcome. Splicing Enhancers are cis-acting elements that promote exon inclusion while Splicing Silencers are sequence motifs that promote exon skipping. These types of regulatory sequences can be found in exons (ESE and ESS) or in introns (ISE and ISS). Adapted from (Scotti et al., 2016).

Accessory components can also link splicing with other processes such as transcription or chromatin modification As most of the introns are co-transcriptionally spliced (Brugiolo et al., 2013 ; Tilgner et al., 2012), the elongation rate of the polymerase can play a role altering splice site choices (Kornblihtt, 2007). The timing of appearance of the splicing-relevant sequences generates a kinetic competition between the splicing factors recognizing them. Additionally, the CTD domain of Pol II has been described to interact with several splicing factors (McCracken et al., 1997), also coupling transcription and splicing. Moreover, epigenetic modifications can recruit splicing factors to the pre-mRNA through bridging factors (Luco et al., 2011) and certain epigenetic changes can

modulate the nucleosome positioning or the polymerase elongation rate, also affecting alternative splicing (Iannone et al., 2015; Kornblihtt et al., 2013).

Evolution and physiological function of Alternative Splicing

The high prevalence of alternative splicing in higher eukaryotes provides cells with the capacity to expand their proteome without increasing the size of the genome and allows for complex regulation of the gene expression. A correlation has been reported between organism complexity and alternative splicing: only 25% of *Caenorhabditis elegans* genes display alternative splicing, 45% in *Drosophila melanogaster,* 63% in *Mus musculus* and over 88% in *Homo* sapiens in the 88% (based on Ensembl annotation, not on RNA-sequencing data) (Lee et al., 2015).

While the benefits of alternative splicing in terms of protein diversity and regulatory potential are obvious, understanding the immediate benefits that alternative splicing could have brought to cells when it was first acquired are less clear. One model is that ancestors of spliceosomal introns existed since early evolution, as “virus-like” genetic elements. Those mobile elements, progenitors of the self-catalytic group II introns, would have been present in the genome of the α-proteobacteria, progenitor of the mitochondria, during the process of endosymbiosis (this line of thought is supported by the presence of small numbers of group II introns in the

genome of bacteria). Comparative studies point to the possibility that the last eukaryotic common ancestor (LECA) had an intron rich genome, implying it had a splicing system that could be mostly based on self-catalytic introns (Koonin et al., 2013; Rogozin et al., 2012 ; Roy et al., 2009) Spliceosomal introns in LECA would have evolved and thrived from the autocatalytic group II introns present in the genome of the α-proteobacteria. It has been hypothesized that the process of intronization of the genome will have triggered key events for the evolution of the eukaryotic cells, such as the appearance of a nuclear membrane, the linearity of chromosomes, the origin of telomers and telomerases or the ubiquitination system (Koonin, 2006 ; Martin et al., 2006).

With time, the evolution of efficient trans-acting factors, capable of removing those “group II –like” introns, would have released the evolutionary pressure from the intronic sequences, allowing a large degree of degeneration on its sequence. The degeneration of the cis-elements of the autocatalytic introns, together with the evolution of trans-acting factors, allowed the evolution of a progressively more complex spliceosome machinery (Irimia & Roy, 2014; Koonin, 2006). The function of this first rudimentary spliceosome could have been the control of gene expression by silencing genes through intron retention, a mechanism still prevalent in certain organisms (Irimia & Roy, 2014).

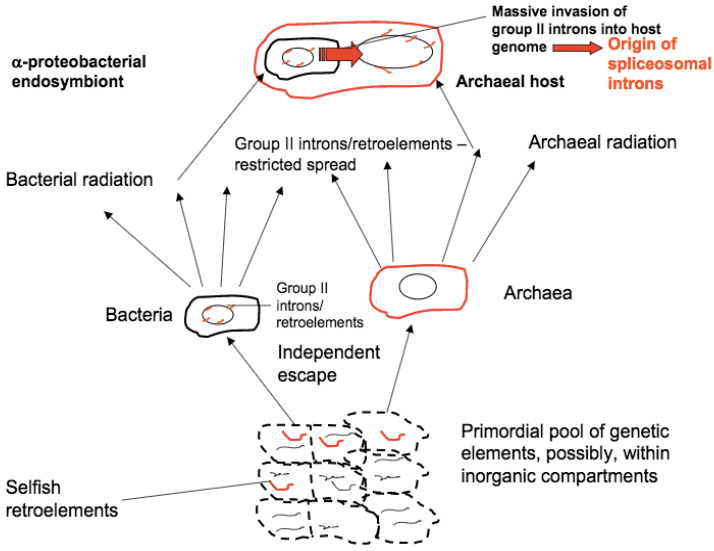


Figure 5 Evolution of Spliceosomal Introns. Representation of the hypothetical origin of the spliceosomal introns from a primordial pool of genetic elements. Adapted from (Koonin, 2006).

Despite the commonalities between the splicing in all eukaryotic branches, divergences in certain evolutionary linages have occurred. A striking example is the prevalence of exon skipping in metazoa, while intron retention remains the most common type of alternative splicing in the rest of eukaryotes (McGuire et al., 2008). Likely related to this, most animal cassette exons have a number of nucleotides multiple of 3, so its inclusion or skipping does not perturb the open reading frame, facilitating the generation of protein variants containing or lacking specific domains (Gilbert, 1978; Xing et al., 2005).

Alternative exons can influence mRNA stability, or code for functional domains, protein-protein interaction domains, localization signals or premature stop codons. For example, FAS pre-mRNA contains a cassette exon 6 encoding a trans-membrane domain. Exon 6 inclusion leads to an mRNA that codes for a membrane-bound form of the receptor that facilitates apoptosis, while FAS exon 6 skipping generates a soluble isoform that is secreted, repressing programmed cell death by squelching the Fas ligand in the extracellular milieu (Cascino et al., 1996). Different isoforms of a protein can have different functions; in fact some studies suggest that different isoforms are so different between them as if they were different genes (Yang et al., 2016).

There are also tissue-specific splicing programs. An example is PTB/nPTB, the Polypyrimidine Tract Binding Protein and its close paralog nPTB; nPTB and PTB display an almost mutually-exclusive expression pattern. While nPTB is expressed mainly in post-mitotic neurons, PTB is expressed in neural precursors and most other cell types (Ashiya et al., 1997; Kikuchi et al., 2000; Markovtsov et al., 2000; Polydorides et al., 2000). PTB controls its own protein levels by inducing the skipping of exon 11 in its pre-mRNA, leading to a premature stop codon (Wollerton et al., 2004). PTB also modulates nPTB splicing in non-neuronal cells, leading to exon skipping and generating an mRNA containing premature stop codons which is subject to NMD (Wollerton et al., 2004);

therefore, PTB modulates its own protein levels and prevents nPTB expression. PTB can repress splicing either by preventing the binding of spliceosome (Oberstrass et al., 2005) or by affecting early events in intron or exon definition (Izquierdo et al., 2005; Sharma et al., 2005; Sharma et al., 2008). During the process of neuron maturation there is an early switch between the expression of PTB and nPTB. This switch activates a network of spliced isoforms in the developing neurons, involving a set of exons that are more sensitive to PTB repression (like c-src N1 exon and exon 8A of CaV1.2 (Boutz et al., 2007; Tang et al., 2011). As nPTB-mediated repression is weaker than PTB, nPTB will maintain the repression of splicing only in a subset of those exons repressed by PTB. This suggests that PTB down-regulation is a major factor in establishing the splicing-mediated neuronal differentiation program. Consistent with this concept, PTB knock down can confer neuronal phenotypic features to cells in culture or other origin (Xue et al., 2013).

One of the most recent and exciting pieces of research about tissue specific alternative splicing comes from the study of neural microexons. Microexons are small exons, from 3 to 27 nucleotides, that code for short stretches of amino acids (1-9), which can modulate the function of interacting domains on the surface of proteins, often neurogenesis-related. Strikingly, the flanking sequences of microexons, their tissue-specific regulation and developmental alternative splicing are the

most conserved among all alternative splicing types/programs in vertebrates. Deregulated splicing of some of these microexons has been associated with several mental illnesses like autism or intellectual disability (Irimia, Weatheritt, et al., 2014). Most of these microexons are regulated by the protein nSR100, which is downregulated in autism spectrum disorder. Knockout mice for nSR100 display impaired nervous system development and broad changes in the alternative splicing of microexons in neurons. Restoration of a 6-nt microexon splicing in the gene Unc13b reverts the growth defects on nSR100 mutant primary neurons (Quesnel-Vallieres et al., 2015). The discovery of this alternative splicing program opens the possibility of studying a whole new set of splicing events that are not only related to the development of the nervous system but also linked to mental disabilities.

As it can be inferred from the above, the complexity of alternative splicing not only allows for tissue-specific splicing but also for developmentally regulated splicing. The expression of certain protein isoforms is limited to specific moments during development, like embryonic PKM. The PKM gene encodes pyruvate kinase, the enzyme that catalyzes the synthesis of ATP from ADP and phosphoenolpyruvate, one of the last steps of the glycolytic pathway. PKM pre-mRNA has two mutually exclusive exons: inclusion of exon 9 generates the PKM1 isoform, while inclusion of exon 10 generates

PKM2. PKM1 is expressed in adult tissues and activates oxidative phosphorylation, while PKM2 is expressed in embryonic tissues and promotes anaerobic glycolysis. Importantly, a splicing switch from PKM1 to PKM2 occurs in certain tumors, conferring them the capacity to carry out aerobic glycolysis and an energetic advantage that can explain the Warburg effect (Christofk et al., 2008; Warburg, 1956). It has also been reported that c-Myc, a well-known oncogene, upregulates PTB, hnRNPA1 and hnRNPA2, three regulators of PKM splicing (Christofk et al., 2008; C. J. David et al., 2010), highlighting the importance of understanding splicing regulatory networks in cancer (see below).

Cancer and Alternative Splicing

The cooperative effects in splice site recognition and alternative splicing respond to the need for tight regulation, but they also make splicing susceptible to be altered in disease. Examples include beta-thalassemia, spinal muscular atrophy, Duchenne muscular dystrophy, retinitis pigmentosa, frontotemporal dementia, cystic fibrosis and cancer (Cooper et al., 2009). Cancer cells capitalize on alterations of gene expression with a positive impact on cell proliferation, evading apoptosis, invading other tissues, etc. Since splicing is a major mechanism of gene regulation subject to tight control, it is not surprising that a plethora of studies link alteration in alternative splicing and cancer (Charles J. David et al., 2010; Kaida et al., 2012)

Cancers were originally considered a homogeneous mass of proliferative cells isolated from their environment. Today cancers are considered complex tissues, with their own cellular organization and distinctive molecular profiles (reviewed in (Hanahan et al., 2011)). Carcinogenesis is a stepwise process by which premalignant cells accumulate alterations to overcome the different control mechanisms that maintain tissue homeostasis (Hanahan et al., 2011). In fact, due to its progressive nature, the process of carcinogenesis can be seen as a Darwinian evolution in which pre-malignant cells compete against the regulatory pathways that control tissue homeostasis; the acquisition of pernicious mutations in

cancer cells is a constant tug of war against the organism's control mechanisms. Those cancer cells acquiring capacities to hijack the organism defenses will thrive, proliferate and be positively selected; in contrast, premalignant cells unable to bypass the control defenses will be eliminated. The progressive acquisition of tumoral capacities will lead to the development of malignant neoplasias that will evolve into a developed cancer.

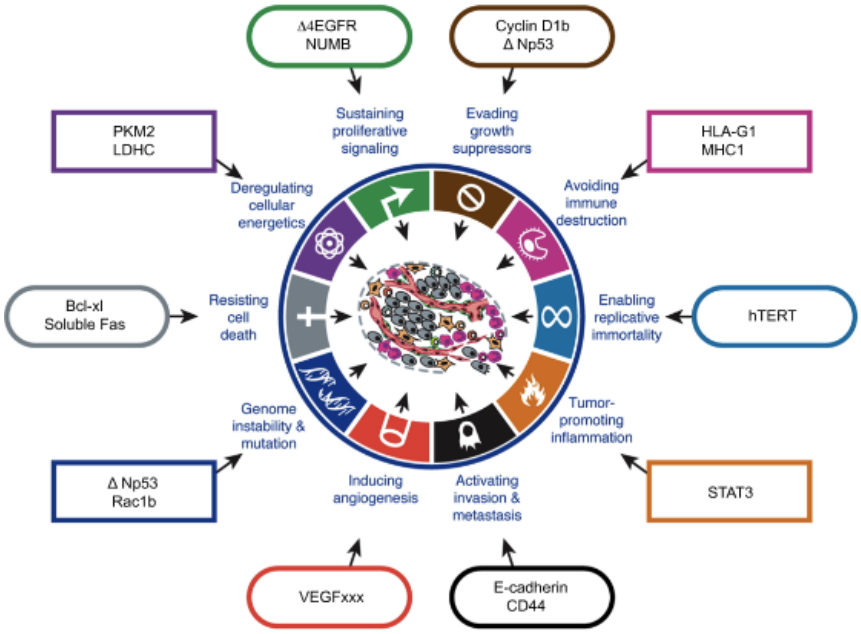


Figure 6 Examples of Alternative Splicing contributing to the different Hallmarks of Cancer. The Hallmarks of cancer are a compendium of pathways that play crucial roles in tumorigenesis and tumor progression. We present examples of genes, belonging to these pathways, whose alternative splicing has been related to cancer. Modified from (Hanahan et al., 2011).

Although the process of carcinogenesis requires uncontrolled proliferation of cells, other cellular, biochemical and molecular factors are common to almost all cancer types and are important for their development. Cancer hallmarks include the proliferation of cells (sustained proliferation, evasion of growth suppression, resistance to cell death and replicative immortality), the sustained growth and dissemination of the tumor mass through the body (invasiveness, angiogenesis, and metabolic alterations), the interplay between tumor and immune system (avoidance of immune destruction and tumor promoting inflammation) and cancer cell susceptibility to mutations (genomic instability) (Hanahan et al., 2000, 2011)

Each cancer hallmark is a compendium of biological processes controlling a specific aspect of cell biology. Alterations of these pathways can lead to the progressive accumulation of cancer-associated traits. In fact, deregulated splicing events have been related with different cancer-associated pathways (Fig.6, Hallmarks of alternative splicing in cancer). Alternative splicing alterations associated with tumors can be due to alterations in splicing factors or to alterations in sequences important for the splicing events.

Although in general the complex phenotypes of cancer cells can not be explained by splicing alterations in one unique gene, several examples of misspliced genes that account for important cancer traits have been reported. An important

example of misspliced genes that significantly contribute to cancer proliferation is the intracellular inhibitor of the Notch pathway NUMB. NUMB gene contains a cassette exon 9 whose inclusion is increased in several types of malignancies, including breast, lung and bladder (Misquitta-Ali et al., 2011; Sebestyén et al., 2015). Experiments in cultured cells showed that the expression of one or the other isoform controls the clonogenic capacity of cancer cells, an indirect measure of their proliferative capacity (Bechara et al., 2013; Misquitta-Ali et al., 2011).

Splicing factors as oncogenes and tumor suppressors

One example of how alterations in a splicing factor contribute to cancer is RBM10, a splicing factor that is mutated in several types of cancer including breast, bladder and among the most frequently mutated proteins in lung adenocarcinoma (e.g. in 8% of LUAD tumors (Collisson et al., 2014)). Like other splicing factors the positional binding of RBM10 can have different effects on different events (Bechara et al., 2013) but its effect on NUMB exon 9 splicing has been characterized (Bechara et al., 2013; Hernandez et al., 2016). Several RBM10 mutants have been reported to display impaired capacity to regulate NUMB exon 9 splicing, which can contribute (probably along with other alternative splicing events) to the cell proliferation effects.

Another illustrative example of the complex interconnection between splicing and cancer is the SR protein SRSF1. Considered as a potent oncogene, SRSF1 is upregulated in different human tumors (Karni et al., 2007), and its overexpression promotes the transformation of mammary epithelial cells and rodent fibroblasts (Anczuków et al., 2015; Karni et al., 2007). Analogously to what happens with RBM10 and NUMB, SRSF1 modulates CASC4 exon 9 inclusion, which can partially recapitulate the transformation activities of SRSF1 (Anczukow et al., 2012). Upon SRSF1 overexpression several of its target genes are misspliced: BIN1 is spliced to a non-antitumor isoform; MNK2 splicing is altered, the levels of the antitumor isoform MNK2a are reduced, preventing p38 translocation into the nucleus but still enhancing eIF4E phosphorylation (linking the miss-regulation of splicing and translational control in cancer) (Maimon et al., 2014); in addition, S6K1 isoform 2 is generated, again recapitulating SRSF1 transforming capacities (Karni et al., 2007). SRSF1 oncogenic properties are also related with the mTOR pathway (Karni et al., 2008) and with a constitutively active isoform of RON, that plays a role in the invasiveness of cancer cells (Ghigna et al., 2005). SRSF1 is a good example of how complex the relationship between cancer and splicing can be, with alterations in splicing factors altering multiple cancer-relevant processes and hallmarks.

Lung adenocarcinomas illustrate well the impact of alternative splicing regulation in cancer progression, with alterations in both splicing regulators and splicing events and with a direct connection between altered regulators and altered events. Studies from the Cancer Genome Atlas Research Network sequencing 230 lung adenocarcinoma samples revealed that of the 18 genes significantly mutated, two of them encode splicing factors: RBM10, mutated in 8% of the samples and U2AF1, mutated in 3% of the tumors (Collisson et al., 2014). RBM10 and U2AF35 (encoded by U2AF1) implicated in 3'SS recognition (Bechara et al., 2013; Merendino et al., 1999; Wu et al., 1999; Zorio et al., 1999). 3'SS recognition seems to be indeed linked with proliferation control, because other factors that also contribute to this process, like SF3B1 and ZRSR2, have been found mutated in hematological malignancies (reviewed in (Yoshida et al., 2014)). Alteration of specific splicing events have also been linked to lung adenocarcinoma; MET splicing was altered in 4% of the samples in the Cancer Genome study (Collisson et al., 2014); while, using alternative splicing microarrays and a different cohort of patients alterations in splicing events in lung cancer have been reported for VEGFA, MACF1, APP and NUMB (Misquitta-Ali et al., 2011). Of relevance, a link between the mutational state of RBM10, a frequently mutated regulator, and NUMB, an altered splicing event, has been documented (Bechara et al., 2013) also as part of this research (Hernandez et al., 2016).

The interplay between cancer and aberrant splicing is also illustrated by the fact that some antitumor drugs modulate splicing. Screenings of natural compounds derived from bacterial fermentation products led to the identification of FR901464, Pladienolides, GEX1A and Thailanstatines, antitumour compounds that target the spliceosome both *in vitro* and *in vivo* (Liu et al., 2013; Miller-Wideman et al., 1992; Nakajima et al., 1996; Sakai et al., 2004). The interest on these compounds is not only therapeutic because they can be also useful tools to study the splicing reaction. Spliceostatin A, a product derived from fermentation products of *Pseudomonas*, has been shown to bind to SF3b, and its effects on splicing are similar to those observed by the knock-down of SR155; among those genes, there were several cell proliferation related genes, such as Aurora A kinase (Corrionero et al., 2011). This points and other data towards the antitumor effect of the drug being mediated by its effect on perturbing the fidelity of splice site recognition.

Other therapeutic approaches to correct splicing-related diseases include the use of modified antisense oligonucleotides (AONs) for correcting aberrant splicing variants. Important progress has been made for example in the development of AONs for the treatment of Spinal Muscular Atrophy (SMA), a disease caused by inactivation of the SMN1 gene. A second copy of the gene cannot provide function because exon 7 is mainly skipped, generating a

protein isoform that is not stable (Cho et al., 2010). Exon 7 of SMN1 and SMN2 differs only by a single nucleotide, C/U in position +6. This unique difference converts an ESE in SMN1 to an ESS in SMN2 by generating a binding site for the repressor hnRNP A1, resulting in the skipping of the exon 7 (Kashima et al., 2003). Results from clinical trials using AONs to block an intronic splicing repressor element (enhancing SMN2 exon 6 inclusion) are extremely encouraging [http://ir.ionispharma.com/phoenix.zhtml?c=222170&p=irol-newsArticle&ID=2191319].

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J. C., Mann, M., & Lamond, A. I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO Journal, 19*(23), 6569-6581. doi:10.1093/emboj/19.23.6569

Alekseyenko, A. V., Kim, N., & Lee, C. J. (2007). Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA, 13*(5), 661-670. doi:10.1261/rna.325107

Alioto, T. S. (2007). U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Research, 35*(Database issue), D110-115. doi:10.1093/nar/gkl796

Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shirole, N. H., . . . Krainer, A. R. (2015). SRSF1-Regulated Alternative Splicing in Breast Cancer. *Molecular Cell, 60*(1), 105-117. doi:10.1016/j.molcel.2015.09.005

Anczukow, O., Rosenberg, A. Z., Akerman, M., Das, S., Zhan, L., Karni, R., . . . Krainer, A. R. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature Structural & Molecular Biology, 19*(2), 220-228. doi:10.1038/nsmb.2207

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., . . . Olson, E. N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell, 160*(4), 595-606. doi:10.1016/j.cell.2015.01.009

Ashiya, M., & Grabowski, P. J. (1997). A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA, 3*(9), 996-1015.

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation By a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type Iii. *The Journal of experimental medicine, 79*(2), 137-158. doi:10.1084/jem.79.2.137

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature, 226*(5252), 1209-1211.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell, 116*(2), 281-297.

Basu, M. K., Makalowski, W., Rogozin, I. B., & Koonin, E. V. (2008). U12 intron positions are more strongly conserved

between animals and plants than U2 intron positions. *Biol Direct, 3*, 19. doi:10.1186/1745-6150-3-19

Beadle, G. W., & Tatum, E. L. (1941). Genetic Control of Biochemical Reactions in Neurospora. *Proceedings of the National Academy of Sciences of the United States of America, 27*(11), 499-506. doi:10.1086/281267

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E., & Valcárcel, J. (2013). RBM5, 6, and 10 Differentially Regulate NUMB Alternative Splicing to Control Cancer Cell Proliferation. *Molecular Cell, 52*(5), 720-733. doi:10.1016/j.molcel.2013.11.010

Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America, 74*(8), 3171-3175. doi:10.1073/pnas.74.8.3171

Berglund, J. A., Rosbash, M., & Schultz, S. C. (2001). Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA, 7*(5), 682-691.

Berk, A. J., & Sharp, P. A. (1978). Structure of the adenovirus 2 early mRNAs. *Cell, 14*(3), 695-711. doi:10.1016/0092-8674(78)90252-0

Boutz, P. L., Stoilov, P., Li, Q., Lin, C. H., Chawla, G., Ostrow, K., . . . Black, D. L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes & Development, 21*(13), 1636-1652. doi:10.1101/gad.1558107

Brachet, J. (1954). Effects of ribonuclease on the metabolism of living root-tip cells. *Nature, 174*(4436), 876-877.

Brachet, J. (1955). Action of ribonuclease and ribonucleic acid on living amoebae. *Nature, 175*(4463), 851-853.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., & Willard, H. F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature, 349*(6304), 38-44. doi:10.1038/349038a0

Brugiolo, M., Herzel, L., & Neugebauer, K. M. (2013). Counting on co-transcriptional splicing. *F1000Prime Rep, 5*, 9. doi:10.12703/P5-9

Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y. M., & Baralle, F. E. (2004). hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Research, 32*(14), 4224-4236. doi:10.1093/nar/gkh752

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., . . . Gustincich, S. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature, 491*(7424), 454-457. doi:10.1038/nature11508

Cascino, I., Papoff, G., De Maria, R., Testi, R., & Ruberti, G. (1996). Fas/Apo-1 (CD95) receptor lacking the intracytoplasmic signaling domain protects tumor cells from Fas-mediated apoptosis. *Journal of Immunology, 156*(1), 13-17.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., . . . Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell, 147*(2), 358-369. doi:10.1016/j.cell.2011.09.028

Chan, S. P., & Cheng, S. C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *Journal of Biological Chemistry, 280*(35), 31190-31199. doi:10.1074/jbc.M505060200

Chan, S. P., Kao, D. I., Tsai, W. Y., & Cheng, S. C. (2003). The Prp19p-associated complex in spliceosome activation. *Science, 302*(5643), 279-282. doi:10.1126/science.1086602

Cho, S., & Dreyfuss, G. (2010). A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity. *Genes & Development, 24*(5), 438-442. doi:10.1101/gad.1884910

Chow, L. T., Roberts, J. M., Lewis, J. B., & Broker, T. R. (1977). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. *Cell, 11*(4), 819-836. doi:10.1016/0092-8674(77)90294-X

Christofk, H. R., Vander Heiden, M. G., Harris, M. H., Ramanathan, A., Gerszten, R. E., Wei, R., . . . Cantley, L. C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature, 452*(7184), 230-233. doi:10.1038/nature06734

Chung, S., McLean, M. R., & Rymond, B. C. (1999). Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA, 5*(8), 1042-1054.

Collisson, E. A., Campbell, J. D., Brooks, A. N., Berger, A. H., Lee, W., Chmielecki, J., . . . Tsao, M.-S. (2014). Comprehensive

molecular profiling of lung adenocarcinoma. *Nature, 511*(7511), 543-550. doi:10.1038/nature13385

Cooper, T. a., Wan, L., & Dreyfuss, G. (2009). RNA and disease. *Cell, 136*(4), 777-793. doi:10.1016/j.cell.2009.02.011

Corrionero, A., Minana, B., & Valcarcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes & Development, 25*(5), 445-459. doi:10.1101/gad.2014311

Crick, F. (1958). On Protein Synthesis. *The Symposia of the Society for Experimental Biology*, 138-166.

Crick, F. H., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature, 192*, 1227-1232. doi:10.1038/1921227a0

Crispino, J. D., Blencowe, B. J., & Sharp, P. A. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science, 265*(5180), 1866-1869.

David, C. J., Boyne, A. R., Millhouse, S. R., & Manley, J. L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes & Development, 25*(9), 972-983. doi:10.1101/gad.2038011

David, C. J., Chen, M., Assanah, M., Canoll, P., & Manley, J. L. (2010). HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature, 463*(7279), 364-368. doi:10.1038/nature08697

David, C. J., & Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & Development, 24*(21), 2343-2364. doi:10.1101/gad.1973010

Davila Lopez, M., Rosenblad, M. A., & Samuelsson, T. (2008). Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. *Nucleic Acids Research, 36*(9), 3001-3010. doi:10.1093/nar/gkn142

De Conti, L., Skoko, N., Buratti, E., & Baralle, M. (2012). Complexities of 5'splice site definition: implications in clinical analyses. *RNA Biol, 9*(6), 911-923. doi:10.4161/rna.20386

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., . . . Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under

physiological conditions. *Molecular and Cellular Biology, 26*(14), 5528-5543. doi:10.1128/MCB.00582-06

Del Gatto-Konczak, F., Bourgeois, C. F., Le Guiner, C., Kister, L., Gesnel, M. C., Stevenin, J., & Breathnach, R. (2000). The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Molecular and Cellular Biology, 20*(17), 6287-6299.

Dlakic, M., & Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA, 17*(5), 799-808. doi:10.1261/rna.2396011

Dvinge, H., Kim, E., Abdel-Wahab, O., & Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews Cancer*. doi:10.1038/nrc.2016.51

Fica, S. M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., . . . Piccirilli, J. a. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature, 503*(7475), 229-234. doi:10.1038/nature12734

Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., . . . Piccirilli, J. A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature, 503*(7475), 229-234. doi:10.1038/nature12734

Forch, P., Puig, O., Kedersha, N., Martinez, C., Granneman, S., Seraphin, B., . . . Valcarcel, J. (2000). The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Molecular Cell, 6*(5), 1089-1098.

Galej, W. P., Oubridge, C., Newman, A. J., & Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature, 493*(7434), 638-643. doi:10.1038/nature11843

Garapin, A. C., Cami, B., Roskam, W., Kourilsky, P., Le Pennec, J. P., Perrin, F., . . . Chambon, P. (1978). Electron microscopy and restriction enzyme mapping reveal additional intervening sequences in the chicken ovalbumin split gene. *Cell, 14*(3), 629-639.

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P. M., . . . Biamonti, G. (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Molecular Cell, 20*(6), 881-890. doi:10.1016/j.molcel.2005.10.026

Gilbert, W. (1978). Why genes in pieces? *Nature, 271*(5645), 501.

Goldstein, L., & Plaut, W. (1955). Direct Evidence for Nuclear Synthesis of Cytoplasmic Ribose Nucleic Acid. *Proc Natl Acad Sci U S A, 41*(11), 874-880.

Gozani, O., Potashkin, J., & Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Molecular and Cellular Biology, 18*(8), 4752-4760.

Grainger, R. J., & Beggs, J. D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA, 11*(5), 533-557. doi:10.1261/rna.2220705

Grote, M., Wolf, E., Will, C. L., Lemm, I., Agafonov, D. E., Schomburg, A., . . . Luhrmann, R. (2010). Molecular architecture of the human Prp19/CDC5L complex. *Molecular and Cellular Biology, 30*(9), 2105-2119. doi:10.1128/MCB.01505-09

Hall, S. L., & Padgett, R. A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *Journal of Molecular Biology, 239*(3), 357-365. doi:10.1006/jmbi.1994.1377

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell, 100*(1), 57-70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Hernandez, J., Bechara, E., Schlesinger, D., Delgado, J., Serrano, L., & Valcarcel, J. (2016). Tumor suppressor properties of the splicing regulatory factor RBM10. *RNA biology, 13*(4), 466-472. doi:10.1080/15476286.2016.1144004

Hooke, R. (1665). Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses, with Observations and Inquiries Thereupon. .

Iannone, C., Pohl, A., Papasaikas, P., Soronellas, D., Vicent, G. P., Beato, M., & ValcaRcel, J. (2015). Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. *RNA, 21*(3), 360-374. doi:10.1261/rna.048843.114

Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell, 147*(4), 789-802. doi:10.1016/j.cell.2011.10.002

Irimia, M., & Roy, S. W. (2014). Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol, 6*(6). doi:10.1101/cshperspect.a016071

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., . . . Blencowe, B. J. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell, 159*(7), 1511-1523. doi:10.1016/j.cell.2014.11.035

Izquierdo, J. M., Majos, N., Bonnal, S., Martinez, C., Castelo, R., Guigo, R., . . . Valcarcel, J. (2005). Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Molecular Cell, 19*(4), 475-484. doi:10.1016/j.molcel.2005.06.015

Jackson, R. J., & Standart, N. (1990). Do the poly(A) tail and 3' untranslated region control mRNA translation? *Cell, 62*(1), 15-24.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3*(3), 318-356. doi:10.1016/S0022-2836(61)80072-7

Jeyaraj, S., O'Brien, D. M., & Chandler, D. S. (2009). MDM2 and MDM4 splicing: an integral part of the cancer spliceome. *Front Biosci (Landmark Ed), 14*, 2647-2656.

Kaida, D., Schneider-Poetsch, T., & Yoshida, M. (2012). Splicing in oncogenesis and tumor suppression. *Cancer Science, 103*(9), 1611-1616. doi:10.1111/j.1349-7006.2012.02356.x

Kandels-Lewis, S., & Seraphin, B. (1993). Involvement of U6 snRNA in 5' splice site selection. *Science, 262*(5142), 2035-2039.

Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., & Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Structural & Molecular Biology, 14*(3), 185-193. doi:10.1038/nsmb1209

Karni, R., Hippo, Y., Lowe, S. W., & Krainer, A. R. (2008). The splicing-factor oncoprotein SF2/ASF activates mTORC1. *Proc Natl Acad Sci U S A, 105*(40), 15323-15327. doi:10.1073/pnas.0801376105

Kashima, T., & Manley, J. L. (2003). A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nature Genetics, 34*(4), 460-463. doi:10.1038/ng1207

Kikuchi, T., Ichikawa, M., Arai, J., Tateiwa, H., Fu, L., Higuchi, K., & Yoshimura, N. (2000). Molecular cloning and characterization of a new neuron-specific homologue of rat polypyrimidine tract binding protein. *J Biochem, 128*(5), 811-821.

Kim, S. H., & Lin, R. J. (1996). Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-

mRNA splicing. *Molecular and Cellular Biology, 16*(12), 6810-6819.

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., & Chrousos, G. P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal, 3*(107), ra8. doi:10.1126/scisignal.2000568

Konarska, M. M., Vilardell, J., & Query, C. C. (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Molecular Cell, 21*(4), 543-553. doi:10.1016/j.molcel.2006.01.017

Kondo, Y., Oubridge, C., van Roon, A. M., & Nagai, K. (2015). Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *Elife, 4*. doi:10.7554/eLife.04986

Koodathingal, P., Novak, T., Piccirilli, J. A., & Staley, J. P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. *Molecular Cell, 39*(3), 385-395. doi:10.1016/j.molcel.2010.07.014

Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct, 1*, 22. doi:10.1186/1745-6150-1-22

Koonin, E. V., Csuros, M., & Rogozin, I. B. (2013). Whence genes in pieces: reconstruction of the exon-intron gene structures of the last eukaryotic common ancestor and other ancestral eukaryotes. *Wiley Interdiscip Rev RNA, 4*(1), 93-105. doi:10.1002/wrna.1143

Koonin, E. V., & Martin, W. (2005). On the origin of genomes and cells within inorganic compartments. *Trends in Genetics, 21*(12), 647-654. doi:10.1016/j.tig.2005.09.006

Kornblihtt, A. R. (2007). Coupling transcription and alternative splicing. *Advances in Experimental Medicine and Biology, 623*, 175-189.

Kornblihtt, A. R., Schor, I. E., Allo, M., Dujardin, G., Petrillo, E., & Munoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Reviews: Molecular Cell Biology, 14*(3), 153-165. doi:10.1038/nrm3525

Kramer, A., & Utans, U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *EMBO Journal, 10*(6), 1503-1509.

Lambowitz, A. M., & Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol, 3*(8), a003616. doi:10.1101/cshperspect.a003616

Leder, P., Tilghman, S. M., Tiemeier, D. C., Polsky, F. I., Seidman, J. G., Edgell, M. H., . . . Norman, B. (1978). The cloning of mouse globin and surrounding gene sequences in bacteriophage lambda. *Cold Spring Harb Symp Quant Biol, 42 Pt 2*, 915-920.

Lee, Y., & Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual Review of Biochemistry, 84*, 291-323. doi:10.1146/annurev-biochem-060614-034316

Lengyel, P., Speyer, J. F., & Ochoa, S. (1961). Synthetic polynucleotides and the amino acid code. *Proc Natl Acad Sci U S A, 47*, 1936-1942.

Lerner, M. R., & Steitz, J. A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences of the United States of America, 76*(11), 5495-5499. doi:10.1073/pnas.76.11.5495

Lesser, C. F., & Guthrie, C. (1993). Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science, 262*(5142), 1982-1988.

Levene, P. A. (1910). On the biochemistry of nucleic acids. *Journal of the American Chemical Society, 32*, 231-240. doi:10.1016/B978-0-12-205350-4.50020-2

Levene, P. A. (1912). Structure of Thymus Nucleic Acid.

Lin, C. F., Mount, S. M., Jarmolowski, A., & Makalowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evolutionary Biology, 10*, 47. doi:10.1186/1471-2148-10-47

Liu, X., Biswas, S., Tang, G. L., & Cheng, Y. Q. (2013). Isolation and characterization of spliceostatin B, a new analogue of FR901464, from Pseudomonas sp. No. 2663. *J Antibiot (Tokyo), 66*(9), 555-558. doi:10.1038/ja.2013.38

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., & Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell, 144*(1), 16-26. doi:10.1016/j.cell.2010.11.056

Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., Smrzka, O. W., Wutz, A., . . . Barlow, D. P. (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nature Genetics, 25*(1), 19-21. doi:10.1038/75546

Maimon, A., Mogilevsky, M., Shilo, A., Golan-Gerstl, R., Obiedat, A., Ben-Hur, V., . . . Karni, R. (2014). Mnk2 alternative splicing modulates the p38-MAPK pathway and impacts Ras-induced transformation. *Cell Rep, 7*(2), 501-513. doi:10.1016/j.celrep.2014.03.041

Makarova, O. V., Makarov, E. M., Urlaub, H., Will, C. L., Gentzel, M., Wilm, M., & Luhrmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO Journal, 23*(12), 2381-2391. doi:10.1038/sj.emboj.7600241

Markovtsov, V., Nikolic, J. M., Goldman, J. A., Turck, C. W., Chou, M. Y., & Black, D. L. (2000). Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Molecular and Cellular Biology, 20*(20), 7463-7479.

Martin, W., & Koonin, E. V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature, 440*(7080), 41-45. doi:10.1038/nature04531

Mayr, E. (1982). The Growth of Biological Thought: Diversity, Evolution, and Inheritance

.

Mazzarello, P. (1999). A unifying concept: the history of cell theory. *Nature Cell Biology, 1*(1), E13-15. doi:10.1038/8964

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., . . . Bentley, D. L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature, 385*(6614), 357-361. doi:10.1038/385357a0

McGuire, A. M., Pearson, M. D., Neafsey, D. E., & Galagan, J. E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol, 9*(3), R50. doi:10.1186/gb-2008-9-3-r50

McKie, A. B., McHale, J. C., Keen, T. J., Tarttelin, E. E., Goliath, R., van Lith-Verhoeven, J. J., . . . Inglehearn, C. F. (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Human Molecular Genetics, 10*(15), 1555-1562.

Mendel, G. (1865). Experiments in Plant Hybridization. *Journal of the Royal Horticultural Society, IV*(1865), 3-47.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., & Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature, 402*(6763), 838-841. doi:10.1038/45602

Miller-Wideman, M., Makkar, N., Tran, M., Isaac, B., Biest, N., & Stonard, R. (1992). Herboxidiene, a new herbicidal substance from Streptomyces chromofuscus A7847. Taxonomy, fermentation, isolation, physico-chemical and biological properties. *J Antibiot (Tokyo), 45*(6), 914-921.

Misquitta-Ali, C. M., Cheng, E., O'Hanlon, D., Liu, N., McGlade, C. J., Tsao, M. S., & Blencowe, B. J. (2011). Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. *Molecular and Cellular Biology, 31*(1), 138-150. doi:10.1128/MCB.00709-10

Nakajima, H., Hori, Y., Terano, H., Okuhara, M., Manda, T., Matsumoto, S., & Shimomura, K. (1996). New antitumor substances, FR901463, FR901464 and FR901465. II. Activities against experimental tumors in mice and mechanism of action. *J Antibiot (Tokyo), 49*(12), 1204-1211.

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, T., Trupin, J., Rottman, F., & O'Neal, C. (1965). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the United States of America, 53*(5), 1161-1168. doi:10.1073/pnas.53.5.1161

Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A, 47*, 1588-1602.

Nishimura, S., Jacob, T. M., & Khorana, H. G. (1964). Synthetic deoxyribopolynucleotides as templates for ribonucleic acid polymerase: the formation and characterization of a ribopolynucleotide with a repeating trinucleotide sequence. *Pnas, 52*(6), 1494-1501.

Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., . . . Allain, F. H. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science, 309*(5743), 2054-2057. doi:10.1126/science.1114066

Papasaikas, P., & Valcárcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends in Biochemical Sciences, 41*(1), 33-45. doi:10.1016/j.tibs.2015.11.003

Polydorides, A. D., Okano, H. J., Yang, Y. Y., Stefani, G., & Darnell, R. B. (2000). A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc Natl Acad Sci U S A, 97*(12), 6350-6355. doi:10.1073/pnas.110128397

Pyle, A. M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys, 37*, 317-336. doi:10.1146/annurev.biophys.37.032807.125908

Quesada, V., Conde, L., Villamor, N., Ordonez, G. R., Jares, P., Bassaganyas, L., . . . Lopez-Otin, C. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics, 44*(1), 47-52. doi:10.1038/ng.1032

Quesada, V., Ramsay, A. J., & Lopez-Otin, C. (2012). Chronic lymphocytic leukemia with SF3B1 mutation. *N Engl J Med, 366*(26), 2530. doi:10.1056/NEJMc1204033

Quesnel-Vallieres, M., Irimia, M., Cordes, S. P., & Blencowe, B. J. (2015). Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes & Development, 29*(7), 746-759. doi:10.1101/gad.256115.114

Rogozin, I. B., Carmel, L., Csuros, M., & Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biol Direct, 7*, 11. doi:10.1186/1745-6150-7-11

Roy, S. W., & Irimia, M. (2009). Splicing in the eukaryotic ancestor: form, function and dysfunction. *Trends in Ecology & Evolution, 24*(8), 447-455. doi:10.1016/j.tree.2009.04.005

Ruskin, B., Zamore, P. D., & Green, M. R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell, 52*(2), 207-219.

Sakabe, N. J., & de Souza, S. J. (2007). Sequence features responsible for intron retention in human. *BMC Genomics, 8*, 59. doi:10.1186/1471-2164-8-59

Sakai, T., Sameshima, T., Matsufuji, M., Kawamura, N., Dobashi, K., & Mizui, Y. (2004). Pladienolides, new substances from culture of Streptomyces platensis Mer-11107. I. Taxonomy, fermentation, isolation and screening. *J Antibiot (Tokyo), 57*(3), 173-179.

Schneider, C., Will, C. L., Makarova, O. V., Makarov, E. M., & Luhrmann, R. (2002). Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Molecular and Cellular Biology, 22*(10), 3219-3229.

Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews: Genetics, 17*(1), 19-32. doi:10.1038/nrg.2015.3

Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M. A., . . . Eyras, E. (2015). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils

novel cancer-relevant splicing networks. *bioRxiv*, 023010-023010. doi:10.1101/023010

Sharma, S., Falick, A. M., & Black, D. L. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Molecular Cell, 19*(4), 485-496. doi:10.1016/j.molcel.2005.07.014

Sharma, S., Kohlstaedt, L. A., Damianov, A., Rio, D. C., & Black, D. L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nature Structural & Molecular Biology, 15*(2), 183-191. doi:10.1038/nsmb.1375

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R., & Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research, 34*(14), 3955-3967. doi:10.1093/nar/gkl556

Smith, D. J., Konarska, M. M., & Query, C. C. (2009). Insights into branch nucleophile positioning and activation from an orthogonal pre-mRNA splicing system in yeast. *Molecular Cell, 34*(3), 333-343. doi:10.1016/j.molcel.2009.03.012

Smith, D. J., Query, C. C., & Konarska, M. M. (2008). "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Molecular Cell, 30*(6), 657-666. doi:10.1016/j.molcel.2008.04.013

Spizzo, R., Nicoloso, M. S., Croce, C. M., & Calin, G. A. (2009). SnapShot: MicroRNAs in Cancer. *Cell, 137*(3), 586-586 e581. doi:10.1016/j.cell.2009.04.040

Staley, J. P., & Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Molecular Cell, 3*(1), 55-64.

Steitz, T. A., & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A, 90*(14), 6498-6502.

Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., & Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Molecular Cell, 9*(1), 31-44.

Sugnet, C. W., Kent, W. J., Ares, M., Jr., & Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput*, 66-77.

Tang, Z. Z., Sharma, S., Zheng, S., Chawla, G., Nikolic, J., & Black, D. L. (2011). Regulation of the mutually exclusive exons 8a and 8 in the CaV1.2 calcium channel transcript by

polypyrimidine tract-binding protein. *Journal of Biological Chemistry, 286*(12), 10007-10016. doi:10.1074/jbc.M110.208116

Tardiff, D. F., & Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. *RNA, 12*(6), 968-979. doi:10.1261/rna.50506

Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V., & Leder, P. (1978). Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. *Proc Natl Acad Sci U S A, 75*(2), 725-729.

Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakrabortty, S., Djebali, S., . . . Guigo, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Research, 22*(9), 1616-1625. doi:10.1101/gr.134445.111

Tseng, C. K., & Cheng, S. C. (2008). Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science, 320*(5884), 1782-1784. doi:10.1126/science.1158993

Tseng, C. K., Liu, H. L., & Cheng, S. C. (2011). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA, 17*(1), 145-154. doi:10.1261/rna.2459611

Turunen, J. J., Niemela, E. H., Verma, B., & Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA, 4*(1), 61-76. doi:10.1002/wrna.1141

Valadkhan, S., & Jaladat, Y. (2010). The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. *Proteomics, 10*(22), 4128-4141. doi:10.1002/pmic.201000354

van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P. B., de Bruijn, E., . . . Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol, 15*(1), R6. doi:10.1186/gb-2014-15-1-r6

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell, 136*(4), 701-718. doi:10.1016/j.cell.2009.02.009

Wang, L., Lawrence, M. S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., . . . Wu, C. J. (2011). SF3B1 and other novel

cancer genes in chronic lymphocytic leukemia. *N Engl J Med, 365*(26), 2497-2506. doi:10.1056/NEJMoa1109016

Wang, Q., Hobbs, K., Lynn, B., & Rymond, B. C. (2003). The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts. *Journal of Biological Chemistry, 278*(10), 7875-7883. doi:10.1074/jbc.M210839200

Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N. M., Rot, G., . . . Ule, J. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biology, 8*(10), e1000530. doi:10.1371/journal.pbio.1000530

Warburg, O. (1956). On respiratory impairment in cancer cells. *Science, 124*(3215), 269-270.

Weiss, R. (1998). Viral RNA-dependent DNA polymerase RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Reviews in Medical Virology, 8*(1), 3-11.

Wilkie, G. S., Dickson, K. S., & Gray, N. K. (2003). Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends in Biochemical Sciences, 28*(4), 182-188. doi:10.1016/S0968-0004(03)00051-3

Will, C. L., Schneider, C., MacMillan, A. M., Katopodis, N. F., Neubauer, G., Wilm, M., . . . Query, C. C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO Journal, 20*(16), 4536-4546. doi:10.1093/emboj/20.16.4536

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

Wu, S., Romfo, C. M., Nilsen, T. W., & Green, M. R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature, 402*(6763), 832-835. doi:10.1038/45590

Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., & Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature, 389*(6652), 745-749. doi:10.1038/39631

Xing, Y., & Lee, C. J. (2005). Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. *PLoS Genetics, 1*(3), e34. doi:10.1371/journal.pgen.0010034

Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., . . . Fu, X. D. (2013). Direct conversion of fibroblasts to neurons

by reprogramming PTB-regulated microRNA circuits. *Cell, 152*(1-2), 82-96. doi:10.1016/j.cell.2012.11.045

Yan, C., Wan, R., Bai, R., Huang, G., & Shi, Y. (2016). Structure of a yeast activated spliceosome at 3.5 A resolution. *Science, 353*(6302), 904-911. doi:10.1126/science.aag0291

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G. M., Hao, T., Richardson, A., . . . Vidal, M. (2016). Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell, 164*(4), 805-817. doi:10.1016/j.cell.2016.01.029

Yoshida, K., & Ogawa, S. (2014). Splicing factor mutations and cancer. *Wiley Interdiscip Rev RNA, 5*(4), 445-459. doi:10.1002/wrna.1222

Zaher, H. S., & Green, R. (2009). Fidelity at the molecular level: lessons from protein synthesis. *Cell, 136*(4), 746-762. doi:10.1016/j.cell.2009.01.036

Zahler, A. M., Neugebauer, K. M., Lane, W. S., & Roth, M. B. (1993). Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science, 260*(5105), 219-222.

Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I., & Green, M. R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci U S A, 89*(18), 8769-8773.

Zorio, D. A., & Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature, 402*(6763), 835-838. doi:10.1038/45597

1. First Manuscript

Tumor suppressor properties of the splicing regulatory factor RBM10.

Hernández J, Bechara E, Schlesinger D, Delgado J, Serrano L, Valcárcel J. Tumor suppressor properties of the splicing regulatory factor RBM10. RNA Biol. 2016 Apr 2;13(4):466–72. DOI: 10.1080/15476286.2016.1144004

2. Second Manuscript

Mechanism of repression of NUMB exon 9 alternative splicing by RBM10

Mechanism of repression of NUMB exon 9 alternative splicing by RBM10

Jordi Hernández, Martin Rübbelke, Cedrik Magis, Maria Chatzou, Cedrik Notredame, Michael Sattler, Juan Valcárcel

Jordi Hernández performed all the experiments and analyses, except those of Figure 4 and 2E (see below) and wrote the manuscript under Juan Valcárcel's supervision.
The experiments and analyses of Figure 4 were performed by Martin Rübbelke under the supervision of Michael Sattler.
The analyses of Figure 2 E were proposed by Jordi Hernández and carried out by Cedrik Magis and Maria Chatzou under the supervision of Cedrik Notredame.

Introduction

Due to its central role in eukaryotic gene regulation, alternative splicing is a major mechanism for controlling cellular differentiation, homeostasis and growth (Barash et al., 2010; Wahl et al., 2009). Miss-regulation of alternative splicing contributes to tumor initiation and progression by generating protein isoforms that play a role in the different aspects of the tumorigenic process, including cell-proliferation, apoptosis, metastasis and angiogenesis (David et al., 2010; Oltean et al., 2014). Multiple splicing factors have been reported to be up- or down-regulated or mutated in different malignancies. For instance in breast cancer, SRSF1 has been reported to be up-regulated and function as a potent oncogene (Anczuków et al., 2015; Karni et al., 2007), while the levels of RBM5, a splicing regulator of apoptosis-related genes (Bonnal et al., 2008), have been shown to be altered in both directions (Oh et al., 1999; Oh et al., 2002; Rintala-Maki et al., 2007). In lung adenocarcinoma, the most common type of lung cancer, genes encoding two splicing factors, U2AF1 and RBM10, have been found among the most frequently mutated (Collisson et al., 2014).

Genome-wide studies have documented the high frequency of alternative splicing alterations in cancer (Danan-Gotthold et al., 2015; Dvinge et al., 2015; Sebestyén et al., 2015)). In

lung adenocarcinoma, one of the most frequently altered splicing events is the increased inclusion of NUMB exon 9 (Misquitta-Ali et al., 2011). The NUMB isoform corresponding to exon 9 inclusion (NUMB-PRR^{L}) has been associated with increased cell proliferation (Bechara et al., 2013; Misquitta-Ali et al., 2011) and active Notch signaling, a key pathway for the formation and maintenance of KrasG12V-driven non-small cell lung carcinomas (Maraver et al., 2012).

Alternative splicing of NUMB is regulated by RBM10, a splicing factor with tumor suppressor properties (Hernandez et al., 2016), frequently mutated in several types of tumors, including lung (Imielinski et al., 2012), colorectal (Giannakis et al., 2016) and bladder cancer (Nordentoft et al., 2014). Mutations on RBM10 have been associated with cellular hyper-proliferation, both *in vitro* and in mouse xenografts, a phenotype that can be reversed by enforced expression of the NUMB-PRR^{S} isoform, arguing that NUMB is a key target of RBM10 in the regulation of cell proliferation(Bechara et al., 2013; Hernandez et al., 2016).

RBM10 has two close related proteins, RBM5 (which shares 50% of amino acid sequence identity) and RBM6 (30% of identity). RBM5 has also been implicated in lung cancer, where its protein levels are down-regulated (Oh et al., 2002), and in other carcinomas (Angeloni, 2007), including (Zhao et al., 2012) and breast (Edamatsu et al., 2000). RBM5 has

been shown to affect late events in spliceosome assembly (Bonnal et al., 2008) and control the splicing switch of apoptosis related genes such as caspase-2 (Fushimi et al., 2008) or FAS (Bonnal et al., 2008). While less studied, RBM6 levels are also altered in breast cancer (Rintala-Maki et al., 2007).

RBM5, RBM6 and RBM10 are highly similar in amino acid conservation and also in domain composition and have been shown to share a common set of target splicing events, although their effects can be antagonistic (Bechara et al., 2013). Their sequences include two RNA recognition motifs (RRM), two zinc fingers (ZnF), a glycine patch (G-patch) (Aravind et al. 1999) and an octamer of aromatic repeats (OCRE) domain (Callebaut et al., 2005). In RBM5 the OCRE domain has been reported as a key protein-protein interaction domain for the regulation of FAS exon 6 alternative splicing (Bonnal et al., 2012). The solution structure of the second RRM domain of RBM5 bound to CU and GA rich sequences (Song et al., 2012) has been used to model the structure of RBM10 RRM2 This domain can bind to NUMB exon 9 polypyrimidine tract *in vitro* (Hernandez et al., 2016), and this RBM10 binding site was also identified in cells in (Bechara et al., 2013).

Various pieces of evidence point towards the importance of the RRM2 domain in RBM10. First, RBM10 has been shown

to lose its tumor-suppressor properties in certain RRM2 mutants (Bechara et al., 2013). Second, RRM2 is a highly conserved domain (Fig 2). Third, two natural RBM10 isoforms that differ in the presence or absence of a valine residue in position 354, located in the second alpha helix of the RRM2 domain (Bechara et al., 2013; Tessier et al., 2015) regulate NUMB exon 9 splicing, while an oncogenic mutant of the protein, V354E identified in lung cancer cells, is defective in NUMB splicing regulation(Bechara et al., 2013; Hernandez et al., 2016). Interestingly the three different RRM2 domains (-354V, V354 and V354E) are able to bind to RNA *in vitro*, suggesting that their difference in function is not due to distinct capacities to bind to its target pre-mRNA.

Despite its importance in lung cancer and tumor progression, the mechanism by which RBM10 mutants lose their function remains poorly understood. Here we combine our previous knowledge of RBM10 with bioinformatic analysis, evolutionary data and cancer associated mutations to test the effects of RBM10 RRM2 mutants on NUMB exon 9 alternative splicing. NMR structures for the RRM2 domain of both wild-type isoforms and the V354E mutant, shows that there are no major conformational differences between these variants. Co-immunoprecipitation of RBM10 (V354 and V354E) identified PRPF19 as an interactor which is lost in the V354E mutant suggesting that it may be a partner of RBM10-mediated splicing inhibition. Taken together our results suggest that the

RRM2 domain of RBM10 is highly conserved and resilient to mutations, and the oncogenic mutant V354E, located in this domain, loses its interaction with PRPF19, pointing towards a possible mechanism of action for RBM10.

Results

Amino acid conservation and accumulation of cancer-associated mutations in RBM 10, 5 and 6

Accumulation of mutations in cancer can provide insights into functions relevant for tumor progression. To analyze the distribution of cancer-associated mutations in RBM10 -as well as in the related factors RBM5 and RBM6-, we used publicly available datasets (cBioportal) to localize mutations reported in all cancer types, across the domain structure of these proteins.

While missense mutations represent more than 75% of the mutations found in RBM5 and RBM6 (79% and 77%, respectively), missense mutations represent only 57% of all reported mutations in RBM10, suggesting that loss of function by introduction of premature termination codons is a relatively common mechanism for RBM10 inactivation in cancer, in agreement with its proposed function as a tumor suppressor (Bechara et al., 2013; Hernandez et al., 2016).

105 (RBM5), 173 (RBM6) and 148 (RBM10) missense mutations (Supplementary Table 1) were mapped to the different protein regions and domains of these polypeptides. RRM2 was the most frequently mutated defined domain of RBM10 (consistent with the detection of proto-oncogenic mutations (Bechara et al., 2013; Hernandez et al., 2016) and

RBM6, while RRM1 (closely followed by RRM2) was the most frequently mutated defined domain in RBM5 (Figure 1). After normalizing by domain size, ZnF2 is proportionally the most frequently mutated domain in RBM10 and RBM6, suggesting an important function relevant for the control of tumor progression (Supplementary Table 2).

Next we focused on RBM10 and analyzed the distribution of all mutations (including missense, non-sense, frame-shift insertions and deletions, and splice site mutations) from all cancer types and compared it to the expected frequency of mutations if they were distributed proportionally to domain length. The domain with the highest ratio of mutations over what would be expected from a uniform distribution was the OCRE domain, followed by the ZnF2 (highly mutated for its size, as also observed for missense mutations (Figure 1, Supplementary Table 2), and the RRM1 (Figures 2A and B). Remarkably, while RRM2 is the most frequently missense-mutated domain of RBM10 (Figure 1), the frequency of all types of mutations combined is lower than expected in this domain (Figure 2B), indicating that missense mutations are the most common type of mutations in this domain in cancer cells. This is of special interest considering that 40% of all non-missense mutations are located before the beginning of RRM2, in the first 25% of the protein (Supplementary Figure 1), suggesting a positive selection of early truncates of RBM10 in tumors, possibly related to the observation that

truncating mutants often act as dominant negatives regarding the function of RBM10 in the regulation of NUMB alternative splicing {Hernandez, 2016 #74}.

Given the high frequency of mutation of RBM10 in lung adenocarcinomas (Imielinski et al, 2012), we carried out a similar analysis of mutation distribution in these tumors (Figures 2C and D). The distribution of the 36 mutations (including missense, non-sense, frame-shift insertions and deletions, and splice mutations) followed similar enrichment profiles, with mutations in the RRM2 (and, particularly the G-patch) domains appearing at lower frequencies than expected (Figure 2D) and with missense mutations before RRM2 being less frequent than expected (e.g. only 1 out of 9 mutations in RRM1 was a missense mutation). Mutations in the OCRE domain, in contrast, were particularly enriched (Figure 2D), as seen for all cancer types, where the mutations in the OCRE domain are more abundant than what is expected from a uniform distribution (Fig 2 B). In RBM5, the OCRE domain plays a role in splicing regulation and mediates binding to several splicing factors (Bonnal et al., 2008). These results suggest that the OCRE domain of RBM10 plays a particularly important role in the progression of lung adenocarcinomas.

To complement these studies, we carried out two additional analyses. The first was to study the phylogenetic conservation of RBM10, retrieving from ENSEMBL all known

orthologues with sequence identity above 70% (38 sequences from the tetrapod superclass) and aligning them using T-Coffee software (Magis et al., 2014; Notredame et al., 2000). Considering only point substitutions, 1006 sequence variations were detected, mapping to different regions of the protein (Figure 2E, upper scheme. As expected, variation was more frequent in non-annotated regions, while ZnF2 and RRM2 were the least variable domains. The most variable annotated domains correspond to the ZnF1, RRM1 and OCRE domains. These results argue for significant functional differences between the two RRM domains as well as between the two ZnF domains. Of interest, a region of high conservation was detected between amino acids 630 and 741, which is 36% identical to a BAG domain (Doong et al., 2002) (see Discussion).

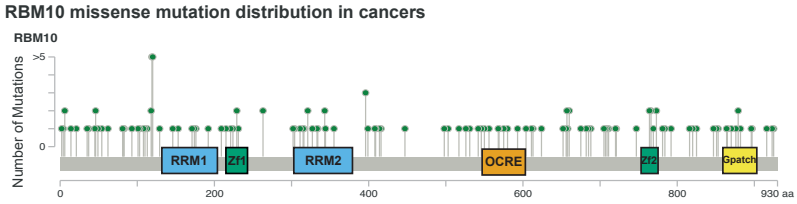
The second analysis (the lower scheme of Figure 2E) consists of an *in silico* prediction of the predicted deleterious effect of missense mutations at each RBM10 position using PROVEAN, an algorithm that predicts functional effects of protein variants based on orthologous alignments. For each amino acid position of the protein, the algorithm generates 19 variants (corresponding to the other possible 19 amino acids) and aligns them to all the RBM10 orthologues found on NCBI. We average the score of the 19 variants on each position to get an Average Predicted Disruptive Score for each residue of the protein. Any variation of the protein with a score lower

than -2.5 is predicted to be disruptive for the protein function. As expected, we found that all RBM10 mutations that impaired NUMB alternative splicing regulation had scores below -2.5 (data not shown). A good overall anti-correlation was found between this score and the degree of conservation at each position (Figure 2E, compare lower and upper schemes). Interestingly, the OCRE domain was predicted to be most sensitive to mutations, in concordance with its higher frequency of mutations (Figure 2 B).

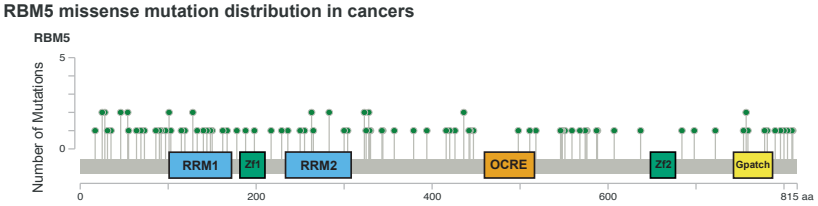
Taken together, the combination of phylogenetic analysis and cancer-associated mutations distribution allow us to identify domains especially sensitive to mutations and argue that the OCRE domain is highly sensitive (Figure 2E lower panel), in agreement with the high frequency of mutations found in cancer samples (Figure 2D). RRM2 is not only the domain with most missense mutations (Figure 1D), but is also highly conserved during evolution (Figure 2E upper panel), arguing for an important function (e.g. control of cell proliferation, (Bechara et al., 2013; Hernandez et al., 2016)), which is altered in cancer.

Figure 1

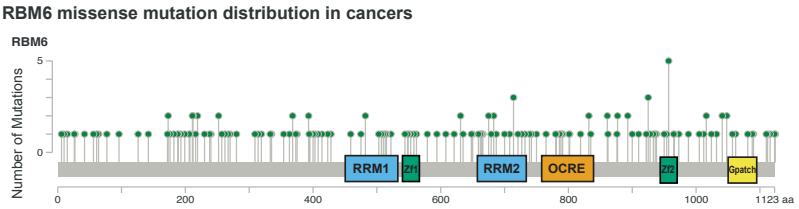
A



B



C



D

Number of missense mutations across the different domains of RBM5, RBM6 and RBM10

	RRM1	ZnF1	RRM2	OCRE	ZnF2	G-patch	Disordered
RBM5	**13**	2	10	2	0	6	66
RBM6	5	5	**15**	9	5	5	84
RBM10	8	6	**14**	9	9	9	93

Figure 1. Cancer-associated missense mutations in RBM5, RBM6 and RBM10. Lollipop representation of the location of cancer-associated missense mutations in RBM10 (A), RBM5 (B) and RBM6 (C) along the linear polypeptide sequence, with key identifiable domains indicated.(D) Quantification of cancer-associated missense mutations across the different domains of RBM5, RBM6 and RBM10. Data were collected from cBioportal containing information from 28 different tumor types, from 147 different studies.

Figure 2

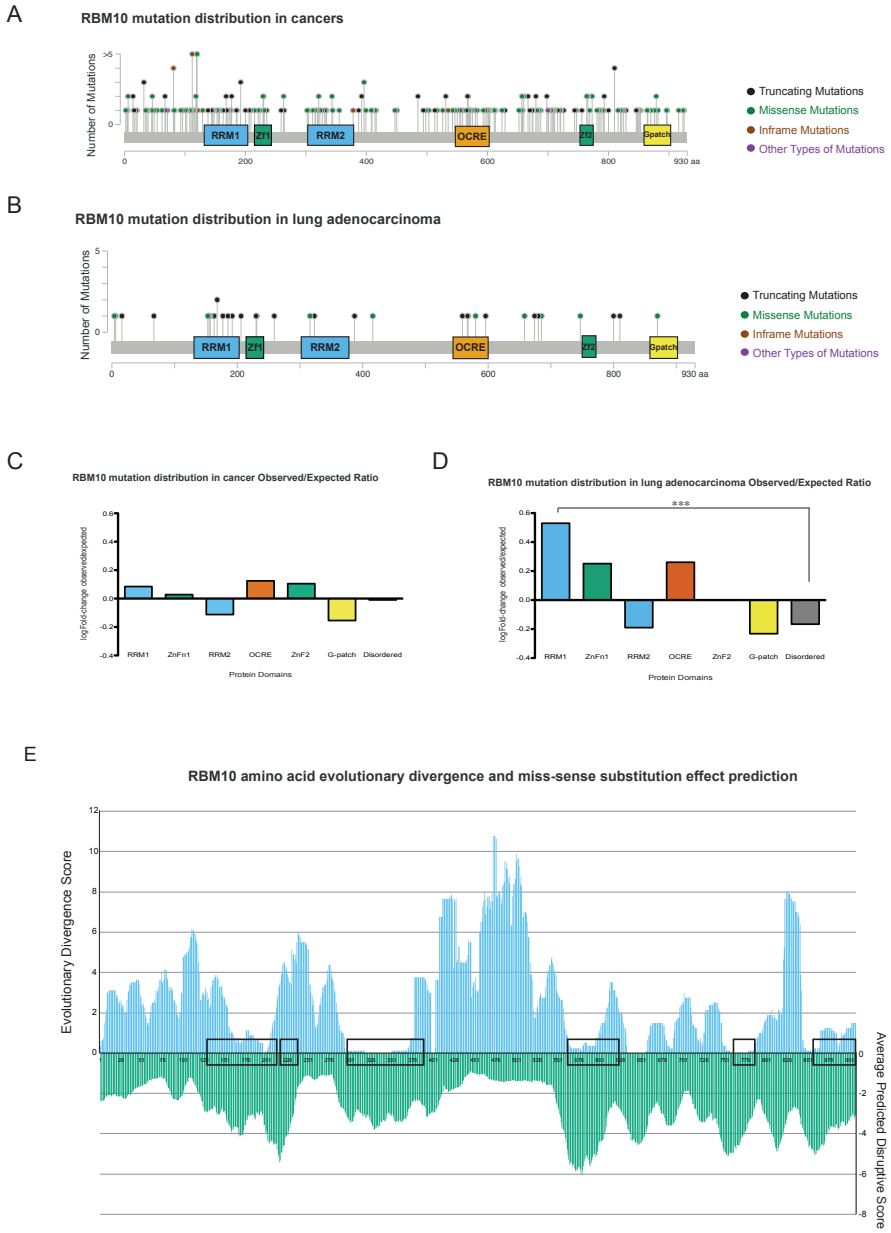


Figure 2. Distribution of all cancer-associated mutations and evolutionary conservation of RBM10. (A,B)) Lollipop representation of all cancer-associated (A) or lung adenocarcinoma-associated (B) mutations reported in cBioportal for RBM10. Black: Truncating. Green: Missense. Brown: In-frame. Purple: Others. (C, D) Ratio of Observed/Expected mutations across RBM10 domains for all cancer-associated (C) or lung adenocarcinoma-associated (D) mutations. Expected number of mutations estimated on the basis of domain length, assuming uniform distribution of the mutations across the protein (E) Estimation of evolutionary divergence for each amino acid in RBM10.

Upper part represents the Evolutionary Divergence Score obtained for each amino acid position of RBM10, averaged to its surrounding 20 residues. Lower part represents the Average predicted Disruptive score for each position, averaged to its surrounding 20 residues.

Effect of RBM10 RRM2 mutants in NUMB alternative splicing regulation

Being the most mutated domain of RBM10, and due to its known implications in the tumor-suppressor properties of the protein (Bechara et al., 2013; Hernandez et al., 2016) we decided to investigate the effect of several cancer-associated mutants of RBM10 RRM2 in the regulation of NUMB exon 9 alternative splicing.

HEK-293 cells were co-transfected with a NUMB minigene (RG6-NUMB) comprising NUMB alternative exon 9 and flanking intronic sequence (100nt upstream and 50nt downstream) between constitutive exons of chicken tropomyosin (Orengo et al., 2006) and increasing amounts of vectors expressing wild type or mutant RBM10. 24h post-transfection, cells were collected, RNA isolated and retro-transcribed using oligo-dT and random primers, and subsequently PCR amplified with vector-specific primers, and the ratio between NUMB exon 9 inclusion/skipping quantified by capillary electrophoresis.

As reported in HeLa cells (Hernandez et al., 2016), overexpression of both RBM10 WT isoforms (354V and -354) promoted NUMB exon 9 skipping to similar levels (Figures

3B-C, compare lanes 1-7), arguing that the presence or absence of a valine at position 354 does not alter the activity of RBM10 in the regulation of NUMB exon 9 splicing. As previously reported (Hernandez et al., 2016), titration of the oncogenic mutant V354E failed to cause efficient exon skipping, while V354D displayed reduced activity (Figures 3B-C, lanes 14-19), suggesting that the disruption of RBM10 regulatory function is not simply explained by the presence of a negative charge. Titration of mutant I316F, detected in lung adenocarcinoma (Imielinski et al., 2012), also failed to modulate NUMB exon 9 splicing (Figures 3B-C, lanes 8-10), while mutant E380K, found on endometrium cancer (Supplementary table 3), displayed reduced activity, similar to V354D (Figures 3B-C, lanes 20-22). In contrast, overexpression of the colorectal cancer mutant V333I (Supplementary table 3) promoted NUMB exon 9 skipping as the wild type protein (Figures 3B-C, lanes 11-13).

Figure 3

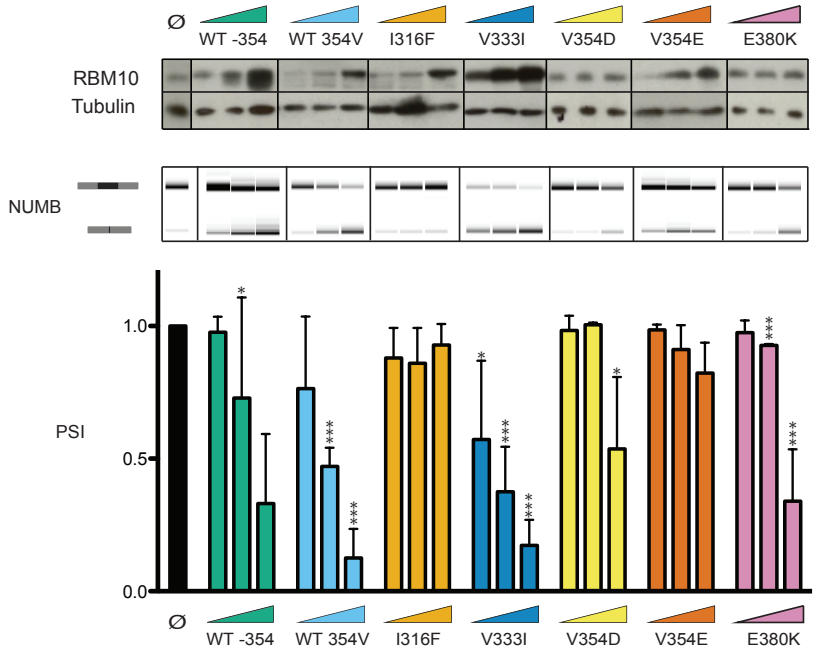


Figure3. Activity of RBM10 RRM2 cancer-associated mutants on NUMB exon 9 repression. Co-transfection experiments in HEK-293 cells of RG6-NUMB reporter and vectors expressing different mutants for RBM10 RRM2, as indicated, were followed by protein and RNA isolation and analysis of RBM10 protein expression by Western blot (upper panel, **β**-tubulin was used as a loading control) and NUMB exon 9 inclusion/skipping by RT-PCR (middle panel) and PSI values (average and standard deviation for 4 replicates, lower panel; statistical significance was addressed by student's t-test, the number of * represent the p-value. If p-value <0.05 then * if p-value <0.01 then ** if p-value <0.001 then ***

All together, these results identify cancer-associated mutations in RBM10 RRM2 (V354E, E380K and I316F) that are relevant for the function of the protein in alternative splicing regulation of NUMB and therefore facilitate cancer cell growth, while other mutations (e.g. V333I) retain function.

Structural characterization of the RRM2 domain by NMR

Previous structural modeling indicated that substitution of valine 354 by glutamic acid (V354E) was unlikely to affect the structure of the RRM2 domain of RBM10 or to alter its RNA binding properties, because V354 was predicted to reside in one alpha-helix in the opposite side of the RRM beta-sheet RNA binding surface (Hernandez et al., 2016).

To characterize the effect of the extra valine in more detail, structures of RBM10 RRM2 -354 and V354 were solved by solution state NMR. Although the fold of the RRM2 -354 structure is similar to the PDB deposited structure 2M2B (Serrano P. and collaborators), there is not a perfect overlay between both. NMR structures of RRM2 -354 and RRM2 354V are overall very similar and, as predicted, the additional V354, which is located on the second alpha helix, is pointing towards the solvent. This second alpha helix is straight in the RRM2 -354 structure and slightly bent away from the core of the domain in the RRM2 354V structure. In addition, the N- and C-termini of RRM2, which are part of the beta-sheet, are slightly different. (Figure 4B, a). The ^{1}H-^{15}N correlation spectra of the 354V variant and the V354E mutant show only minor local chemical shift differences around the mutation site, indicating no effect on the domain structure.

To identify residues affected by RNA binding, we carried out NMR titration experiments using an RNA corresponding to

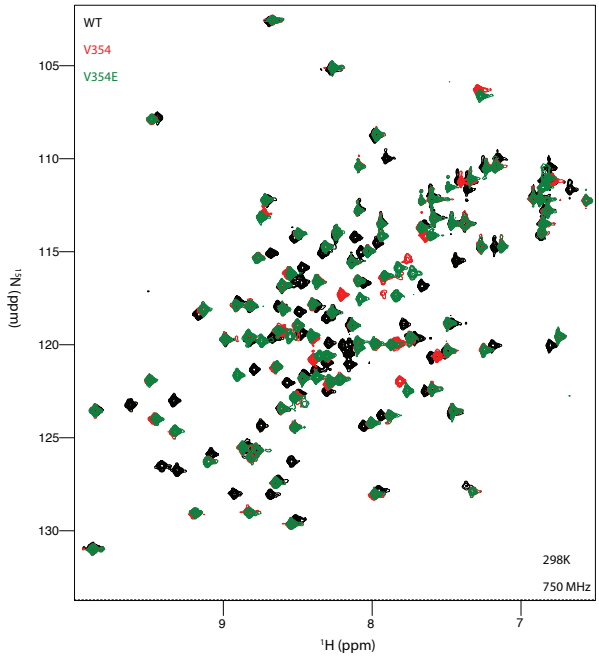
NUMB exon 9 polypyrimidine-tract. In the RRM2 -354, regions with CSP were located in the loop between the beta-strands 2 and 3, the beta-sheet and alpha helix 2 (Figure 4B, b, darker regions). In the RRM2 354V, the CSP pattern is similar to the -V354 variant, but the alpha helix 2 shows no changes in chemical shift upon RNA binding (Figure 4B, c, more intense red regions). It is not clear if the CSP of alpha helix 2 in the variant without valine 354 are caused by direct contacts with the RNA. In the RRM2 V354E mutant, the CSP pattern was the same as for RRM2 V354 (Figure 4B, d, lighter green regions), suggesting no difference in RNA binding between the RRM2 354V and V354E.

Isothermal Titration Calorimetry experiments were performed to measure the thermodynamic properties of the different RRM2 domains and their RNA-binding dissociation constant. Thermodynamic parameters are summarized at the bottom of Figure 4C. RRM2 -354 displays the highest affinity (Kd=15.53nM), while the affinities of RRM2 354V (Kd=55.87) and RRM2 354V (Kd=62.17) are very similar. This is consistent with the differences in NMR titration experiments observed between the -354 variant and the V354 and V354E variants. These results are in qualitative agreement with those previously obtained using gel retardation assays (Hernandez et al., 2016) and argue that failure of the lung cancer RBM10 V354E mutant to regulate NUMB splicing is unlikely due to a defect in RNA binding, consistent with the

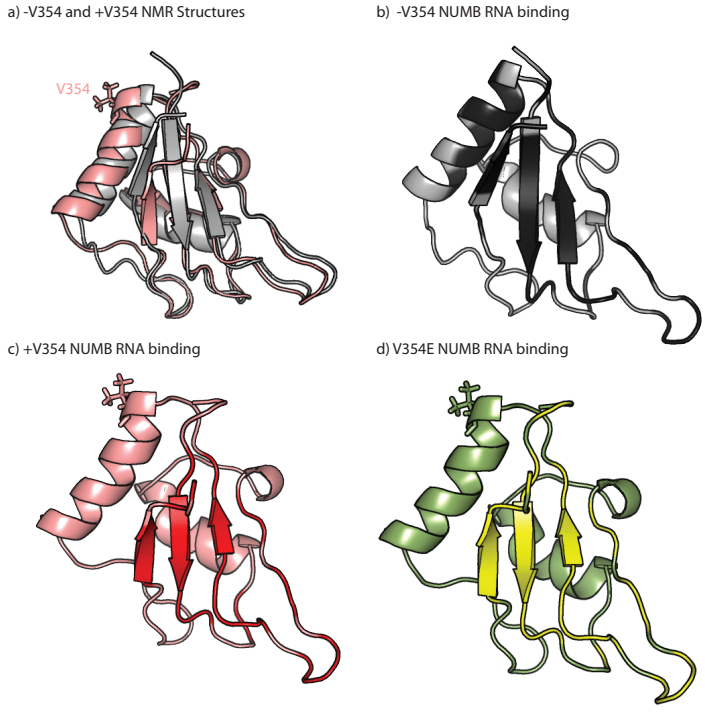
location of 354V/E side chain outside the RNA binding interface.

Figure 4

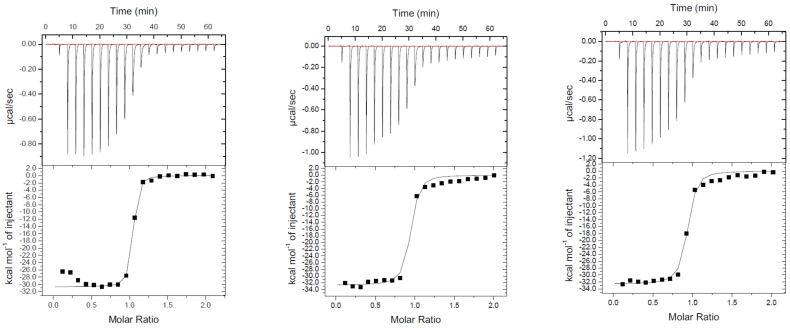
A


WT
V354
V354E
105
110
115
120
125
130
^{15}N (ppm)
298K
750 MHz
9
8
7
^{1}H (ppm)

B


a) -V354 and +V354 NMR Structures
b) -V354 NUMB RNA binding
V354
c) +V354 NUMB RNA binding
d) V354E NUMB RNA binding

C



	-354V	V354	V354E
KD	16.53 +/- 4.21 nM	55.87 +/- 21.41 nM	62.11 +/- 17.17 nM
N	0.997 +/- 0.003	0.903 +/- 0.015	0.894 +/- 0.08
ΔH	-3.065E4+/- 268.3 cal/mol	-3.272E4 +/- 557.4 cal/mol	-3.258E4 +/- 513.5 cal/mol
ΔS	-67.2 cal/mol/deg	-76.6 cal/mol/deg	-76.3 cal/mol/deg

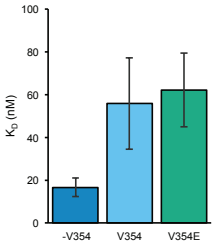


Figure 4. Structure and properties of RBM10 RRM2 domain. (A) ^{1}H-^{15}N NMR spectra of RBM10 -354V (Black), RBM10 V354 (Red) and RBM10 V354E (Green). (B) RRM2 Structures and RNA binding surfaces: a) Solution NMR structures of RBM10 RRM2 isoforms -V354 (grey) and +V354 (salmon). b) Location of NUMB RNA (3' splice site region of intron 8) binding site in the RRM2 –V354 variant. Residues with a chemical shift perturbation larger than 0.05 ppm upon NUMB RNA binding are highlighted in black on the -V354 solution NMR structure. c) NUMB RNA binding site in the RRM2 +V354 variant. Residues with a chemical shift perturbation larger than 0.05 ppm upon NUMB RNA binding are highlighted in red on the +V354 solution NMR structure. d) NUMB RNA binding site of the RRM2 V354E mutant. Residues with a chemical shift perturbation larger than 0.05 ppm upon NUMB RNA binding are highlighted in yellow on the +V354 solution NMR structure. C) Isothermal Calorimetry binding curves for RBM10 RRM2 (-354V -left-, V354 -centre- and V354E -right) binding to NUMB RNA. Thermodynamical values for these interactions are represented on the bottom Table. Estimated K_D values are presented in the bottom graphic (bars represent the fitting error).

Identification of RBM10 interactors by pull-down and mass-spectrometry

To explain the functional difference in NUMB splicing regulation between RBM10 354V and the mutant V354E (Figure 3), without significant differences in overall domain structure (Fig 4 A) or the affinity with which they bind to NUMB intron 8 3' splice site region (Figure 4C), we decided to look for protein partners that differentially associate with the wild type and V354E mutant (i.e. factors that would fail to bind to RBM10 because of the negatively charged glutamic acid in the RRM2 or that would associate specifically with this mutant). For this, we generated stable HEK293 cell lines that expressed moderate levels of epitope-tagged RBM10 354V or V354E. A HEK293 cell line transfected with the empty vector was used as a control. Whole cell extracts from these cell lines where treated with RNase and DNase, and RBM10 and its co-interacting partners where pulled-down using resin-immobilized anti-tag antibodies. Biological triplicates of these precipitates where analyzed by mass-spectrometry, and the data analyzed using MSstats software.

545 proteins were identified in RBM10 354V pull-downs and 533 in RBM10 V354E pull downs. Figures 5A and B display the proteins identified as volcano plots representing the fold enrichment and statistical significance of the identified polypeptides. High confidence candidates with substantial and statistically consistent enrichment are highlighted in blue

(except RBM10, which is highly enriched, as expected, and represented in orange). 15 interactors of RBM10 354V passed the high confidence threshold (log2 Fold-change >1 and p-value <0.01)(Table 1), while 11 interactors were identified for RBM10 V354E using the same criteria (Table 2).

RBM10 354V interactors included (12/15) spliceosomal components, with a notable enrichment in U2 snRNP-associated proteins. 5 of these RBM10 interactors had been previously reported in a large scale yeast-2-hybrid interaction screening (Hegele et al., 2012) (RBM10, SF3B4, U2SURP, DHX15 and PRPF19) although only 3 had been validated by Co-IP in that work (RBM10 self-interaction, DHX15 and U2SURP) (Hegele et al., 2012). Our results thus reveal a clear relationship of RBM10 with U2 components that may have implications for the splicing repression mechanism of RBM10 (see Discussion).

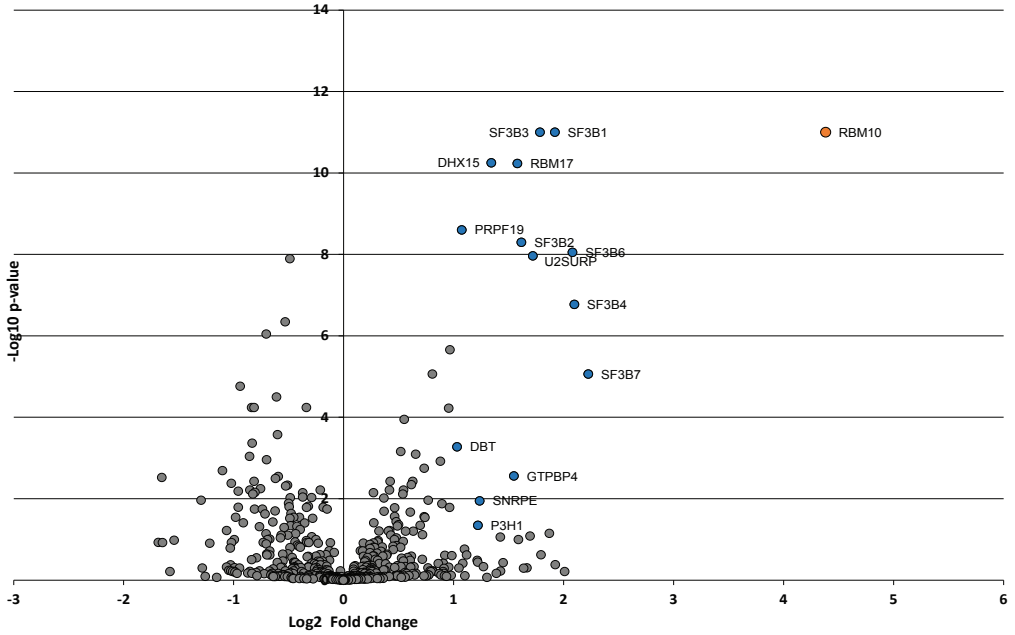
Comparison between the list of RBM10 354V interactors and the oncogenic mutant V354E (Table 2) revealed three factors that were significantly more abundant in precipitates of the wild-type protein: U2SURP, SF3B3 and SF3B6, as well as several other factors that were enriched in precipitates of the wild-type protein but that were not enriched in precipitates of the mutant protein compared to control precipitates: GTBP4, DHX15, SNRPE, P3H1, DBT and -interestingly- PRPF19.

To validate these results, we carried RBM10 co-immunoprecipitation experiments in extracts from A549 lung adenocarcinoma cells (which express only RBM10 V354E) HEK 293 cells (mainly expressing RBM10 -354) and HeLa cells (mainly expressing RBM10 354V). After treatment with RNase and DNase, precipitation of RBM10 with specific antibodies and normalization for comparable amounts of RBM10, western blots were carried out to detect the presence of RBM10 interactors. These experiments validated interactions of RBM10 with SF3B3, U5-116, U2AF65, U2AF35 and PRPF19 (Figure 5C), but did not validate interactions with DHX15, PTBP1 or SF3B6 (Figure 5C). Interestingly the mutant RBM10 V354E did not interact with PRPF19 (Figure 5C) suggesting that interaction with PRPF19 may be compromised by the presence of glutamic acid at position 354 within the RRM2 of RBM10.

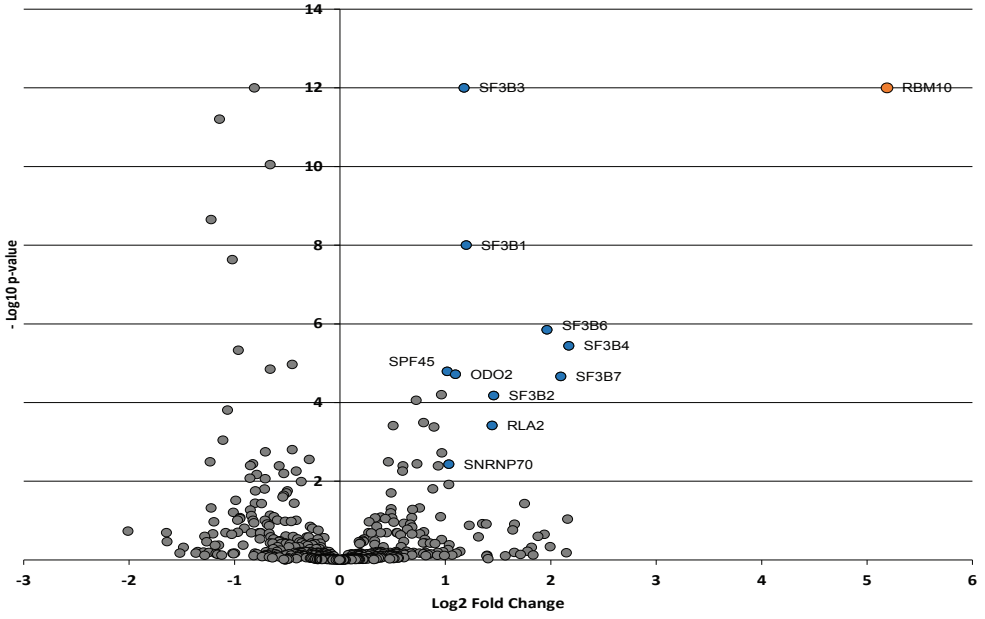
Taken together, our co-immunoprecipitation and mass-spectrometry data reveal that RBM10 interacts with core U2 snRNP components and PRP19, and that the lung cancer mutant V354E mutant, fails to interact with PRPF19, a feature possibly linked to its failure to regulate NUMB alternative splicing and to inhibit cell proliferation.

Figure 5

A


-Log10 p-value
Log2 Fold Change
SF3B3
SF3B1
RBM10
DHX15
RBM17
PRPF19
SF3B2
SF3B6
U2SURP
SF3B4
SF3B7
DBT
GTPBP4
SNRPE
P3H1

B


- Log10 p-value
Log2 Fold Change
SF3B3
RBM10
SF3B1
SF3B6
SF3B4
SPF45
ODO2
SF3B7
SF3B2
RLA2
SNRNP70

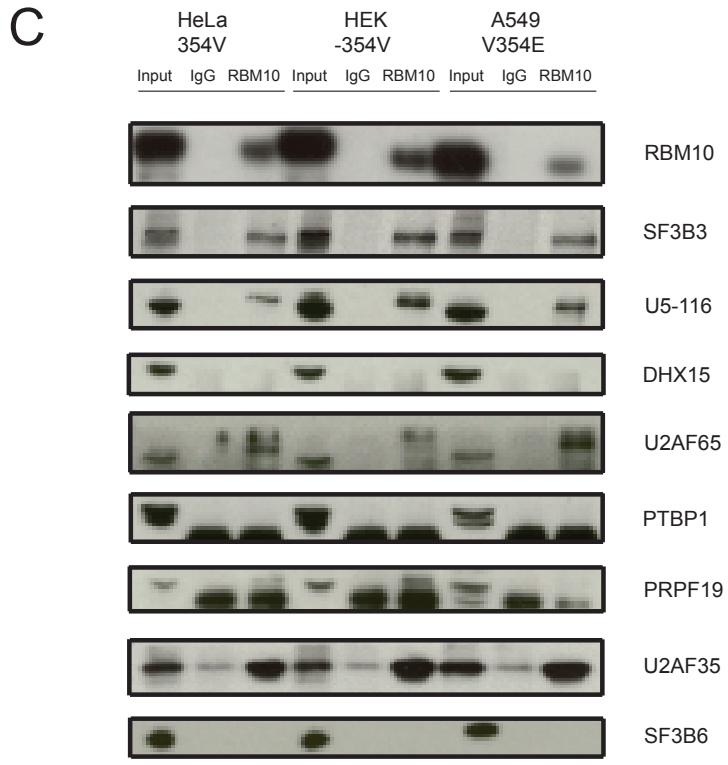


Figure 5. Identification of RBM10 interactors by co-immunoprecipitation and mass-spectrometry analysis. (A) Volcano plot representing the -log10-pvalue versus the Log2 Fold Change of peptides identified by the mass spectrometry in co-precipitates of RBM10 V354 versus peptides identified in the negative control. Peptides with a Log2 Fold Change > 1 and a p-value < 0.01 are represented in blue (high confidence interactors). RBM10 is marked in orange. (B) Volcano plot representing the -log10-pvalue versus the Log2 Fold Change of the peptides identified by mass spectrometry in co-precipitates of RBM10 V354E versus the peptides identified in the negative control. Peptides with a Log2 Fold Change > 1 and a p-value < 0.01 are represented in blue (high confidence interactors). RBM10 is marked in orange. (C) Western blot analysis of putative RBM10 interactors in co-immunoprecipitation experiments using HeLa (expresses 354V variant), HEK-293 (expresses -354 variant) and A549 (expresses V354E variant) cells. Input corresponds to 3% of the volume of extract used for the IP, while 1/3 of IP with IgG (negative control) or anti-RBM10 antibodies were loaded on the gel. The antibodies used for probing the blots for the presence of the indicated proteins are shown in the right column.

Table 1

Protein Name	Fold Change	p-value	Complex	Former interactors	PD-MS
RBM10	4.38	1.00E-11	A complex	Direct Interaction	
SF3B7	2.22	8.68E-06	U2 snRNA associated protein		
SF3B4	2.10	1.69E-07	U2 snRNA associated protein	Direct Interaction	
SF3B6	2.08	8.90E-09	U2 snRNA associated protein		*
SF3B1	1.92	1.00E-11	U2 snRNA associated protein		
SF3B3	1.79	1.00E-11	U2 snRNA associated protein		*
U2SURP	1.72	1.08E-08	U2 related	Direct Interaction	*
SF3B2	1.62	5.06E-09	U2 snRNA associated protein		
RBM17	1.58	5.87E-11	U2 related		
GTPBP4	1.55	0.00277	Ribosome Biogenesis		^
DHX15	1.34	5.63E-11	Splicing	Direct Interaction	^
SNRPE	1.24	0.01131	Tri-snRNPs		^
P3H1	1.22	0.04501	Collagen Biosynthesis		^
PRPF19	1.07	2.51E-09	PRP19 complex	Direct Interaction	^
DBT	1.03	0.00053	Aliphatic amino acids metabolism		^

Table 1. RBM10 V354 interactors identified by Mass-spectrometry. * Statistically enriched in RBM10 V354 relative to RBM10 V354E. ^ Statistically enriched in RBM10 V354 relative to the control.

Table 2

Protein Name	Fold Change	p-value	Complex	Former interactors	PD-MS
RBM10	5.19	1.00E-12	A complex	Direct Interaction	
SF3B7	2.10	2.15E-05	U2 snRNA associated protein		
SF3B4	2.17	3.59E-06	U2 snRNA associated protein	Direct Interaction	
SF3B6	1.96	1.40E-06	U2 snRNA associated protein		*
SF3B2	1.46	6.57E-05	U2 snRNA associated protein		
RLA2	1.44	0.000378932	Ribosomal protein		v
SF3B1	1.20	9.92E-09	U2 snRNA associated protein		
SF3B3	1.18	1.00E-12	U2 snRNA associated protein		*
ODO2	1.10	1.90E-05	Mitochondrial protein		v
SNRNP70	1.03	0.003664398	U1 Sm protein		v

Table 2. RBM10 V354E interactors identified by Mass-spectrometry. *Statistically enriched in RBM10 V354 relative to RBM10 V354E. v Statistically enriched in RBM10 V354E relative to the control

Materials and Methods

Mutation Distribution Analysis in RBMs

The data for the analysis of the distribution of RBMs mutation was downloaded from the cBiportal webpage. Redundant entries were removed using a homemade R script. The number of mutations according to tumor types and its distribution was calculated using the entries from RBMs Unique Mutants List (Supplementary Table 1). The expected number of mutants in each domain was calculated using the same total number of reported mutants (256 mutations for RBM10 in all tumor types and 36 mutations for LUAD) expecting a uniform distribution depending strictly on domain length.

The evolutionary conservation analysis was performed on a dataset of 1-to-1 RBM10 orthologs extracted from Ensembl version 75 (Flicek et al., 2014). We considered only tetrapod superclass orthologs with a minimum of 70% identity when compared to the human sequence of RBM10 protein. The resulting datasets of 38 RBM10 orthologs was multiply aligned using T-Coffee version 11.0 (Magis et al., 2014). The accuracy of the resulting Multiple Sequence Alignment (MSA) was evaluated using the consistency score given by T-Coffee. From the previously generated MSA, using a homemade PERL script, we generated a list of all punctual

amino acid sequence variations (1006 in total) using the human RBM10 sequence as a reference; insertion and deletion were not taken into account in the list of sequence variations. An evolutionary profile was then generated using a normalized (by the number of sequences) count of variations per each column within the MSA.

The mutational predictive effect was established using the PROVEAN (Yongwook Choi, 2012; Y. Choi et al., 2012) webserver, which belongs to the SIFT group of mutation effect predictors. PROVEAN works by giving a sequence and a list of variants. Once ROVEAN gets a protein sequence, its homologs are searched against the NCBI database using BLAST and clustered using CD-HIT. Based on the selected homologs, the PROVEAN scores are computed for each of the variants provided. The score thresholds for the prediction are variants with a score equal to or below -2.5 to be considered "deleterious" and with a score above -2.5 "neutral."

What we did is to provide to the webserver a list of each amino acid in each position mutating to all possible 19 amino acids and asked PROVEAN to score the effect of the corresponding mutation. For each position we then averaged the effect to get a better feeling how sensitive is a certain position to a mutation. Finally we used a sliding window of 20

amino acids over the sequence to know which regions of the protein are more acceptable to mutations and which not.

Cell culture and transfection conditions

HeLa, HEK-293 and A549 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS, Gibco) plus 50units/ml penicillin and 50ml/ml streptomycin. Stable HEK-293 cell lines, expressing taged-RBM10 proteins, where maintained in the same media as HEK-293 cells but supplemented with 0.5mg/ml of Geneticin (Gibco).

Transfections were carried out using Lipofectamine 2000 (Invitrogen) and all the down-stream processing of the sample (RNA isolation, RT-PCR, PCR with splicing specific primers and capillar electrophoresis) were performed as in (Hernandez et al.).

RRM2 structure production and structure characterization

The NMR structures of the two RRM2 (Residues 300-383) variants, with and without V354, were produced in E.Coli BL21(DE3). Structural calculations where performed using a standard NOE based approach. All spectra were recorded on a Bruker AVIII 600 MHz spectrometer with cryogenic triple resonance gradient probe or an AVIII 750 MHz spectrometer

with triple resonance gradient probe. Proteins were measured in a 20mM sodium phosphate at pH 6.5, containing 50mM sodium chloride, 1mM DTT and 5% D_2O.

All spectra were processed with NMRPipe/Draw (Delaglio et al., 1995) and analyzed using CcpNmr analysis (Vranken et al., 2005). Cyana 3 (Schmidt et al., 2015) was used for structure calculation.

RNA titrations

Both variants of the RRM2, with and without V354, and the V354E mutant were titrated with a NUMB gene-derived 12mer RNA (5'-UUGUCUGCUCCC-3'). ^{1}H-^{15}N-correlation spectra for different protein RNA-ratios (1:0.3, 1:0.6, 1:1, 1:1.5) were recorded. From the final titration points, carried out at a 1.5x excess of RNA over protein, chemical shift perturbation (CSP) between the free protein (free) and the protein-RNA-complex (RNA) were calculated using the following formula, where δ is the proton (H) or nitrogen (N) chemical shift:

$$CSP = \sqrt{(\delta_{H\,free} - \delta_{H\,RNA})^2 + \left(\frac{(\delta_{N\,free} - \delta_{N\,RNA})}{10}\right)^2}$$

ITC experiments

ITC experiments were performed on a Malvern MicroCal ITC 200 system. Protein and RNA were dissolved in 20mM sodium phosphate, 50mM sodium chloride, 5mM beta-mercaptoethanol buffer at pH 6.5. The cell of the ITC machine was filled with 20μM of the NUMB RNA oligo and the syringe with 200μM of the protein construct comprising all three RNA binding domains (Residues 125-383). The measurement was performed at 298 K and a total amount of 40μL was injected in 20 injections to the 200μL sample in the cell. ITC binding curves were recorded for the -354, V354 and V354E variants.

RBM10 variant purification, Mass-spectrometry analysis and Pull-down assays

RBM10 -354V and RBM10 V354E were cloned in p-DEST26 including a 8 amino acids Strep-tag (IBA Biosciences) in the C-terminal end of the protein. Three different HEK-293 stable cell lines were generated, one expressing RBM10 -354V Strep-tag, one expressing RBM10 V354E Strep-tag and one transformed with the empty p-DEST26 vector. The stable cell lines were maintained in culture with 0.5mg/ml of Geneticin (Gibco).

The co-purification of tagged-RBM10 and interactors was performed using Strep-Tactin Superflow high capacity columns. Cells from two confluent 150mm diameter Tissue Culture Dish (FALCON) were pulled together for each biological replicate, lysed by repeated cycles of freezing in liquid nitrogen and thawing in a 37ºC water bath (repeating the cycle 5 times). Washes and elutions were performed as indicated in the columns instructions with the commercial buffers from Iba-livesciences. The experiments were performed in biological triplicates and the second eluate of each condition was collected and submitted to mass spectrometry analyses (pull-down assays were performed following the Strep-Tactin Superflow high capacity columns manufacturer's instructions).

For mass-spectrometry analyses, immuno-precipitated proteins were washed with ammonium bicarbonate (3x, 200mM NH4HCO3) and re-suspended in 60µl of 6M Urea + 200mM NH4HCO3. Then the sample was reduced by adding 10µl of 10mM DTT + 200mM NH4HCO3 (1h, 37ºC) and alkylated by adding 10µl of 20mM IAA + 200mM NH4HCO3 (30 minutes, room temperature). Samples were diluted prior trypsinization (overnight, 37ºC) and the digestion reaction was stopped with formic acid (5% final concentration). Tryptic peptides were desalted with C18 columns and re-suspended in 10µL of H2O + 0.1% formic acid. 4.5µL of each peptide mixture was analyzed using a LTQ-Orbitrap Velos Pro mass

spectrometer (Thermo Fisher Scientific, San Jose, USA) coupled to a nano-LC (Proxeon, Odense, Denmark) equipped with a reversed-phase chromatography 2-cm C18 pre-column (Acclaim PepMap-100, Thermo; 100 µm i.d., 5µm), and a 25-cm C18 analytical column (Nikkyo Technos, 75µm i.d., 3µm). Chromatographic gradients started at 3% buffer B with a flow rate of 300nL/min and gradually increased to 7% buffer B in 1 minute and to 35% buffer B in 60 minutes. After each analysis, the column was washed for 10 minutes with 90% buffer B (Buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile). The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.5kV and source temperature at 200°C. Ultramark 1621 was used for external calibration of the FT mass analyzer prior the analyses. The background polysiloxane ion signal at m/z 445.1200 was used as lock mass. The instrument was operated in data-dependent acquisition (DDA) mode, and full MS scans with 1 microscan at resolution of 60 000 were used over a mass range of m/z 350−1500 with detection in the Orbitrap. Auto gain control (AGC) was set to 106, dynamic exclusion was set at 60 s, and the charge-state filter disqualifying singly charged peptides for fragmentation was activated. Following each survey scan, the 10 most intense ions with multiple charged ions above a threshold ion count of 5000 were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via collision-induced dissociation (CID) were acquired in the

linear ion trap, AGC was set to 3·104 and isolation window of 2.0 m/z, activation time of 30ms, and maximum injection time of 250ms were used. All data were acquired with Xcalibur software v2.2. Acquired data were analyzed using the Proteome Discoverer software suite (v1.4, Thermo Fisher Scientific), and the Mascot search engine (v2.5, Matrix Science) was used for peptide identification. Data were searched against the human protein database derived from the SwissProt database plus common contaminants (April 2016; 20,200 sequences). A precursor ion mass tolerance of 7 ppm was used, and up to three missed cleavages were allowed. The fragment ion mass tolerance was set to 0.5Da, and oxidation (M), and acetylation (Protein N-term) were defined as variable modifications, whereas carbamidomethylation (C) was set as fixed modification. The identified peptides were filtered by FDR < 5%. Protein-protein interactors were assessed with the MSstats (R package for the analysis of mass spec data).

Pull-down experiments in HEK-293, HeLa and A549 were performed in whole cell extracts. Cells were lysed with lysis buffer (TrisHCL pH7.4 50mM, NaCl 150mM, NP40 0.5%, DTT 1mM and protease inhibitor cocktail (Complete, Roche)), and the extract was incubated night with 15µl of magnetic beads previously bound to 2µg RBM10 antibody (Sigma, HP17034972). The RBM10 antibody was bound to the magnetic beads by incubating them together for 40 minutes in rotation at room temperature, the beads bound to the

antibody were precipitated with a magnetic rack, the supernatant was removed and the beads washed three times with lysis buffer. After overnight incubation in lysis buffer, the beads were recovered with a magnetic rack, washed three times with twice their volume of PBS-0.1% Tween-20 and finally eluted from the beads with 4xSDS.

The eluate was boiled for 4' and the samples fractionated by electrophoresis on SDS-containing polyacrylamide gels. Homemade polyacrylamide gels contained a stacking upper-part of 4% poly-acrylamide and a separator of variable polyacrylamide concentration (depending on the proteins of interest). Prestained protein ladder (PageRuler Prestain, Life Technologies) was loaded next to the samples to determine polypeptides molecular masses. The fractionation was carried out in a Mini-PROTEAN Tetra Cell System (BioRad) in SDS-PAGE running buffer (25mM Tris base, 192mM glycine and 0.1% SDS), with variable time and voltage conditions. After proteins were separated by size, they were transferred to a nitrocellulose membrane (Amersham Protran 0.2 NC) in a wet transfer system (BioRad, MiniTransBlot cell) with transfer buffer (25mM Tris base, 192mM glycine and 20% ethanol) at 400mA for 90 minutes. Some eluates where fractionated in NuPAGE 4-12% Bis-Tris pre-casted gels (Invitrogen) following the manufacturer's instructions, and then transferred to nitrocellulose as explained above.

After transfer, membranes were blocked with 10% skim-powder milk (Central Lechera Asturiana) diluted in PBS-0.1%Tween at room temperature for 40 minutes. Incubation with primary antibodies was carried out either for 1 hour at room temperature or over night at 4ºC. The concentration used of the different primary antibodies was variable, and were diluted in PBS-0.1%Tween to a concentration recommended by the manufacturer. After primary antibody incubation, membranes were washed three times for 10 minutes in PBS-0.1Tween and incubated for 1 hour with secondary antibody (NA931V/NA934V GE healthcare) diluted 1:10.000 in PBS-0.1%Tween.

The antibodies used were: RBM10 antibody (Sigma, HP17034972); SF3B3 (provided kindly by Lurhman's lab (from Max Planck Institute); U5-116 (ProteinTech Europe, 10208-1-AP); DHX15 (Abcam, ab70454); U2AF35 (Zuo et al., 1996) U2AF65 (MC3, (Gama-Carvalho et al., 1997)); PTB (PTB Ab (19108) (4ws)); PRPF19 (Bethyl, A300-101A); SF3B6 (NOVUS, NBP1-87430).

Discussion

The related aims of our study were 1) to provide an overview of current knowledge about the mutational landscape of RBM10 in tumors, adding also a phylogenetic perspective and 2) advance our understanding of the mechanisms by which RBM10 represses NUMB exon 9 splicing. Our results show that RBM10 RRM2 domain is the second most conserved domain, but also the one that comparatively accumulates the highest proportion of missense mutations (Figures 1 and 2). We have tested a battery of cancer-associated mutants in RBM10 RRM2 and found a range of effects on the regulation of NUMB exon 9 alternative splicing (Figure 3). Focusing on the RBM10 oncogenic mutant V354E, structural data of RBM10 RRM2 -354, 354V and V354E shows no major differences between the three domains and, consistently all three RRM2 domains bind to RNA with similar affinity (Figure 4). Finally, a proteomic analysis of RBM10 interactors showed that while wild-type RBM10 can interact with PRPF19, the oncogenic mutant V354E cannot, suggesting a possible functional partner of RBM10 relevant for splicing repression (Figure 5).

Amino acid conservation and cancer mutation distribution in RBM10

Our analysis of the distribution of mutations in RBM10 points towards the importance of the RRM2 domain as key for RBM10 function (Figure 1D). While RBM10 is a splicing

factor with a relatively high ratio of cancer-associated truncating mutations (only 57% of cancer-associated RBM10 mutants are missense), 40% are located N-terminal of the RRM2. Additionally, the RRM2 domain is relatively well conserved across RBM10 orthologues (Figure 2 E). Taken together, these observations suggest that the RRM2 plays a key role in RBM10 function, with positive selection for missense mutants within the domain and for mutants lacking the domain in cancer, suggesting that alterations in this domain play a role in cancer progression.

When the cancer mutation distribution analysis was combined with phylogenetic conservation, it became obvious that the two RRM domains display very different patterns, with RRM1 being less conserved and less mutated than RRM2. The results of Figure 4 and previous studies (Hernandez et al., 2016) argue that RRM2 domain can bind to NUMB exon 9 polypyrimidine tract, an *in vivo* RBM10 target (Bechara et al., 2013). It is therefore conceivable that RRM2 directs RBM10 binding to its target sequence, while RRM1 provides more general avidity for RNA binding without contributing to the binding specificity of RBM10, or that it provides some other function that is less evolutionarily conserved and less critical for the tumor suppressor properties of the protein.

It is also worth to mention that the region between amino acids 630 and 741 is not only significantly conserved, but also

enriched in cancer-associated mutations (of which 17/30 are missense mutations). Multiple alignment detects 36% sequence homology to a BAG domain, opening the possibility that this region harbors a non-annotated domain of this type (Doong et al., 2002). BAG domains have been characterized in BAG-family proteins, which interact though their BAG domain with the ATPase domain of Hsc70 and Hsp70 heat shock proteins to regulate their chaperoning activity (Reviewed in (Doong et al., 2002)). The analysis of truncation mutants suggest that this region of RBM10 can be relevant for function in cancer.

RRM2 mutants

We have tested the effects of a variety of RBM10-RRM2 cancer-associated mutants on NUMB exon 9 alternative splicing regulation. There range from very strong splicing inhibition (V333I), similar to the wild-type RBM10 (RBM10 - 354 and 354V), to reduced inhibitory capacity (V354D and E380K) or a complete loss of functionality (V354E and I316F). These results either suggest that some RRM2 mutations accumulate as passenger amino acid changes, or that these mutations are relevant for other activities of RBM10 RRM2 in cancer, e.g. causing misregulation of other alternative splicing events or other aspects of RNA biology important for the progression of specific tumors.

We have focused our attention on mutant V354E because it disrupts RBM10 function, it is located in a position that varies between RBM10 natural isoforms (that contain or not valine at this position) and it does not prevent RBM10 from binding RNA (Figure 4 and Hernández et al 2016). The NMR structures of RBM10 WT -354, WT 354V showed that there are no major structural differences between these variants and the ^{1}H-^{15}N-correlation spectrum of V354E indicates that the mutation does not influence the structure (Figure 4), as already proposed by structural modeling (Hernandez et al., 2016). In fact, even the RNA binding affinity of the three RRM2 domains, assessed by mobility shift and isothermal titration calorimetry, and supported also by NMR titration experiments, is similar suggesting that loss of regulatory function is not due to reduced RNA binding (Figure 4 D).

Given the dual function of RRM domains in RNA binding and protein-protein interactions (REF), we considered the possibility that, the loss of function in the V354E mutant was due to alterations in the profile of protein partners. Our co-IP followed by mass-spectrometry experiments, both using epitope-tagged versions of RBM10 and endogenous protein variants in different cellular models, suggest that this may be the case (Figure 5), as V354E displays reduced interaction with certain protein partners, particularly with PRPF19.

Although PRPF19 has classically been described as a member of the B-act complex, it was also reported to be recruited through an interaction with U2AF early-on in spliceosome assembly (David et al., 2011), suggesting the existence of a non-canonical spliceosomal assembly pathway.

One of the first steps during the canonical spliceosome assembly is the formation of the E-complex, where U2AF will bind to the poly-pyrimidine tract of the 3'SS and the down-stream exon; some splicing inhibitors like Sxl and PTB have been reported to inhibit the splicing reaction by preventing U2AF from binding to the 3'SS. Once the E-complex has properly formed U2 will be recruited to the branch point region, displacing SF1 from its binding. The binding of U2 to the intron will be stabilized by several U2-associated proteins, and the A-complex will be formed. Afterwards, the tri-snRNP, together with the PRPF19 complex will be recruited. Although the PRPF19 complex has classically been described as playing a role in the B-active stage, there are several reports suggesting it may be playing a role also in previous stages of the splicing reaction (Chung et al., 1999; Tardiff et al., 2006; Wang et al., 2003). It has also been reported that PRP19 interacts with U2AF65 and the phosphorylated CAP of Pol II during co-transcriptional splicing (David et al., 2011).

The would like to propose a model for RBM10 splicing repression by which, instead of just competing with U2AF for binding to the PPT, RBM10 will be interacting with U2 associated proteins (Figure 5 C) in a repressive complex that, although formed by several U2 members would be unable to promote the splicing reaction. In this model PRP19 would be recruited in this “RBM10-represive complex” bringing the tri-snRNP to early to the reaction generating an abortive complex. While the mutant V354E, due to its lack of interaction with PRP19 will not be able to inhibit the splicing.

Figure 6

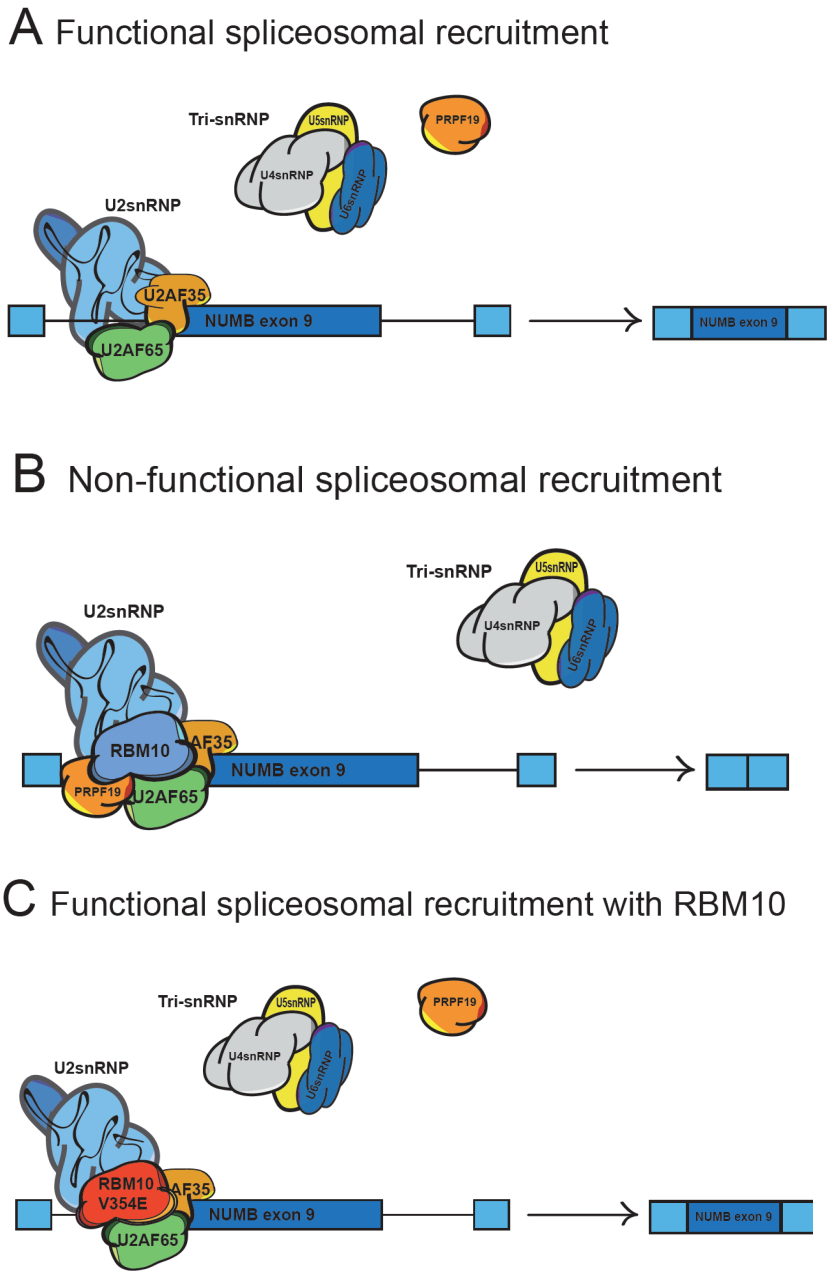


Figure 6. Proposed model for the mechanism of NUMB exon 9 repression by RBM10. RBM10 binding to NUMB intron 8 3' splice site helps to recruit U2 snRNP and PRP19 components in a configuration that prevents further assembly of the tri-snRNP and subsequent steps of spliceosome assembly. The V354E mutant fails to interact with PRP19, allowing normal steps in tri-snRNP assembly and therefore allowing activation of the 3' splice site.

Supplementary Material

Supplementary Table 1: Unique RBM mutations list.

https://www.dropbox.com/sh/1qr36dae9ppe1rr/AACYvGamQ8o1j3qvC8JzGzR4a?dl=0

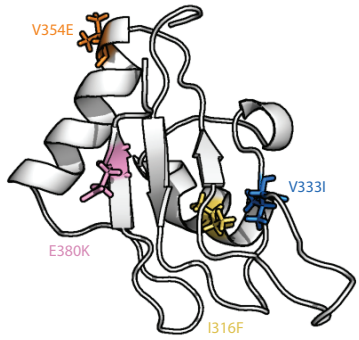
Supplementary Table 2: Frequency of missense mutations normalized by domain length.

	RRM1	ZnF1	RRM2	OCRE	ZnF2	Gpatch	Non-annotated
RBM5	0.163	0.069	0.119	0.034	0.000	0.130	0.181
RBM6	0.063	0.179	0.183	0.106	0.200	0.109	0.109
RBM10	0.100	0.200	0.167	0.150	0.360	0.196	0.154

Supplementary Table 2. Number of missense mutation in each domain normalized by the domain's length.

Supplementary Table 3: RBM10 mutants

A

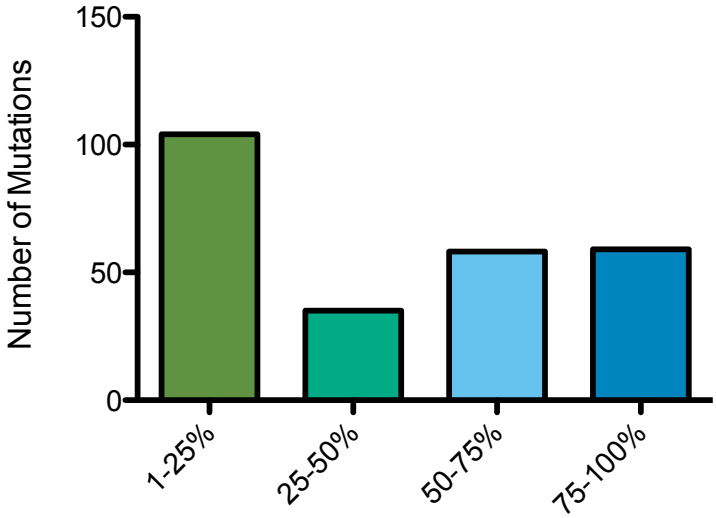


Mutation Code	Origin	ID	Tissue Found
WT 354V	Swiss-Prot	P98175-1	
WT -354	Swiss-Prot	P98175-2	
V354E	Bechara et al 2013	NA	A549 Lung Adenocarcinoma
V354D	Hernández et al 2016	NA	Tailored mutation
I316F	cBioportal	LUAD-S01357	Lung Adenocarcinoma
V333I	TCGA, cBioportal	TCGA-AA-3811-01	Colorectal Cancer
E380K	Cosmic	COSM1121716	Endometrial Cancer

(A) Representation of RBM10 RRM2. Mutated amino acids in cancer-associated RBM10 variants, used in Figure 3, are highlighted.

Supplementary Table 3. RBM 10 wild type (WT) variants and missense mutants tested in functional alternative splicing assays (Figure 3). ID corresponds to a unique identifier of the mutant in their reported collections (Swiss-Prot for the natural isoforms, TCGA, cBioportal, Cosmic or publications for the mutants). Tissue Found column indicates the source of the mutant reported.

Supplementary Figure 1



Supplementary Figure 1. Number of Mutations across RBM10 length. Total number of mutations (including all types) where calculated for each quarter of the protein.

Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shirole, N. H., . . . Krainer, A. R. (2015). SRSF1-Regulated Alternative Splicing in Breast Cancer. *Molecular Cell, 60*(1), 105-117. doi:10.1016/j.molcel.2015.09.005

Angeloni, D. (2007). Molecular analysis of deletions in human chromosome 3p21 and the role of resident cancer genes in disease. *Brief Funct Genomic Proteomic, 6*(1), 19-39. doi:10.1093/bfgp/elm007

Aravind, L., & Koonin, E. V. (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends in Biochemical Sciences, 24*(9), 342-344.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., . . . Frey, B. J. (2010). Deciphering the splicing code. *Nature, 465*(7294), 53-59. doi:10.1038/nature09000

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E., & Valcárcel, J. (2013). RBM5, 6, and 10 Differentially Regulate NUMB Alternative Splicing to Control Cancer Cell Proliferation. *Molecular Cell, 52*(5), 720-733. doi:10.1016/j.molcel.2013.11.010

Bonnal, S., Martínez, C., Förch, P., Bachi, A., Wilm, M., & Valcárcel, J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Molecular Cell, 32*(1), 81-95. doi:10.1016/j.molcel.2008.08.008

Bonnal, S., Vigevani, L., & Valcárcel, J. (2012). The spliceosome as a target of novel antitumour drugs. *Nature reviews. Drug discovery, 11*(11), 847-859. doi:10.1038/nrd3823

Callebaut, I., & Mornon, J. P. (2005). OCRE: a novel domain made of imperfect, aromatic-rich octamer repeats. *Bioinformatics, 21*(6), 699-702. doi:10.1093/bioinformatics/bti065

Choi, Y. (2012). A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, 414-417

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One, 7*(10), e46688. doi:10.1371/journal.pone.0046688

Chung, S., McLean, M. R., & Rymond, B. C. (1999). Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA, 5*(8), 1042-1054.

Collisson, E. A., Campbell, J. D., Brooks, A. N., Berger, A. H., Lee, W., Chmielecki, J., . . . Tsao, M.-S. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature, 511*(7511), 543-550. doi:10.1038/nature13385

Danan-Gotthold, M., Golan-Gerstl, R., Eisenberg, E., Meir, K., Karni, R., & Levanon, E. Y. (2015). Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Research, 43*(10), 5130-5144. doi:10.1093/nar/gkv210

David, C. J., Boyne, A. R., Millhouse, S. R., & Manley, J. L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes and Development, 25*(9), 972-982. doi:10.1101/gad.2038011

David, C. J., & Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & Development, 24*(21), 2343-2364. doi:10.1101/gad.1973010

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR, 6*(3), 277-293.

Doong, H., Vrailas, A., & Kohn, E. C. (2002). What's in the 'BAG'?--A functional domain analysis of the BAG-family proteins. *Cancer Letters, 188*(1-2), 25-32.

Dvinge, H., & Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. *Genome Med, 7*(1), 45. doi:10.1186/s13073-015-0168-9

Edamatsu, H., Kaziro, Y., & Itoh, H. (2000). LUCA15, a putative tumour suppressor gene encoding an RNA-binding nuclear protein, is down-regulated in ras-transformed Rat-1 cells. *Genes to Cells, 5*(10), 849-858.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., . . . Searle, S. M. (2014). Ensembl 2014. *Nucleic Acids Research, 42*(Database issue), D749-755. doi:10.1093/nar/gkt1196

Fushimi, K., Ray, P., Kar, A., Wang, L., Sutherland, L. C., & Wu, J. Y. (2008). Up-regulation of the proapoptotic caspase 2 splicing isoform by a candidate tumor suppressor, RBM5. *Proc Natl Acad Sci U S A, 105*(41), 15708-15713. doi:10.1073/pnas.0805569105

Gama-Carvalho, M., Krauss, R. D., Chiang, L., Valcarcel, J., Green, M. R., & Carmo-Fonseca, M. (1997). Targeting of

U2AF65 to sites of active splicing in the nucleus. *Journal of Cell Biology, 137*(5), 975-987.

Giannakis, M., Mu, X. J., Shukla, S. A., Qian, Z. R., Cohen, O., Nishihara, R., . . . Garraway, L. A. (2016). Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep*. doi:10.1016/j.celrep.2016.03.075

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., . . . Stelzl, U. (2012). Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Molecular Cell, 45*(4), 567-580. doi:10.1016/j.molcel.2011.12.034

Hernandez, J., Bechara, E., Schlesinger, D., Delgado, J., Serrano, L., & Valcarcel, J. (2016). Tumor suppressor properties of the splicing regulatory factor RBM10. *RNA biology, 13*(4), 466-472. doi:10.1080/15476286.2016.1144004

Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., . . . Meyerson, M. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell, 150*(6), 1107-1120. doi:10.1016/j.cell.2012.08.029

Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., & Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Structural & Molecular Biology, 14*(3), 185-193. doi:10.1038/nsmb1209

Magis, C., Taly, J. F., Bussotti, G., Chang, J. M., Di Tommaso, P., Erb, I., . . . Notredame, C. (2014). T-Coffee: Tree-based consistency objective function for alignment evaluation. *Methods Mol Biol, 1079*, 117-129. doi:10.1007/978-1-62703-646-7_7

Maraver, A., Fernandez-Marcos, P. J., Herranz, D., Canamero, M., Munoz-Martin, M., Gomez-Lopez, G., . . . Serrano, M. (2012). Therapeutic effect of gamma-secretase inhibition in KrasG12V-driven non-small cell lung carcinoma by derepression of DUSP1 and inhibition of ERK. *Cancer Cell, 22*(2), 222-234. doi:10.1016/j.ccr.2012.06.014

Misquitta-Ali, C. M., Cheng, E., O'Hanlon, D., Liu, N., McGlade, C. J., Tsao, M. S., & Blencowe, B. J. (2011). Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. *Molecular and Cellular Biology, 31*(1), 138-150. doi:10.1128/MCB.00709-10

Nordentoft, I., Lamy, P., Birkenkamp-Demtroder, K., Shumansky, K., Vang, S., Hornshoj, H., . . . Orntoft, T. F. (2014). Mutational context and diverse clonal development in early

and late bladder cancer. *Cell Rep, 7*(5), 1649-1663. doi:10.1016/j.celrep.2014.04.038

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology, 302*(1), 205-217. doi:10.1006/jmbi.2000.4042

Oh, J. J., Grosshans, D. R., Wong, S. G., & Slamon, D. J. (1999). Identification of differentially expressed genes associated with HER-2/neu overexpression in human breast cancer cells. *Nucleic Acids Research, 27*(20), 4008-4017.

Oh, J. J., West, A. R., Fishbein, M. C., & Slamon, D. J. (2002). A candidate tumor suppressor gene, H37, from the human lung cancer tumor suppressor locus 3p21.3. *Cancer Res, 62*(11), 3207-3213.

Oltean, S., & Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene, 33*(46), 5311-5318. doi:10.1038/onc.2013.533

Orengo, J. P., Bundman, D., & Cooper, T. a. (2006). A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Research, 34*(22), e148-e148. doi:10.1093/nar/gkl967

Rintala-Maki, N. D., Goard, C. A., Langdon, C. E., Wall, V. E., Traulsen, K. E., Morin, C. D., . . . Sutherland, L. C. (2007). Expression of RBM5-related factors in primary breast tissue. *Journal of Cellular Biochemistry, 100*(6), 1440-1458. doi:10.1002/jcb.21134

Schmidt, E., & Guntert, P. (2015). Automated structure determination from NMR spectra. *Methods Mol Biol, 1261*, 303-329. doi:10.1007/978-1-4939-2230-7_16

Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M. A., . . . Eyras, E. (2015). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *bioRxiv*, 023010-023010. doi:10.1101/023010

Song, Z., Wu, P., Ji, P., Zhang, J., Gong, Q., Wu, J., & Shi, Y. (2012). Solution structure of the second RRM domain of RBM5 and its unusual binding characters for different RNA targets. *Biochemistry, 51*(33), 6667-6678. doi:10.1021/bi300539t

Tardiff, D. F., & Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. *RNA, 12*(6), 968-979. doi:10.1261/rna.50506

Tessier, S. J., Loiselle, J. J., McBain, A., Pullen, C., Koenderink, B. W., Roy, J. G., & Sutherland, L. C. (2015). Insight into the role of alternative splicing within the RBM10v1 exon 10 tandem donor site. *BMC Res Notes, 8*, 46. doi:10.1186/s13104-015-0983-5

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., . . . Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins, 59*(4), 687-696. doi:10.1002/prot.20449

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell, 136*(4), 701-718. doi:10.1016/j.cell.2009.02.009

Wang, Q., Hobbs, K., Lynn, B., & Rymond, B. C. (2003). The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts. *Journal of Biological Chemistry, 278*(10), 7875-7883. doi:10.1074/jbc.M210839200

Zhao, L., Li, R., Shao, C., Li, P., Liu, J., & Wang, K. (2012). 3p21.3 tumor suppressor gene RBM5 inhibits growth of human prostate cancer PC-3 cells through apoptosis. *World J Surg Oncol, 10*, 247. doi:10.1186/1477-7819-10-247

Zuo, P., & Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes & Development, 10*(11), 1356-1368.

3. Third Manuscript

NUMB alternative splicing regulation in lung tumors

NUMB alternative splicing regulation in lung tumors

Jordi Hernández, Pablo J. Fernandez, Miguel Rovira, Jessica Vitos, Caterina Coli, Panagiotis Papasaikas, Manuel Serrano, Juan Valcárcel

Jordi Hernández performed all the experiments and analyses, except those of Figures 1, 2 and 5 (see below) and wrote the manuscript under Juan Valcárcel's supervision.
The analyses of Figure 1 were performed by Caterina Coli (CRG), as proposed by Jordi Hernández.
The analyses of Figure 2 were carried out by Caterina Coli and Panagiotis Papasaikas (CRG), as proposed by Jordi Hernández.
The experiment of Figure 5 A was performed by Jessica Vitos and Jordi Hernández.
The experiment of Figure 5 B was carried out jointly by Pablo J. Fernandez and Jordi Hernández under the supervision of Manuel Serrano and Juan Valcárcel.
The experiment of Figure 5 C was performed by Miguel Rovira under the supervision of Manuel Serrano.

Introduction

Alternative splicing is a major regulatory mechanism in eukaryotic gene expression, and its alterations have been linked with a plethora of various diseases, including retinitis pigmentosa, spinal muscular atrophy and cancer (Blencowe, 2006; Cooper et al., 2009)

The relationships between alternative splicing and human pathologies are multiple. In some diseases, a particular splicing factor is frequently mutated, e.g. Prp8 in retinitis pigmentosa (McKie et al., 2001). Other illnesses are caused by the miss-regulation of a single splicing event, like spinal muscular atrophy (SMA), where the SMN2 gene cannot replace the function of the mutated SMN1 gene cannot be provided by the SMN2 gene because exon 7 is mostly skipped due to a single nucleotide difference with SMN1 exon 7 (Cho et al., 2010). Some splicing factors can act as oncogenes, like SRSF1 (Anczuków et al., 2015; Karni et al., 2007), or as tumor suppressors, like RBM10 (Hernández et al). Splicing factors can be mutated in certain cancers, like SF3B1 in leukemia (Quesada, Conde, et al., 2012; Quesada, Ramsay, et al., 2012), or their levels altered, like RBM5 in breast cancers (Oh et al., 1999) (Oh et al., 2002; Rintala-Maki et al., 2007).

Genome-wide studies have reported alternative splicing events commonly altered in cancer (Danan-Gotthold et al., 2015; Dvinge et al., 2015; Sebestyén et al., 2015). For instance, in lung tumors NUMB exon 9 inclusion is increased when compared to healthy lung tissue (Misquitta-Ali et al., 2011; Sebestyén et al., 2015)). NUMB exon 9 codes for 49 amino acids, located in a proline-rich region (PRR) in the middle part of the protein. The isoform lacking NUMB exon 9 (NUMB-PRR^S) plays an important role in cellular differentiation and in the development of the nervous system (Bani-Yaghoub et al., 2007; Dho et al., 1999; Toriya et al., 2006; Verdi et al., 1999). In contrast, the NUMB isoform that includes exon 9 (NUMB-PRR^L) promotes cell proliferation (Bechara et al., 2013; Dho et al., 1999; Toriya et al., 2006; Verdi et al., 1999), and has been associated with an active Notch pathway (Bechara et al., 2013; Frise et al., 1996; Misquitta-Ali et al., 2011), which is important for cancer progression (reviewed in (Roy et al., 2007)) .Thus, the inhibition of the Notch pathway using gamma-secretase inhibitors has therapeutic effects in mice models of KrasG12V-driven non-small cell lung carcinomas (Maraver et al., 2012)

Lung cancer is a leading cause of death world-wide, causing over a million human deaths per year (Collisson et al., 2014). Thus, the identification of novel therapeutic targets is of

paramount importance for reducing lung cancer-related mortality.

The modulation of alternative splicing events holds the promise of novel therapeutic tools (reviewed in (Havens et al., 2013)) One interesting strategy is the use of modified antisense oligonucleotides (AONs). AONs are short synthetic nucleic acids capable of base pairing with a target sequence; typically they include chemical modifications aimed to increase their stability, affinity for RNA or efficient delivery (reviewed in (Havens et al., 2016)). Important progress has been made in the treatment of SMA using AONs, which have reached clinical trials. (http://ir.ionispharma.com/phoenix.zhtml?c=222170&p=irol-newsArticle&ID=2191319). SMA is characterized by the death of motor neurons caused by the loss of SMN protein (Lefebvre et al., 1995), the amount of residual SMN protein correlating with the severity of the disease, whose effects range from muscular weakness to paralysis (Cho et al.). AONs targeting an intronic splicing silencer enhance inclusion of SMN2 exon 7, providing therapeutic levels of SMN protein (Hua et al., 2010; Hua et al., 2011).

In this report we have attempted to identify AONs able to modulate NUMB exon 9 alternative splicing such that they inhibit proliferation of lung adenocarcinoma cells, as a first

step towards the development of a potential therapeutic tool for reducing tumor progression.

Results

Increased NUMB exon 9 inclusion in different types of tumors

Increased NUMB exon 9 inclusion has been reported as one of the most common splicing alterations in lung cancer (Misquitta-Ali et al.; Sebestyén et al., 2015). Increased NUMB exon 9 inclusion has been shown to increase the proliferative capacity of cancer cells *in vitro* (Bechara et al., 2013). To further analyze tumor types displaying increased NUMB exon 9 inclusion we took advantage of data from (Sebestyén et al., 2015) that analyzed RNA-sequencing data from 11 tumor types from the TCGA consortium.

By comparing the levels of NUMB exon 9 inclusion (PSI) in paired tumor/non-tumor samples, we detected significant increases in Breast Invasive Carcinoma, Liver Hepatocellular Carcinoma, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma and Prostate Adenocarcinoma (Figure 1, compare violin plots representing the distribution of NUMB exon 9 inclusion (Y axis) for 11 tumor types (X axis) in healthy (blue) vs cancerous (orange) tissue from the same patient; PSI calculation and statistical analyses were performed as in (Sebestyén et al., 2015).

Therefore, in 5 out of the 11 tumor types analyzed, NUMB exon 9 inclusion is significantly increased when compared to

healthy tissue, suggesting that the misregulation of NUMB alternative splicing is a relatively common alteration in cancer. This is consistent with the observation that NUMB-PRR^{L}, the NUMB protein isoform encoded by the mRNA resulting from exon 9 inclusion, is associated with cell proliferation *in vitro* (Bechara et al., 2013; Misquitta-Ali et al., 2011), and in tumor xenografts *in vivo* (Bechara et al., 2013).

Figure 1

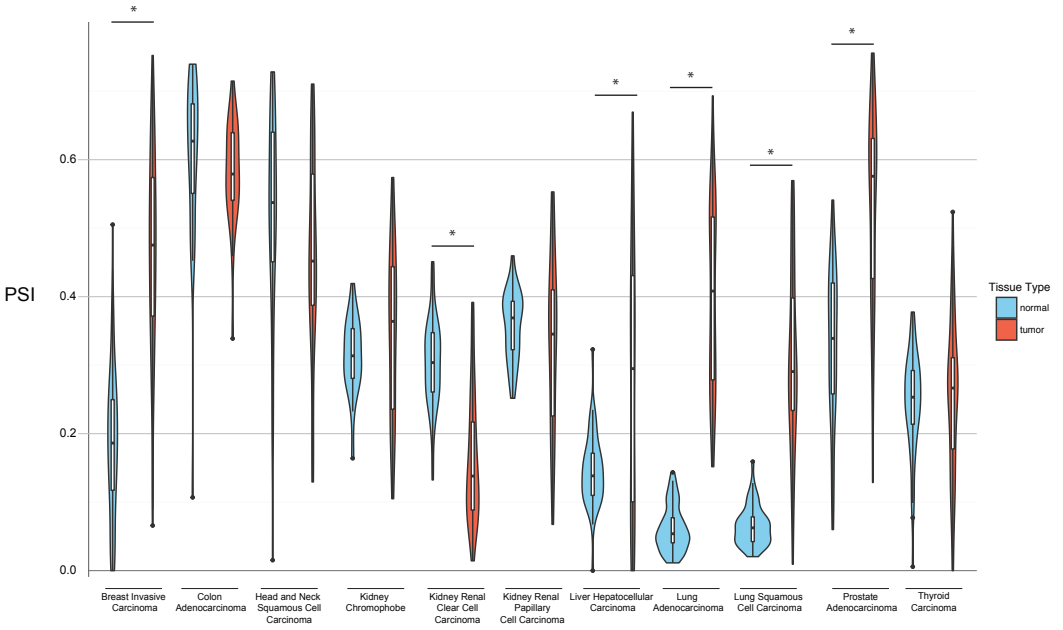


Figure 1. Distribution of NUMB exon 9 PSI values across different tumor types. Violin plot analysis of NUMB exon 9 PSI distribution in paired (blue: normal tissue, red: tumor) samples across 11 tumor types. The analysis was carried out using data published in (Sebestyén et al.). Statistical significance was addressed with Mann Whitney U test, statistical significance is represented by *.

Identification of possible NUMB exon 9 alternative splicing regulators

Despite its prominent role in controlling cell proliferation (Bechara et al., 2013; Misquitta-Ali et al., 2011) NUMB has been classically described as playing a role in the differentiation of the nervous System (Bani-Yaghoub et al., 2007; Dho et al., 1999; Toriya et al., 2006; Verdi et al., 1999). Therefore we considered of interest to assess how is the distribution of NUMB exon 9 alternative splicing across the different human tissues. For that we took advantage of the RNA-sequencing data from the GTEx dataset (Consortium, 2015). We used SANJUAN pipeline to calculate NUMB exon 9 PSI (percentage of exon inclusion) across the different tissues present in the GTEx (Figure 2) and counted the FPKM reads of NUMB, to see its expression levels. As a rule of thumb, NUMB exon 9 is generally skipped (PSI values are generally low), with the exception of mesoderm-derived tissues where the PSI is slightly higher. It is of special relevance to mention that, although deriving from the mesoderm lineage, lung displays very low levels of NUMB exon 9 inclusion. Interestingly, NUMB FPKM are highest in lung. This is of special relevance because of the increased inclusion of NUMB exon 9 in lung adenocarcinomas (Misquitta-Ali et al., 2011). It is also worth to mention that the levels the of NUMB exon 9 inclusion are extremely low in ectoderm-derived tissue (specially nervous tissue), most probably related to the classic role of NUMB in nerve cell

differentiation (Bani-Yaghoub et al., 2007; Dho et al., 1999; Toriya et al., 2006; Verdi et al., 1999).

We performed a network based analysis on this data set to identify genes, whose expression correlates with NUMB exon 9 inclusion. Only genes whose expression varies across different tissues were used. This analysis revealed a list of potential regulators (Table 1). Among these potential regulators we were particularly interested in SRSF9 (Table 1), because of two reasons. First, an analysis of correlations between changes in expression of splicing regulators and changes in NUMB alternative splicing in a variety of tumors also revealed a potential functional relationship between SRSF and NUMB splicing (Sebestyén et al. 2015). Second, bioinformatic analysis of putative binding sites for splicing regulators revealed a potential SRSF9 binding site located in a region close to the 3' end of the exon that subsequent experiments revealed that it displays a potent ESE activity (see below, Figure 3).

Figure 2

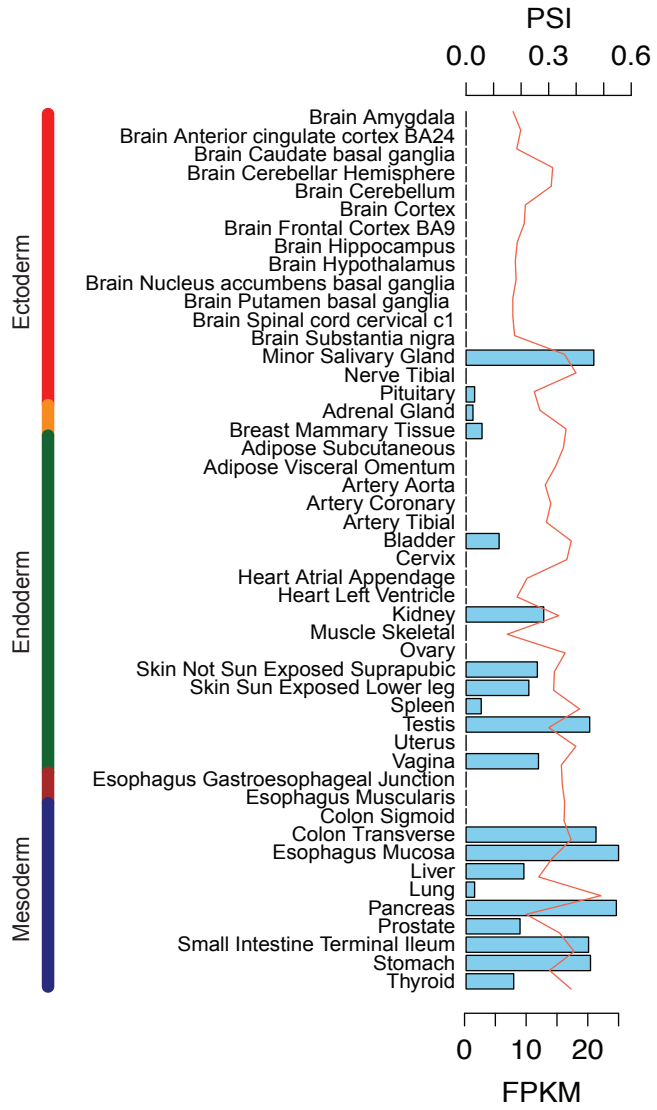


Figure 2. Distribution of NUMB exon 9 PSI values across human tissues. RNA sequencing data from the GTEx project was analyzed using SANJUAN pipeline to estimate NUMB exon 9 PSI values (blue bars) in different tissues, grouped by embryonic layer of origin. The levels of RNA expression (FPKM, red line) were also obtained from the same dataset.

Table 1

Gene Name	Correlation with NUMB exon 9 PSI
DHX32	0.79
PPIL2	0.71
PTBP3	0.67
WIBG	0.65
DDX52	0.64
LARP4B	0.64
SRSF9	0.61
CNOT4	-0.60
SRP14	-0.61
SNRNP27	-0.62
CWC26	-0.64
IFIT5	-0.65
DNAJC8	-0.70
WBP4	-0.73

Table 1. Genes whose gene expression (mRNA levels) correlates with NUMB exon 9 inclusion in normal human tissues. Estimates of NUMB exon 9 PSI and putative regulator gene expression were extracted from GTEx datasets.

Identification of NUMB exon 9 Exonic Splicing Enhancers and blockage by the use of Modified Antisense Oligonucleotides

Our previous work to modulate NUMB exon 9 inclusion in lung cancer A549 cells (as well as in ovarian cancer HeLa and breast cancer MCF7 cell lines), 2'-O-methyl, phosphothioate-modified Antisense Oligonucleotides (AONs) complementary to a region of the pre-mRNA encompassing the 5' splice site (SS) were used to promote exon 9 skipping (Bechara et al., 2013). To further explore other exonic regions that could be effectively targeted by AONs to induce NUMB exon 9 skipping, we performed a systematic scanning of NUMB exon 9 using 21 nucleotide-AONs complementary to consecutive, overlapping regions of the exon (a total of 24 AONs covering the 144nt exon) (Figure 3A). We also included AONs complementary to regions of the pre-mRNA encompassing the 3' SS or the 5' SS, the branch point (BP) region or a Random sequence (RND) as controls.

A stably transformed HEK-293T cell line expressing a reporter NUMB exon 9 minigene (RG6-NUMB, (Bechara et al., 2013; Hernandez et al., 2016)) was transfected with 100nM of individual AONs. When transfected at 0.2ng/µl, the reporter RG6-NUMB displays a high level of NUMB exon 9 inclusion (approximately 0.9 PSI), which facilitates the detection of any switch towards skipping (the effect caused by the AONs). 24 hours after transfection, RNA was isolated,

retro-transcribed and amplified with primers specific for detecting NUMB exon 9 inclusion/skipping in the minigene. The results of Figure 3A, were NUMB exon 9 inclusion (PSI) was calculated for three technical replicates for each AON, show that cells transfected with the control RND oligo displayed almost full exon 9 inclusion (PSI 0.93, first panel from the left in the lower row, while transfection with other AONs induced a wide range of exon skipping effects. Blocking of the BP region had a very limited effect (0.87, first panel from the left in the upper row), while blocking the 3'SS or the 5'SS led to increased levels of exon skipping (PSI of 0.61 and 0.30, respectively). The significant remaining levels of exon inclusion, however, indicated that the AONs were not capable of preventing the binding of the splicing machinery.

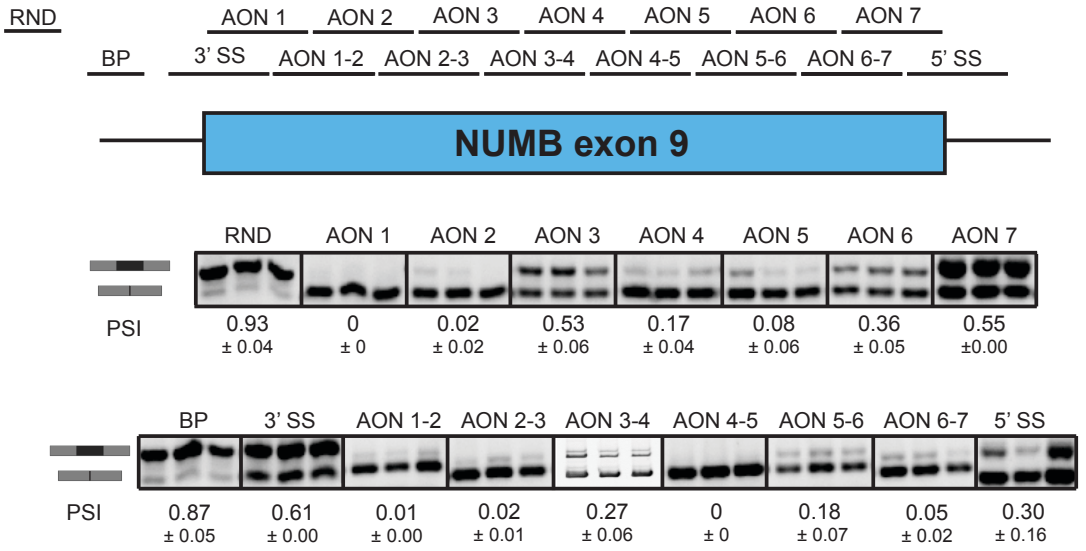
Remarkably, however, many of the AONs complementary to exonic sequences induced higher levels of exon skipping, suggesting that the exon is dense in regulatory (Exonic Splicing Enhancer, ESE) regions. In particular, three regions stand out:

a) The one covered by AON 1, AON 2 and AON 1-2 and AON 2-3 (inducing virtually complete skipping: upper row, second and third panels, lower row, third and fourth panels). The more limited effects of the 3' SS AON and AON 3-4 delineates the boundary of this ESE between nucleotides A (+4) and C (+30) of the exon

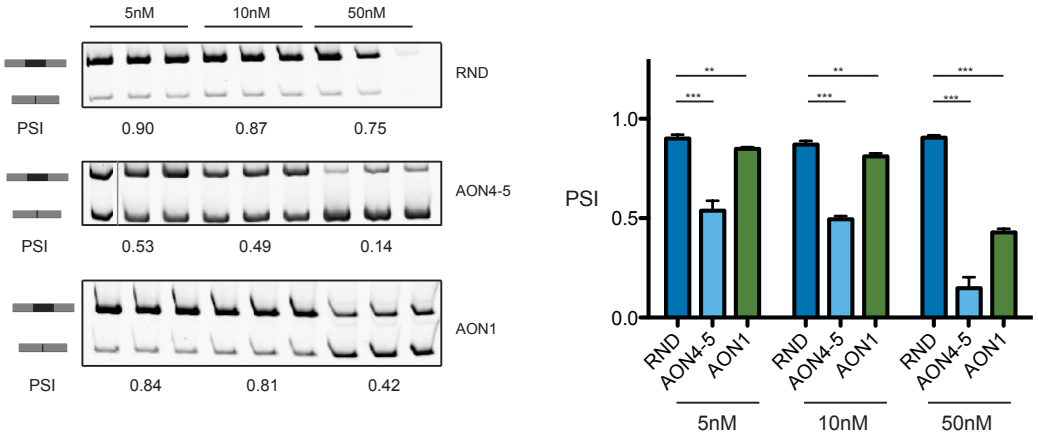
b) The one covered by AON 4, AON 5 and AON 4-5, inducing also very significant levels of skipping (83 to nearly 100%: upper row, panels 5-6; lower row, panel 6). The boundaries of this ESE are likely to be between residues C (+72) and T (+92).

c) The one covered by AON 6-7 (lower row, panel 8). The more limited effects of the flanking AONs 6 and 7 suggests that the boundaries of this ESE are contained within nucleotides G (+114) and C (+134).

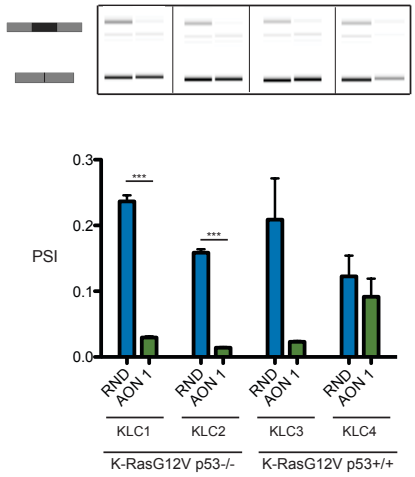
Figure 3


RND
AON 1
AON 2
AON 3
AON 4
AON 5
AON 6
AON 7
BP
3' SS
AON 1-2
AON 2-3
AON 3-4
AON 4-5
AON 5-6
AON 6-7
5' SS
NUMB exon 9
RND AON 1 AON 2 AON 3 AON 4 AON 5 AON 6 AON 7
PSI
0.93 ± 0.04
0 ± 0
0.02 ± 0.02
0.53 ± 0.06
0.17 ± 0.04
0.08 ± 0.06
0.36 ± 0.05
0.55 ±0.00
BP 3' SS AON 1-2 AON 2-3 AON 3-4 AON 4-5 AON 5-6 AON 6-7 5' SS
PSI
0.87 ± 0.05
0.61 ± 0.00
0.01 ± 0.00
0.02 ± 0.01
0.27 ± 0.06
0 ± 0
0.18 ± 0.07
0.05 ± 0.02
0.30 ± 0.16

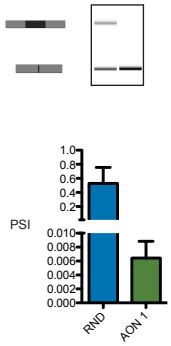
B


5nM
10nM
50nM
RND
PSI
0.90
0.87
0.75
AON4-5
PSI
0.53
0.49
0.14
AON1
PSI
0.84
0.81
0.42
PSI
1.0
0.5
0.0
**
***
RND
AON4-5
AON1
5nM
10nM
50nM

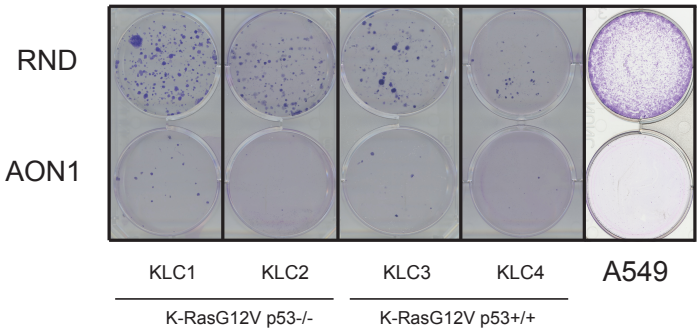
C


PSI
0.3
0.2
0.1
0.0
***
***
RND
AON 1
KLC1
KLC2
KLC3
KLC4
K-RasG12V p53-/-
K-RasG12V p53+/+

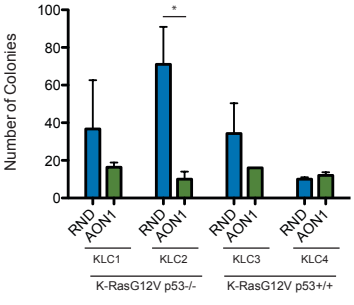
D

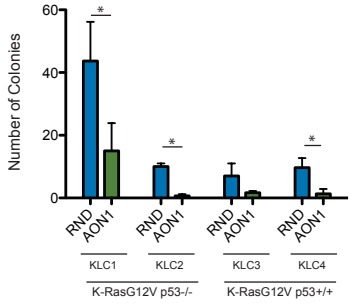

PSI
1.0
0.8
0.6
0.4
0.2
0.010
0.008
0.006
0.004
0.002
0.000
RND
AON 1

E


RND
AON1
KLC1
KLC2
KLC3
KLC4
A549
K-RasG12V p53-/-
K-RasG12V p53+/+

F


Number of Colonies
100
80
60
40
20
0
*
RND
AON1
KLC1
KLC2
KLC3
KLC4
K-RasG12V p53-/-
K-RasG12V p53+/+


Number of Colonies
60
40
20
0
*
*
*
RND
AON1
KLC1
KLC2
KLC3
KLC4
K-RasG12V p53-/-
K-RasG12V p53+/+

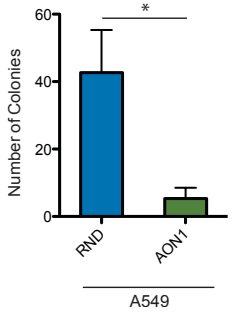

Number of Colonies
60
40
20
0
*
RND
AON1
A549

Figure 3. Identification of Exonic Splicing Enhancers (ESEs) in NUMB exon 9. (A) "Exonic walk" with Antisense Oligonucleotides (AONs). AONs complementary to the exonic and neighbouring regions indicated on top were transfected into HEK-293 cells stably expressing a RG6-NUMB minigene and the patterns of NUMB exon 9 inclusion/skipping analysed by RT-PCR Values of average PSI and standard deviation for the three independent replicas are indicated. (B) Titration of AONs RND (control), AON4-5 and AON1 in human A549 lung cancer cell line at the indicated concentrations. Results of RT-PCR as in (A) are shown in the left panel and quantified in the right panel. Student's t-test was used for determining the statistical significance value. P-value<0.05 *, p-value<0.001 ***(C) Effect of AON1 on NUMB exon 9 alternative splicing in mouse lung cancer cells. Top: results of RT-PCR. Bottom: quantification of PSI values, as in (B). The cell lines utilized and their p53 status are indicated. (D) Effect of AON1 on NUMB exon 9 alternative splicing in A549 lung cancer cells (E) Clonogenic assays of the indicated lung cancer cell lines after treatment with AON RND or AON1. Clones were detected by crystal violet staining 7 days after transfection. (F) Quantification of clonogenic assays of the indicated mouse lung cancer cells after treatment with 100nM (left panel) or 50nM (middle panel) of AONs. Results for human lung cancer A549 cells are shown on the right. Student's t-test was used for determining the statistical significance value. P-value<0.05 *.

Having identified strong ESEs for NUMB exon 9 inclusion, we wanted next to test if these sequences were also functional in the endogenous NUMB transcripts in a lung adenocarcinoma cell line, A549 that displays high levels of NUMB exon 9 inclusion (Misquitta-Ali et al., 2011), and where NUMB splicing controls cell proliferation (Bechara et al., 2013). Three different concentrations of AONs 1 and AON 4-5 (5, 10 and 50nM) were transfected into A549 cells, and 24h after the transfection, RNA was isolated and the endogenous levels of NUMB exon 9 inclusion/skipping analyzed by RT-PCR. The results of Figure 2B show that, indeed, titratable effects on exon skipping were detected for both AON1 and AON 4-5 even at 5nM, progressively increasing at higher AON concentrations until skipping products were majority (despite

the likely presence of non-transfected cells in the culture), while no effect was detectable with the RND oligo (Figure 3B, compare lanes 1-3 with 4-6 and 7-9 of the left panels, quantification of a minimum of three biological replicas is represented in the right part of the Figure).

We next decided to test the effect of the AONs in mouse cancer cells, for several reasons. First, we wanted to test whether mouse cancer cells respond to the AONs, indicating that the function of the ESEs is also maintained, which would be helpful for experiments using genetic mouse models (see below). Second, we wanted to test if the modulation of NUMB alternative splicing also affect the proliferative capacity of mice lung cancer cells. Third, we wanted to test if the possible effect of the AONs in mouse cancer cells proliferation was or not mediated by p53, because NUMB was shown to control p53 activity by preventing its degradation (Colaluca et al., 2008). For that we used a battery of cancer cell lines derived from a KRAS-G12V lung adenocarcinoma mouse model, with and without p53 (Ambrogio et al., 2014); While the sequence of the second ESE (targeted by AON4-5)was not totally conserved, the sequence of the 5' most ESE (targeted by the AON1) is 100% conserved between human and mice. For these reasons, we transfected AON1 at a concentration of 100nM in four different lung adenocarcinoma KRAS-G12V-derived mice cell lines (KLC1-4), in parallel with A549 cells, isolated RNA 24 hours after transfection and the levels of

NUMB exon 9 inclusion/skipping measured by capillary electrophoresis.

The upper panel of Figure 3C shows a representative image of the electropherograms corresponding to the NUMB isoforms in different cell lines treated with random oligo or AON1, while the lower panel shows a quantification of the PSI values provided by the capillary electrophoresis software for three technical replicas of the experiments. The results show that although NUMB exon 9 skipping is very prominent in these cells, the AONs further decrease PSI values. The significance of the effects, however, is more prominent in $p53^{-/-}$ cells (compare lanes 1-4 with 5-8). The Figure 3D represents the same experiment done in parallel for the human LUAD cell line A549 (Figure 3D).

We also tested if modulation of NUMB exon 9 alternative splicing had an effect on the clonogenic capacity of the mouse adenocarcinoma cancer cells. Figure 2E shows representative wells from the colony formation experiment. The upper well of each panel was transfected with 100nM of RND AON while the lower well was transfected with 100nM of AON1 (Figure 3E). Quantification of biological triplicates indicates a reduction in colony formation upon transfection with AON1 for three of the cell lines (KLC1-KLC3) and little effect on the KCL4 cell line, which has a limited clonogenic capacity even under control conditions (Figure 3F). KCL4 was

also the cell line that did not show an increase in exon 9 skipping upon transfection with AON1 (Figure 3C), arguing the effects on colony formation indeed correlate with changes in NUMB exon 9 alternative splicing. Transfection of mice cells with reduced amounts of AON1 displayed the same profile (Figure 3F right).

Characterization of 5' ESE (ESE1) region and potential cognate trans-acting regulators

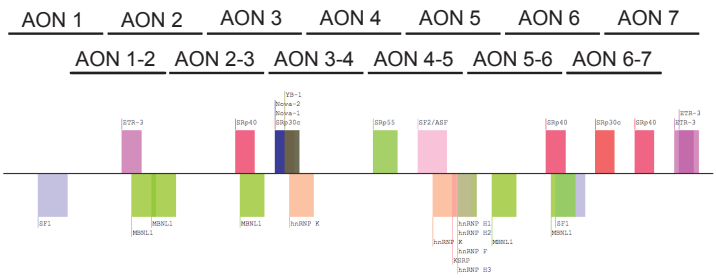
Several splicing factors have been shown to regulate NUMB exon 9 alternative splicing, including RBM10 and QKI (Zong et al., 2014) (Bechara et al., 2013; Hernandez et al., 2016). RBM10 is thought to mediate its effect through its binding to NUMB intron 8 polypyrimidine tract (Bechara et al., 2013), while QKI has two different binding sites, one in the intron and one in the 5’ most region of the exon (Zong et al., 2014).

To identify which splicing factors may be acting though the ESE1 element, we performed a bioinformatic analysis to identify potential binding motifs for splicing regulators. Our analysis revealed a potential binding site for SF1/BBP, the Branchpoint Binding Protein in the region from +8 to +16 of NUMB exon 9 (CACTGACTC). We were intrigued by the possibility that SF1/BBP performs a splicing regulatory function when bound to an exon (while it has never been reported that SF1/BBP acts through binding to exons, 33% of its binding sites are exonic (Corioni et al., 2011). Although no

CLIP-tags for SF1 were identified corresponding to our ESE1 (data not shown (Corioni et al., 2011), we generated two mutants of the RG6-NUMB minigene: Mut1, which is predicted to disrupt the SF1/BBP binding site (CACTGACTC > CACTCACCT), and Mut2, which is predicted to have a milder effect (CACTGATTC > CACCGACTC) (Figure 4A). We performed co-transfection experiments in HEK-293 cells where increasing concentrations of a SF1/BBP expression vector were co-transfected with a constant concentration of the RG6-NUMB reporter. 48h after transfection, RNA and protein samples were collected and the fraction of exon 9 inclusion (PSI) quantified by RT-PCR and capillary electrophoresis (Figure 4B, lower and right panels), while the levels of SF1/BBP overexpression were estimated by western blot using a specific antibody (Figure 4B, upper left panels). While mutant Mut1 slightly increased exon 9 skipping, Mut2 did not show any effect (Figure 4B, compare lanes 1, 5 and 9 in left lower panel and in right panel). Overexpression of SF1/BBP showed a slight tendency towards increased exon 9 inclusion for the three reporter minigenes, but the effects were not quantitatively significant (Figure 4B). We conclude that despite the presence of a putative binding site for SF1/BBP, this protein is unlikely to mediate the effects of the ESE1 element.

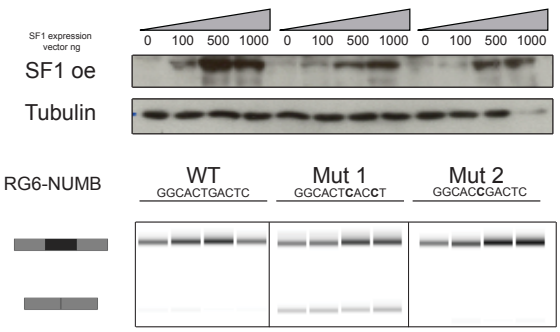
Figure 4

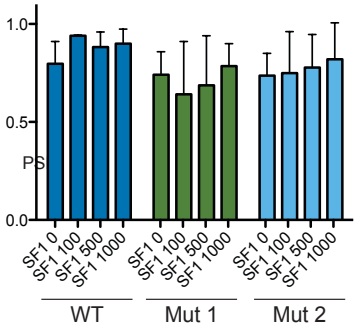
A



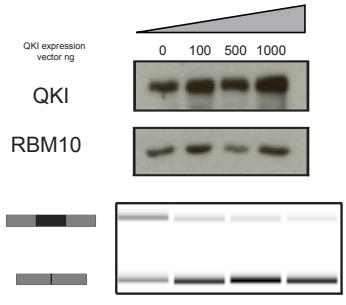
	RG6-NUMB minigene sequences	Predicted effect
WT	GGCACTGACTC	Predicted SF1 binding
Mut	GGCACTCACCT	SF1 binding abolished
Mut	GGCACCGACTC	Reduced SF1 binding

B





C



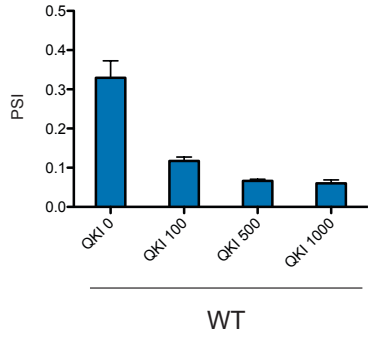


Figure 4. Analysis of putative regulators of NUMB exon 9 acting through ESE1. (A) Bioinformatic analysis of putative binding motifs for cognate trans acting splicing factors binding to different regions of NUMB exon 9 using SpliceAid2 software. The regions of the exon complementary to the different AONs utilized in the study are indicated. The location of putative binding motifs is indicated by colored blocks. The table shows the mutations introduced in the RG6-NUMB reporter aiming at reducing SF1 binding, as predicted according to (Corioni et al.2011) and SpliceAid2. (B) Effects of SF1 overexpression in HEK-293 cells co-transfected with different versions of the RG6-NUMB minigene. Upper left panel: western blot analysis of SF1 overexpression (oe) upon transfection of the indicated amounts of expression plasmid. Western blot analysis of tubulin were used as loading control. Lower left panel: RT-PCR analysis of NUMB exon 9 inclusion/skipping isoforms. Right panel: quantification of the RT-PCR results. (C) Effects of QKI overexpression in HEK-293 cells co-transfected with RG6-NUMB minigene. Analysis were carried out as in (B).

We realized that the SF1/BBP motif (ACUNAC) (Corioni et al.) is very similar to the consensus QKI binding site (ACUAAY), which has been described to regulate NUMB alternative splicing in lung cancer (Zong et al.). Therefore we decided to test the effect of QKI overexpression in our minigene reporter (Figure 4C and 4D). The results show that QKI overexpression (Figure 4C, left upper panel) led to increased levels of exon 9 skipping in the wild type reporter (Figure 4C, left lower panel and quantification in the right panel), consistent with previous results (Zong et al., 2014). We conclude that an excess of QKI causes NUMB exon 9 skipping and therefore QKI is also unlikely to be the trans-acting factor that promotes NUMB exon 9 inclusion through the ESE1 element.

Taken together, our results suggest that despite the presence of putative binding motifs within ESE1, neither SF1/BBP not

QKI are likely to be the factors that mediate the effects of this enhancer that are efficiently antagonized by AON1.

Modulation of NUMB alternative splicing in vivo

We next wanted to test if the strong splicing modulation of NUMB exon 9 by AON1 observed in cells in culture (Figure 3) could be recapitulated *in vivo*.

For that, we performed intratracheal administration of 100µl of AON1 at a concentration of 3100ng/µl (an approximate concentration of 12µg of AON1/gram of mouse weight) in wild-type BL6 adult male mice. 78h after the administration the animals were sacrificed and lung and liver samples were collected. RNA was isolated from the tissues and endogenous NUMB exon 9 PSI values analyzed by RT-PCR (Figure 5A). The administration of AON1 in the lung of healthy mice led to a significant reduction in NUMB PSI values (Figure 5A), while it did not have detectable effects in the liver (Supplementary Figure 1).

Two main conclusions can be drawn from this experiment. First, the levels of exon 9 inclusion in endogenous NUMB transcripts in the lungs of healthy mice are very low (around 3%). Despite of this, AON1 was capable, when administrated with PBS directly into the lung, to induce a significant further increase in exon skipping, suggesting that AON1 could be an

interesting reagent to test in tumor samples, which display higher levels of NUMB exon 9 inclusion.

Figure 5

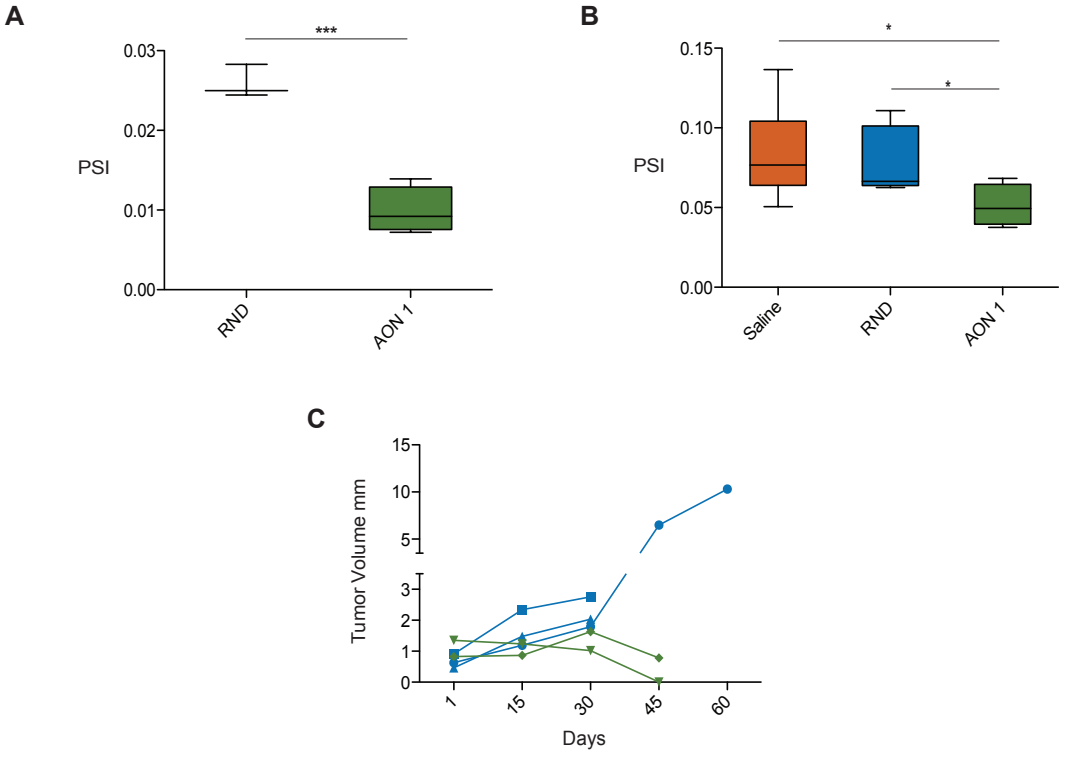


Figure 5. Effect of NUMB exon 9 alternative splicing modulation by AON1 on lung tumor progression. (A) Intratracheal administration of AON1 in BL6 young adult mice induces NUMB exon 9 skipping in healthy lung tissue. Box plot PSI values for lung samples upon administration of RND, 3 mice, (control AON) or AON1, 4 mice, are represented. Student's t-test was used for determining the statistical significance p-value<0.001 *** (B) Treatment of KRAS-G12V mouse lung tumors with a single administration of AON1 induces NUMB exon 9 skipping. PSI values from tumor samples upon administration of saline (7 tumors for saline), RND (control AON, 6 tumors) or AON1 (4 tumors) are analyzed as in (B) Student's t-test was used for determining the statistical significance p-value<0.05 *(C) Reduction on the volume of mouse KRAS-G12V tumors upon weekly administration of AON1. Tumor volume (measured by PET-SCAN) over time is indicated for mice treated with AON1 (blue) or RND (green). Each line corresponds to a single tumor.

Modulation of NUMB alternative splicing in tumors

The possibility to modulate NUMB exon 9 alternative splicing *in vivo* (Figure 5A), prompted us to test if AON1 could modulate NUMB alternative splicing in lung tumors. For this, we used a genetic mouse model of KRAS-G12V driven non-small cell lung carcinoma (Guerra et al., 2003). Young adult mice were treated with tamoxifen around 6 months of age to induce the activation of Cre, which by removing an engineered premature STOP codon before the oncogenic KRAS-G12V, leads to expression of the oncogenic version of the protein. 6 to 8 months after the administration of tamoxifen, mice started to develop lung adenomas. At that point mice were treated with a single intratracheal administration of either AON1, RND (using an approximate concentration of 12μg of AON1/gram of weight of the mice) or saline buffer and sacrificed 3 weeks after. Lung samples were extracted and tumors were micro-dissected for their posterior analysis, collecting also non-tumoral lung samples. Samples were processed as in the previous experiment with healthy mice. The results of Figure 5B show 1) that the levels of NUMB exon 9 inclusion in tumors treated with control oligos or saline buffer were more than 3 fold higher than in healthy tissue (compare PSI values of RND between Figures 5A and 5B) the PSI values of non-tumoral tissue in KRAS-G12V - activated mice were also higher and comparable to the tumors (Supplementary Figure 2), which may be explained by the existence of premalignant lesions induced by oncogenic

KRAS activation in the apparently non-tumoral tissue; and 2) the administration of AON1 was able to reduce significantly the levels of NUMB exon 9 inclusion in the treated tumors (Figure 5B, compare lanes 1-2 and 3). No significant differences between administrating saline buffer or the RND AON were observed, indicating that the control AON does not alter the ratios of the NUMB isoforms.

Taken together, the results suggest that a single intratracheal administration of AON1 is capable of significantly reducing NUMB exon 9 inclusion in KRAS-G12V derived tumors, suggesting that it could also have an impact in the growth of these tumors. The change in NUMB splicing is maintained at least during 3 weeks after the administration, suggesting the possibility of long-term effects. These results were obtained for 4 tumors, coming from 2 AON1 treated mice, 5 tumors coming from 3 mice treated with AON RND and 6 tumors coming from 2 saline treated mice. Experiments are in progress to increase the statistical significance of these observations.

The effective splicing modulation in lung tumors though the direct administration of this AON1 encouraged us to evaluate the effect of AON1 in tumor progression. We used the same lung cancer mouse model described above (KRAS-G12V) and 4-5 months after the tamoxifen administration, when the tumors were detectable by micro-CT (PET-scan), mice were

treated with either AON1 or RND. Weekly intranasal administrations of 100µl of AON (at a concentration of 3100ng/µl) were performed to each mice (regardless of their weight). Tumor growth was followed every two weeks by PET-SCAN. The results of the follow up on tumor growth are summarized in Figure 5C. Although still preliminary due to the limited number of mice per group (3 mice treated with the RND AON and 2 with the AON1), our data suggest that the treatment of lung tumors in a KRAS-V12D mouse with AON1 can have therapeutic effects for preventing tumor growth. 45 days after starting the treatment, the size of the AON1-treated tumors was drastically reduced (with total remission of the tumor in one case) while the control tumors continue to grow. Further experiments are under way to assess the statistical significance of these initial observations.

Materials and Methods

NUMB exon 9 alternative splicing in cancer samples

Our analysis of NUMB exon 9 PSI distribution in cancer samples was derived from (Sebestyén et al.). We compared NUMB exon 9 PSI across 11 different tumor types using only paired tumor/non-tumor samples. Mann Whitney U test was used to assess differences in PSI distribution between of healthy and tumor samples; correction of multiple testing was performed with Benjamini-Hochberg.

NUMB exon 9 PSI distribution across tissues

RNA sequencing data obtained by the GTEX consortium (The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans) was analyzed using the SANJUAN pipeline, obtaining the PSI value for NUMB exon 9 in each tissue. SANJUAN is a package for splicing analysis that identifies and annotates differentially-used competing junctions and differentially retained introns between pairs of conditions. SANJUAN does not use existing annotated splicing events, therefore it can identify novel splicing events absent from transcript structure databases. NUMB gene expression (FPKM) was also determined using data from the GTEx consortium.

Network-based analysis was performed for identifying potential regulators of NUMB exon 9 splicing (as in Papasaikas et al 2016). The list of potential NUMB exon 9 regulators was obtained from the genes whose gene expression had the highest or the lowest correlation with NUMB exon 9 PSI.

AONs design and NUMB exon 9 splicing modulation

We designed 2'-O-methyl phosphothioate-modified Antisense Oligonucleotides of 21 nucleotides in length and reverse-complementary to NUMB exon 9 sequence (a scheme can be seen in Figure 3A and Figure 4A), which were synthesized by Sigma-Aldrich. Two batches of AONs were ordered, one of 8mg and one of 16mg. AONs were ordered dry, resuspended in PBS and stored at -20ºC.

Cells Culture and Transfection

All cell lines used in this article where cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS, Gibco) and antibiotics (50units/ml penicillin and 50ml/ ml streptomycin).

Cell transfection, RNA extraction and processing (including reverse-transcription, PCR amplification and isoform quantification by capillary electrophoresis) were performed as

in (Hernandez et al.). PSI quantification was done either by capillar electrophoresis (Hernandez et al. 2016) or by electrophoretic separation of the PCR products in 6% acrylamide gels. In the latter case, SyberGreen was used for band detection, images were acquired with Geldoc (Biorad) and the band quantification was performed with FIJI software (Schindelin et al., 2012).

For clonogenic assays, 20.000 mice cells were cultured in 6-well plates (2000 A549 cells) and transfected using Lipofectamine RNAiMAX (Invitrogen) with 100nM of AONs in following the manufacturer's instructions. Cells were maintained in culture for 7 days, then media was removed and cells were fixed with 1ml of methanol and stained with crystal violet overnight. Afterwards, the wells were washed with water and let to dry. Colony quantification was performed using FIJI software and a homemade R script.

Prediction of splicing factor binding sites

SpliceAid2 web tool was used for predicting splicing factors binding sites across NUMB exon 9 (Piva et al., 2012) (http://193.206.120.249/splicing_tissue.html).

Genetic constructs

The vector for overexpressing SF1 was kindly provided by Angela Kramer (University of Geneva) (SF1-HL1 cloned in pEGFP-C1 (Arning et al., 1996)) and QKI overexpressing vector (QKI-5) was kindly provided by Jingyi Hu (Shanghai Institutes for Biological Sciences) (Zong et al., 2014)

Mut1 and Mut2 versions of RG6-NUMB minigenes (Bechara et al., 2013; Hernandez et al., 2016) were generated using QuickChange Lighting kit (from Agilent technologies).

Mice experiments

Wild-type BL6 male mice, older than 12 weeks were used for testing the effect of AONs in vivo in healthy lungs. AONs were resuspended at a concentration of 3100ng/µl, and a volume of 100ul was intratracheally administered to the mice (regardless of their weight). Three days after the administration, mice were sacrificed and samples were collected. Tissue was disaggregated using glass-beads (Glass-beads acid-washed, 425-600um, Sigma) and a bite-biter (Mini-Beadbeater, Biospec Products), resuspended in homogenization buffer from Maxwell 16 LEV simple RNA tissue kit (Promega). After tissue disruption, RNA was isolated from the samples using Maxwell robot following

Maxwell 16 LEV simple RNA tissue kit manufacturer's instructions.

RERTert, K-rasLSLG12V mice (Guerra et al.) Were used for testing the effect of AONs on tumors. 22 week old mice were injected intraperitoneally with tamoxifen (1mg/mice) to induce tumor formation. Once tumors could be detected by micro-CT PET-SCAN (6-8 months) a single administration of AON was performed (3100ng/µl AONs in saline buffer, 100ul/mice) intratracheally. After 3 weeks, mice were sacrificed and tissue samples were collected. Lung tumors were micro-dissected and healthy and tumor tissues were processed in parallel as explained for healthy mice tissue (see above).

For the experiments of tumor progression, the same mouse model (RERTert, K-rasLSLG12V mice) and protocol was used. Once mice displayed tumors, they were administered the same dose of AON once a week for two months via intranasal administration. Tumor growth was followed by micro-CT PET-SCAN every other week for two months.

Discussion

In this study we further document how the increased inclusion of NUMB exon 9 is a common misregulated event in different types of tumors (Figure 1). We present an exhaustive scanning of NUMB exon 9 regulatory regions using AONs (Figures 2A and 2B) and show how the specific blockage of ESE regions reduces the colony formation capacity of a variety of lung cancer cell lines from human and mouse (Figures 2E and 2F). We also show how the intratracheal administration of AON1, an antisense oligonucleotide that efficiently blocks NUMB exon 9 inclusion in vitro, is also able to promote NUMB exon 9 inclusion in healthy lung and lung adenomas of mice (Figure 4).

The results presented here may represent a first step in the regulation of NUMB alternative splicing in lung tumors for therapeutic purposes. Further experiments evaluating the direct effect of AON1 in tumor growth *in vivo* and the molecular mechanisms behind this alternative splicing switch are ongoing.

In vitro modulation of NUMB exon 9 alternative splicing: identification and characterization of ESE1 region

The levels of inclusion of NUMB exon 9 are increased in several tumor types when compared to healthy tissue (Figure 1 and (Sebestyén et al., 2015)) and it is also known that the

NUMB-PRR^{L} isoform encoded by the mRNA that includes exon 9 promotes cell proliferation (Bechara et al., 2013; Misquitta-Ali et al., 2011). The rationale and final goal of this study was to try to correct NUMB alternative splicing in cancer tissues as a way to prevent abnormal cell-proliferation. For doing so we performed an exhaustive scanning of NUMB exon 9 using AONs to identify regions in the exon that promote inclusion and therefore act as Exonic Splicing Enhancers (ESE).

The results of our scanning, monitoring effects on transcripts derived from a minigene under conditions that display high levels of exon 9 inclusion, indicated that blocking of most of the regions of the exon promote skipping to a certain extent (Figure 2A). This suggests a high density of ESEs in this exon, although not every regulatory motif may be active in endogenous NUMB transcripts of healthy or cancerous cells (Figure 2B). However, at least two of these regions had a conserved function in cancer cells, as both AON1 (blocking a putative enhancer (ESE1) at the 5' end of the exon) and AON4-5 (blocking a putative enhancer (ESE2) in the middle region of the exon) efficiently induced exon skipping in endogenous transcripts of cancer cell lines, the effects of AON1 being conserved in mice, consistent with 100% conservation of the AON1 complementary sequence between human and mouse (Figure 2C).

The blockage of ESE1 by AON1 correlated with reduced colony formation capacity of cultured lung cancer cells, from both human and mouse (Figure 2E). Our battery of mouse cells included both p53 proficient and knock-out cells; the status of p53 did not modify the effect of AON1 in either splicing modulation or colony formation capacity (although the colony formation capacity of the p53 proficient cells was not so much impaired as in the p53 knock-out cells). However, as NUMB helps preventing the degradation of p53 (Colaluca et al., 2008), its splicing modulation could have different effects in the long-term progression of tumors depending on their p53 status.

Although bioinformatic analyses suggested the possibility that the effects of ESE1 could be mediated by QKI and/or SF1/BBP, our experimental results do not support these hypotheses, and further work is needed to identify the relevant trans-acting factors that promote NUMB exon 9 inclusion.

It is also relevant to mention the effect of the AON6-7. An AON that is located very close to the 3' end of NUMB exon 9. It displays a strong regulatory effect (PSI 0.05) and none of its neighboring AONs do modulate splicing so strongly. Moreover, there is a predicted binding site for SRSF9 (also known as SRp30c) in that area (Figure 4A), and SRSF9 is one of the genes whose expression levels correlates

positively with NUMB exon 9 inclusion in healthy tissue (Figure 2B) and its levels have also been reported to correlate with NUMB exon 9 alternative splicing in cancer (Sebestyén et al., 2015). All this together suggests a potential role for SRSF9 in the modulation of NUMB exon 9 alternative splicing though its predicted binding site in the region complementary to AON6-7.

In vivo modulation of NUMB exon 9 alternative splicing

AON1 had significant effects in promoting NUMB exon 9 skipping and in the colony formation capacity of lung cancer cells (Figure 2). The next logical step was to evaluate the effect of AON1 *in vivo.* We decided to administer the AON1 (diluted in saline buffer) directly into the lung of the mice; other routes of administration where tested (intravenous) but the results were not satisfactory (data not shown).

Intratracheal administration of AON1 in wild-type BL6 mice resulted in higher levels of NUMB exon 9 exon skipping in the lung (Figure 4A). The same approach was used with the non-small cell lung cancer model KRAS-G12V, which also resulted in a reduction of NUMB exon 9 PSI (Figure 4B). A reduction of NUMB PSI could be observed 3 weeks after a single AON1 administration, illustrating the strong regulatory potential of AON1 and the persistence of their effects *in vivo*.

This is consistent with other observations using AONs with similar chemical modifications (Hua et al., 2010; Hua et al., 2011). We have ongoing experiments assessing in this KRAS-G12V mouse model treated with AON1 whether changes can be observed in tumor progression *in vivo,* by X-ray Computer Tomography Scans. Although still preliminary, our results in Figure 5C show how the weekly administration of AON1 reduces tumor progression of lung cancer in KRAS-G12V mouse model.

The use of genetically engineered mouse models has been a milestone in oncology research; this mouse model in particular has been already used for investigating new therapies for lung adenocarcinomas by the use of gamma-secretase inhibitors (Maraver et al.). However, mouse KRAS-G12V derived tumors are not a total phenocopy of human adenocarcinomas; while mice lung tumors display a 3-fold increase in NUMB exon 9 PSI (Figure 4B and Supplementary Figure 2), human LUAD samples display more than 4-fold PSI increase (Figure 1). Because of this difference in PSI increase, we are currently working with orthotopic lung tumor models, using A549 cancer cells, to test the effect of AON1 in a more “human-like” environment.

Taken together, our data suggest that *in vivo* modulation of NUMB exon 9 alternative splicing in lung tumors is possible by the use of AONs. Further experiments will determine if this

change in NUMB splicing is enough to reduce tumor growth *in vivo.* If this would be the case, it could represent a new therapeutic tool in oncology based upon the direct modulation of a single alternative splicing.

The modulation of alternative splicing events, for controlling cancer cell proliferation, has been reported before: MNK2 alternative splicing has been modulated *in vitro* by the use of AONs, reducing the oncogenic properties of the cells (Maimon et al. 2014). PKM has been modulated *in vitro* in a panel of glioblastoma cells (Wang et al., 2012); STAT3 alternative splicing has been modulated in xenografted tumors (by intratumor injection of morpholinos) reducing tumor burden (Zammarchi et al., 2011). AONs conjugated with lipid nano-particles, administrated systemically, modulate Bcl alternative splicing and control xenograft tumors growth and its metastasis (Bauman et al., 2010). However, up to our knowledge this is the first time that naked AONs have been directly administrated into mouse tumor models modifying splicing *in vivo* and compromising the tumor growth. Our results represent a break-through in the field of cancer treatment by the modulation of alternative splicing, validating a new strategy for targeting cancer proliferation: the *in vivo* modulation of misregulated alternative splicing cancer associated events in endogenous cancers by the administration of naked AONs.

Ambrogio, C., Carmona, F. J., Vidal, A., Falcone, M., Nieto, P., Romero, O. A., . . . Villanueva, A. (2014). Modeling lung cancer evolution and preclinical response by orthotopic mouse allografts. *Cancer Research, 74*(21), 5978-5988. doi:10.1158/0008-5472.CAN-14-1606

Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shirole, N. H., . . . Krainer, A. R. (2015). SRSF1-Regulated Alternative Splicing in Breast Cancer. *Molecular Cell, 60*(1), 105-117. doi:10.1016/j.molcel.2015.09.005

Arning, S., Gruter, P., Bilbe, G., & Kramer, A. (1996). Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA, 2*(8), 794-810.

Bani-Yaghoub, M., Kubu, C. J., Cowling, R., Rochira, J., Nikopoulos, G. N., Bellum, S., & Verdi, J. M. (2007). A switch in numb isoforms is a critical step in cortical development. *Developmental Dynamics, 236*(3), 696-705. doi:10.1002/dvdy.21072

Bauman, J. A., Li, S. D., Yang, A., Huang, L., & Kole, R. (2010). Anti-tumor activity of splice-switching oligonucleotides. *Nucleic Acids Research, 38*(22), 8348-8356. doi:10.1093/nar/gkq731

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E., & Valcárcel, J. (2013). RBM5, 6, and 10 Differentially Regulate NUMB Alternative Splicing to Control Cancer Cell Proliferation. *Molecular Cell, 52*(5), 720-733. doi:10.1016/j.molcel.2013.11.010

Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell, 126*(1), 37-47. doi:10.1016/j.cell.2006.06.023

Cho, S., & Dreyfuss, G. (2010). A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity. *Genes & Development, 24*(5), 438-442. doi:10.1101/gad.1884910

Colaluca, I. N., Tosoni, D., Nuciforo, P., Senic-Matuglia, F., Galimberti, V., Viale, G., . . . Di Fiore, P. P. (2008). NUMB controls p53 tumour suppressor activity. *Nature, 451*(7174), 76-80. doi:10.1038/nature06412

Collisson, E. A., Campbell, J. D., Brooks, A. N., Berger, A. H., Lee, W., Chmielecki, J., . . . Tsao, M.-S. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature, 511*(7511), 543-550. doi:10.1038/nature13385

Consortium, G. T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene

regulation in humans. *Science, 348*(6235), 648-660. doi:10.1126/science.1262110

Cooper, T. a., Wan, L., & Dreyfuss, G. (2009). RNA and disease. *Cell, 136*(4), 777-793. doi:10.1016/j.cell.2009.02.011

Corioni, M., Antih, N., Tanackovic, G., Zavolan, M., & Krämer, A. (2011). Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Research, 39*(5), 1868-1879. doi:10.1093/nar/gkq1042

Danan-Gotthold, M., Golan-Gerstl, R., Eisenberg, E., Meir, K., Karni, R., & Levanon, E. Y. (2015). Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res, 43*(10), 5130-5144. doi:10.1093/nar/gkv210

Dho, S. E., French, M. B., Stacy, A., McGlade, C. J., & Woods, S. A. (1999). CELL BIOLOGY AND METABOLISM : Characterization of Four Mammalian Numb Protein Isoforms : IDENTIFICATION OF CYTOPLASMIC AND MEMBRANE-ASSOCIATED VARIANTS OF THE PHOSPHOTYROSINE BINDING DOMAIN Characterization of Four Mammalian Numb Protein Isoforms. doi:10.1074/jbc.274.46.33097

Dvinge, H., & Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. *Genome Med, 7*(1), 45. doi:10.1186/s13073-015-0168-9

Frise, E., Knoblich, J. A., Younger-Shepherd, S., Jan, L. Y., & Jan, Y. N. (1996). The Drosophila Numb protein inhibits signaling of the Notch receptor during cell-cell interaction in sensory organ lineage. *Proc Natl Acad Sci U S A, 93*(21), 11925-11932.

Guerra, C., Mijimolle, N., Dhawahir, A., Dubus, P., Barradas, M., Serrano, M., . . . Barbacid, M. (2003). Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. *Cancer Cell, 4*(2), 111-120.

Havens, M. A., Duelli, D. M., & Hastings, M. L. (2013). Targeting RNA splicing for disease therapy. *Wiley Interdiscip Rev RNA, 4*(3), 247-266. doi:10.1002/wrna.1158

Havens, M. A., & Hastings, M. L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Research, 44*(14), 6549-6563. doi:10.1093/nar/gkw533

Hernandez, J., Bechara, E., Schlesinger, D., Delgado, J., Serrano, L., & Valcarcel, J. (2016). Tumor suppressor properties of

the splicing regulatory factor RBM10. *RNA biology, 13*(4), 466-472. doi:10.1080/15476286.2016.1144004

Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M. a., Bennett, C. F., & Krainer, A. R. (2010). Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes & Development, 24*(15), 1634-1644. doi:10.1101/gad.1941310

Hua, Y., Sahashi, K., Rigo, F., Hung, G., Horev, G., Bennett, C. F., & Krainer, A. R. (2011). Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature, 478*(7367), 123-126. doi:10.1038/nature10485

Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., & Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Structural & Molecular Biology, 14*(3), 185-193. doi:10.1038/nsmb1209

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., . . . et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell, 80*(1), 155-165.

Maimon, A., Mogilevsky, M., Shilo, A., Golan-Gerstl, R., Obiedat, A., Ben-Hur, V., . . . Karni, R. (2014). Mnk2 alternative splicing modulates the p38-MAPK pathway and impacts Ras-induced transformation. *Cell Rep, 7*(2), 501-513. doi:10.1016/j.celrep.2014.03.041

Maraver, A., Fernandez-Marcos, P. J., Herranz, D., Canamero, M., Munoz-Martin, M., Gomez-Lopez, G., . . . Serrano, M. (2012). Therapeutic effect of gamma-secretase inhibition in KrasG12V-driven non-small cell lung carcinoma by derepression of DUSP1 and inhibition of ERK. *Cancer Cell, 22*(2), 222-234. doi:10.1016/j.ccr.2012.06.014

McKie, A. B., McHale, J. C., Keen, T. J., Tarttelin, E. E., Goliath, R., van Lith-Verhoeven, J. J., . . . Inglehearn, C. F. (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Human Molecular Genetics, 10*(15), 1555-1562.

Misquitta-Ali, C. M., Cheng, E., O'Hanlon, D., Liu, N., McGlade, C. J., Tsao, M. S., & Blencowe, B. J. (2011). Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. *Molecular and Cellular Biology, 31*(1), 138-150. doi:10.1128/MCB.00709-10

Oh, J. J., Grosshans, D. R., Wong, S. G., & Slamon, D. J. (1999). Identification of differentially expressed genes associated

with HER-2/neu overexpression in human breast cancer cells. *Nucleic Acids Res, 27*(20), 4008-4017.

Oh, J. J., West, A. R., Fishbein, M. C., & Slamon, D. J. (2002). A candidate tumor suppressor gene, H37, from the human lung cancer tumor suppressor locus 3p21.3. *Cancer Res, 62*(11), 3207-3213.

Piva, F., Giulietti, M., Burini, A. B., & Principato, G. (2012). SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Human Mutation, 33*(1), 81-85. doi:10.1002/humu.21609

Quesada, V., Conde, L., Villamor, N., Ordonez, G. R., Jares, P., Bassaganyas, L., . . . Lopez-Otin, C. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics, 44*(1), 47-52. doi:10.1038/ng.1032

Quesada, V., Ramsay, A. J., & Lopez-Otin, C. (2012). Chronic lymphocytic leukemia with SF3B1 mutation. *N Engl J Med, 366*(26), 2530. doi:10.1056/NEJMc1204033

Rintala-Maki, N. D., Goard, C. A., Langdon, C. E., Wall, V. E., Traulsen, K. E., Morin, C. D., . . . Sutherland, L. C. (2007). Expression of RBM5-related factors in primary breast tissue. *Journal of Cellular Biochemistry, 100*(6), 1440-1458. doi:10.1002/jcb.21134

Roy, M., Pear, W. S., & Aster, J. C. (2007). The multifaceted role of Notch in cancer. *Current Opinion in Genetics & Development, 17*(1), 52-59. doi:10.1016/j.gde.2006.12.001

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., . . . Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods, 9*(7), 676-682. doi:10.1038/nmeth.2019

Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M. A., . . . Eyras, E. (2015). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *bioRxiv*, 023010-023010. doi:10.1101/023010

Toriya, M., Tokunaga, A., Sawamoto, K., Nakao, K., & Okano, H. (2006). Distinct functions of human numb isoforms revealed by misexpression in the neural stem cell lineage in the Drosophila larval brain. *Developmental Neuroscience, 28*(1-2), 142-155. doi:10.1159/000090760

Verdi, J. M., Bashirullah, a., Goldhawk, D. E., Kubu, C. J., Jamali, M., Meakin, S. O., & Lipshitz, H. D. (1999). Distinct human NUMB isoforms regulate differentiation vs. proliferation in the neuronal lineage. *Proceedings of the National Academy*

of Sciences of the United States of America, 96(18), 10472-10476. doi:10.1073/pnas.96.18.10472

Wang, Z., Jeon, H. Y., Rigo, F., Bennett, C. F., & Krainer, A. R. (2012). Manipulation of PK-M mutually exclusive alternative splicing by antisense oligonucleotides. *Open Biol, 2*(10), 120133. doi:10.1098/rsob.120133

Zammarchi, F., de Stanchina, E., Bournazou, E., Supakorndej, T., Martires, K., Riedel, E., . . . Cartegni, L. (2011). Antitumorigenic potential of STAT3 alternative splicing modulation. *Proc Natl Acad Sci U S A, 108*(43), 17779-17784. doi:10.1073/pnas.1108482108

Zong, F. Y., Fu, X., Wei, W. J., Luo, Y. G., Heiner, M., Cao, L. J., . . . Hui, J. (2014). The RNA-Binding Protein QKI Suppresses Cancer-Associated Aberrant Splicing. *PLoS Genetics, 10*(4). doi:10.1371/journal.pgen.1004289

General Discussion

The aim of this PhD thesis was the study of the regulation of NUMB exon 9 alternative splicing, with an emphasis on the impact that RBM10 mutations found in cancer can have on this regulation (first two manuscripts). As a follow up of our mechanistic studies, we investigated ways in which NUMB splicing can be modulated *in vivo* to reduce tumor growth in models of lung cancer (third manuscript). Our data confirm and expand the relevance of NUMB exon 9 alternative splicing for controlling cell proliferation and tumor growth, provide initial insights into the mechanisms of splicing regulation by RBM10 and suggest possible tools for therapeutic use in cancer.

The first two parts of the thesis focus on RBM10, a *bona fide* regulator of NUMB alternative splicing (Bechara et al. 2013; Hernandez et al. 2016). We have characterized several cancer-associated RBM10 mutants and how they modulate NUMB alternative splicing (first two manuscripts). Work in collaboration with the groups of Luis Serrano (CRG) and Michael Sattler (Hemlholtz Zentrum, Munich) has helped to provide insights into the structure of the RRM2 domain of RBM10 in its two natural variant forms (V354 and -354V). Taken together, our data suggest that there are no major structural differences between the two natural isoforms of the RRM2 domain (firsts two manuscripts). However for the -354V isoform, CSP experiments detect a differential perturbation in the second alpha-helix in the presence of RNA. This perturbation (not present in either the V354 or the V354E) correlates with slightly higher affinity of variant -354V for NUMB exon 9 polypyrimidine tract and may reflect a conformational restraint depending on the presence or absence of V354, which influences RNA binding at the beta-

sheet surface. On the other hand, the location of residue 354 in the outside surface of the alpha-helix, suggests that the dramatic difference in activity between V354 and V354E RBM10, without differences in RNA binding, is likely to be related to the involvement of V354 in a surface engaged either in interactions with other factors or in intramolecular interactions.

We used an overexpression system to identify RBM10 interactors for wild type RBM10 354V and oncogenic RBM10 V354E (second manuscript). The mutant V354E apparently fails to interact with PRPF19. Our preliminary results suggest that RBM10 needs to interact with PRPF19 for inhibiting NUMB exon 9 splicing, although further experiments are needed to prove this point. While RBM10 -as many other splicing regulators- has been proposed to play a role in the initial stages of spliceosome assembly, PRPF19 has classically been described as a crucial factor for later stages of the splicing reaction (Ajuh et al., 2000; Papasaikas et al. 2016; Wahl et al. 2009), although some reports argued that it can also play a role in initial stages (David et al. 2011). If PRPF19 would actually be needed for RBM10 to inhibit splicing, we envision two different mechanisms: 1) RBM10 interacts with PRPF19 (which itself interacts with the tri-snRNP complex) and brings it to the splicing reaction too early, generating an abortive complex, unable to remove NUMB intron 8. In this model, PRPF19 will join the spliceosome too early to be functional; and 2) RBM10 acts as a local squelcher of PRP19, whereby due to its affinity for NUMB exon 9 polypyrimidine tract, the local concentration of RBM10 would be high and could sequester PRPF19, In this model, U2 snRNP could displace RBM10 binding to the polypyrimidine tract but RBM10, by maintaining its interaction with PRPF19, would prevent PRP19 from joining in subsequent steps of spliceosome assembly. Future

experiments will hopefully distinguish between these models or suggest alternative mechanisms.

We also characterized the distribution of RBM10 mutants in cancer, where we detected a positive selection for early truncating mutants (second manuscript), correlating with our results of overexpression of truncating mutants (first manuscript). Our data illustrate how the regulation of NUMB exon 9 splicing by RBM10 overexpression is progressively lost as RBM10 is progressively shortened (first manuscript). In fact, in cancer samples in which RBM10 is mutated, there is a higher accumulation of non-missense mutations in the first 25% of the protein, which, according to our truncating mutant results, would play a dominant negative role on NUMB exon 9 alternative splicing, promoting exon inclusion and therefore the pro-proliferative isoform)(first manuscript). The early truncation mutants may act as driver mutations for tumor progression through their effects on the splicing of NUMB and well as other genes.

Taken together, our results show that RBM10 is a tumor-suppressor with a tendency for accumulating truncating mutations and that its function in the inhibition of splicing is more complex than simply blocking access of the splicing machinery to particular splice sites, most likely involving interactions with other factors, including PRPF19 (first and second manuscripts). In contrast with other splicing factors (such as Sxl (Valcarcel et al. 1993)) that act by competing with U2AF for binding to a polypyrimidine tract, RBM10 seems to interact with U2AF and many (if not all) U2 snRNP components but fail to recruit them to intron 8 3' splice site region in a functional form. Our co-immunoprecipitation experiments (and previous reports from others (Hegele et al. 2012)) show that RBM10 interacts with several U2 associated splicing factors, including the U2AF heterodimer, with the

PRP19 complex and some members of the tri-snRNP. Our study with mutant V354E also suggests that, in this particular mutant, some of the members of this “RBM10-mediated repressive complex” are lost and consequently RBM10 loses its inhibitory potential. One could imagine an “U2-like” complex including RBM10 that actively inhibits the splicing reaction, either by preventing the recruitment of key factors, by sequestering some other necessary proteins or by orchestrating or preventing conformational rearrangements that lead to abortive spliceosome recruitment.

The last part of the Thesis is focused on how to modulate NUMB exon 9 alternative splicing in tumors (third manuscript). We carried out a systematic screening with AONs along NUMB exon 9 to identify functional regulatory regions. Three putative ESE regions were identified. The function of at least two of these ESE was conserved in cancer cells and one of them was conserved also in mice (ESE1). We tested the effect of AON1 to block ESE1 *in vitro*, in both human and mouse lung cancer cells. AON1 efficiently promotes NUMB exon 9 skipping and decreases the colony formation capacity of the cells. Furthermore, AON1 is capable of modifying NUMB exon 9 alternative splicing in the lung of healthy mice and in lung tumors. AON1 was tested in a genetic mouse model of lung cancer (KRAS-G12V) where the direct administration into the lung of AON1 modified NUMB splicing in the tumors and reduced tumor growth. Although still preliminary, our data confirms the important potential of *in vivo* modulation of NUMB exon 9 alternative splicing in tumors (third manuscript).

The results presented here may represent first steps towards new therapeutic for lung cancer. The next steps include a) an increase in the number of mice and tumors analyzed, b) explore the effect of the AON1 on orthotopic tumor models

and c) repeat these experiments with different doses of AON1 to evaluate how much is needed for its therapeutic effects. Of course, toxicological studies and research on the side effects of this compound would be mandatory.

Another interesting possible future area of work for the project would be the treatment of other types of cancer in which NUMB/Notch signaling is important for cancer progression, e.g. those that display abnormal NUMB exon 9 inclusion. Those would include breast, liver and bladder cancer (Misquitta-Ali et al., 2011; Sebestyén et al., 2015). The administration of AON1 to the liver could be achieved by intravenous injection, while specific delivery systems for breast and cancer would need to be developed.

The rationale behind our approach can be (and has been) applied to other splicing events that control key cancer associated traits that are altered in cancer. For example CASC4 exon 9 inclusion mediates the transformation effects of SRSF1(Anczuków et al. 2012; Anczukow et al. 2015). There are different approaches for modulating alternative splicing with therapeutic goals. Several antitumour drugs have been reported to act through the modulation of alterative splicing (including bacterial fermentation products and their derivatives (Liu et al., 2013; Miller-Wideman et al., 1992; Nakajima et al., 1996; Sakai et al., 2004). However, the effects of the inhibition of the splicing by these compounds are broad and difficult to control. More specific approaches include the use of morpholinos to target specific genes (Zammarchi et al., 2011) or the use of AONs. Although the use of morpholinos in xenograft tumors has been reported (Zammarchi et al., 2011), our data show that direct delivery of AONs to the target organ is an effective way to modulate alternative splicing in tumors in genetic mouse models. The effective modulation of cancer-associated splicing events in

tumors using AONs is becoming a powerful concept for novel tumor therapies.

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J. C., Mann, M., & Lamond, A. I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO Journal, 19*(23), 6569-6581. doi:10.1093/emboj/19.23.6569

Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shirole, N. H., . . . Krainer, A. R. (2015). SRSF1-Regulated Alternative Splicing in Breast Cancer. *Molecular Cell, 60*(1), 105-117. doi:10.1016/j.molcel.2015.09.005

Anczukow, O., Rosenberg, A. Z., Akerman, M., Das, S., Zhan, L., Karni, R., . . . Krainer, A. R. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature Structural & Molecular Biology, 19*(2), 220-228. doi:10.1038/nsmb.2207

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E., & Valcárcel, J. (2013). RBM5, 6, and 10 Differentially Regulate NUMB Alternative Splicing to Control Cancer Cell Proliferation. *Molecular Cell, 52*(5), 720-733. doi:10.1016/j.molcel.2013.11.010

David, C. J., Boyne, A. R., Millhouse, S. R., & Manley, J. L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes and Development, 25*(9), 972-982. doi:10.1101/gad.2038011

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., . . . Stelzl, U. (2012). Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Molecular Cell, 45*(4), 567-580. doi:10.1016/j.molcel.2011.12.034

Hernandez, J., Bechara, E., Schlesinger, D., Delgado, J., Serrano, L., & Valcarcel, J. (2016). Tumor suppressor properties of the splicing regulatory factor RBM10. *RNA biology, 13*(4), 466-472. doi:10.1080/15476286.2016.1144004

Liu, X., Biswas, S., Tang, G. L., & Cheng, Y. Q. (2013). Isolation and characterization of spliceostatin B, a new analogue of FR901464, from Pseudomonas sp. No. 2663. *J Antibiot (Tokyo), 66*(9), 555-558. doi:10.1038/ja.2013.38

Miller-Wideman, M., Makkar, N., Tran, M., Isaac, B., Biest, N., & Stonard, R. (1992). Herboxidiene, a new herbicidal substance from Streptomyces chromofuscus A7847. Taxonomy, fermentation, isolation, physico-chemical and biological properties. *J Antibiot (Tokyo), 45*(6), 914-921.

Misquitta-Ali, C. M., Cheng, E., O'Hanlon, D., Liu, N., McGlade, C. J., Tsao, M. S., & Blencowe, B. J. (2011). Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. *Molecular and Cellular Biology, 31*(1), 138-150. doi:10.1128/MCB.00709-10

Nakajima, H., Hori, Y., Terano, H., Okuhara, M., Manda, T., Matsumoto, S., & Shimomura, K. (1996). New antitumor substances, FR901463, FR901464 and FR901465. II. Activities against experimental tumors in mice and mechanism of action. *J Antibiot (Tokyo), 49*(12), 1204-1211.

Papasaikas, P., & Valcárcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends in Biochemical Sciences, 41*(1), 33-45. doi:10.1016/j.tibs.2015.11.003

Sakai, T., Sameshima, T., Matsufuji, M., Kawamura, N., Dobashi, K., & Mizui, Y. (2004). Pladienolides, new substances from culture of Streptomyces platensis Mer-11107. I. Taxonomy, fermentation, isolation and screening. *J Antibiot (Tokyo), 57*(3), 173-179.

Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M. A., . . . Eyras, E. (2015). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *bioRxiv*, 023010-023010. doi:10.1101/023010

Valcarcel, J., Singh, R., Zamore, P. D., & Green, M. R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature, 362*(6416), 171-175. doi:10.1038/362171a0

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell, 136*(4), 701-718. doi:10.1016/j.cell.2009.02.009

Zammarchi, F., de Stanchina, E., Bournazou, E., Supakorndej, T., Martires, K., Riedel, E., . . . Cartegni, L. (2011). Antitumorigenic potential of STAT3 alternative splicing modulation. *Proc Natl Acad Sci U S A, 108*(43), 17779-17784. doi:10.1073/pnas.1108482108