SYSTEMATIC FUNCTIONAL ANALYSES OF SPLICEOSOMAL COMPONENTS REVEAL NOVEL MECHANISMS OF ALTERNATIVE SPLICING REGULATION

Juan Ramón Tejedor Vaquero

DOCTORAL THESIS UPF - 2014

THESIS DIRECTOR

Dr. Juan Valcárcel Juárez

Gene Regulation, Stem Cells and Cancer Program
Center for Genomic Regulation (CRG), Barcelona

A todos aquellos a los que he tenido la suerte

de encontrar en el camino

ABSTRACT

Alternative splicing is an essential regulatory layer of gene expression that expands the coding potential of the genome in multicellular organisms. The spliceosome -the sophisticated machinery involved in intron removal- allows versatile regulation of gene expression programs. The splicing process relies on the dynamic interplay between hundreds of components of the spliceosome, and the steps at which the complex process of the splicing reaction can be regulated remain largely unknown.

The main objective of this thesis has been to develop high-throughput approaches to systematically identify novel regulators of alternative splicing, as well as to study the mechanisms by which they modulate splice site choice. We have identified a variety of regulators of Fas/CD95 alternative splicing within and outside of the splicing machinery and provide novel insights into connections between iron homeostasis and alternative splicing regulation. Using computational networks, we carried out a systematic functional analysis of the spliceosome components and their regulatory potential. Our results reveal the extensive regulatory plasticity of core spliceosome components throughout its assembly process. They also identified links between alternative splicing and iron homeostasis, providing a mechanism by which iron modulates alternative splicing through regulation of the RNA binding properties of a Zinc knuckle domain in the SR regulatory protein SRSF7.

The results of this thesis highlight the value of high throughput technologies and network analyses to study complex molecular mechanisms, and unveils novel functional connections between the splicing machinery and other cellular processes.

RESUMEN

El procesamiento alternativo del pre-ARNm constituye uno de los pilares esenciales en la regulación de la expresión génica y expande la capacidad codificadora del genoma en organismos multicelulares. El Espliceosoma – la maquinaria encargada de la eliminación alternativa de los intrones- permite una regulación multifacética de los programas genéticos en el interior de la célula. El proceso de corte y empalme se sustenta en la interacción dinámica de cientos de componentes del Espliceosoma, y los distintos niveles de regulación de la compleja reacción de splicing permanecen aun sin descubrir.

El objetivo principal de esta tesis se ha centrado en el desarrollo de tecnologías sistematicas de alto cribado para identificar reguladores potenciales del procesamiento alternativo del pre-ARNm, así como los mecanismos implicados en su regulación. Hemos identificado una gran variedad de reguladores del procesamiento alternativo de Fas/CD95, tanto componentes esenciales del espliceosoma como factores implicados en otros procesos biologicos, y hemos observado una conexión inédita entre la regulación del splicing alternativo y el proceso de homeostasis modulado por hierro. Mediante el uso de redes computacionales, hemos llevado a cabo un análisis sistemático y funcional de los componentes del Espliceosoma y hemos identificado el potencial regulador de los mismos en la reacción de corte y empalme. Nuestros resultados reflejan una inmensa plasticidad de los factores esenciales del Espliceosoma a lo largo de toda a reacción de ensamblaje. Además, hemos conseguido

identificar el mecanismo potencial por el cual la homeostasis del hierro ejerce su función en splicing alternativo a través de la modulación de la actividad de unión a RNA -mediada por un dominio de unión a cinc- en la proteína reguladora de splicing SRSF7.

Los resultados de esta tesis enfatizan la relevancia de las tecnologías emergentes de alto cribado y el análisis de redes computacionales en el estudio de complejos mecanismos moleculares, y desvelan nuevas conexiones funcionales entre la maquinaria de splicing y otros procesos celulares.

PREFACE
THE RNA ODYSSEY

Every scientific revolution has been linked with remarkable technological advances, and conversely, major technical achievements have been led by novel concepts or ideas. Every turn in this close relationship between science and technology have allowed a better understanding of the universe around us. And if we have seen further than other men it's because we have stood on the shoulders of giants.

The XX century has been a prolific period for many scientific disciplines, and particularly, the starting point of the emerging molecular biology field. Major advancements in the early 20's and 30's in the fields of X-ray diffraction, electron microscopy, electrophoresis and ultracentrifugation techniques allowed scientist to characterize in detail the world of the colloids, chemical compounds whose structure and function were not well understood. The most fundamental level of life was composed by four groups of macromolecules: carbohydrates, lipids, nucleic acids (DNA and RNA) and proteins, being the interplay between the latter ones the main focus for the studies carried over the following years.

The first evidence that demonstrated the existence of a precise relationship between genes and metabolic enzymes was provided in 1941 by the work of George Beadle and Edward Tatum in Neurospora crassa (Beadle *et al.*, 1941), which led them to postulate the one gene, one enzyme hypothesis. In 1944, the Avery–MacLeod–McCarty experiment (Avery *et al.*, 1944) suggested that DNA, rather than proteins, may be the hereditably material of bacteria, and further work carried out by Alfred Hershey

and Martha Chase in bacteriophages (Hershey *et al.*, 1952) confirmed that result, demonstrating that DNA was, in fact, the genetic material present in our genes. A milestone work by Linus Pauling in 1949 (Pauling *et al.*, 1949) revealed the first proof of a human disease caused by an abnormal protein, and demonstrated that Mendelian inheritance determined the specific physical properties of the polypeptides. All those evidences set up the dawn of the molecular genetics but it wasn't until the early 50's when truly modern molecular biology emerged to scene.

In 1950, Erwin Chargaff published the Chargaff rules, stating that from any given specie or cell, the DNA content of pyrimidine and purine bases shows an equimolar ratio distribution (Vischer *et al.*, 1948; Chargaff, 1950). In 1952, Rosalind Franklin and Maurice Wilkins performed X-ray crystallography studies of DNA, providing crucial information about its structure (Franklin *et al.*, 1953). And with the previous background, James Watson and Francis Crick proposed in 1953 the helical, double-stranded and antiparallel model for the DNA molecule (Watson *et al.*, 1953). Implicit also in this model, the base pairing structure that they postulated led them to hypothesize about possible copying mechanisms of the genetic material. Further studies focused the attention of the enzymatic activity responsible for the perpetuation of the DNA molecule, and in 1956, Kornberg et al (Lehman *et al.*, 1958) isolated the first known DNA polymerase (DNA pol I) able to replicate DNA. On the other hand, Meselson and Stahl performed one of the most elegant experiments in molecular biology (Meselson *et al.*, 1958) to demonstrate the semiconservative replication previously suggested by Watson and Crick, in which each DNA strand could act as a

template for the synthesis of a new strand, setting up the basis for the mechanism of genetic transmission.

On the RNA side, evidences from the work of Caspersson (Caspersson 1950) and Brachet (Brachet, 1954, 1955) in the early fifties, shed some light into the role of RNA in protein synthesis. At that time, it was shown by Goldstein and Plaut that RNA was synthetized in the nucleus (Goldstein *et al.*, 1955) and that it could contain some protein coding information, acting in the same way as the inheritable material in viruses (Gierer *et al.*, 1956, Fraenkel *et al.*, 1955). Even though its function remained a mystery for the next few years, its cellular distribution was studied in detail. RNA was located in different compartments: partly in the nucleus (Goldstein *et al.*, 1955), partly in some cytoplasmic particles called microsomes (Palade, 1955), and partly as a soluble cytoplasmic fraction (Zamecnik *et al.*, 1958).

In an enormous effort to compile these experimental evidences into novel concepts, Francis Crick postulated in the late fifties the central dogma of molecular biology, in which a unidirectional flow of information from nucleic acids to proteins was hypothetically irreversible. DNA -the macromolecule containing all the information of our genomes- has to be transcribed into RNA -the passenger molecule- which, in turn, can be translated into proteins, “and once the information have passed into protein, it cannot get out again”. (Crick, 1958). Although the primordial central dogma resulted to be incomplete, and provided the RNA molecule with a merely carrier function in the world of the *“microsomal particles*”, it was a great stimulus to imagination in order to fill the gap between the theory and the experiments. Crick himself recognized the lack of proof for

some of the hypothesis, and predicted a fruitful expansion of the knowledge in the molecular biology field within the following years. But probably no one in the 50's would predict that the most underestimated biopolymer would have such as many variable roles nowadays.

The same year that Crick announced the central dogma, some evidences related to new RNA functions arose in the field. In 1958, Hoagland *et al.* discovered the RNA molecule from the soluble fraction involved in protein synthesis, and they termed it transfer RNA (Hoagland *et al.*, 1958). On the other hand, Richard B. Roberts proposed the term ribosome to refer to the previously identified microsomal particles. The RNA particle forming the ribosomes was called rRNA, and together with the tRNA, constitute the machinery for protein synthesis (Warner *et al.* 1963)

In 1961, Jacob and Monod led a seminal work regarding the control of the genetic material (Jacob *et al.*, 1961). The discovery of the E.coli Lac operon provided the first example of a transcriptional regulation system involving genes and proteins. They also suggested the presence of an intermediate RNA molecule that linked the information encoded in DNA and proteins, which they called messenger RNA (mRNA).

Cracking the genetic code focused much of the attention of the scientist at that time. With all the players on the field (DNA, and different kinds of RNA), was only a matter of time to understand how the information from DNA was transmitted into proteins. Some technical advances accelerated this process. On one hand, the discovery of the RNA polymerase by Hurwitz in 1960, and posterior isolation (Furth *et al.*, 1962) allowed the generation of synthetic

RNA molecules that could act as artificial mRNAs. On the other, the development of novel in vitro translation systems facilitated the translation of those artificial RNAs into proteins.

In 1961, Brenner et al (Crick *et al.*, 1961) demonstrated that codons consist on 3 DNA bases by introducing proflaving-inducing mutations in the rIIB gene from T4 bacteriophage. Also in that year, Nirenberg and Matthaei were the first to elucidate the nature of the codon (Nirenberg *et al.*, 1961). By translating a poly uracil RNA sequence in cell free extracts they obtained a polypeptide containing only the amino acid phenylalanine. Further work from Severo Ochoa (Lengyel *et al.*, 1961), Nirenberg (Nirenberg *et al.*, 1965) and Khorana's laboratories (Nishimura *et al.*, 1964) elucidated the rest of the genetic code. The discovery of the tRNA structure by Robert W. Holley in 1965 (Holley *et al.*, 1965) put the cherry on the pie on those meticulous studies. The genetic code was finally cracked.

Despite the substantial contribution towards the mechanisms of translation and the flow of information from our genomes to our proteomes, there were still many things to discover in the RNA field. After the tranquility generated with the revelation of the genetic code, the pillars of the central dogma of molecular biology were endangered with the discovery of reverse transcriptase in 1970 (Temin *et al.*, 1970; Baltimore, 1970). This enzyme was able to generate complementary DNA from an RNA template, and thus, transfer the genetic information encoded in the RNA to the DNA counterpart. The same year of the breakthrough that paved the way for the modern molecular biology era, Crick recognized that novel forms of genetic flow could arise (Crick, 1970). Rather than a

strict rule, the central dogma was now a flexible and evolvable entity, as evolvable as the novel and emerging RNA processing field.

The first evidence for the processing of protein coding RNAs was found in 1971 with the discovery of the mRNA Polyadenylation mechanism (Edmonds *et al.*, 1971). Stretches of adenosines were added at the 3' end of mRNAs, coating them with information that was absent in the DNA molecule and protecting them against the action of nucleases. In 1975, a second RNA processing mechanism arose. The addition of a methylated guanosine residue linked to the 5' end of mRNA was reported as a characteristic feature of mRNAs that function in eukaryotic cells (Wei *et al.*, 1975; Furuichi *et al.*, 1975). The enzymatic activity required for the addition and maintenance of the universal CCA sequence on the 3' end of tRNA molecules was also isolated in this period (Carre *et al.*, 1970). But the most striking discovery regarding RNA processing mechanisms came with the observation that mRNAs were often smaller than the DNA sequences that encode them. In 1977, a series of publications by the groups of Richard Roberts and Phillip Sharp (Chow *et al.*, 1977; Berget *et al.*, 1977; Berk *et al.*, 1988) demonstrated that the information coded in our genes was discontinuous, and sequences retained in the mRNA molecule (exons) were isolated between non coding RNA sequences (introns) that were removed from the premature RNA transcripts. This unexpected breakthrough represented the birthmark of the RNA splicing era.

Whereas a novel kind of small RNA molecules comprising between 150 and 200 nucleotides were accidentally identified by Weinberg

and Penman since 1968 (Weinberg *et al.*, 1968), its real function remained a mystery until the decade of the 80's. In 1980 Lerner and Steitz identified the molecular identity of the factors recognized by antibodies present in the serum of Lupus erythematous patients (Lerner *et al.*, 1980). A ribonucleoprotein (RNP) particle composed of Sm proteins and several small RNAs was immunoprecipitated with the patients autoimmune sera, and detailed examination of the pull down components led them to characterize 5 classes of small RNAs termed U1, U2, U4, U5 and U6 due to their high content in uridine residues. Further studies demonstrated a direct involvement of those highly conserved snRNPs in the splicing of heterogeneous nuclear RNAs (hnRNA), suggesting also a mechanism of exon recognition between U1 snRNP and the splice junctions of those pre mRNA molecules (Lerner *et al.*, 1980; Padgett *et al.*, 1983). And the discovery of the first alternative splicing events in the human genome (Early *et al.*, 1980; Rosenfeld *et al.*, 1981) brought up a novel and striking concept: multiple proteins can arise from the information encoded in one single gene.

But while most of the efforts were focused now in the characterization of the mechanism involved in the complex splicing reaction, another stunning discovery thrilled the RNA biology field. At the time in which was thought that enzymes were the only known biological catalysts, Thomas Cech and Sidney Altman identified the first evidence for a biochemical catalytic reaction driven by an RNA molecule: the ribozyme. While Tom Cech observed that the intron from Tetrahymena thermophila could be spliced out in the absence of any added cell extract (Zaug *et al.*, 1980), Sidney Altman (Kole *et al.*, 1980) found that the essential component from the RNase-P particle responsible for the

maturation of tRNAs was in fact an RNA molecule itself. These experiments represented a big step in RNA biology, as conferred the RNA molecule with an active role in cellular processes.

These new concepts contributed to the expansion of the RNA world hypothesis, previously proposed by Carl Woese (Woese, 1967), Francis Crick (Crick, 1968) and Leslie Orgel (Orgel, 1968) in the sixties and finally coined by Walter Gilbert in 1986 (Gilbert, 1986). In this scenario, RNA may have played a critical role in prebiotic evolution at the time before the molecules with more specialized functions came to dominate the coding of biological information and catalysis. Although this hypothesis resulted impossible to reproduce by direct empirical conditions, further experimental work in the following years demonstrated the role of the RNA molecule as the central player of many biochemical reactions, suggesting the idea of a critical role of RNA in the mechanisms of life.

Pioneer studies carried by Carol Greider and Elizabeth Blackburn in 1985 demonstrated the role of Telomerase, a ribonucleoprotein particle involved in the maintenance of telomere length (Greider *et al.*, 1985). The discovery of the telomerase gave an answer to the problem of a compensatory mechanism for telomere shortening, previously proposed by Alexey Olovnikov (Olovnikov, 1973), and the involvement of telomerase RNA component in cell aging and cancer was demonstrated some years later (Feng *et al.*, 1995),

On the other hand, a novel mechanism for information flow arose in 1989 with the discovery of the RNA editing process, suggesting that cells may stray again from central dogma (Covello *et al.*, 1989; Gualberto *et al.*, 1989). RNA molecules can undergo nucleobase modifications in their specific nucleotide sequence after being

transcribed from RNA polymerase II, and thus, potentially affect the amino acid sequence of the encoding protein, differing from the one predicted by the genomic DNA sequence.

Further technological advances and the discovery of novel experimental techniques – as the discovery of the polymerase chain reaction method by Kary Mullis (Saiki *et al.*, 1985) - facilitated the work of the modern molecular biologist in the 90's decade. And the development of an original combinatorial chemistry technique called SELEX (acronym of Systematic Evolution of Ligands by Exponential Enrichment) allowed the identification of novel RNA-Protein interactions by in vitro selection of RNAs (or also DNA molecules) that bind specific proteins or target ligands (Tuerk *et al.*, 1990; Ellington *et al.*, 1990). The method proved to be of extremely relevance for the identification of RNA or DNA binding motifs by its protein interacting factors, and has generated much of the information regarding motif analysis that scientists use nowadays.

Moreover, the discovery of the RNA interference mechanism by Andrew Fire and Craig Mello shed some light into the complex regulatory networks that govern gene expression programs in living cells (Fire *et al.*, 1998; Hamilton *et al.*, 1999). Up to now, gene silencing methods induced by microRNAs or small interfering RNA are one of the most powerful techniques to modulate and study the functional roles of protein coding genes in different organisms, and constituted a basic pillar in the development of novel high throughput and genome wide studies.

Although the most astonishing result regarding the functional role of RNA in biochemical reactions was the fact that the ribosome -the cellular factory that manufacture proteins- was a truly ribozyme.

Previous observations suggested that the catalytic activity required for peptide bond formation should reside in the rRNA fraction, as extensively deproteinized ribosomal subunits could still catalyze the translation reaction (Noller *et al.*, 1992). Further structural work carried out by the groups of Scott Strobel and Thomas Steitz demonstrated the crucial role of the rRNA component at the catalytic center of the ribosome (Ban *et al.*, 2000; Muth *et al.*, 2000; Nissen *et al.*, 2000), where proteins were the structural blocks that support and stabilize it. In fact, the rRNA sequences that conform the ribosomal active site share an extremely high degree of conservation across different species or even biological kingdoms, suggesting that RNA catalysis of biochemical reactions was a feature of the last common ancestor since the rise of life on earth.

The new millennium ushered a wave of change in the way that scientists conceived molecular biology and the interpretation of biological data, setting the basis for the modern bioinformatics era. The release of the first draft of the human genome in 2001 constituted one of the largest collaborative project worldwide (International Human Genome Sequencing Consortium 2001), and the further announcement of the full draft of the human genome in 2004 (International Human Genome Sequencing Consortium, 2004) expanded enormously the knowledge about the information encoded in our genes. Surprisingly, protein coding genes represented only a little fraction of the information present in our genomes, and a waste land of unexplored DNA territories – noted as junk DNA- coped much of the attention over the following years.

Additional advances in mass spectrometry, X-ray crystallography and nuclear magnetic resonance technologies allowed a precise

characterization of the proteins and factors involved in intricate biochemical reactions. In this period, a plethora of studies focused on the comprehensive analysis of the structure and complexity of the ribosome (Wimberly *et al.*, 2000; Schluenzen *et al.*, 2000; Yusupov *et al.*, 2001) and the spliceosome (Rappsilber *et al.*, 2002; Zhou *et al.*, 2002; Makarov *et al.*, 2002) helped to elucidate the mechanism of action of the two most complex ribonucleoprotein machineries within the cell.

Technical improvements of the microarray platforms developed during the 90's (Schena *et al.*, 1995), together with the emergence of next generation sequencing techniques (Schendure *et al.*, 2008), allowed a detailed characterization of transcriptomes and global gene expression patterns of cells, tissues, and organisms, as well as in vivo genome wide profiling of DNA or RNA binding proteins by ChIP on ChIP (Ren *et al.*, 2000), ChIP-seq (Johnson *et al.*, 2007) or CLIP-seq (Licatalosi *et al.*, 2008) methods respectively. But many other studies with direct implications in human health, as the search of novel genetic variants in human populations (International HapMap Consortium, 2007; 1000 Genomes Project Consortium, 2012), de novo sequencing of cancer genomes (International Cancer Genome Consortium, 2010) or the genetic basis of rare disorders, took also advantage of these novel procedures.

And the combination of all those previous technological advances ended up with the publication of the final draft of the Encyclopedia of DNA Elements (ENCODE) in an enormous collaborative effort to determine the role of all functional elements of the human genome (ENCODE Project Consortium, 2012). The ENCODE public

research project - intended as a follow-up to the Human Genome Project – aimed to identify regulatory elements such as gene promoters, transcriptional regulatory sequences, histone modifications, chromatin structures, in vivo RNA binding proteins and transcription factor binding sites and all possible landscapes of transcription in a genome wide manner, in multiple cell lines. As major conclusions, much of the functional non-coding DNA is transcribed and probably involved in the regulation of the expression of coding genes, and those genes contain multiple regulatory sites located at near and distant positions, suggesting that gene regulation is far more complex than was previously believed.

The aforementioned, as well as many other studies, have helped to understand the central role of RNA in living organisms. To date, more than 20 classes of RNA molecules are currently described, including coding and non-coding RNA examples with several biochemical functions. Messenger RNA (**mRNA**), which convey genetic information from DNA to the ribosome. Ribosomal RNA (**rRNA**), the structural and functional component of the ribosome. Transfer RNA (**tRNA**), which serves as the physical link between the nucleotide sequence of nucleic acids and the amino acid sequence of proteins. Small nucleolar RNA (**snoRNA**) implicated in chemical modifications of other rRNAs, tRNAs and snRNAs. Small nuclear RNA (**snRNA**) involved in the processing of pre-mRNAs. Small cajal body specific RNA (**scaRNA**), which guide the modification of spliceosomal RNAs U1, U2, U4, U5 and U12. Guide RNA (**gRNA**), involved in RNA editing. **Ribonuclease P**, a ribozyme that cleaves off a precursor sequence of RNA on tRNA molecules. **Ribonuclease MRP**, involved in precursor rRNA

processing. **Y RNA** necessary for DNA replication through interactions with chromatin and initiation proteins. **Telomerase RNA**, required for telomere extension. Antisense RNA (**aRNA**), complementary to messenger RNAs. **CRISPR RNA**, which guide Cas proteins to silence exogenous genetic material. Long non-coding RNA (**lncRNA**), with several functions in gene regulation. **Micro RNA** and small interfering RNA (**siRNAs**), involved in transcriptional and post-transcriptional regulation of gene expression. Piwi-interacting RNA (**piRNA**), related to epigenetic and post-transcriptional gene silencing of retrotransposons and other genetic elements in germ line cells. **Trans-actin siRNA**, that represses gene expression through post-transcriptional gene silencing. Repeat associated siRNA (**RasiRNA**) involved in the RNA interference (RNAi) pathway. **7SK RNA**, which regulates transcription by controlling the positive transcription elongation factor P-TEFb. Circular RNA (**circRNA**) formed by covalently closed continuous loop, and potential gene regulators. **Retrotransposons,** which are transposable elements via RNA intermediates. Vault RNA **(vRNA)**, linked to drug resistance. And many more examples for which its function remains to be elucidated.

Since the discovery of the nucleic acids by Friedrich Miescher in 1868 (Miescher, 1871; Dahm, 2010), we have been honored to witness the exponential growth of the molecular biology field. Every single step has settled the foundations for the next generations of scientist, and those in turn, will pave the way for the next ones, in an endless cycle of knowledge acquisition. The scientists of the future will know that, contrary to the one gene – one enzyme theory, many proteins can be generated from a single gene. They

will learn about the relevance of RNA and the fact that proteins are not the only catalyzers of biochemical reactions within the cell. And they will probably rebel against central dogmas established in our era. However, even with the privilege of knowledge, science will not be a bed of roses for them. The increasing amounts of biological information generated nowadays will require profound changes in our actual analytical methods, and data interpretation will imply interdisciplinary collaboration between experts from different fields. But hopefully, the cocktail resulting from the combination of innovative experimental and bioinformatics approaches will facilitate the discovery of novel mechanisms involved in the complex reaction of life. It seems that far from over, the RNA Odyssey has just begun…

"Those of us among the thousands of us who did not discover the structure of DNA, the universal genetic code, or the first proteins that can and do control genes can still justifiably enjoy our achievements and be secure in the knowledge that we took part in a great and continuing enterprise." *JAMES DARNELL*

References

1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature. *491*(7422):56-65.

Avery OT, Macleod CM, McCarty M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of a transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med. *79*(2):137-58.

Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. (1970). Nature. *226*(5252):1209-11.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science. *289*(5481):905-20.

Beadle GW, Tatum EL. (1941). Genetic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci U S A. *27*(11):499-506.

Berget SM, Moore C, Sharp PA. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A. *74*(8):3171-5.

Berk AJ, Sharp PA. (1978). Structure of the adenovirus 2 early mRNAs. Cell. *14*(3):695-711.

Brachet J. (1954). Effects of ribonuclease on the metabolism of living root-tip cells. Nature. *174*(4436):876-7.

Brachet J. (1955). Action of ribonuclease and ribonucleic acid on living amoebae. Nature. *175*(4463):851-3.

Carre DS, Litvak S, Chapeville F. (1970). Purification and properties of Escherichia coli CTP (ATP)-tRNA nucleotidyltransferase. Biochim Biophys Acta. *224(*2):371-81.

Caspersson TO. (1950). Cell Growth and Cell Function: A Cytochemical Study. W. W. Norton & Company. 185 pp

Chargaff E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia. *6*(6):201-9.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell. *12*(1):1-8.

Chow LT, Roberts JM, Lewis JB, Broker TR. (1977). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell. *11*(4):819-36.

Covello PS, Gray MW. (1989). RNA editing in plant mitochondria. Nature. *341*(6243):662-6.

Crick F. (1970). Central dogma of molecular biology. Nature. *227*(5258):561-3.

Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. (1961) General nature of the genetic code for proteins. Nature. *192*:1227-32.

Crick FH. (1958). On protein synthesis. Symp Soc Exp Biol. *12*:138-63.

Crick FH. (1968). The origin of the genetic code. J Mol Biol. *38*(3):367-79.

Dahm R. (2005). Friedrich Miescher and the discovery of DNA. Dev Biol. *278*(2):274-88.

Early P, Rogers J, Davis M, Calame K, Bond M, Wall R, Hood L. (1980). Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. Cell. *20*(2):313-9.

Edmonds M, Vaughan MH, Jr., Nakazato H. (1971). Polyadenylic Acid Sequences in the Heterogeneous Nuclear RNA and Rapidly-Labeled Polyribosomal RNA of HeLa Cells: Possible Evidence for a Precursor Relationship. Proc Natl Acad Sci U S A. *68*(6): 1336–1340.

Ellington AD, Szostak JW. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature. *346*(6287):818-22.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature. *489*(7414):57-74.

F. Miescher. (1871). Ueber die chemische Zusammensetzung der Eiterzellen. Med.-Chem. Unters. 4

Feng J, Funk WD, Wang SS, Weinrich SL, Avilion AA, Chiu CP, Adams RR, Chang E, Allsopp RC, Yu J, *et al.* (1995). The RNA component of human telomerase. Science. *269*(5228):1236-41.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. *391*(6669):806-11.

Fraenkel-Conrat H, Williams RC. (1955). Reconstitution of active tobacco mosaic virus form its inactive protein and nucleic acid components. Proc Natl Acad Sci U S A. *41*(10):690-8.

Franklin RE, Gosling RG. (1953). Molecular configuration in sodium thymonucleate. Nature. *171*(4356):740-1.

Furth JJ, Hurwitz J, Anders M. (1962). The role of deoxyribonucleic acid in ribonucleic acid synthesis. I. The purification and properties of ribonucleic acid polymerase. J Biol Chem. *237*:2611-9.

Furuichi Y, Muthukrishnan S, Shatkin AJ. (1975). 5'-Terminal m-7G(5')ppp(5')G-m-p in vivo: identification in reovirus genome RNA. Proc Natl Acad Sci U S A. *72*(2):742-5.

Gierer A, Schramm G. (1956). Infectivity of ribonucleic acid from tobacco mosaic virus. Nature. *177*(4511):702-3.

Gilbert W. (1986). Origin of life: The RNA world. Nature *319*, 618.

Goldstein L, Plaut W. (1955) Direct evidence for nuclear synthesis of cytoplasmic ribose nucleic. Proc Natl Acad Sci U S A. *41*(11):874-80.

Greider CW, Blackburn EH. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell. *43*(2 Pt 1):405-13.

Gualberto JM, Lamattina L, Bonnard G, Weil JH, Grienenberger JM. (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. Nature. *341*(6243):660-2.

Hamilton AJ, Baulcombe DC. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science. *286*(5441):950-2.

Hershey AD, Chase M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol. *36*(1):39-56.

Hoagland MB, Stephenson ML, Scott JF, Hecht LI, Zamecnik PC. (1958). A soluble ribonucleic acid intermediate in protein synthesis. J Biol Chem. *231(*1):241-57.

Holley RW, Apgar J, Eeverett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A. (1965). Structure of a ribonucleic acid. Science. *147*(3664):1462-5.

International Cancer Genome Consortium. (2010). International network of cancer genome projects. Nature. *464*(7291):993-8.

International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature. *449*(7164):851-61.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. Nature. *409*(6822):860-921.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. Nature. *431*(7011):931-45.

Jacob F, Monod J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 318-56.

Johnson DS, Mortazavi A, Myers RM, Wold B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science. *316*(5830):1497-502.

Kole R, Baer MF, Stark BC, Altman S. (1980). E. coli RNAase P has a required RNA component. Cell. *19*(4):881-7.

Lehman IR, Bessman MJ, Simms ES, Kornberg A. (1958). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. J Biol Chem. *233*(1):163-70.

Lengyel P, Speyer JF, Ochoa S. (1961). Synthetic polynucleotides and the amino acid code. Proc Natl Acad Sci U S A. *47*(12): 1936–1942.

Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. (1980). Are snRNPs involved in splicing?. Nature. *283*(5743):220-4.

Lerner MR, Steitz JA. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. Proc Natl Acad Sci U S A. *76*(11):5495-9.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. *456*(7221):464-9.

M Nirenberg, P Leder, M Bernfield, R Brimacombe, J Trupin, F Rottman, C O'Neal. (1965). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. Proc Natl Acad Sci U S A. *53*(5): 1161–1168.

Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M, Lührmann R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science. *298*(5601):2205-8.

Matthaei JH, Nirenberg MW. (1961). Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts. Proc Natl Acad Sci U S A. *47*:1580-8.

Meselson M, Stahl FW. (1958). The replication of DNA in ESCHERICHIA COLI. Proc Natl Acad Sci U S A. *44*(7):671-82.

Muth GW, Ortoleva-Donnelly L, Strobel SA. (2000). A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. Science. *289*(5481):947-50.

Nirenberg MW, Matthaei JH. (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*:1588-602.

Nishimura S, Jacob TM, Khorana HG (1964). Synthetic deoxyribopolynucleotides as templates for ribonucleic acid polymerase: the formation and characterization of a ribopolynucleotide with a repeating trinucleotide sequence. Proc Natl Acad Sci U S A. *52*(6): 1494–1501.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science. *289*(5481):920-30.

Noller HF, Hoffarth V, Zimniak L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. Science. *256*(5062):1416-9.

Olovnikov AM. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. J Theor Biol. *41*(1):181-90.

Orgel LE. (1968). Evolution of the genetic apparatus. J Mol Biol. *38*(3):381-93.

Padgett RA, Mount SM, Steitz JA, Sharp PA. (1983). Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. Cell. *35*(1):101-7.

Palade GE. (1955). A small particulate component of the cytoplasm. J Biophys Biochem Cytol. *1*(1):59-68.

Pauling L, Itano HA, Singer SJ., Wells IC (1949). Sickle cell anemia a molecular disease. Science. 1949. *110*(2865):543-8.

Rappsilber J, Ryder U, Lamond AI, Mann M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res. *12*(8):1231-45.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. (2000). Genome-wide location and function of DNA binding proteins. Science. *290*(5500):2306-9.

Rosenfeld MG, Amara SG, Roos BA, Ong ES, Evans RM. (1981). Altered expression of the calcitonin gene associated with RNA polymorphism. Nature. *290*(5801):63-5.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. *230*(4732):1350-4.

Schena M, Shalon D, Davis RW, Brown PO. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. *270*(5235):467-70.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell. *102*(5):615-23.

Shendure J, Ji H. (2008). Next-generation DNA sequencing. Nat Biotechnol. *26*(10):1135-45.

Temin HM, Mizutani S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature. *226*(5252):1211-3.

Tuerk C, Gold L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. *249*(4968):505-10.

Vischer E, Chargaff E. (1948). The separation and quantitative estimation of purines and pyrimidines in minute amounts. J Biol Chem. *176*(2):703-14.

Warner JR, Knopf PM, Rich A. (1963). A multiple ribosomal structure in protein synthesis. Proc Natl Acad Sci U S A. *49*:122-9.

Watson JD, Crick FH. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. *171*(4356):737-8.

Wei CM, Gershowitz A, Moss B. (1975). Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. Cell. *4*(4):379-86.

Weinberg RA, Penman S. (1968). Small molecular weight monodisperse nuclear RNA. J Mol Biol. *38*(3):289-304.

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. (2000). Structure of the 30S ribosomal subunit. Nature. *407*(6802):327-39.

Woese CR. (1967). The genetic code: the molecular basis for genetic expression. p. 186. Harper & Row

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. (2001). Crystal structure of the ribosome at 5.5 A resolution. Science. *292*(5518):883-96.

Zamecnik PC, Stephenson ML, Hecht LI. (1958). Intermediate Reactions in Amino Acid Incorporation. Proc Natl Acad Sci U S A. *44*(2):73-8.

Zaug AJ, Cech TR. (1980). In vitro splicing of the ribosomal RNA precursor in nuclei of Tetrahymena. Cell. *19*(2):331-8.

Zhou Z, Licklider LJ, Gygi SP, Reed R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature. *419*(6903):182-5.

KEY WORDS

Alternative splicing

Anti-tumor drugs

Fas/CD95

Iron

Pre-mRNA splicing

Spliceosome

Regulatory networks

RNA binding

SR proteins

Zinc Finger

ABBREVIATIONS

3' - Three prime

5' - Five prime

A - Adenosine

Ab - Antibody

ARN - Ácido ribonucleico

aRNA - Antisense RNA

AS - Alternative splicing

ASE - Alternative splicing events

ATCC - American type culture collection

ATP - Adenosine triphosphate

B act - Spliceosome B active complex

bp - Base pairs

BP - Branch point

BSA - Bovine Serum Albumin

C - Cytidine

C-terminal - Carboxi-terminal

cDNA - mRNA complementary DNA

CE - Cassette exon

CF - Cystic fibrosis

ChIP - Chromatin immunoprecipitation

circRNA - Circular RNA

CLIP - Cross linking IP

Clk - Cdc2-like kinase

CMV - cytomegalovirus

CNRS - Centre national de la recherche scientifique

CoA - Coenzyme A

CRISPR - Clustered regularly interespaced short palindromic repeats

CSHL - Cold Spring Harbor Laboratory

CTD - Carboxi-terminal domain

DMEM - Dulbecco's modified Eagle's medium

DNA - Deoxyribonucleic acid

dNTPs - Deoxynucleotides

DTT - Dithiothreitol

EDTA - Ethylenediamine tetraacetic acid

EJC - Exon junction complex

EMBL - European Molecular Biology Laboratory

EMSA - Electrophoretic mobility shift assay

ENCODE - Encyclopedia of DNA elements

ESE - Exonic splicing enhancer

ESS - Exonic splicing silencer

EST - Expressed sequencing tags

FACS - Fluorescence-activated cell sorting

FDR - False discovery rate

G - Guanosine

GO - Gene ontology

gRNA - Guide RNA

hnRNA - Heterogeneous nuclear RNA

HTCE - High throughput capillary electrophoresis.

IgM - Immunoglobulin M

IP - Immunoprecipitation

IPTG - Isopropyl 1-thio-B-D-galactopyranoside

IR - Intron retention

IRE - Iron responsive element

ISE - Intronic splicing enhancer

ISS - Intronic splicing silencer

K - Lysine

KB - Kilobase

Kda - Kilodalton

lncRNA - Long non-coding RNA

MAD - Median average deviation

MDS - Myelodysplastic syndrome

ME - Mutually exclusive exon

miRNA - Micro RNA

MRF - Markov random field

mRNA - Messenger RNA

mRNP - Messenger ribonucleoparticle

Mut - Mutant

N-terminal - Amino terminal

NMD - Nonsense-mediated decay

nt - Nucleotide

NTC - Nineteen complex

OH - Hydroxide group

PAGE - Polyacrylamide gel electrophoresis

PBS - Phosphate buffered saline

PCA - Principal component analysis

PCR - Polymerase chain reaction

piRNA - Piwi-interacting RNA

PK - Proteinase K

PMSF - Phenylmetahnesulfony lfluoride

Pol - Polymerase

PPI - Protein-protein interaction

PPT - Polypirimidine tract

pre-mRNA - Premature mRNA

PSI - Percent spliced In

pTEFb - Positive transcription elongation factor b

PVDF - Polyvinylidene difluoride

PWM - Position weight matrix

pY tract - Polypyrimidine tract

RasiRNA - Repeat associated siRNA

RIPA - Radio-immuno-preciptation assay

RNA - Ribonucleic acid

RNAi - RNA interference

rRNA - Ribosomal RNA

RT-PCR - Reverse-transcription polymerase chain reaction

scaRNA - Small cajal body-specific RNA

SDS - Sodium dodecyl sulfate

SELEX - Systematic Evolution of Ligands by Exponential Enrichment

Seq - Sequencing

siRNA - Small interfering RNA

SMN - Survival of motor neuron

snoRNA - Small nucleolar RNA

snRNA - Small nuclear RNA

snRNP -Small nuclear ribonucleoprotein particle

SR - Serine/Arginine-rich

ss - Splice site

T - Thymidine

TCA - tyricarboxylic acid cycle

TMG - Trimethylguanosine

tRNA - Transfer RNA

TTP - Tristetraproline

U - Uracil

U2AF - U2 auxiliary factor

UGM - Undirected graphical model

vRNA - Vault RNA

WB - Western blot

WT - Wild type

Y RNA - Class of small non coding RNA

Z - Z-score

Zn - Zinc

TABLE OF CONTENTS

GENERAL INTRODUCTION

The origins of splicing

Forty years ago, after the discovery of the DNA structure and the identification of the genetic code, there was the general assumption that the transfer of information from genes to proteins was universal for all forms of life. At the time gene regulation was studied in relatively simple bacterial organisms and it was believed that the genes were structured as a contiguous string of base pairs from which transcripts would emerge during the transcription process and mRNA translation collinearly occur. In other words, there was a one-to-one correspondence of nucleotide bases between the DNA and the mRNA transcribed from a given gene locus. Although these notions were largely accurate for the prokaryotic kingdom, those concepts were not consistent with further observations made in eukaryotic organisms (Sharp, 2005).

On one hand, complex organisms contain much larger and complex genomes, and the hypothesis that such genomes could encode simple bacteria-type genes seemed unlikely. On the other hand, cellular RNAs tended to be longer in the nucleus compared to their counterparts in the cytoplasm, even though both contained the universal cap structure and the polyadenylation sequence at their 5' and 3' ends respectively. A new paradigm emerged in 1977, with the discovery of the split genes and the RNA splicing process.

The research groups led by Phillip Sharp and Richard Roberts identified a series of RNA molecules –termed mosaics – each of which contained sequences from non-contiguous sites in the adenoviral genome after viral infection of mammalian cells (Berget

et al., 1977; Chow *et al.*, 1977). By directly comparing the mRNA and its corresponding nuclear DNA, they identified DNA fragments that were removed through processing of the longer RNA precursor (Fig Intro 1). Such processing involved the removal of internal sequences, known as introns, and splicing together of the flanking coding parts, known as exons. This puzzling discovery offered an explanation to the obsevation that the long nuclear RNA and the shorter cytoplasmic mRNA could have the same cap and poly (A) tail termini, as well as the differences in length between nuclear and cytoplasmic RNAs. These novel insights marked the dawn of the era of RNA splicing.

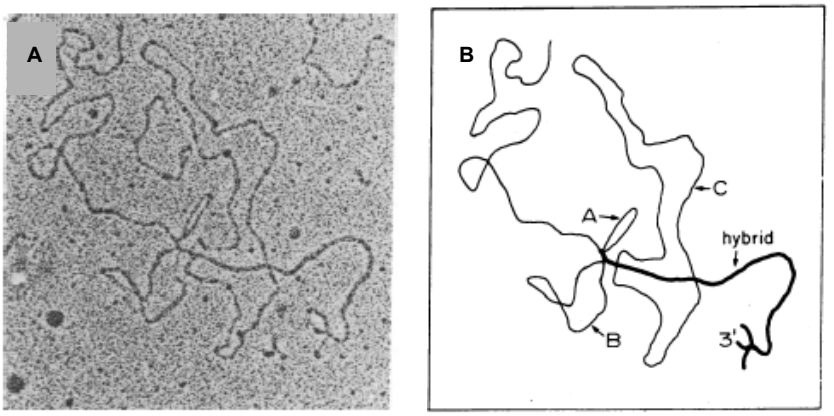

Figure Intro 1 – Electron micrograph evidence for split genes. An example of a hybrid between single-stranded EcoRI A DNA and hexon RNA is shown in **A** and diagrammed in **B**. The hybrid region is indicated by a heavy line; loops A, B and C (single stranded unhybridized DNA are joined by hybrid regions resulting from annealing of upstream DNA sequences to the 5´tail of hexon mRNA. (Modified from Berget *et al.*, 1977).

RNA splicing and Alternative splicing

In the intricate process involved in the control of the genetic flow of information, several steps are required to ensure the faithful and precise expression of different genetic programs. Pre-mRNA splicing is the molecular process by which introns are removed from nascent pre-mRNA transcripts in order to generate the mature messenger RNA (Fig Intro 2). The final mRNA consists of the remaining exonic sequences, which are sequentially connected through the splicing process and serve as templates for protein synthesis by the translation machinery. (Green, 1986; Wahl *et al.*, 2009)

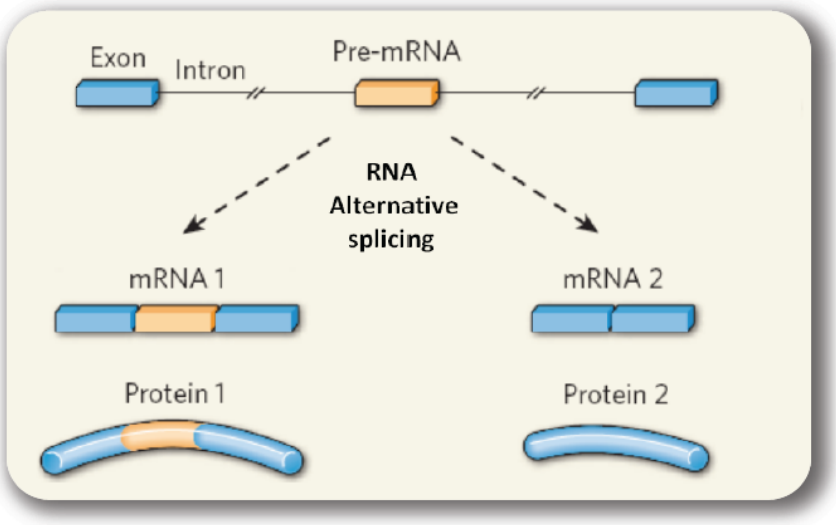

Figure Intro 2 – Alternative pre-mRNA splicing. Genomic DNA sequences are transcribed as messenger RNA precursors (pre-mRNA) containing exons and introns that can be processed by alternative pathways to generate different mRNAs encoding distinct proteins. (Modified from Tejedor and Valcarcel, 2010).

Interestingly, alternative patterns of splicing generated from a single pre-mRNA molecule can yield different mRNA species from a single gene locus (Fig Intro 2). The first evidence for endogenous cellular genes that could undergo alternative splicing was reported shortly after the discovery of the splicing process itself, involving a member of the immunoglobulin superfamily, the IgM gene (Early *et al.*, 1980). Subsequently, several other studies identified differential patterns of exon inclusion in other genes, with potential implications for health and disease (Rosenfeld *et al.*, 1981; Treisman *et al.*, 1983).

The combination of different exons in the mature mRNA can give rise to different interpretations of the genetic information present in the original DNA molecule. Given the surprising similarity in the overall number of genes between mammals and worms or insects, it has been argued that the occurrence of multiple mRNA transcripts within single genes may be an important contributor to the expansion of proteome diversity in higher eukaryotes (Maniatis and Tasic, 2002; Nilsen and Graveley, 2010).

Alternative splicing and genome complexity

Since the very first attempts to estimate the number of genes in different living organisms, just after the discovery of the genetic code, the number of predicted human coding genes has grown steadily smaller. Early estimations based on expressed sequencing tags (ESTs) implied that a high number of protein coding genes would be required to explain the proteome diversity in mammals

and other higher eukaryote species (Pertea *et al.*, 2010). But right after the publication of the human genome project the estimated number of human protein coding genes decreased dramatically, suggesting that other mechanisms involved in genome complexity must exist (Fig Intro 3A).

In fact, direct comparison between the numbers of predicted genes across different species indicated that the amount of coding genes did not vary dramatically between vertebrates and invertebrates, or even between plant and animal species (Fig Intro 3B), However, the complexity – or the number of messengers that each of those genes could generate – was surprisingly different.

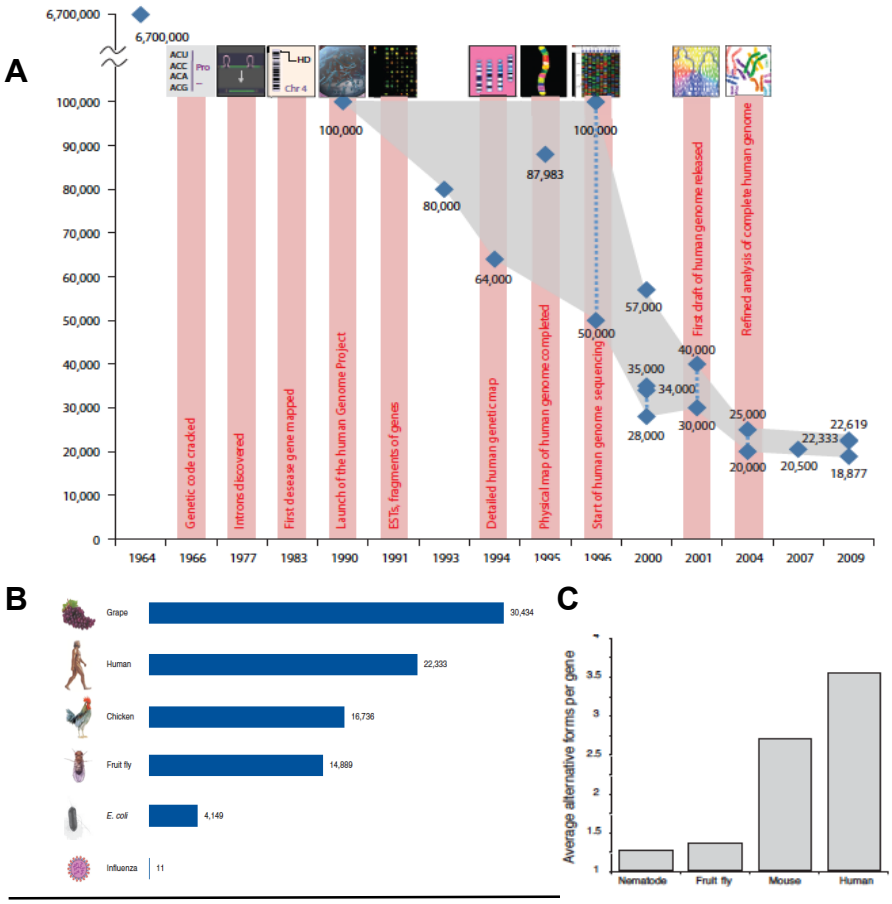

Figure Intro 3 – Alternative splicing can contribute to the functional diversification of genomic information. **A)** Evolution of human gene number counts, relative to human genome-related milestones. Individual estimates of the human gene count are shown as blue diamonds. The range of estimates at different times is shown by the two vertical blue dotted lines. **B)** The number of genes in a species bears little relation to its size or to intuitive measures of complexity. **C)** Average number of alternative splice forms per gene for each organism. (Extracted from Pertea and Salzberg, 2010; and Kim *et al.*, 2004)

In average, mammalian genomes are able to express more transcript isoforms per gene, implying increased protein coding potential from a single gene locus (Fig Intro 3C). Very recent work carried out by the groups of Burge and Blencowe (Merkin *et al.*, 2013; Barbosa-Morais *et al.*, 2013, respectively) demonstrated the contribution of alternative splicing to the establishment of tissue-specific gene expression programs across different species. Increased splicing complexity emerged as a general evolutionary strategy adopted by vertebrates, reporting frequent conversion between alternative and constitutive splicing. Moreover, the higher prevalence of alternative splicing in primates could be functionally linked to the extent of neuronal tissue development observed in those species, which raises the interesting possibility that alternative splicing diversification is associated with the acquisition of progressively complex cognitive brain functions.

Global impact of alternative splicing

The substantial progress in the design and application of novel and sensitive genome-wide techniques has allowed the global study of alternative splicing at a genomic scale. The development of custom

exon-junction microarray platforms represented a milestone in the large-scale profiling of alternative splicing events. Initial studies carried out with high-density oligonucleotide arrays (Johnson *et al.*, 2003; Pan *et al.*, 2004) concluded that at least 74% of the multi-exon-containing genes could undergo alternative splicing. Further studies confirmed the global impact of alternative splicing in humans, but also in other species such as mouse or chimp (Srinivasan *et al.*, 2005; Calarco *et al.*, 2007), and the overall impact of splicing factors or environmental stimuli on global splicing regulation (Hung *et al.*, 2007, Makeyev *et al.*, 2007, Pleiss *et al.*, 2007). Current implementation of next generation sequencing (NGS) techniques increased the sensitivity and resolution of transcriptome studies and the latest estimations from Wang *et al.* (2008), Wilhelm *et al.* (2008) and Pan *et al.* (2008) indicate that more than 90% of human genes are alternatively spliced. This observation highlights the relevance of alternative splicing in many cellular processes and its extensive contribution to the generation of proteomic complexity in different contexts, such as cell differentiation, cell transformation, cell reprogramming or apoptosis (Maniatis and Tasic, 2002; Nilsen and Graveley 2010, Stamm *et al.*, 2005. Schwerk and Schulze-Osthoff, 2005; David and Manley, 2010; Gabut *et al.*, 2011; Han *et al.*, 2013).

The splicing community has made steady efforts to generate computational tools to predict regulatory sequences important for alternative splicing regulation (Fairbrother *et al.*, 2002; Wang *et al.*, 2004; Zhang and Chasin, 2004) depending on the cell type or cellular environmental conditions. A seminal contribution to the understanding of the splicing code was provided by Barash *et al.* (2010). By computing a large number of parameters generated by

the splicing community over decades, and using transcriptomic data generated in different tissues, these authors were able to predict with high accuracy the extent of alternative splicing for a given exon depending in a subset of tissue types (see Preview by Tejedor and Valcárcel on page 49). Further refinement of the splicing code, including new parameters generated in multiple cell lines, will help to provide a more precise view of tissue-specific alternative splicing in the following years.

The circuits of splicing regulation are integrated in the larger network of gene regulation, influencing -and being influenced by- other processes such as chromatin remodeling, transcription, splicing, polyadenylation, RNA export or translation (Blencowe, 2006; Auboeuf *et al.*, 2007). On one hand, alternative splicing serves as a quality control filter for the modulation of the levels of gene expression, as alternative transcripts including premature termination codons can be degraded by the nonsense-mediated decay (NMD) machinery (Maquat, 2004). On the other, alternative splicing can increase the proteins structural and functional diversity by differential inclusion of specific domains (Nilsen and Graveley, 2010). Recent work carried out by Merkin *et al.* (2012) showed that differentially included exons among different tissues often correspond to protein sequences enriched in phosphorylation sites, suggesting that -together with kinase amounts- splicing modulation can control the extent of protein modification in a tissue specific manner. Furthermore, the extent of specie-specific alternative splicing is frequently observed in genes encoding nucleic acid binding proteins, and overall, the modulated exons correspond to unstructured regions of proteins with potential implications in the

establishment of novel protein-protein interactions (Barbosa-Morais *et al.*, 2013, Ellis *et al.*, 2012; Buljan *et al.*, 2012).

Patterns of alternative splicing

A variety of exon / intron decisions and combinations can generate multiple classes of alternative splicing patterns. The most common pattern of alternative splicing is known as cassette exon skipping, in which an internal, typically short exon, is spliced out together with its flanking introns. This form of alternative splicing regulation accounts for nearly 40% of the splicing events observed in higher eukaryotes (Alekseyenko *et al.*, 2007; Sugnet *et al.*, 2004). Interestingly, the prevalence of exon skipping increases in a gradual manner in the eukaryotic evolutionary tree (Kim *et al.*, 2008), suggesting that exon skipping is the class of alternative splicing events that contributes the most to phenotypic complexity in vertebrate species.

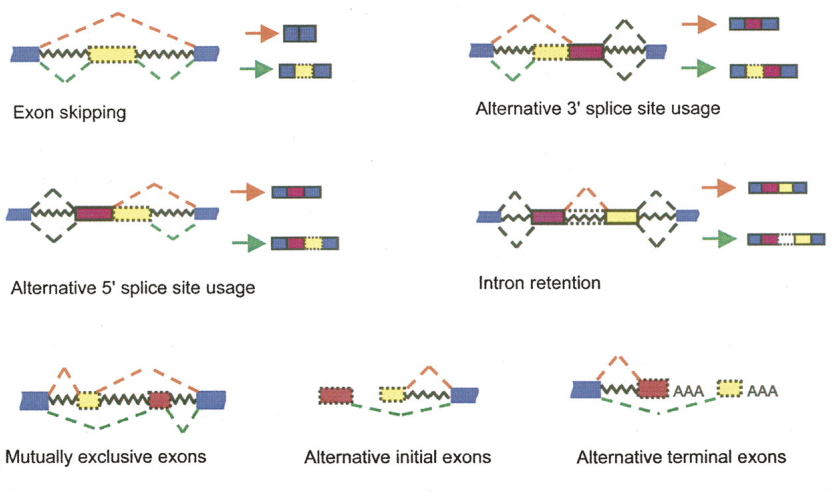

Figure Intro 4 – Types of alternative splicing. Schematic representation of the most frequent patterns of alternative splicing events.

The second most common type of alternative splicing regulation is alternative 3' splice site and 5' splice site usage, which occur when two different splice sites are recognized at one end of an exon. Selection of alternative 5' or 3' splice sites account for 18.5% and 8% of splicing regulation, respectively, often involving changes in coding sequences, sometimes as limited as a single codon (Blencowe, 2006).

The third most abundant class of regulation is intron retention, in which an intron is preserved in the mature mRNA transcript. While these events are apparently rare in the animal kingdom (Alekseyenko *et al.*, 2007; Kim *et al.*, 2008; Sakabe *et al.*, 2007; Sugnet *et al.*, 2004), they represent is the most prevalent type of alternative splicing event in plants, fungi and protozoa.

Another category of alternative splicing events includes complex alternative splicing patterns such as mutually exclusive exons, alternative first exons and alternative terminal exons (Ast, 2004; Kim *et al.*, 2008; Black, 2003).

Cis-acting elements involved in RNA splicing

Correct processing of pre-mRNA molecules requires precise definition of the location of exon-intron boundaries. The fidelity of the splicing phenomenon relies on the presence of cis-acting elements and trans-acting factors involved in the identification of

the intronic and exonic edges. Three essential sequences are required for the splicing reaction to occur (Figure Intro 5). These sequences are well defined in budding yeast introns, but less so in higher eukaryotes, where some splicing signals can be highly variable, complicating their identification by computational-based tools, including comparative methods (Green, 1986; Corvelo *et al.*, 2010).

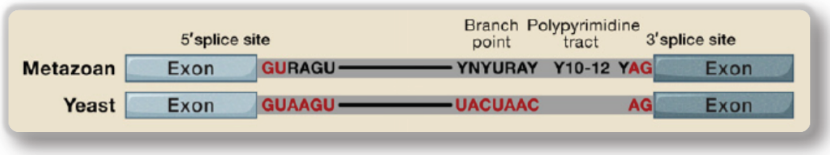

Figure Intro 5 - Cis acting elements involved in RNA splicing. The three conserved signals that enable recognition of RNA by the spliceosome include: the exon–intron junctions at the 5′ and 3′ ends of introns (the 5′ splice site (5′ SS) and 3′ splice site (3’ SS) and the branch site sequence located upstream of the 3′ SS. Higher eukaryotes contain, in addition, a polypyrimidine tract (PPT) located between the 3′ SS and the branch site (from Wahl *et al.*, 2009).

The 5’ end of the intron –known as the 5’ splice site or donor site – contains a highly conserved GU dinucleotide sequence followed by a less conserved consensus sequence AG/GURAGU (R= A or G nucleotide, / = exon-intron boundary) (Horowitz and Krainer, 1994, Roca *et al.*, 2013). On the other side of the intron, the intronic 3’ end –known as 3’ splice site or acceptor site - is characterized by the presence of the AG dinucleotide, which in higher eukaryotes is usually preceded by an upstream pyrimidine-rich nucleotide, the Polypyrimidine Tract (Ruskin et al 1984, Reed and Maniatis 1985). The polypyrimidine tract is composed of pyrimidine stretches of different lengths, with a general correspondence between 3’ ss strength and the abundance of uridine residues in the region

(Ruskin and Green, 1985; Reed, 1989; Coolidge *et al.*, 1997). Finally, the branch point region is defined by the presence of a conserved adenosine (the branch point nucleotide –which forms a 2'-5' phosphodiester bond with the 5' end of the intron after the first catalytic step of the splicing reaction, see below-), flanked by a highly degenerated sequence YNYURAY (Y=U or C, R= G or A, the underlined A indicates the branched adenosine) (Zhang, 1998), and is generally located within 18-40 upstream of the 3' splice site (Ruskin et al 1984; Reed and Maniatis 1985)

The splicing reaction, at a glance

The process of intron removal consists of two transesterification reactions that occur in two sequential steps (Figure Intro 6). In the first step of the splicing reaction, the 2'-OH group from the branch point Adenosine attacks the phosphate group between the 5' splice site Guanosine, and the last nucleotide of the exon. As a result, the phosphodiester bond between the 3' end of the exon and the 5' end of the intron is excised, leading to the formation of an intermediate lariat containing a 2'-5' phosphodiester bond between the 2' –OH or the adenosine ribose ring and the 5' phosphate at the intronic 5' end (Konarska *et al.*, 1985; Ruskin and Green 1985). In the second catalytic step, the free 3'-OH from the excised exon carries a nucleophilic attack on the phosphate bond linking the last nucleotide of the intron (the guanosine) with the first nucleotide of the second exon. This leads to ligation of the exons to form the mature mRNA and the release of the intron, still in a lariat structure (Ruskin et al, 1985; Konarska *et al.*, 1985; Moore and Sharp 1993;).

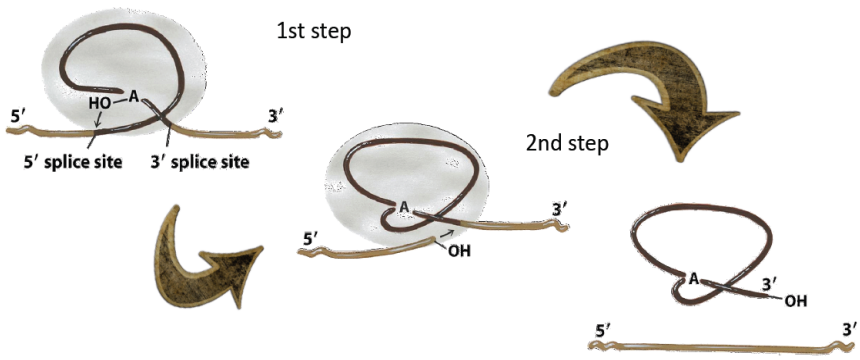

Figure Intro 6 – The two trans-esterification steps of the RNA splicing reaction. The first catalytic reaction occurs by a nucleophilic attack of the 2'-OH of the branch point adenosine to the 5' splice site phosphate. In the second catalytic step, the 3'-OH of the 5' free exon attacks the 3' splice site phosphate, resulting in exon-exon ligation and the release of the intron lariat.

A number of RNA molecules, including the group I and group II introns, have the potential to undergo self-splicing (Michel *et al.*, 2009). Group II self splicing introns are able to perform a chemical reaction with high degree of similarity to the pre-mRNA splicing process, including the generation of lariat intermediates and very similar coordination of metal ions by equivalent RNA structures (Fica *et al.*, 2013). This suggests that the complex pre-mRNA splicing reaction evolved ultimately from Group II introns. As nuclear pre-mRNAs have lost their potential to undergo self-splicing, they require the concerted action of a sophisticated ribonucleoprotein machinery – the spliceosome- to remove intronic sequences from nascent transcripts.

The spliceosome

The spliceosome is one of the most complicated ribonucleoprotein machineries within the cell. It is composed of five snRNPs (U1, U2, U4, U5 and U6) and a plethora of more than 200 protein factors that control the progression of the reaction during the splicing cycle. Contrary to the ribosome, which is a stable ribonucleoprotein complex, the spliceosome is a highly dynamic entity, and requires a high degree of molecular plasticity to achieve its final goal. On one hand, the pre-mRNA substrate plays a crucial role in the formation of an active site for the splicing reaction. On the other, different ribonucleoprotein components contribute to formation of an active catalytic site, and progression through the reaction is mediated by the recruitment and release of a multitude of factors involved in the formation of RNA-RNA, RNA-protein and protein-protein interactions that will help to recognize splice sites and shape the catalytic center of the spliceosome (reviewed in Wahl et a., 2009).

The biogenesis of the modular components of the spliceosome is a complex process and requires several maturation steps. Each of the snRNPs involved in the splicing reaction contain a highly structured uridine-rich snRNA, a common set of Sm proteins (Lsm for U6) and a specific set of protein components (Wahl *et al.*, 2009). Most of the snRNAs (U1, U2, U4 and U5) are transcribed by RNA polymerase II, and are characterized by the presence of a 5'-trimethylguanosine cap (TMG), a 3' stem-loop structure important for the generation of the snRNA's correct 3' end and an Sm protein binding site. In contrast, U6 snRNA is transcribed from RNA polymerase III, contains a mono-methylphosphate cap and is

bound by Lsm core proteins, which form a 7 protein ring around the snRNA similar to that of Sm proteins on the other spliceosomal snRNAs (Matera *et al.*, 2007).

Prior to becoming functional components of the spliceosome, snRNAs must undergo a complex maturation process which involves capping of the snRNAs and export to the cytoplasm, Sm core protein assembly -mediated by the SMN complex – in which Sm proteins B/B', D1, D2, D3, E, F, and G form a ring around the Sm site (Raker *et al.*, 1996)- m7G cap trimethylation (Mouaikel *et al.*, 2002) and 3' end trimming by 3' 5' exonucleases (Huang *et al.*, 1999). The trimethylguanosine cap and the Sm core promote the nuclear targeting of the assembled snRNP particle and localization in Cajal bodies, which is an essential step for the assembly of some snRNP specific proteins and snRNP editing by introducing site-specific modifications, such as methylation and pseudouridylation (Will and Luhrmann, 2001; Matera *et al.*, 2007).

On average, proteins represent more than two-thirds of the spliceosomal mass, indicating a crucial role of these factors in the rearrangements required for splicing catalysis. Recent studies focusing on spliceosomal protein purification at defined stages of the reaction have identified a highly dynamic exchange of proteins from one step to the next during the splicing process (Behzadnia *et al.*, 2007; Bessonov *et al.*, 2008; Deckert *et al.*, 2006; Hartmuth *et al.*, 2002; Jurica *et al.*, 2002; Makarov *et al.*, 2002). These transitions shed light into the workings of the spliceosome during the course of the splicing reaction.

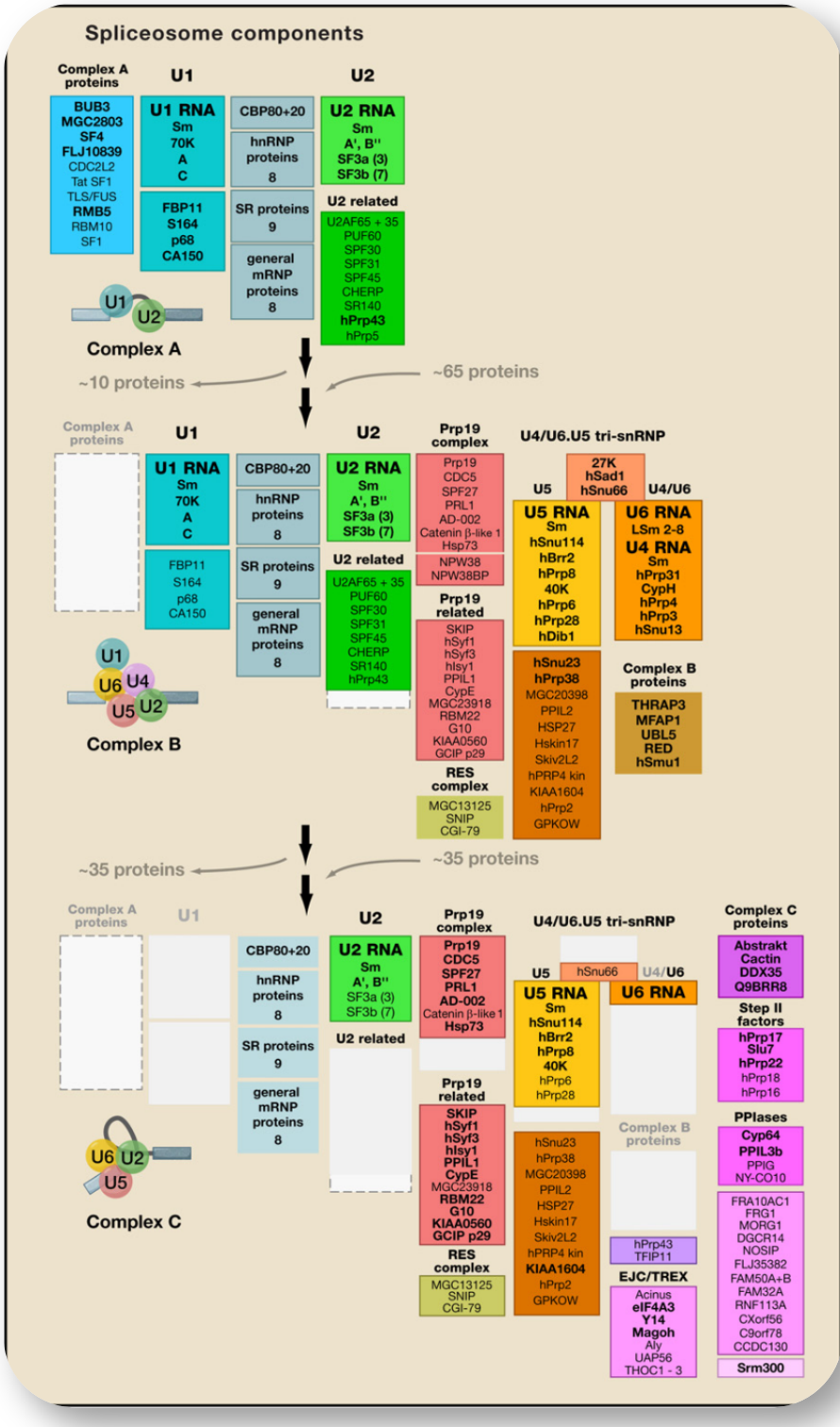
Spliceosome components
Complex A proteins
BUB3
MGC2803
SF4
FLJ10839
CDC2L2
Tat SF1
TLS/FUS
RMB5
RBM10
SF1
U1
U1 RNA
Sm
70K
A
C
FBP11
S164
p68
CA150
CBP80+20
hnRNP proteins
8
SR proteins
9
general mRNP proteins
8
U2
U2 RNA
Sm
A', B''
SF3a (3)
SF3b (7)
U2 related
U2AF65 + 35
PUF60
SPF30
SPF31
SPF45
CHERP
SR140
hPrp43
hPrp5
Complex A
~10 proteins
~65 proteins
Prp19 complex
Prp19
CDC5
SPF27
PRL1
AD-002
Catenin β-like 1
Hsp73
NPW38
NPW38BP
Prp19 related
SKIP
hSyf1
hSyf3
hIsy1
PPIL1
CypE
MGC23918
RBM22
G10
KIAA0560
GCIP p29
RES complex
MGC13125
SNIP
CGI-79
U4/U6.U5 tri-snRNP
27K
hSad1
hSnu66
U5
U4/U6
U5 RNA
Sm
hSnu114
hBrr2
hPrp8
40K
hPrp6
hPrp28
hDib1
U6 RNA
LSm 2-8
U4 RNA
Sm
hPrp31
CypH
hPrp4
hPrp3
hSnu13
hSnu23
hPrp38
MGC20398
PPIL2
HSP27
Hskin17
Skiv2L2
hPRP4 kin
KIAA1604
hPrp2
GPKOW
Complex B proteins
THRAP3
MFAP1
UBL5
RED
hSmu1
Complex B
~35 proteins
~35 proteins
Complex C
Complex C proteins
Abstrakt
Cactin
DDX35
Q9BRR8
Step II factors
hPrp17
Slu7
hPrp22
hPrp18
hPrp16
PPIases
Cyp64
PPIL3b
PPIG
NY-CO10
FRA10AC1
FRG1
MORG1
DGCR14
NOSIP
FLJ35382
FAM50A+B
FAM32A
RNF113A
CXorf56
C9orf78
CCDC130
Srm300
EJC/TREX
Acinus
eIF4A3
Y14
Magoh
Aly
UAP56
THOC1 - 3
hPrp43
TFIP11

Figure Intro 7 – Compositional Dynamics of Human A, B, and C Spliceosomal Complexes. Protein composition of the human A complex (Behzadnia *et al.*, 2007), the human B complex (Bessonov *et al.*, 2008), and the human C complex (Bessonov *et al.*, 2008) as determined by mass spectrometry from highly purified complexes. Proteins are grouped according to snRNP association, function, presence in a stable heteromeric complex, or association with a particular spliceosomal complex, as indicated. (Extracted from Wahl *et al.*, 2009)

One key conclusion extracted from those studies was the fact that various groups of spliceosomal proteins were present in the spliceosome throughout the splicing cycle, and also, that a large number of auxiliary spliceosomal proteins were recruited during the sequential spliceosomal transitions. In addition, a large number of spliceosomal proteins were destabilized at each stage of the splicing process, revealing a high rate of spliceosome remodeling during the splicing cycle (see Fig Intro 7).

The spliceosome assembly cycle

As mentioned above, the splicing machinery assembles on each RNA substrate in a stepwise manner. Although a full spliceosomal particle composed by the five snRNPs can be purified under low salt conditions (Stevens *et al.*, 2002), in vitro and in vivo experiments strongly suggest the consecutive assembly of the spliceosomal components through the splicing cycle (Konarska and Sharp, 1986; Pikienly *et al.*, 1986; Lacadie and Rosbash, 2005; Agafonov *et al.*, 2011). Interestingly, the mechanism is highly conserved in all intron-containing organisms, from yeast to vertebrates, suggesting a common ancestral origin for the splicing process (Ast, 2004).

Spliceosome assembly includes the following steps. The pre-mRNA 5' splice site is recognized by the U1 snRNP particle through protein-RNA and RNA-RNA interactions that do not require ATP hydrolysis. The 5' end of U1 snRNA can establish base-pairing interactions with the last three nucleotides of the exon and the first six of the intron (Lerner *et al.*, 1980; Mount *et al.*, 1983; reviewed in Madhani and Guthrie, 2004), while the U1-specific protein U1C contacts the 5' splice site region and may stabilize U1-5'ss base pairing (Du and Rosbash, 2002; Heinrichs *et al.*, 1990; Pomeranz Krummel *et al.*, 2009). In addition, the earliest assembly phase of the spliceosome requires recognition of the branch point and 3' splice site. In budding yeast the branch point is recognized by the Branch Point Binding protein (BBP), while in higher eukaryotes SF1/BBP binds to the branch point cooperatively with the U2 auxiliary factor (U2AF), a heterodimer of 65 and 35 Kda subunits that recognize, respectively, the polypyrimidine tract and the 3' splice site AG (Kramer and Utans, 1991; Berglund *et al.*, 1997; Ruskin *et al.*, 1988; Zamore et al, 1992; Zhang et al, 1992; Merendino *et al.*, 1999; Wu et al, 1999; Zorio and Blumenthal, 1999). Recognition of the 5' splice site by U1 snRNP and of the 3' end of the intron by SF1/BBP – U2AF, and likely molecular links between these complexes, can commit splice sites to undergo the splicing reaction (Légrain et al, 1988; Séraphin and Rosbash, 1989). This complex is therefore known as commitment, or E complex.

Subsequent to the formation of pre-spliceosomal E complex, U2 snRNP engages in base pairing interactions with the pre-mRNA branch point sequence, a process that requires the hydrolysis of ATP, leading to the release of the SF1/BBP protein and the

formation of the spliceosomal A complex. The most characteristic feature of this RNA-RNA duplex is the presence of the unpaired branch adenosine, that bulges out the helix facilitating the nucleophylic attack of its 2' OH group on the 5' splice site (Query et al., 1994)

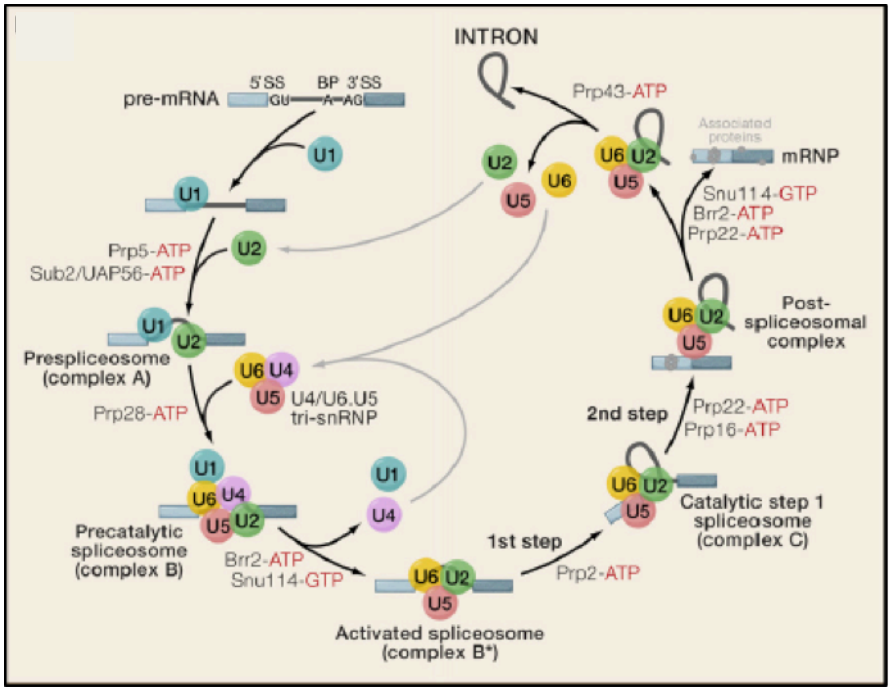

Figure Intro 8 - The stepwise assembly of spliceosomal snRNPs. A group of evolutionarily conserved DExD/H-type RNA-dependent ATPases / helicases act at specific steps of the splicing cycle to catalyze RNA-RNA rearrangements and also RNP remodeling events. (Extracted from Wahl *et al.*, 2009)

Right after A complex formation, the tri-snRNP particle (composed of the U4/U6 di snRNP and the U5 snRNP) is recruited as a preassembled U4/U6. U5 snRNP, which leads to formation of B complex (Konarska and Sharp, 1987). Even though all the necessary components for the splicing reaction are present,

spliceosome complex B remains catalytically inactive and requires major conformational and compositional rearrangements prior to the first trans-esterification step of the reaction. During spliceosome activation, U1 and U4 snRNPs are destabilized and released from the complex, forming the activated spliceosome – also termed B* or B act complex- (Bessonov *et al.*, 2008) and characterized by the hyperphosphorylation of SF3B1 (Wang *et al.*, 1998). Then, after a series of intramolecular RNA-RNA rearrangements between the U2 and U6 snRNAs (Sun and Manley, 1995) and the U6 snRNA and the 5' splice site (Lamond *et al.*, 1988; Wassarman and Steitz, 1992), the active spliceosome undergoes the first catalytic step of splicing, resulting in spliceosomal C complex formation. The PRP19 associated complex (NTC) is also recruited to the reaction at this stage, and is required for stable association of the different snRNPs and in conformational changes that must occur during spliceosome activation (Hogg *et al.*, 2010). Several other spliceosomal rearrangements occur prior to the second catalytic step (Konarska *et al.*, 2006). After the second catalytic step, SF3B1 is dephosphorylated (Shi *et al.*, 2006), the spliceosome is disassembled, and the final mRNA molecule is released in the form of an mRNP. Intermediate molecules and products such as the intron lariats are degraded by exonucleases and the U2, U5 and U6 snRNPs are recycled for further rounds of splicing (Wahl *et al.*, 2009).

Combinatorial control of alternative splicing

To overcome the problem of the highly degenerated splicing signals –and to take advantage of it to achieve regulation- ,a variety of additional parameters allow successful splice site recognition. Splice site strength, splicing enhancers and silencers, exon-intron organization and RNA secondary structure are some of the features involved in the recognition of the proper splice sites in each biological condition.

Splice site strength is generally estimated by the potential base pairing of U1 and U2 snRNAs with the 5' splice site or the branch site, respectively, the length and pyrimidine content of the polypyrimidine tract, and the sequences surrounding the 3' splice site (Horowitz and Krainer, 1994; Ruskin and Green, 1985; Zhang 1998). Perfect base pairing of U1 snRNP with the 5' splice site is a good indicator of strong splice site recognition, although there are some exceptions to this rule. On one hand, shifts in the U1 snRNA-5' splice site base pairing can help to recognize non-canonical 5' splice sites (Roca and Krainer, 2009). Moreover, U1 snRNP does not seem to be required for the splicing reaction in the presence of an excess of SR proteins (Crispino *et al.*, 1994; 1996). On the other hand, extended U1 base pairing could halt further remodeling of the spliceosome and the establishment of the proper interactions required for different steps of the splicing cycle (Chiou *et al.*, 2013). Therefore splice site recognition requires a fine balance between precise molecular recognition and dynamic interactions along the spliceosome assembly pathway. Weak 5' splice sites can be

properly recognized by the presence of uridine-rich stretches 3' of the 5' splice site and by trans acting factors TIA-1/TIAR which, through interactions with U1C, help to recruit U1 snRNP (Förch *et al.*, 2000; 2002; Wang *et al.*, 2010).

Binding of trans-acting factors to the polypyrimidine tract or the AG acceptor site increase the efficiency of the splicing reaction. (Merendino *et al.*, 1999; Wu *et al.*, 1999; Zamore *et al.*, 1992; Zorio and Blumenthal, 1999). The recognition of the AG dinucleotide at the 3' splice site by U2AF35 is required for U2AF65 binding to polypyrimidine tracts with poor pyrimidine content (Merendino *et al.*, 1999; Wu *et al.*, 1999; Zorio and Blumenthal, 1999). Other factors related to U2AF65, including PUF60 and RBM39, can bind to this regions and promote the splicing of introns with weak 3' splice site signals (Corsini *et al.*, 1999; Hastings *et al.*, 2007; Dowhan *et al.*, 2005).

On the other hand, the presence of other cis regulatory sequences, as well as the relative concentrations of their cognate trans-acting factors, can have a direct effect in the modulation of the splice site choice.

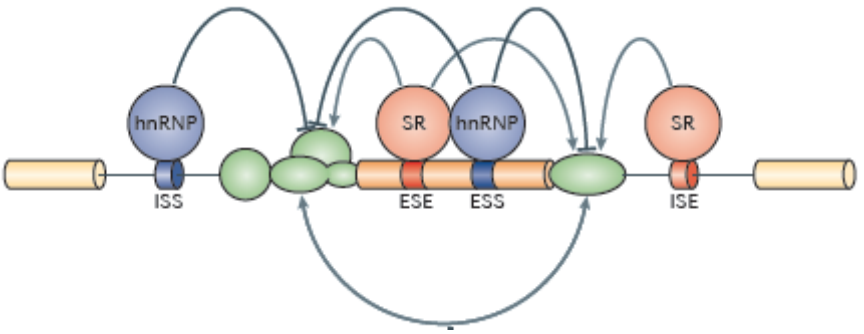

Figure Intro 9 – Splicing enhancers and silencers. Splicing enhancers are usually bound by SR proteins, which interact with spliceosomal components to stabilize their binding to neighboring splice sites. Binding of SR proteins to ESEs or ISEs stimulates the binding of U2AF in the 3' splice site and/or U1 snRNP in the 3' splice site. Conversely, hnRNP proteins tend to bind exon and intron splicing silencers, having opposite effects on alternative splicing regulation. (Extracted from Kornblihtt *et al.*, 2013).

The recognition and recruitment of the splicing machinery is highly dependent on the presence of splicing enhancers (exonic splicing enhancers or ESEs, intronic splicing enhancers or ISEs), while splicing silencers (exonic splicing silencer or ESSs, intronic splicing silencer or ISSs) can interfere with the selection of particular splice sites (Matlin *et al.*, 2005; Chen and Manley, 2009). These regulatory sequences are bound by two main families of non-snRNP RNA-binding proteins, the SR and hnRNP proteins, and the activities of these groups of factors have been the focus of the vast majority of studies regarding alternative splicing regulation (Cartegni *et al.*, 2002; Long and Caceres, 2009; Martinez-Contreras *et al.*, 2007).

The exon-intron organization plays also an important role in the final outcome of the splice site choice. While in higher eukaryotes introns tend to be much larger than exons, in lower eukaryotes, introns are significantly smaller and exons are substantially larger, suggesting differential mechanisms for splicing regulation. In the case of short intronic sequences, the spliceosome recognizes the intron as a unit, through the process known as intron definition (Fox-Walsh *et al.*, 2005, de Conti *et al.*, 2013). On the other side, small exons are recognized among the vast sea of intronic sequences by the identification of the splice site signals across the

exon through the process of exon definition (Berget *et al.*, 1995; de Conti *et al.*, 2013). The molecular determinants of splice site pairing mediating intron or exon definition remain largely unknown, although the protein Prp40 has been proposed to play such a role in intron definition in budding yeast (Abovich and Rosbash., 1997).

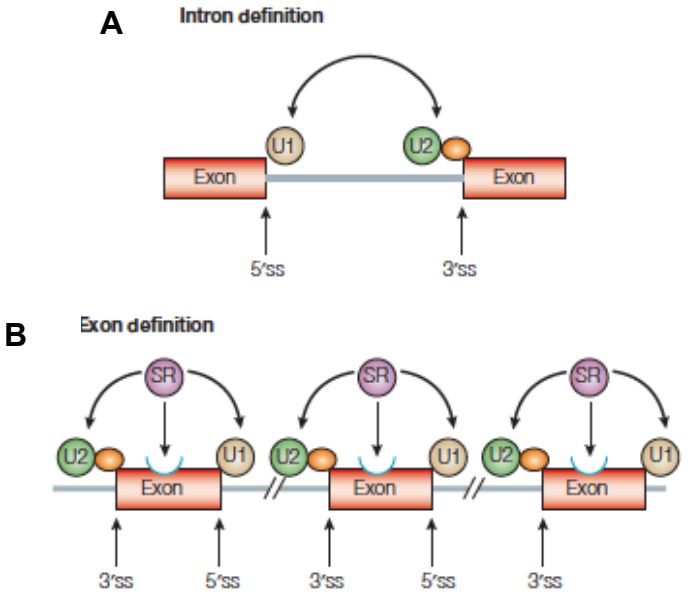

Figure Intro 10 - Exon and intron definition processes. **A**) The intron definition process involves the binding of U1 snRNP to the upstream 5′ splice site and U2AF and U2 snRNP to the downstream polypyrimidine tract and branch site of the same intron. Intron definition selects pairs of splice sites located on both ends of the same intron. **B**) In the exon definition process, SR proteins bind to exonic splicing enhancers and recruit U1 to the downstream 5′ss and the splicing factor U2AF to the upstream polypyrimidine tract and the 3′ss, as well as U2 to the branch site. Exon definition promotes the formation of a 'cross-exon' recognition complex, recruiting the basal splicing machinery to the splice sites flanking the exon (extracted from Ast, 2004).

In addition, RNA secondary structures can also affect the combinatorial control of alternative splicing. On one hand, they can hinder the accessibility of splicing factors to functional sequences, like exon splicing enhancers or particular splice sites, by sequestering them in stem loop structures (Hiller *et al.*, 2007). Moreover, the presence of different RNA secondary structures may modify the relative distance of the splicing signals, and could alter their ability to modulate splice site choice (Buratti and Baralle, 2004; Meyer *et al.*, 2011; Plass *et al.*, 2012; Graveley, 2005; Raker *et al.*, 2009)

Transcription, chromatin and alternative splicing

Increasing evidence indicates that alternative splicing is not only regulated by the relative abundance of splicing factors but, similarly to other RNA processing events, (i.e capping, cleavage and polyadenylation and editing), can be controlled also by the transcription and chromatin remodeling machineries (Maniatis and Reed., 2002; Montes *et al.*, 2012; Kornblihtt *et al.*, 2013).

Evidence indicates that pre-mRNA splicing often occurs co-transcriptionally (Beyer and Osheim, 1988; Pandya-Jones and Black, 2009; Khodor *et al.*, 2011; Tilgner *et al.*, 2012;)- Several factors can be involved in the regulation of co-transcriptional splicing, including promoter architecture (Cramer *et al.*, 1997), transcription factors (Kornblihtt *et al.*, 2004), chromatin remodelers (Batsché *et al.*, 2006) and factors affecting chromatin structure (Luco *et al.*, 2010). To explain the effect of the coupling between

the transcriptional and splicing machineries, two mechanisms have been proposed: The recruitment model and the kinetic model.

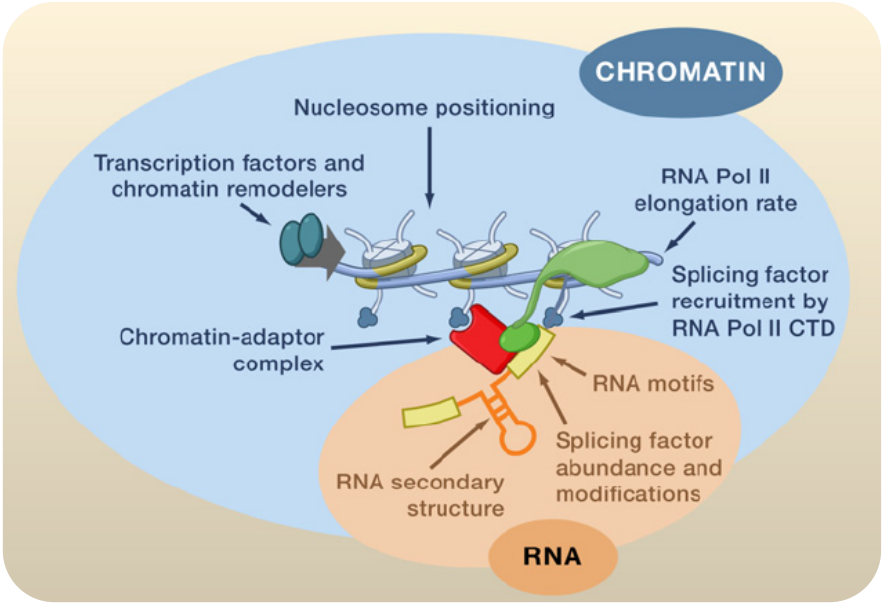

Figure Intro 11 - An Integrated Model for the Regulation of Alternative Splicing. Differential alternative splicing patterns are determined by a combination of parameters including *cis*-acting RNA regulatory elements together with transcriptional and chromatin properties that modulate the recruitment of splicing factors to the pre- mRNA. (Extracted from Luco *et al.*, 2011).

The recruitment model proposes that the recruitment of splicing factors to transcription sites is mediated by elements of the transcriptional machinery. A clear example of association of splicing factors with transcription factors is the carboxy terminal domain (CTD) of the RNA polymerase II (Muñoz *et al.*, 2010). Several studies have identified different mechanisms by which the

CTD of RNA Polymerase II is able to recruit different transcription and splicing factors involved in splicing regulation (de la Mata and Kornblihtt, 2006; Monsalve *et al.*, 2000; Huang *et al.*, 2013). More recent reports have documented a role for different chromatin marks in the recruitment of RNA processing factors, acting as a bridge between chromatin and regulatory factors that affect the regulation of nascent pre-mRNA, suggesting that epigenetic and splicing regulation can be inter-connected (Luco *et al.*, 2011, Iannone *et al.*, 2013).

The kinetic model of co-transcriptional splicing suggests that the rates of RNA Pol II-mediated elongation can influence the outcome of splicing by affecting the time at which splice sites and regulatory sequences emerge from the nascent pre-mRNA during transcription. In such scenario, slight modifications in Pol II elongation rates mediated either by sequences that induce Pol II pausing (Roberts *et al.*, 1998; Close *et al.*, 2012), or by drugs that modify the chromatin or elongation factor status (Nogues *et al.*, 2003; Ip *et al.*, 2011) can alter the outcome of alternative splicing. The most direct support of the kinetic coupling has come from the use of slow RNA Pol II mutants with reduced elongation rate in vitro and in vivo (Chen *et al.*, 1996; Boireau *et al.*, 2007), which lead generally to higher levels of alternative exon inclusion, supporting the idea that slow RNA synthesis provides a more ample window for the first of two competing splice sites to engage in the splicing reaction and therefore lead to its selection.

Splicing and disease

Alternative splicing contributes to the generation of a highly dynamic human proteome through tight coordination of splicing events, but it also could make cells vulnerable if the mechanisms governing splicing regulation are not properly controlled. Mutations in cis-acting sequences and in trans-acting factors have been shown to contribute to human disease. An unexpectedly large fraction of exonic mutations display a primary pathogenic effect on splicing, and single point mutations affecting splice sites can account for approximately 10% of the genetic diseases (Wang and Cooper, 2007; Cooper and Dreyfuss, 2009). Moreover, normal genetic variation significantly contributes to disease severity and susceptibility by affecting splicing efficiency and several missense and silent mutations located in exonic or intronic regions can lead to disease phenotypes by the generation of aberrant alternative splicing patterns in relevant genes. In fact, it was hypothesized that more than half of known hereditary disease-causing mutations could disrupt splicing by this mechanism (Lopez-Bigas *et al.*, 2006)

Many studies have focused on the role of cis-acting mutations in the susceptibility or severity of disease, as in the case of the Duchenne syndrome, which is caused by mutations in the intron 78 of the dystrophin gene. The disruption of a splicing silencer generates an aberrant transcript with premature stop codons, reducing the amount of functional protein within the cell (Disset *et al.*, 2006). Another example of disease causing mutations affecting splicing patterns is found in cystic fibrosis, in which the combination of different mutations in the CFTR gene with the presence of polymorphic regions influences the severity of the disease. (Buratti

et al., 2004). A disruption in the balance between Tau isoforms has been observed in Frontotemporal dementia patients carrying splicing mutations in the microtubule-associated protein tau gene MAPT (Liu and Gong, 2008). Also, a silent mutation in the LMNA gene that enhances the use of a cryptic 5' splice site leads to the production of a negative dominant protein that induces aberrant nuclear morphology and genome instability, and has been linked with the premature aging phenotype observed in the Hutchinson-Gilford Progeria Syndrome (Lopez-Mejia *et al.*, 2011).

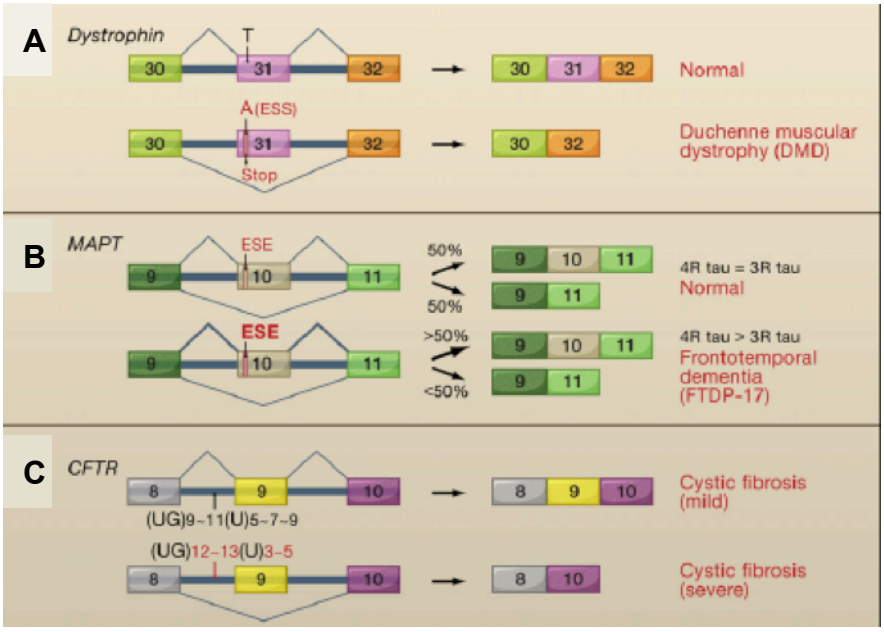

Figure Intro 12 – Disrupted splicing patterns caused by cis acting mutations. **A**) A T/A substitution in exon 31 of the *dystrophin* gene simultaneously creates a premature termination codon (STOP) and an ESS, leading to exon 31 skipping. (**B**) Mutations within and downstream of exon 10 of the *MAPT* gene encoding the tau protein affect splicing regulatory elements and disrupt the normal 1:1 ratio of mRNAs including or excluding exon 10. (**C**) Polymorphic (UG)m(U)n tracts within the 3' splice site of the *CFTR* gene exon 9 influence the extent of exon 9 inclusion and the level of full-length functional protein, modifying the

severity of cystic fibrosis (CF). (Extracted from Cooper and Dreyfuss, 2009).

On the other hand, mutations and / or differential expression of trans acting factors could lead to disease, and even though mutation of core spliceosomal components could be anticipated to compromise cell viability, various disorders have been associated with mutations in the core spliceosomal machinery. In the case of Retinitis pigmentosa, mutations in several core spliceosome factors (PRP3, PRP8, PRP31, PRP9) have been linked with retinal degeneration and blindness, and interestingly, only photoreceptor neurons seem to be affected by the haploinsufficiency of these essential components (Mordes *et al.*, 2006). The inactivation of the SMN1 gene, involved in snRNP processing, leads to spinal muscular atrophy, a motor neuron neurodegenerative disease. (Lefebvre *et al.*, 1995). Finally, recent studies have documented frequent mutations in the core spliceosomal machinery in cancer, including Myelodysplastic Syndromes (MDS) and Chronic Lymphocytic Leukemia (CLL), in which mutations in splicing factors are present in 15 % (CLL) to up to 70 % (MDS) of the patients (Yoshida *et al.*, 2011; Quesada *et al.*, 2011)

In fact, complex diseases as cancer show a high repertoire of splicing defects, caused by direct mutations in spliceosomal components or by aberrant expression of splicing factors. As an example, disruption in the expression of SR proteins (SRSF1, SRSF6) and hnRNP proteins (PTB, HNRNPA1, HNRNPA2B1), has been correlated with the prognosis of multiple cancers (Anczukow *et al.*, 2012; Karni *et al.*, 2007; David *et al.*, 2010)

The discovery of novel disease-causing mutations in RNA splicing has raised interest about the spliceosome as a new therapeutic target. In addition, our increased understanding of RNA biology and chemistry is providing new RNA-based tools for developing therapeutics. Hopefully, the combination of bioinformatics tools and high throughput technologies will help to decipher a novel disease-related splicing code able to predict disease causing or disease modifying mutations in the near future.

References

Abovich N, Rosbash M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell. *89*(3):403-12.

Agafonov DE, Deckert J, Wolf E, Odenwälder P, Bessonov S, Will CL, Urlaub H, Lührmann R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. Mol Cell Biol. *31*(13):2667-82.

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*, 661–670.

Anczuków O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, Muthuswamy SK, Krainer AR. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. Nat Struct Mol Biol. *19*(2):220-8.

Ast, G. How did alternative splicing evolve? (2004) Nature Rev. Genet. *5,* 773–782..

Auboeuf D, Batsché E, Dutertre M, Muchardt C, O'Malley BW. (2007). Coregulators: transducing signal from transcription to alternative splicing. Trends Endocrinol Metab. *18*(3):122-9.

Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. (2010). Deciphering the splicing code. Nature. *465*(7294):53-9.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science. *338*(6114):1587-93.

Batsché, E., Yaniv, M. & Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nature Struct. Mol. Biol. *13*, 22–29.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J. *26*, 1737–1748.

Berget SM, Moore C, Sharp PA. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A. *74*(8):3171-5.

Berget SM. (1995). Exon recognition in vertebrate splicing. J Biol Chem. *270*(6):2411-4.

Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell. *89*(5):781-7.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846–850.

Beyer AL, Osheim YN. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. *2*(6):754-65.

Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. (2003). Annu. Rev. Biochem. *72*, 291–336.

Blencowe BJ. (2006). Alternative splicing: new insights from global analyses. Cell. *126*(1):37-47.

Boireau S, Maiuri P, Basyuk E, de la Mata M, Knezevich A, Pradet-Balade B, Bäcker V, Kornblihtt A, Marcello A, Bertrand E. (2007). The transcriptional cycle of HIV-1 in real-time and live cells. J Cell Biol. *179*(2):291-304.

Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, Babu MM. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol Cell. *46*(6):871-83.

Buratti E, Baralle FE. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol. *24*(24):10505-14.

Buratti E, Brindisi A, Pagani F, Baralle FE. (2004). Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: a functional link with disease penetrance. Am J Hum Genet. *74*(6):1322-5.

Calarco JA, Xing Y, Cáceres M, Calarco JP, Xiao X, Pan Q, Lee C, Preuss TM, Blencowe BJ. (2007). Global analysis of alternative splicing differences between humans and chimpanzees. Genes Dev. *21*(22):2963-75.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. *31*(13):3568-71.

Chen M, Manley JL. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol. *10*(11):741-54.

Chen, Y., Chafin, D., Price, D. H. & Greenleaf, A. L. (1996). Drosophila RNA polymerase II mutants that affect transcription elongation. J. Biol. Chem. *271*, 5993–5999.

Chiou NT, Shankarling G, Lynch KW. (2013). hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly. Mol Cell. *49*(5):972-82.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell. *12*(1):1-8.

Chow LT, Roberts JM, Lewis JB, Broker TR. (1977). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell. *11*(4):819-36.

Close P, East P, Dirac-Svejstrup AB, Hartmann H, Heron M, Maslen S, Chariot A, Söding J, Skehel M, Svejstrup JQ. (2012). DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. Nature. *484*(7394):386-9.

Coolidge CJ, Seely RJ, Patton JG. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. Nucleic Acids Res. *25*(4):888-96.

Cooper TA, Wan L, Dreyfuss G. (2009). RNA and disease. Cell. *136*(4):777-93.

Corsini L, Bonnal S, Basquin J, Hothorn M, Scheffzek K, Valcárcel J, Sattler M. (2007) U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nat Struct Mol Biol. *14*(7):620-9.

Corvelo A, Hallegger M, Smith CW, Eyras E. (2010). Genome-wide association between branch point properties and alternative splicing. PLoS Comput Biol. *6*(11):e1001016.

Cramer, P., Pesce, C. G., Baralle, F. E. &?Kornblihtt, A. R. (1997). Functional association between promoter structure and transcript alternative splicing. Proc. Natl Acad. Sci. USA. *94*, 11456–11460

Crispino JD, Blencowe BJ, Sharp PA. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. Science. *265*(5180):1866-9.

Crispino JD, Mermoud JE, Lamond AI, Sharp PA. (1996). Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. RNA. *2*(7):664-73.

David CJ, Manley JL. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev. *24*(21):2343-64.

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*, 364–368

De Conti L, Baralle M, Buratti E. (2013). Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA. *4*(1):49-60

de la Mata, M. & Kornblihtt, A. R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. Nature Struct. Mol. Biol. *13*, 973–980.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol. Cell. Biol. *26*, 5528–5543.

Disset A, Bourgeois CF, Benmalek N, Claustres M, Stevenin J, Tuffery-Giraud S. (2006). An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements. Hum Mol Genet. *15*(6):999-1013.

Dowhan DH, Hong EP, Auboeuf D, Dennis AP, Wilson MM, Berget SM, O'Malley BW. (2005). Steroid hormone receptor coactivation and

alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. Mol Cell. *17*(3):429-39.

Du H, Rosbash M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. Nature. *419*(6902):86-90.

Early P, Rogers J, Davis M, Calame K, Bond M, Wall R, Hood L. (1980). Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. Cell. *20*(2):313-9.

Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, Calarco JA, Wang X, Pan Q, O'Hanlon D, Kim PM, Wrana JL, Blencowe BJ. Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell. *46*(6):884-92.

Fairbrother WG, Yeh RF, Sharp PA, Burge CB. (2002). Predictive identification of exonic splicing enhancers in human genes. Science. *297*(5583):1007-13.

Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. (2013). RNA catalyses nuclear pre-mRNA splicing. Nature. *503*(7475):229-34.

Förch P, Puig O, Kedersha N, Martínez C, Granneman S, Séraphin B, Anderson P, Valcárcel J. (2000). The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. Mol Cell. *6*(5):1089-98.

Förch P, Puig O, Martínez C, Séraphin B, Valcárcel J. (2002). The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. EMBO J. *21*(24):6882-92.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A. *102*(45):16176-81.

Gabut M, Samavarchi-Tehrani P, Wang X, Slobodeniuc V, O'Hanlon D, Sung HK, Alvarez M, Talukder S, Pan Q, Mazzoni EO, Nedelec S, Wichterle H, Woltjen K, Hughes TR, Zandstra PW, Nagy A, Wrana JL, Blencowe BJ. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell. *147*(1):132-46.

Graveley BR. (2005). Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell. *123*(1):65-73.

Green MR. (1986). Pre-mRNA splicing. Annu Rev Genet. *20*:671-708.

Han H, Irimia M, Ross PJ, Sung HK, Alipanahi B, David L, Golipour A, Gabut M, Michael IP, Nachman EN, Wang E, Trcka D, Thompson T,

O'Hanlon D, Slobodeniuc V, Barbosa-Morais NL, Burge CB, Moffat J, Frey BJ, Nagy A, Ellis J, Wrana JL, Blencowe BJ. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature. *498*(7453):241-5.

Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc. Natl. Acad. Sci. USA. *99*, 16719–16724.

Hastings ML, Allemand E, Duelli DM, Myers MP, Krainer AR. (2007). Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). PLoS One. *2*(6):e538.

Heinrichs V, Bach M, Winkelmann G, Lührmann R. (1990). U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. Science. *247*(4938):69-72.

Hiller M, Zhang Z, Backofen R, Stamm S. (2007). Pre-mRNA secondary structures influence exon recognition. PLoS Genet. *3*(11):e204.

Hogg R, McGrail JC, O'Keefe RT. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. Biochem Soc Trans. *38*(4):1110-5.

Horowitz DS, Krainer AR. (1994). Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. Trends Genet. *10*(3):100-6.

Huang Q, Pederson T. (1999). A human U2 RNA mutant stalled in 3' end processing is impaired in nuclear import. Nucleic Acids Res. *27*(4):1025-31.

Huang Y, Li W, Yao X, Lin QJ, Yin JW, Liang Y, Heiner M, Tian B, Hui J, Wang G. (2012). Mediator complex regulates alternative mRNA processing via the MED23 subunit. Mol Cell. *45*(4):459-69.

Hung LH, Heiner M, Hui J, Schreiner S, Benes V, Bindereif A. (2008). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. RNA. *14*(2):284-96.

Iannone C, Valcárcel J. (2013). Chromatin's thread to alternative splicing regulation. Chromosoma. *122*(6):465-74.

Ip JY, Schmidt D, Pan Q, Ramani AK, Fraser AG, Odom DT, Blencowe BJ. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. Genome Res. *21*(3):390-401.

Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. (2003). Genome-

wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science. *302*(5653):2141-4.

Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002). Purification and characterization of native spliceosomes suitable for three- dimensional structural analysis. RNA. *8*, 426–439.

Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol. *14*(3):185-93.

Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT 2nd, Rosbash M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. *25*(23):2502-12.

Kim H, Klein R, Majewski J, Ott J. (2004). Estimating rates of alternative splicing in mammals and invertebrates. Nat Genet. *36*(9):915-6;

Kim, E., Goren, A. & Ast, G. (2008). Alternative splicing: current perspectives. Bioessays. *30*, 38–47.

Konarska MM, Grabowski PJ, Padgett RA, Sharp PA. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature. *313*(6003):552-7.

Konarska MM, Sharp PA. (1986). Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. Cell. *46*(6):845-55.

Konarska MM, Sharp PA. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell. *49*(6):763-74.

Konarska MM, Vilardell J, Query CC. (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. Mol Cell. *21*(4):543-53.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol.*14*(3):153-65.

Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Muñoz, M. J. & Nogués, G. (2004). Multiple links between transcription and splicing. RNA. *10*, 1489–1498

Krämer A, Utans U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. EMBO J. *10*(6):1503-9.

Lacadie SA, Rosbash M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell. *19*(1):65-75.

Lamond AI, Konarska MM, Grabowski PJ, Sharp PA. (1988). Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. Proc Natl Acad Sci U S A. *85*(2):411-5.

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell. *80*(1):155-65.

Legrain P, Seraphin B, Rosbash M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. Mol Cell Biol. *8*(9):3755-60.

Lerner MR, Boyle JA. Mount SM, Wolin SL, Steitz JA. (1980). Are snRNPs involved in splicing?. Nature. *283*(5743):220-4.

Liu, F., and Gong, C.X. (2008). Tau exon 10 alternative splicing and tauopa- thies. Mol Neurodegeneration. *3*, 8.

Long JC, Caceres JF. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem J. *417*(1):15-27.

López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. (2005). Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett. *579*(9):1900-3.

Lopez-Mejia IC, Vautrot V, De Toledo M, Behm-Ansmant I, Bourgeois CF, Navarro CL, Osorio FG, Freije JM, Stévenin J, De Sandre-Giovannoli A, Lopez-Otin C, Lévy N, Branlant C, Tazi J. (2011). A conserved splicing mechanism of the LMNA gene controls premature aging. Hum Mol Genet. *20*(23):4540-55.

Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell.*144*(1):16-26.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. (2010). Regulation of alternative splicing by histone modifications. Science. *327*(5968):996-1000.

Madhani HD, Guthrie C. (1994). Dynamic RNA-RNA interactions in the spliceosome. Annu Rev Genet. *28*:1-26

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science *298*, 2205–2208.

Makeyev EV, Zhang J, Carrasco MA, Maniatis T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell. *27*(3):435-48.

Maniatis T, Tasic B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature. *418*(6894):236-43.

Maniatis, T. & Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499–506

Maquat LE. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol. *5*(2):89-99.

Martinez-Contreras R, Cloutier P, Shkreta L, Fisette JF, Revil T, Chabot B. (2007). hnRNP proteins and splicing control. Adv Exp Med Biol. *623*:123-47.

Matera AG, Terns RM, Terns MP. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol. *8*(3):209-20.

Matlin AJ, Clark F, Smith CW. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol. *6*(5):386-98.

Merendino L, Guth S, Bilbao D, Martínez C, Valcárcel J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature. *402*(6763):838-41.

Merkin J, Russell C, Chen P, Burge CB. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science. 2012 Dec *21*;338(6114):1593-9.

Meyer M, Plass M, Pérez-Valle J, Eyras E, Vilardell J. (2011). Deciphering 3'ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. Mol Cell. *43*(6):1033-9.

Michel F, Costa M, Westhof E. (2009). The ribozyme core of group II introns: a structure in want of partners. Trends Biochem Sci. *34*(4):189-99.

Monsalve M, Wu Z, Adelmant G, Puigserver P, Fan M, Spiegelman BM. (2000). Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. Mol Cell. *6*(2):307-16.

Montes, M., Becerra, S., Sánchez-Álvarez, M. & Suñe, C. (2012). Functional coupling of transcription and splicing. Gene *501*, 104–117.

Moore MJ, Sharp PA. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature. *365*(6444):364-8.

Mordes D, Luo X, Kar A, Kuo D, Xu L, Fushimi K, Yu G, Sternberg P Jr, Wu JY. (2006). Pre-mRNA splicing and retinitis pigmentosa. Mol Vis. *12*:1259-71.

Mouaikel J, Verheggen C, Bertrand E, Tazi J, Bordonné R. (2002) Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. Mol Cell. *9*(4):891-901.

Mount SM, Pettersson I, Hinterberger M, Karmas A, Steitz JA. (1983). The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell. *33*(2):509-18.

Muñoz, M. J., de la Mata, M. & Kornblihtt, A. R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem. Sci. *35*, 497–504.

Nilsen TW, Graveley BR. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature. *463*: 457-463.

Nogués, G., Muñoz, M. J. & Kornblihtt, A. R. (2003). Influence of polymerase II processivity on alternative splicing depends on splice site strength. J. Biol. Chem. *278*, 52166–52171.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. *40*(12):1413-5.

Pan Q, Shai O, Misquitta C, Zhang W, Saltzman AL, Mohammad N, Babak T, Siu H, Hughes TR, Morris QD, Frey BJ, Blencowe BJ. (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol Cell. *16*(6):929-41.

Pandya-Jones A, Black DL. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA. *15*(10):1896-908.

Pertea M, Salzberg SL. (2010). Between a chicken and a grape: estimating the number of human genes. Genome Biol. *11*(5):206.

Pikielny CW, Rymond BC, Rosbash M. (1986). Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. Nature. *324*(6095):341-5.

Plass M, Codony-Servat C, Ferreira PG, Vilardell J, Eyras E. (2012) RNA secondary structure mediates alternative 3'ss selection in Saccharomyces cerevisiae. RNA. *18*(6):1103-15.

Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. (2007). Rapid, transcript-specific changes in splicing in response to environmental stress. Mol Cell. *27*(6):928-37.

Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature. *458*(7237):475-80.

Query CC, Moore MJ, Sharp PA. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev. *8*(5):587-97.

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente DA, Velasco G, Freije JM, Tubío JM, Royo R, Gelpí JL, Orozco M, Pisano DG, Zamora J, Vázquez M, Valencia A, Himmelbauer H, Bayés M, Heath S, Gut M, Gut I, Estivill X, López-Guillermo A, Puente XS, Campo E, López-Otín C. (2011). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet. *44(*1):47-52.

Raker VA, Mironov AA, Gelfand MS, Pervouchine DD. (2009). Modulation of alternative splicing by long-range RNA structures in Drosophila. Nucleic Acids Res. *37*(14):4533-44.

Raker VA, Plessel G, Lührmann R. (1999). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. EMBO J. *15*(9):2256-69.

Reed R, Maniatis T. (1984). Intron sequences involved in lariat formation during pre-mRNA splicing. Cell. *41*(1):95-105.

Reed R. (1989). The organization of 3' splice-site sequences in mammalian introns. Genes Dev. *3*(12B):2113-23.

Roberts, G. C., Gooding, C., Mak, H. Y., Proudfoot, N. J. & Smith, C. W. (1998). Co-transcriptional commitment to alternative splice site selection. Nucleic Acids Res. *26*, 5568–5572.

Roca X, Krainer AR, Eperon IC. (2013). Pick one, but be quick: 5' splice sites and the problems of too many choices. Genes Dev. *27*(2):129-44.

Roca X, Krainer AR. (2009). Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. Nat Struct Mol Biol. *16*(2):176-82.

Rosenfeld MG, Amara SG, Roos BA, Ong ES, Evans RM. (1981). Altered expression of the calcitonin gene associated with RNA polymorphism. Nature. *290*(5801):63-5.

Ruskin B, Green MR. (1985). Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. Nature. *317*(6039):732-4.

Ruskin B, Krainer AR, Maniatis T, Green MR. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell. *38*(1):317-31.

Ruskin B, Zamore PD, Green MR. (1998). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell. *52*(2):207-19.

Sakabe, N. J. & de Souza, S. J. (2007). Sequence features responsible for intron retention in human. BMC Genomics *8*, 59.

Schwerk C, Schulze-Osthoff K. (2005). Regulation of apoptosis by alternative pre-mRNA splicing. Mol Cell. *19*(1):1-13.

Seraphin B, Rosbash M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell. *59*(2):349-58.

Sharp PA. (2005). The discovery of split genes and RNA splicing. Trends Biochem Sci. *30*(6):279-81.

Shi Y, Reddy B, Manley JL. (2006). PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. Mol Cell. *23*(6):819-29.

Srinivasan K, Shiue L, Hayes JD, Centers R, Fitzwater S, Loewen R, Edmondson LR, Bryant J, Smith M, Rommelfanger C, Welch V, Clark TA, Sugnet CW, Howe KJ, Mandel-Gutfreund Y, Ares M Jr. (2005). Detection and measurement of alternative splicing using splicing-sensitive microarrays. Methods. *37*(4):345-59.

Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H. (2005). Function of alternative splicing. Gene. *344*:1-20.

Stevens SW, Ryan DE, Ge HY, Moore RE, Young MK, Lee TD, Abelson J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol 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. *9*, 66–77.

Sun JS, Manley JL. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. Genes Dev. *9*(7):843-54.

Tejedor JR, Valcárcel J. (2010). Gene regulation: Breaking the second genetic code. Nature. *465*(7294):45-6.

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigó R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. *22*(9):1616-25.

Treisman R, Orkin SH, Maniatis T. (1983). Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature. *302*(5909):591-6.

Wahl MC, Will CL, Lührmann R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell. *136*(4):701-18.

Wang C, Chua K, Seghezzi W, Lees E, Gozani O, Reed R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. Genes Dev. *12*(10):1409-14.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature. *456*(7221):470-6.

Wang GS, Cooper TA. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet. *8*(10):749-61.

Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. PLoS Biol. *8*(10):e1000530.

Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. (2004). Systematic identification and analysis of exonic splicing silencers. Cell. *119*(6):831-45.

Wassarman DA, Steitz JA. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science. *257*(5078):1918-25.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. *453*(7199):1239-43.

Will CL, Lührmann R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol. *13*(3):290-301.

Wu S, Romfo CM, Nilsen TW, Green MR. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature. *402*(6763):832-5.

Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. *478*(7367):64-9.

Zamore PD, Patton JG, Green MR. (1992). Cloning and domain structure of the mammalian splicing factor U2AF.Nature. *355*(6361):609-14.

Zhang MQ. (1998). Statistical features of human exons and their flanking regions. Hum Mol Genet. *7*(5):919-32.

Zhang XH, Chasin LA. (2004). Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev. *18*(11):1241-50.

Zorio DA, Blumenthal T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. Nature. *402*(6763):835-8.

PREVIEW

NEWS AND VIEWS

GENE REGULATION: BREAKING THE SECOND GENETIC CODE

J. Ramón Tejedor [1,2*] and Juan Valcárcel [1-3]

[1]Centre de Regulació Genòmica, Dr. Aiguader 88, 08003 Barcelona, Spain.

[2]Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain.

[3]Institució Catalana de Recerca i Estudis Avançats, Passeig Lluis Companys 23, 08010 Barcelona, Spain

Correspondence to: juan.valcarcel@crg.eu

Keywords: pre-mRNA splicing, Spliceosome, alternative splicing, regulatory networks, splicing code.

Nature 465, 45–46 (06 May 2010)
doi:10.1038/465045a
Published online 05 May 2010

http://www.nature.com/nature/journal/v465/n7294/full/465045a.html

Article located in USB Drive \Thesis documents\Preview

Tejedor JR, Valcárcel J. Gene regulation: Breaking the second genetic code. Nature. 2010 May 6; 465(7294): 45-6. DOI: 10.1038/465045a

OBJECTIVES

I. Develop high-throughput methods to systematically identify factors involved in alternative splicing regulation.

II. Identify genome-wide regulators of Fas/CD95 alternative splicing, and the possible links between splicing and other relevant celular processes.

III. Study potential mechanisms of alternative splicing regulation derived from the results of the systematic approaches described above.

RESULTS

CHAPTER I:

FUNCTIONAL SPLICING NETWORK REVEALS EXTENSIVE REGULATORY PLASTICITY OF THE CORE SPLICEOSOMAL MACHINERY

Panagiotis Papasaikas[1*], J. Ramón Tejedor [1,2*], Luisa Vigevani [1,2] and Juan Valcárcel [1-3]

[1]Centre de Regulació Genòmica, Dr. Aiguader 88, 08003 Barcelona, Spain.

[2]Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain.

[3]Institució Catalana de Recerca i Estudis Avançats, Passeig Lluis Companys 23, 08010 Barcelona, Spain

* The first two authors contributed equally to this work.

Correspondence to: juan.valcarcel@crg.eu

Keywords: pre-mRNA splicing, Spliceosome, alternative splicing, regulatory networks, anti-tumor drugs.

The data included in this chapter have been submitted for publication
Manuscript located in USB Drive \Thesis documents\ChapterI_Splicing_networks\Manuscript

Papasaikas P, Tejedor JR, Vigevani L, Valcárcel J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. Mol Cell. 2015 Jan 8; 57(1): 7-22. DOI: 10.1016/j.molcel.2014.10.030

CHAPTER II:

GENOME-WIDE IDENTIFICATION OF FAS/CD95 ALTERNATIVE SPLICING REGULATORS REVEALS LINKS WITH IRON HOMEOSTASIS

J. Ramón Tejedor [1,2*], Panagiotis Papasaikas[1*] and Juan Valcárcel [1-3]

[1]Centre de Regulació Genòmica, Dr. Aiguader 88, 08003 Barcelona, Spain.

[2]Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain.

[3]Institució Catalana de Recerca i Estudis Avançats, Passeig Lluis Companys 23, 08010 Barcelona, Spain

Correspondence to: juan.valcarcel@crg.eu

Keywords: Alternative splicing, Fas/CD95, iron, RNA binding, SR proteins,

The data included in this chapter have been submitted for publication
Manuscript located in USB Drive \Thesis documents\ChapterII_GW_screen\Manuscript

Tejedor JR, Papasaikas P, Valcárcel J. Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. Mol Cell. 2015 Jan 8; 57(1): 23-38. DOI: 10.1016/j.molcel.2014.10.029

GENERAL DISCUSSION

Since the discovery of the RNAi mechanism by Andrew Fire and Craig Mello (Fire *et al.*, 1998), the acquisition of information in functional genomics has expanded exponentially. High-throughput genome-wide screens based on the RNAi silencing technology have proved to be extremely valuable for uncovering novel insights into multiple biologically relevant questions. Additionally, the explosion of bioinformatics analyses, including the recent study of biological networks, made it possible to interpret biological data with unprecedented depth. The synergy between innovative high throughput screening technologies and original computational analyses have allowed us to go beyond state-of-the-art tools for splicing data interpretation, and shed light on several issues with potential implications for future studies in the field. In the following section we will comment on several remarkable topics that were not covered in the discussion of the previous manuscripts.

Chromatin versus Splicing machineries

Results obtained in our genome-wide screen suggested a potential role of chromatin components and transcription factors in Fas/CD95 alternative splicing regulation. In fact, we observed a significant effect of some factors previously identified in the process of transcriptional elongation. Among them, transcriptional coactivators (SKIIP), topoisomerase subunits (TOP2A and TOPBP1), and some chromatin remodeling factors (SWI/SNF or the histone methyl transferase SETD2), showed consistent effects in Fas/CD95 alternative splicing, suggesting the possibility that

modulation of transcription elongation influences Fas/CD95 splicing, as previously reported for other genes (reviewed by Kornblihtt, 2013).

Functional network connections involving chromatin factors are characterized either by interactions between multiple chromatin components (MECP2 and EP300; DNMT1 and SMARCA2; EIF2C2 and KAT2B;), or by the interaction between chromatin and splicing factors in a rather peripheral manner (P29 and SMARCA4; TIA1 and TET1), with very few connections with core spliceosomal components. These peripheral connections may reflect indirect effects on transcriptional kinetics rather than direct effects of the chromatin factors in the splicing process. Nevertheless, two chromatin candidates showed a strong functional interaction with core spliceosomal components: the histone methyl-transferase EHMT2, and the bromodomain protein BRD4.

The strong functional correlation between EHMT2 and SNU114 (U5-116KD), a GTPase that plays a key role in the activation of the spliceosome (Brenner *et al.*, 2005), suggests a direct link between these factors. The methyl-transferase activity of EHMT2 could modulate the mono or di-methylation status of different histones (Tachibana *et al.*, 2001), as H3K9, facilitating the recruitment of protein partners like HP1γ, previously involved in transcription and alternative splicing modulation (Saint-Andre *et al.*, 2011). Densely methylated chromatin regions could affect the processivity of RNA pol II and enhance the recruitment of the tri-snRNP particle to nascent pre-mRNA transcripts of specific gene loci. Yet another possibility would be that the methyl transferase activity of EHMT2 post-translationally modifies splicing factors, including SNU114

Impairment of the methylation status of the GTPase caused by EHMT2 knock down could lead to splicing defects in a subset of introns. Further experiments will be required to discern between these and other possible mechanisms of splicing regulation.

An intriguing anti-correlation was also detected between the bromodomain containing protein BRD4 with BRR2 (U5-200KD) -the helicase that unwinds the U4/U6 RNA interactions during the activation of the spliceosome (Laggerbauer *et al.*, 1998; Maeder *et al.*, 2009)-. BRD4 reads acetylated histones (Wu *et al.*, 2013) and is typically involved in gene activation (Yang *et al.*, 2005) and phosphorylation of the serine 2 of the RNA polymerase II (Zhang *et al.*, 2012). These features suggest a chromatin scenario in which fast transcriptional elongation could compete with proper assembly or activation of the splicing machinery, resulting in exon skipping as predicted by the kinetic model of co-transcriptional splicing regulation (Kornblihtt *et al.*, 2013). Another possibility would be that BRD4 directly or indirectly represses U5-200KD promoter activity.

The remaining pairs of functional interactions are rather peripheral, and could have an impact in different aspects of gene regulation, such as chromatin remodeling (MECP2-EP300; DNMT1-SMARCA2 and EIF2C2-KAT2B), RNA pol II elongation rates (DNMT1-SMARCA2), Spliceosome dynamics (P29-SMARCA4) or DNA-based epigenetic modifications (TET1-TIA1). The anticorrelation between the DNA methyl binding protein MECP2– generally associated with gene repression (Lewis *et al.*, 1992) - and the histone acetyl transferase EP300 – usually involved in gene activation (Eckner *et al.*, 1994)-, suggests that the

antagonistic activities of these factors are also at the basis of their effects on splicing regulation.

The splicing network revealed a significant interaction between the DNA methyl transferase DNMT1 and BRM (SMARCA2), a component of the SWI/SNF chromatin-remodeling complex. Previous studies have reported the role of BRM in alternative splicing regulation (Batsché *et al.*, 2006), suggesting a connection between the chromatin landscape and the phosphorylation / elongation status of the transcribing RNA polymerase through these regions. The positive interaction between BRM and DNMT1 indicates that a parallel, or even a convergent scenario of chromatin and DNA modifications could result in coherent splicing effects. In fact, reports focusing on the role of retinoblastoma tumor suppressor protein Rb in E2F-mediated transcriptional repression (reviewed in Ferreira *et al.*, 2001) suggested a direct implication for DNMT1, BRM and/or other histone deacetylase complexes in specific promoters. If similar complexes would be present in intragenic regions, they could provide the mechanistic basis for changing in splicing detected in the network. It would have been interesting, in this regard, to count on information about the effects of Rb depletion, which unfortunately was not included in our study. Interestingly, mutations in DNMT1 have been identified in Acute Lymphocytic Leukemia patients (Dolnik *et al.*, 2012), and it would be of interest to determine whether, in addition to altered DNA methylation-mediated effects on transcription or genome stability, aberrant splicing patterns could also arise, perhaps related to the effects the recurrent mutations in spliceosomal components described for Chronic Lymphocytic Leukemia (Quesada *et al.*, 2011).

Some reports have identified a direct role for dsRNAs in the modulation of alternative splicing mediated by the RISC components AGO1 and AGO2 (EIF2C2) -involved in gene silencing- and an increase in the H3K9me2 and H3K27me3 histone marks in the areas flanking the alternative spliced region (Alló *et al.*, 2009; Ameyar-Zazoua *et al.* 2012). Surprisingly, the functional association identified in our network, suggested that AGO2 has a positive correlation with the histone acetyltransferase activity of PCAF (KAT2B), which has been previously associated with chromatin remodeling and gene activation (Ogryzko *et al.*, 1996). Moreover, a similar study to that of Allo *et al.* (2012) found that dsRNA-directed alternative splicing of the SMN2 and dystrophin genes required AGO2, but does not involve the associated chromatin changes (H3K9me2 and H3K27me3), suggesting that additional mechanisms of splicing regulation mediated by induced dsRNA could occur (Liu *et al.*, 2012; Li, 2013).

Another functional connection with potential mechanistic implications is established between the spliceosome factor P29 and a component of the SWI/SNF SMARCA4 (BRG1). P29 has been isolated in spliceosomal complex C, and very little is known about its function (Jurica *et al.*, 2002). A preceding study from Tyagi *et al.*, (2009) identified a role for the two subunits of the SWI/SNF complex in the regulation of alternative splicing of Drosophila cells. As a striking conclusion, they observed that traces of the SWI/SNF complex were not associated with chromatin, but rather with the RNA fraction. This result opens up the possibility that the presence of the SWI/SNF complex contains activities that

directly affect late steps factors of the splicing reaction. Interestingly, the fact that SMARCA4 and SMARCA2 are not associated through the same splicing events suggests that they could play independent roles in splice site choice.

And finally, the significant correlation between TIA1 and TET1 proteins suggests an intriguing mechanistic insight. TIA1 enhances exon inclusion by recruiting U1 snRNP to 5' splice sites (Förch *et al.*, 2000; 2002; Wang *et al.*, 2010), while TET proteins play a role in active DNA demethylation by mediating the conversion of the modified genomic base 5-methylcytosine (5mC) into 5-hydroxy-methylcytosine (5hmC). A recent report by Khare *et al.* (2012) showed that the intron-exon boundaries of synaptic genes are differentially marked by this epigenetic modification, especially at their 5' splice sites. The functional link revealed by the network could be based upon coordinated regulation between cytosine hydroxymethylation by TET proteins and the recruitment of TIA-1 to specific 5' splice sites, an interesting lead for further studies.

Despite the novel insights discussed above, it is worth to mention that the splicing network relies in the estimations obtained from a small subset of splicing events. Whether the functional connections between chromatin and splicing factors are representative of the 35 splicing events selected in our study or are valid across a much wider range of alternative splicing regulation remains to be elucidated. However, the reiterative subsampling of events and the resulting 10.000 networks suggest that a core of interactions are likely to be operative in a large range of splice site selection processes.

Functional connection between IK-SMU1

The identification of the "core" functional components as important contributors to splicing regulation, both through the genome-wide screen and the network analysis, highlight interesting areas for further research. In particular, we decided to follow up the IK-SMU1 story due its extremely strong connection in the most conserved core spliceosomal network. The functions of these factors in the spliceosome remain poorly understood. IK was previously characterized as a nuclear protein with alternative RED repeats (Arginine, Glutamic and Aspartic acid, also known as hRED) (Assier *et al.*, 1999). Initial studies concerning the function of IK and SMU1 in Caenorhabditis elegans suggested a potential role of these factors in alternative splicing regulation (Spartz *et al.*, 2004). Interestingly, IK and SMU1 can interact directly and IK protects SMU1 from degradation (Spartz *et al.*, 2004). Further studies proposed a role for IK-SMU1 in the selection of alternative 5' splice sites at late steps of spliceosome assembly (Dassah *et al.*, 2009), with potential implications in Arabidopsis thaliana development (Chung *et al.*, 2009). Our genome-wide data supports this result, although in our case the most overrepresented categories were cassette exons and intron retention events. A possible explanation for these differences is the different prevalence and mechanisms behind different types of alternative splicing events in different organisms. Thus, vertebrates are likely to relay more on exon recognition compared to invertebrates or plants (Ast, 2004).

The developmental defects observed in Arabidopsis mutants suggested a more penetrant phenotype for strains defective in IK,

than for strains defective in SMU1. Interestingly, the apoptosis phenotype observed under conditions of IK-SMU1 knockdown in our assays also indicates a stronger impact of IK deficiency in the activation of cell death. As it was previously shown, IK could stabilize SMU1, and depletion of the former could result in a combined effect associated with more severe cell phenotypes (Spartz *et al.*, 2004; Chung *et al.*, 2009). In-depth analysis of the genes affected upon IK or SMU1 knockdown will help to elucidate possible targets involved in the enhancement of apoptosis. The fact that both factors could interact and stabilize each other in a complex represents the most logic scenario for the tight connection observed in the final network, although we cannot rule out additional mechanisms of action for each of the proteins. Comprehensive proteomic studies identified specific "time" and "location" for these factors within the spliceosome assembly pathway (Hegele *et al.*, 2012; Agafonov *et al.*, 2011). IK and SMU1 form part of the spliceosomal complex B, and they are immediately released after the activation of the spliceosome in the B act complex. The constrained location of these factors within this specific time frame suggests their potential implication in the activation of the spliceosome. The protein-protein interaction network established by IK is greater than that of SMU1. IK can interact with some components of the core splicing machinery involved in spliceosome activation (SNU114), again consistent with a pivotal role for this factor in the transition from B to B act complex (Hegele *et al.*, 2012). In fact our functional network identified stronger connections between this factor and the core spliceosomal components, while SMU1 showed interactions with more peripheral factors or even with components from other snRNPs, suggesting a more fundamental role for IK function in spliceosome activation,

while SMU1 could work at the level of complex B recruitment before their joint release later on in the reaction. A detailed characterization of IK and SMU1 function at the core of the B complex will provide important clues as to their specific roles and how these can be harnessed to modulate splice site selection.

Recent studies have characterized the precise intermolecular rearrangements required for B complex activation, and the implication of particular domains of BRR2 and PRPF8 – two components located at the heart of the catalytic core- in the regulation of this process (Mozzafari-Jovin *et al.*, 2012). A direct substrate competition mediated by the interaction of an intrinsically unstructured element of the PRPF8 Jab1/MPN domain impedes BRR2 U4/U6 unwinding activity, representing a safe lock mechanism prior to spliceosome activation (Mozzafari-Jovin *et al.*, 2013). Spliceosome activation caused by the GTPase SNU114 induces conformational changes in PRPF8 and help to release BRR2 from the repressed state (Brenner *et al.*, 2005). Despite the crucial role of these factors, those studies have been focused in the in vitro characterization of specific protein domains in *in vitro* minimal systems. Other tri-snRNP components are likely to also contribute, and it is conceivable that IK/SMU1 could trigger a domino effect whereby initial interactions between IK and SNU114 could control the spliceosome activation cascade, including the release of the BRR2 unwinding activity. Depending on the levels or activity of IK/SMU1, the kinetics of this cascade could be modified, affecting the timing of the release of the safe-lock mechanism mediated by BRR2-PRPF8 interactions and therefore establishing differences in splice site selection between competing sites.

Further applications for the splicing network technology

In our network study, we have developed a robust and systematic approach to explore alternative splicing regulation. The systematic knockdown of spliceosomal components designed in our custom siRNA library allow us to discover regulators of novel biologically relevant exons in a fast and accurate manner, providing a valuable tool for the splicing community. The use of this methodology is not only restricted to the study of functional connections between spliceosomal components, but it could be helpful to identify splicing regulators of any other alternatively spliced exon of interest. As the RNA extraction and cDNA library generation methods allowed us to extract the information of the full transcriptome, we could essentially analyze any endogenous splicing event –or transcription or polyadenylation or RNA decay- in a given cell line, and could conceivably be adapted also for the study of translational control. The mid to high throughput processivity of the capillary electrophoresis platform is a convenient method for other validations, and could be adapted for overexpression, mutational or drug based screens. Despite the technical adjustments required for data acquisition, the generation of other functional networks could help to understand in depth each of the layers implicated in the regulation of gene expression, as well as the interconnections between those processes involved in the faithful control of the transcriptome in a given cell type.

The power of the splicing network allowed us to reveal novel connections between different physiological responses and the spliceosome machinery, including the link between iron homeostasis and the splicing factor SRSF7 (see below). Furthermore, preliminary experiments using DNA damage-inducing agents or topoisomerase inhibitors have provided us with interesting leads for future studies (data not shown). Functional connections between those treatments (camptothecin, etoposide and doxorrubicin) and splicing factors implicated in transcriptional elongation, such as SRSF2 (Lin *et al.*, 2008) suggest a potential role in the modulation of the cotranscriptional splicing mechanism. Along these lines, most of the connections were peripheral, and the protein acetyl transferase KAT5 – which has been previously involved in the modulation of the DNA damage response (Kaidi and Jackson., 2013) - was a recurrent node in these interactions, suggesting that cell responses to DNA lesions, caused either by direct DNA damage agents or by polymerase stalling, could have a global impact in the regulation of alternative splicing.

Comparisons with previous genome wide screens

Previous genome-wide studies carried out using chromatic reporter systems (Moore *et al.*, 2010) identified several regulators for MCL1 and BCLX alternative splicing. Back to back comparisons with our genome-wide screen indicated the consistently significant capacity of variations in the levels/activity of core splicing components in the modulation of splice site choice. This realization prompted us to

develop the tools for setting up our splicing network approach. However, there was limited overlap between the hits obtained for FAS and for MCL1/BCLX splicing in the two screens in other categories of genes, suggesting divergent regulation by other pathways (data not shown). For example, for MCL1 and BCLX, cell cycle components like AURKA kinase were conspicuous contributors to splicing regulation, while transcription and chromatin remodeling, as well as links with other cellular pathways e.g. iron or energy homeostasis, were prominent hits in FAS alternative splicing. Some of these differences could be due to the type of splicing event analyzed, e.g. FAS is a cassette exon, while BCLX is an alternative 5' splice site, or even to different biases of the experimental methodologies. In fact, our system recapitulates in a faithful manner the endogenous regulatory context of the splicing event, explaining in part the abundant connections with chromatin obtained in our analysis. Moreover, the detection of endogenous transcripts does not rely in other secondary processes -as translation or protein stability- for the interpretation of the data, and thus, represents a direct and robust approach to investigate splicing regulation in a specific physiological context.

Other advantages include the generation of cDNA libraries that can be used to interrogate any endogenous splicing event of interest present in HeLa cells in a fast and accurate manner. Furthermore, the use of barcoded primers coupled with deep sequencing technology could be extended to any other process implicated in the regulation of gene expression in the cell. And the high-throughput potential of the multiplex barcoded approach could be also employed to perform high content drug-based screens or any

combination of biological conditions of interest, allowing relatively low cost analysis of thousands of samples in parallel.

A novel mechanism for iron mediated regulation of gene expression

The intriguing dual role of ACO1 – as an RNA binding protein or as a metabolic enzyme, depending on iron availability- has coped the attention of the iron homeostasis field. Elegant previous work established the role of iron in the modulation of ACO1 RNA binding properties (Gray *et al.*, 1994; Hentze *et al.*, 2010). Interestingly, our screen highlighted a potential role of ACO1 in the modulation of FAS alternative splicing and initially implied the possibility of a direct link between this factor and the splicing machinery. However, our experimental work coupled with bioinformatic predictions of ACO1 binding sites within the FAS gene suggested that the effect of ACO1 in alternative splicing is indirect. Interestingly, our results indicated that, rather than the conventional model based on modulation of ACO1 RNA binding properties, alternative splicing changes were correlated with the intracellular levels of iron (see figure II.25). Further experiments identified SRSF7 as one key factor implicated in this process, and also the role of iron in the modulation of SRSF7 RNA binding properties, thus providing a novel mechanism for alternative splicing regulation by iron homeostasis.

The molecular mechanism by which iron regulates SRSF7-specific splicing patterns could be potentially extended to other RNA binding proteins or transcription factors within the cell. While

SRSF7 was the main candidate amongst several splicing factors containing Zn finger motifs, we cannot rule the possibility that other proteins can also play a role. In fact, some splicing regulators not included in the initial network analysis contain multiple zinc fingers motifs (i.e MBLN1, involved in cell differentiation) that may be affected by iron, as SRSF7 Zinc knuckle is. Further experiments along these lines may shed light on additional effects of iron homeostasis in splicing regulation.

A recent report of RNA binding motifs of many RNA binding proteins identified the putative RNA binding elements of hundreds of factors implicated in different aspects of RNA metabolism (Ray *et al.*, 2013). Our study suggests that modulation of RNA binding efficiency or specificity can arise from variations in the cellular ionic concentrations. Thus, it would be interesting, for example, to use the extensive and systematic RNA compete approach of Ray *et al.* under conditions of different ionic conditions.

Concluding remarks

We believe that the methodology and results summarized in this thesis provide a significance advance for transcriptomics and splicing analyses. In the Fas/CD95 genome-wide screen project, we exploited the single molecule analysis capacities of next generation sequencing technology, and designed a multiplexed barcoded system that allowed us to handle and quantitatively evaluate an unprecedented large number of conditions in parallel compared to previous studies (Smith *et al.*, 2009). This design may find applications in the study of many other molecular or cellular

processes for both basic research and the biotech industry. Another key advantage of our design over reporter-based screens is the direct analysis of endogenous transcripts, offering unique opportunities to capture physiological physiological responses within the cell.

After extensive rounds of hit validation, we provide evidence for the prevalent function of core components of the spliceosome as potential regulators of alternative splicing. Furthermore, we connected physiologically relevant responses -as the modulation of iron homeostasis- with alternative splicing in ways not predicted by previous knowledge of iron homeostasis and RNA metabolism. The splicing network analysis allowed us to delineate the potential splicing regulator (SRSF7) involved in iron-dependent splicing response and identify a plausible regulatory mechanism based upon the modulation of SRSF7 RNA binding properties by iron. These observations provide new models to understand the effects of iron and other cellular perturbations on RNA metabolism.

To conclude, we provide proof of principles of the applicability of novel experimental and computational approaches that can help to decipher novel mechanisms of alternative splicing regulation generally applicable to cultured cells in a variety of physiological, pathological or therapeutic contexts.

References

Agafonov DE, Deckert J, Wolf E, Odenwälder P, Bessonov S, Will CL, Urlaub H, Lührmann R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. Mol Cell Biol. *31*(13):2667-82.

Alló M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyras E, Elela SA, Klinck R, Chabot B, Kornblihtt AR. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing.Nat Struct Mol Biol. *16*(7):717-24.

Ameyar-Zazoua M, Rachez C, Souidi M, Robin P, Fritsch L, Young R, Morozova N, Fenouil R, Descostes N, Andrau JC, Mathieu J, Hamiche A, Ait-Si-Ali S, Muchardt C, Batsché E, Harel-Bellan A. (2012). Argonaute proteins couple chromatin silencing to alternative splicing. Nat Struct Mol Biol. *19*(10):998-1004.

Assier E, Bouzinba-Segard H, Stolzenberg MC, Stephens R, Bardos J, Freemont P, Charron D, Trowsdale J, Rich T. (1999). Isolation, sequencing and expression of RED, a novel human gene encoding an acidic-basic dipeptide repeat. Gene. *230*(2):145-54.

Ast G. (2004). How did alternative splicing evolve?. Nat Rev Genet. *5*(10):773-82.

Batsché E, Yaniv M, Muchardt C.(2006)The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nat Struct Mol Biol. *13*(1):22-9.

Brenner TJ, Guthrie C. (2005). Genetic analysis reveals a role for the C terminus of the Saccharomyces cerevisiae GTPase Snu114 during spliceosome activation. Genetics. *170*(3):1063-80.

Chung T, Wang D, Kim CS, Yadegari R, Larkins BA. (2009). Plant SMU-1 and SMU-2 homologues regulate pre-mRNA splicing and multiple aspects of development. Plant Physiol. *151*(3):1498-512.

Dassah M, Patzek S, Hunt VM, Medina PE, Zahler AM. (2009). A genetic screen for suppressors of a mutated 5' splice site identifies factors associated with later steps of spliceosome assembly. Genetics. *182*(3):725-34.

Dolnik A, Engelmann JC, Scharfenberger-Schmeer M, Mauch J, Kelkenberg-Schade S, Haldemann B, Fries T, Krönke J, Kühn MW, Paschka P, Kayser S, Wolf S, Gaidzik VI, Schlenk RF, Rücker FG,

Döhner H, Lottaz C, Döhner K, Bullinger L. (2012). Commonly altered genomic regions in acute myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. Blood. *120*(18):e83-92.

Eckner R, Ewen ME, Newsome D, Gerdes M, DeCaprio JA, Lawrence JB, Livingston DM. (1994). Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev. *8*(8):869-84.

Ferreira R, Naguibneva I, Pritchard LL, Ait-Si-Ali S, Harel-Bellan A. (2001). The Rb/chromatin connection and epigenetic control: opinion. Oncogene. *20*(24):3128-33.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. *391*(6669):806-11.

Förch P, Puig O, Kedersha N, Martínez C, Granneman S, Séraphin B, Anderson P, Valcárcel J. (2000). The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. Mol Cell. *6*(5):1089-98.

Förch P, Puig O, Martínez C, Séraphin B, Valcárcel J. (2002). The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. EMBO J. *21*(24):6882-92.

Gray NK, Hentze MW. (1994). Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. EMBO J. *13*(16):3882-91.

Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Lührmann R, Stelzl U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell. 2012 Feb *24*;45(4):567-80.

Hentze MW, Muckenthaler MU, Galy B, Camaschella C. (2010). Two to tango: regulation of Mammalian iron metabolism. Cell. *142*(1):24-38.

Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA. *8*(4):426-39.

Kaidi A, Jackson SP. (2013). KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling. Nature. *498*(7452):70-4.

Khare T, Pai S, Koncevicius K, Pal M, Kriukiene E, Liutkeviciute Z, Irimia M, Jia P, Ptak C, Xia M, Tice R, Tochigi M, Moréra S, Nazarians A, Belsham D, Wong AH, Blencowe BJ, Wang SC, Kapranov P, Kustra R, Labrie V, Klimasauskas S, Petronis A. (2012). 5-hmC in the brain is

abundant in synaptic genes and shows differences at the exon-intron boundary. Nat Struct Mol Biol. *19*(10):1037-43.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol.*14*(3):153-65.

Laggerbauer B, Achsel T, Lührmann R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proc Natl Acad Sci U S A. *95*(8):4188-92.

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell. *69*(6):905-14.

Li LC. (2013). Chromatin remodeling by the small RNA machinery in mammalian cells. Epigenetics. *22*;9(1).

Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat Struct Mol Biol. *15*(8):819-26.

Liu J, Hu J, Corey DR. (2012). Expanding the action of duplex RNAs into the nucleus: redirecting alternative splicing. Nucleic Acids Res. *40*(3):1240-50.

Maeder C, Kutach AK, Guthrie C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. Nat Struct Mol Biol. *16*(1):42-8.

Moore MJ, Wang Q, Kennedy CJ, Silver PA. (2010). An alternative splicing network links cell-cycle control to apoptosis. Cell. *142*(4):625-36.

Mozaffari-Jovin S, Santos KF, Hsiao HH, Will CL, Urlaub H, Wahl MC, Lührmann R. (2012). The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. Genes Dev. *26*(21):2422-34.

Mozaffari-Jovin S, Wandersleben T, Santos KF, Will CL, Lührmann R, Wahl MC. (2013). Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. Science. *341*(6141):80-4.

Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell. *87*(5):953-9.

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M,

Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente DA, Velasco G, Freije JM, Tubío JM, Royo R, Gelpí JL, Orozco M, Pisano DG, Zamora J, Vázquez M, Valencia A, Himmelbauer H, Bayés M, Heath S, Gut M, Gut I, Estivill X, López-Guillermo A, Puente XS, Campo E, López-Otín C. (2011). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet. *44*(1):47-52.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, Na H, Irimia M, Matzat LH, Dale RK, Smith SA, Yarosh CA, Kelly SM, Nabet B, Mecenas D, Li W, Laishram RS, Qiao M, Lipshitz HD, Piano F, Corbett AH, Carstens RP, Frey BJ, Anderson RA, Lynch KW, Penalva LO, Lei EP, Fraser AG, Blencowe BJ, Morris QD, Hughes TR. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature. *499*(7457):172-7.

Saint-André V, Batsché E, Rachez C, Muchardt C. (2011). Histone H3 lysine 9 trimethylation and HP1? favor inclusion of alternative exons.Nat Struct Mol Biol.*18*(3):337-44.

Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, Roth FP, Giaever G, Nislow C. (2009). Quantitative phenotyping via deep barcode sequencing. Genome Res. *19*(10):1836-42.

Spartz AK, Herman RK, Shaw JE. (2004). SMU-2 and SMU-1, Caenorhabditis elegans homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice. Mol Cell Biol. *24*(15):6811-23.

Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J Biol Chem. *276*(27):25309-17.

Tyagi A, Ryme J, Brodin D, Ostlund Farrants AK, Visa N. (2009). SWI/SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing. PLoS Genet. *5*(5):e1000470.

Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. PLoS Biol. *8*(10):e1000530.

Wu SY, Lee AY, Lai HT, Zhang H, Chiang CM. (2013). Phospho switch triggers Brd4 chromatin binding and activator recruitment for gene-specific targeting. Mol Cell. *49*(5):843-57.

Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, Zhou Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol Cell. *19*(4):535-45.

Zhang W, Prakash C, Sum C, Gong Y, Li Y, Kwok JJ, Thiessen N, Pettersson S, Jones SJ, Knapp S, Yang H, Chin KC. (2012). Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells. J Biol Chem. *287*(51):43137-55.

ACKNOWLEDGEMENTS

Hace mas de seis años que comencé mis andanzas por el mundo de la ciencia y a lo largo de este tiempo me he encontrado con cientos de personas en el camino que, de alguna u otra manera, han contribuido a dar cuerpo a esta tesis, así como a forjarme como persona. Puede que el azar nos haya situado en la misma senda, pero somos nosotros con nuestras acciones los que vamos adaptándonos a nuestro destino. Sin embargo, aparte de suerte, hay que saber jugar bien las cartas para llegar al final de la partida. Y afortunadamente he estado rodeado de grandes maestros que me han enseñado todas sus bazas ganadoras.

Antes de mostraros mis cartas, me gustaría agradecer a mis padres, Ramón y María del Carmen, y a toda mi familia en general, por todo el apoyo que me han brindado durante este tiempo, y por haber estado preocupados por mi en cada momento que he pasado lejos de ellos. También a mi hermana Sonia, quien la pobre ha tenido que aguantarme durante este tiempo bajo el mismo techo. Quiero que sepáis que nada de esto hubiera sido posible sin vuestra ayuda.

Trio de ases

Puede que mi destino inicial se viera influenciado por la suerte del principiante, ya que tuve la fortuna de contar de primera mano con la ayuda de Alejandro López-Soto, Carlos López-Otín y Juan Valcárcel. A Alejandro le debo toda mi gratitud por introducirme en el mundo de la ciencia y por enseñarme las bases que me han servido para hacer mis pinitos en la carrera investigadora. A

Carlos, por toda su ayuda y sus sabios consejos, así como el impulso inicial que determino mi destino en el CRG. También por su perseverancia como científico, sin duda un modelo a seguir.

Y a Juan, por haber confiado en mi desde el primer momento y haberme dado la oportunidad de hacer el doctorado en este gran oasis llamado CRG. A pesar de los mil y un obstáculos que hemos tenido que superar durante estos arduos proyectos, siempre ha sabido actuar con templanza y a enseñarme que no importa la cantidad de veces que caigas si al final eres capaz de levantarte. Su constante motivación y entrega han hecho posible que saliera adelante y si soy un poco menos ignorante que cuando vine, es en gran parte gracias a él.

A los tres, tened en cuenta que allá donde esté, siempre portare con orgullo esta baza .

Póquer de reyes

A lo largo de estos años he conocido a personas que han ejercido una gran influencia, tanto en mis experimentos como en mis decisiones personales. Estas cartas son Joao Tavanez, Elias Bechara, Jofre Font. Si he aprendido a ser más critico conmigo mismo que cualquier otro es gracias a Joao. Aparte de ser un gran científico, ha sido un gran amigo durante el tiempo que coincidimos en el laboratorio, y sus bromas y su ironía me han ayudado a hacer todo este tiempo mucho mas llevadero.

Elías, el rey de corazones, me ha enseñado a ver el lado positivo de la vida en aquellas situaciones en las que lo das todo por

perdido. Sin duda, la alegría que transmite ha sido determinante para que todos estos años viera los problemas de color de rosa. Ojalá me encuentre muchos mas Elías repartidos por el mundo en el futuro

Y Jofre, una de las primeras personas que conocí tras mi llegada y que me ha hecho sentir como en casa pese a estar tan alejado de mi familia. A los tres, gracias por vuestro apoyo y amistad.

Póquer de reinas

Tambien he tenido la suerte de contar con la ayuda de las chicas de oro, en este caso Anna Corrionero, Maria Paola Paronetto y Camilla Iannone. De Anna aprendí que la perseverancia puede llevarte muy lejos, y pese a sufrir con sus patadas en las espinillas, fue una gran compañera durante este tiempo.

Maria Paola, alma gemela de Elías, también ha contribuido notablemente a formar el científico que llevo dentro. Gracias por todos los cientos de charlas, bromas y regañinas que has tenido que aguantar.

Camilla, mi otra hermana científicamente hablando, quien siempre me ha acompañado a lo largo de este camino y de quien espero, tenga un gran futuro por delante. Aunque la suerte no se ponga de cara, siempre hay que saber sacarse las castañas del fuego, y pese a que quizá ella no sea consciente de ello, su evolución me ha hecho mas fuerte a lo largo de los años.

Y por ultimo, a Marta, mi reina de corazones, por haber aguantado todo este tiempo desde la distancia. Espero que pronto nuestros caminos acaben convergiendo de nuevo.

Full de figuras

El ultimo año ha sido especialmente duro, no solo por la gente que se ha ido marchando del instituto sino por la presión de la tesis, manuscritos, etcétera. Y si ha habido un grupo de personas que me ha ayudado a sobrellevarlo en la medida de lo posible, es el formado por mis compañeros de comedor. Gracias a Rugby princess (Elena), Jordi, Andy, Matteo, Michael, Gaetano, Luisa y a todos los demás que habéis contribuido a que desconectara momentáneamente de la realidad.

Escalera de color

Por otra parte, estos años me han hecho reflexionar sobre lo rápido que pasa el tiempo, y sobre todos los compañeros que han ido pasando por el laboratorio. Gracias a Sophie, no solo por las discusiones científicas, sino por enseñarme que hay batallas por las que merece la pena luchar y ganar. También a Belén, Anna Ribo, y a todos los antiguos y nuevos compañeros que han ido pasando por el laboratorio a lo largo de estos años.

Trios, Parejas y demás

Pido disculpas por no mencionar a todos los que fueron participes de esta historia, pero que sepáis que os estoy muy agradecido a

todos vosotros. (Alessios y Valerias, happy germans (Tom, Hagen and Berni), French connections, PhD students, etc etc etc)

The Joker

Y finalmente, si estos proyectos han salido adelante ha sido gracias a la carta escondida bajo la manga. Gracias a ti Pan, por abrirme los ojos y hacerme ver la ciencia desde otro punto de vista, así como por ser el motivador de mis pensamiento científico con cada una tus magnificas contribuciones. Aunque sea bastante improbable, ojalá pueda encontrar gente tan dedicada como tu en un futuro proximo. Espero que el recuerdo de esta partida perdure para siempre en el tiempo.

Gracias por todo, a todos