Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
SCIENCE BIOMEDICHE E BIOTECNOLOGICHE

CICLO *XXX*

COORDINATORE Prof. Paolo Pinton

Promoter proximal U1snRNP-dependent pre-mRNA processing defects in human coagulation factor 7 deficiency

Settore Scientifico Disciplinare BIO/11

Dottorando
Dott.ssa Petrovic Ivana

_____
(firma)

Tutore
Prof. Pagani Franco

_____
(firma)

Anni 2014/2017

RIASSUNTO

Lo U1 snRNP (small nuclear RiboNuclear Particle) e' una particella fondamentale presente nello spliceosoma ed è coinvolto in diverse fasi del processamento del precursore dell' RNA (pre-mRNA) tra cui lo splicing e la poliadenilazione. Mutazioni nel sito di splicing donatore sugli esoni interni possono ridurre il legame dello U1 snRNP sul pre-mRNA inducendo difetti di splicingcon salto dell'esone. Questi difetti possono essere recuperati da U1 snRNA modificati che si legano al sito del donatore difettoso o alle sequenze introniche a valle. Tuttavia, mutazioni che interessano il sito di splicing 5' localizzato nel primo esone e quindi prossimale al promotore (Promoter Proximal 5'ss o pp 5'ss) sono state poco studiate. In questo lavoro, usando come modello mutazioni localizzate al pp 5 'ss nel fattore di coagulazione 7 (F7), ho studiato come queste varianti associate alla malattia influenzano il processamento del pre-mRNA ed inoltre il potenziale terapeutico di U1 RNA modificati. L'analisi è stata eseguita utilizzando diversi minigeni corrispondenti al F7, transfettati in modo transiente o cromosomicamente integrati nelle cellule T-REx di Flp-In ™ 293, sui quali ho valutato l'effetto delle mutazioni pp 5'ss sullo splicing e sulla selezione di siti di poliadenilazione alternativi. Esperimenti coi minigeni hanno dimostrato che le mutazioni di pp 5'ss, in base alla forza residua del sito donatore, hanno un effetto negativo sullo splicing riducendo drasticamente la quantità di trascritti processati. Tale effetto e' risultato evidente nelle frazioni cellulari cromatiniche, nucleoplasmatiche e citoplasmatiche. Inoltre, nei costrutti che contengono un ulteriore sito di poliadenilazione alternativo, i mutanti al pp 5 'ss hanno promosso l'utilizzo del sito prossimale di poliademilazione. Inoltre, la lunghezza del primo introne accentua questi difetti riducendo ulteriormente lo splicing e la selezione del sito di poliadenilazione prossimale. Questo effetto dipendente dalla lunghezza del primo introne suggerisce che la forza del pp 5' potrebbe anche influenzare la processività della RNAPII. Nei diversi sistemi sperimentali utilizzati, la espressione di U1 modificati ha recuperato i difetti di processamento del pre-mRNA ed, in un minigene competente per lo splicing, ha recuperato l'attività della proteina F7. Attraverso lo studio di varianti di U1 snRNA modificate, e' stao inoltre possibile identificare due elementi strutturali, la proteina 70K e il la quarta struttura a forcina della RNA (stem loop 4), che sono importanti sia per l'aumento dello splicing che per la poliadenilazione. Al contrario la proteina U1-A e' risulta non importante in questi due eventi. Inoltre, in due minigeni associati all'atrofia muscolare spinale e alle condrodisplasie, U1 modificati che si legano ai corrispondenti pp 5'ss erano in grado di aumentare i trascritti. I miei risultati indicano che le mutazioni che colpiscono i pp 5'ss influenzano il legame con lo U1 snRNP e questa particella, principalmente attraverso

il loop 4 e la proteina 70K, è coinvolta nel corretto processamento del pre-mRNA. Le mutazioni di pp 5’ riducono drasticamente i trascritti soggetti a splicing e possono anche attivare potenziali siti di poliadenilazione criptici nel primo introne. Poiché gli snRNA U1 modificati che si legano al sito del donatore difettoso o alle sequenze introniche a valle recuperano in modo efficiente i difetti di base, queste particelle possono rappresentare una nuova strategia terapeutica utile per la correzione delle mutazioni che influenzano i pp 5'ss.

ABSTRACT

U1 snRNP is a core component of the spliceosome involved in different pre-mRNA processing steps, including splicing and polyadenylation. Donor splice site mutations on internal exons reducing U1 snRNP binding on pre-mRNA have been shown to induce splicing defects. These can be recovered by modified U1 snRNAs that bind either at the defective donor site or at downstream intronic sequences. However, very few studies focused on promoter proximal 5' splice site (pp 5'ss) mutations. In this work, using as a model disease-associated mutations at the pp 5'ss in coagulation Factor 7 (F7), I have studied how these variants affect pre-mRNA processing as well as the therapeutic potential of modified U1 RNAs. The analysis was performed using different F7 minigene model systems, either transiently transfected or chromosomally integrated in Flp-In™ 293 T-REx cells where I evaluated the effect of the pp 5'ss mutations on splicing and alternative polyadenylation selection. Minigene experiments showed that the pp 5'ss mutations, according to the residual strength of the donor site, have a negative effect on splicing, severely reducing the amount processed transcripts. This effect was evident in chromatin, nucleoplasmic and cytoplasmic cellular fractions. In addition, in constructs with an additional artificial polyadenylation site, the pp 5' ss mutants promoted the utilization of the proximal polyA site. Interestingly, the length of the first intron increases the defect caused by the pp 5'ss mutations, reducing splicing and promoting the selection of the proximal polyadenylation site. This length-dependent effect suggests that the strength of the pp 5' might also affect RNAPII processivity. Expression of modified U1s rescued the pp 5'ss defects in the different experimental systems, and recovered the F7 protein activity in a splicing competent minigene. Using modified U1 snRNAs variants, I show that two structural elements that constitute the particle, the 70K protein and the stem loop 4 are important both for splicing enhancement and polyadenylation, whereas the U1-A protein is mainly dispensable. Furthermore, in two unrelated minigenes associated to spinal muscular atrophy and chondrodysplasias, modified U1s that bind to their corresponding pp 5'ss were able to increase the spliced transcripts. My results indicate that mutations at pp 5'ss affect U1 snRNP binding and this particle, mainly through stem loop 4 and 70K, is involved in correct processing of the pre-mRNAs. Pp 5'ss mutations severely reduce the splicing transcripts and can also activate potential cryptic polyadenylation sites in the first intron. As modified U1 snRNAs that bind either at the defective donor site or at downstream intronic sequences rescue these defects, these particles represent a novel therapeutic strategy for the correction of mutations that affect the pp 5'ss.

ACKNOWLEDGEMENTS

I wish to thank, first and foremost, my PhD Tutor Franco Pagani, for giving me the opportunity to learn and helping me along the way, and whose guidance and support from the initial to the final level were crucial. During the most difficult times, he gave me moral support and a lots of patience, and without him this dissertation would not have been possible.

I would like to thank all the past and present members of Human Molecular Genetics research group for their friendship and support, and for valuable advice and discussion. Special thanks to Erica, Chiara and Gosia, for guidance and all the knowledge they shared with me; to Giulia and Irving, for all the great discussion and advice; to Mojca, whose wisdom and skills I missed in lab for almost two years; to Federico, Cecilia, and to Michał and Sara without whom this would not have been the same experience.

Thank you Kasia, Sanja and Guido for all the past years, I am lucky to have friends like you, and I owe you big time. Special mention to current and former colleagues from ICGEB Trieste: Vjeko, Jaye, Suruchi, Vale, Karla, Fatemeh and Esra. To each and every one of you I am grateful for adding that one small piece to complete the picture.

I would like to express sincere gratitude for Dr Djordje Fira, my first academic tutor and respected University professor, whose gentle and wise guidance led me to ICGEB, making this thesis possible.

Big thanks to my parents, for always believing in me. Through their own example they showed me the beauty of learning, taught me to focus on opportunities not limits, and gave me the strength and patience not to give up. J & T, your open arms saved me in those critical times when it was insanely hard to get up. Jelena, to you alone I could write lines and lines, and since you are here for me my whole life, the list would be endless. You are the sister that everybody would wish for, and which I am blessed to have. Kostya, your unconditional love, patience and moral support made all the difference in the world. Here I am finding it very difficult to sustain from getting all lyrical, but in fact, what I am the most grateful for cannot be expressed in words.

At the end, my biggest appreciation goes to all of you who did not make it easy for me, because you encouraged me to revise my opinions, and from you I have learned the most. I am honestly thankful for the opportunity to meet new concepts, challenge ideas, and fight for what I felt was correct. You gave meaning to this PhD, and with an honest gratitude - from the bottom of my heart - I dedicate this thesis to you.

Contents

TABLE OF FIGURES

1. INTRODUCTION

1.1 RNA processing

Transcribing information from DNA to precursor RNA is followed by several regulated steps of processing and surveillance, which at the end give a rise to mature, functional molecule (Figure 1-1). Several studies demonstrate extensive coupling of different processes in gene expression, which improves efficiency and specificity of each step (Bentley, 2005; Maniatis & Reed, 2002).

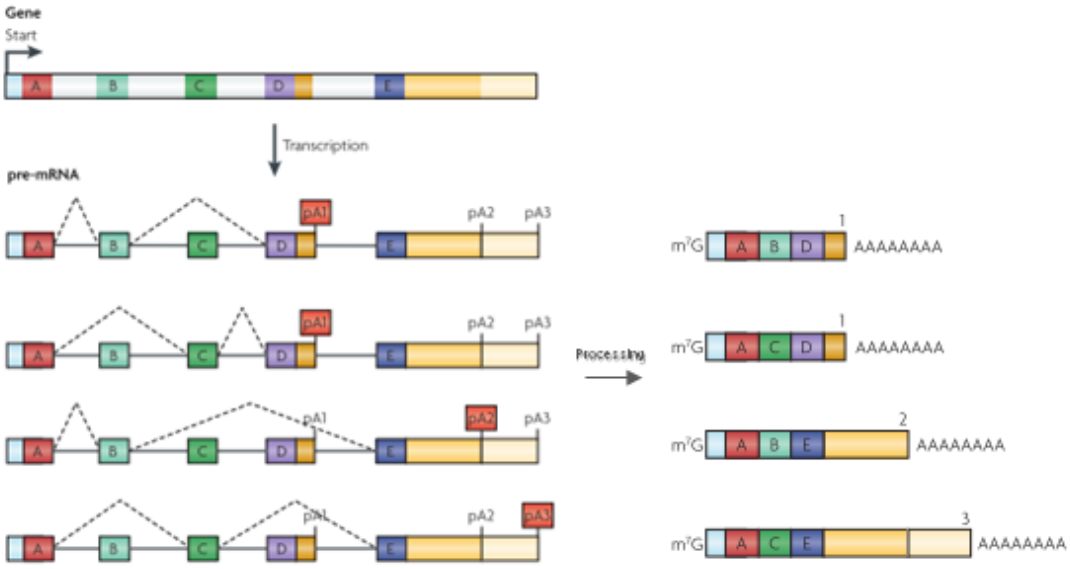

Figure 1-1. Example of gene expression and alternative pre-mRNA processing. Pre-mRNA transcribed from a single gene is alternatively processed to create mRNA isoforms that differ in their coding sequence and/or 3' UTRs (left). Alternative splicing and choice of polyadenyation site (right) generate mature mRNAs that vary in their length and coding potential. [1]

Genomic DNA in eukaryotic cells in transcribed by three DNA-dependent RNA polymerases: RNA polymerase I (RNAPI) is transcribing the ribosomal RNA (rRNA) (Grummt, 2003), RNA polymerase II (RNAPII) transcribes the vast majority of genes, including protein-coding ones (Woychik & Hampsey, 2002), while RNA polymerase III (RNAPIII) transcribes

[1] Modified from (Licatalosi & Darnell, 2010)

genes that encode for short (<400nt) structural RNAs, such as transport (tRNA) and 5S ribosomal RNA (rRNA) (Schramm & Hernandez, 2002). Plants have an additional polymerase IV, which is involved in heterochromatin formation (Herr, Jensen, Dalmay, & Baulcombe, 2005).

Kinetic coupling of RNAPII transcription with RNA processing allows raising the local concentration of messenger RNA (mRNA) processing factors at the vicinity of the nascent transcript (Figure 1-2), and as well might influence the allosterical activation or inhibition of mRNA processing (Bentley, 2005). As a consequence, the transcription elongation can have an effect on RNA folding and assembly of RNA-protein complexes, affecting the choice of alternative processing sites (De La Mata et al., 2003). The major steps of processing precursor RNA include 5' end capping, splicing and 3' end cleavage/polyadenylation, and they are important for maintaining RNA stability, proper transport from nucleus to cytoplasm and promoting of translation (Varani & Nagai, 1998).

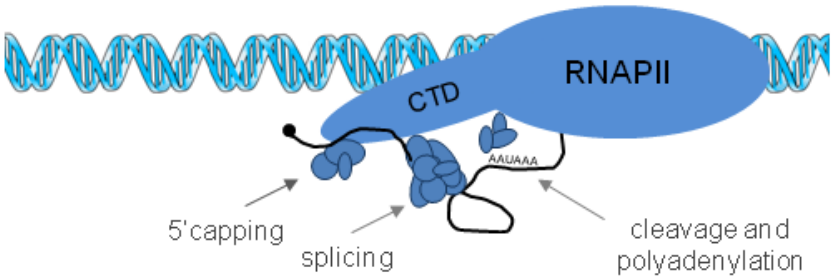

Figure 1-2. Pre-mRNA processing is coupled with transcription. Factors that are necessary for pre-mRNA processing are associated with C-terminal domain (CTD) or RNA polymerase II (RNAPII). Those include factors for 5'capping, spliceosomal components and factors for cleavage of transcript and polyadenylation.

Besides the aforementioned mechanisms of RNA processing that globally affect eukaryotic gene expression, there is another less common process which involves changes in transcript. RNA editing is a phenomenon which results with the sequence of transcribed RNA molecule not matching their DNA template (Benne et al., 1986). There are two different ways in which RNA can be edited: by substitution/conversion of base, or insertion/deletion. RNA editing by substitution requires recognition of specific sequence by enzymes. The best known nucleotide changes are deamination (A-to-I and C-to-U) and amination (U-to-C) reactions (Gott & Emeson, 2000). Insertion/deletion editing starts either post-transcriptionally or co-transcriptionally by creating a base-paired "anchor" duplex between pre-edited mRNA and guide RNA (gRNA) that directs the cleavage by editing endonuclease (Figure 1-3). Depending

on the type of editing, the cleavage is acted upon terminal uridylil-transferase for inserting (Frech & Simpson, 1996), or a U-specific exonuclease, in case of deletion of uridines (Cruz-Reyes & Sollner-Webb, 1996). Cycle is completed by ligating ends of the two mRNA fragments.

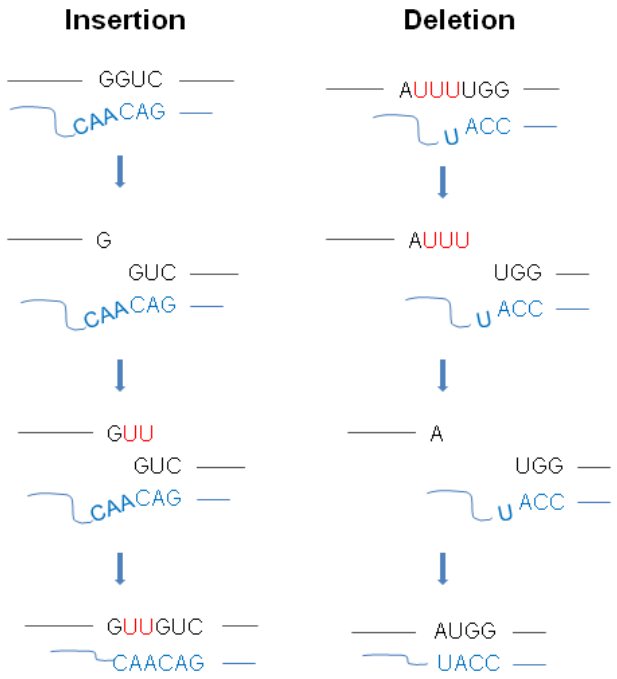

Figure 1-3. General mechanism of RNA editing. Scheme shows catalytic events in insertion and deletion editing. Pre-mRNAs (black strands) are edited 3' to 5' with each gRNA (blue strands) specifying the editing of several sites. Interaction between pre-mRNA and gRNA is regulated through Watson-Crick base pairing (also including G·U base pairs). Editing occurs by series of catalytic steps: endonucleolytic cleavage of the pre-mRNA by an endonuclease occurs upstream of the anchor duplex (8–10 bp) between the pre-mRNA and its 'cognate' gRNA. Us are either added to the 5′ cleavage fragment by a TUTase (insertion editing) or removed by an ExoUase (deletion editing), as specified by the sequence of the gRNA. The resultant 5′ and 3′ mRNA fragments are then ligated by an RNA ligase.[2]

1.1.1 **5' end capping**

The 5'end capping with a methylated guanosine analogue occurs co-transcriptionally. When the nascent transcript created by RNAPII reaches the length of 25-30 nucleotides, 7-methylguanosine is added to its 5'end by three sequential enzymatic reactions (Kim et al., 2004). The first reaction is hydrolysis of the 5' triphosphate end of pre-mRNA to a diphosphate, and is performed with enzyme RNA 5' triphosphatase (RTPase). Second is capping of

[2] Modified from (Stuart et al 2005).

diphosphate with GMP by guanylyltransferase (GTase) to form GpppRNA, and it is followed by (third reaction) its conversion to m^7GpppRNA by S-adenosylmethionine:RNA(guanine-N7)-methyltrasnferase (MTase) (Shuman, 2015). The 5' cap is important for RNA stability (Furuichi, LaFiandra, & Shatkin, 1977), disabling degradation of mRNA by 5' exonucleases (Schwer, Mao, & Shuman, 1998), facilitating splicing (Lewis, Izaurralde, Jarmolowski, McGuigan, & Mattaj, 1996) and transport of spliced mRNA to cytoplasm (Visa, Izaurralde, Ferreira, Daneholt, & Mattaj, 1996). During initiation of translation 5'cap is playing an important role in aligning mRNA on ribosome (Gross et al., 2003).

1.1.2 **3' end cleavage/polyadenylation**

Vast majority of RNAPII generated transcripts contain a poly(A) tail. Processing of 3' ends is consisted of endonulceolitic cleavage of primary transcript, and polyadenylation of the upstream fragment for 70-100 nucleotides in median length (Subtelny, Eichhorn, Chen, Sive, & Bartel, 2014; Wahle & Keller, 1992).

The cleavage of transcript is directed by two elements, one upstream and one located downstream from the cleavage site (Figure 1-4). The element 10-30 nucleotides preceding the site of endonucleolitic cleavage contains conserved hexanucleotide sequence, called polyadenylation signal (PAS): AAUAAA, which is present in ~60% polyadenylated mRNAs (Elkon, Ugalde, & Agami, 2013). However, change of the second A, especially to a U, gives a rise to the most frequent variant in upstream element (USE), AUUAAA, which appears in ~15% of poly(A) sites (Wahle & Keller, 1992). In addition to canonical hexanucleotide AAUAAA and its main variant AUUAAA, nine additional hexamers were indentified and collectively contribute to ~14% of all polyadenylation sites (Beaudoing, Freier, Wyatt, Claverie, & Gautheret, 2000; Elkon et al., 2013).

In addition to the USE, there is a less conserved downstream element required for 3'end cleavage. Downstream element (DSE) contains a short number of U residues - U-rich element (Chou, Chen, & Wilusz, 1994) or, more frequently, contains a consensus like YGUGUUYY, where Y stands for pyrimidine – GU-rich element (Scorilas, 2002).

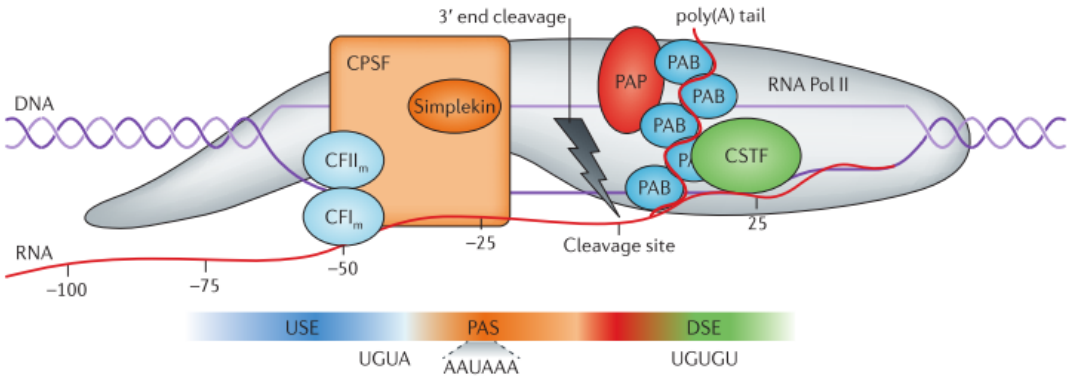

Figure 1-4. Core factors of cleavage and polyadenylation. Trans-acting factors and their corresponding cis-elements are coloured similarly. The key cis-element that dictates cleavage is a 6 nt motif called the poly(A) signal (PAS), the canonical form of which is AAUAAA. In addition to the PAS, nearby U- or GU-rich downstream sequence elements (DSEs) and less well-defined upstream sequence elements (USEs) enhance cleavage efficiency[3]

The cleavage site itself is located between USE and DSE. The distance between AAUAAA and the cleavage site can vary, and sequence that surrounds it is not very conserved. However, the nucleotide at the cleavage site is usually an A. The nucleotide that precedes the A in more than half of the cases is a C (Sheets, Ogg, & Wickens, 1990).

Recognition of polyadenylation signal and downstream element is performed by two large multiprotein complexes, cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CSTF), respectively. CPSF consists of six polypeptides: CPSF4, CPSF2, CPSF1, CPSF3, FIP1L1 and WDR33, from which CPSF1 is recognising the PAS. CSTF has three subunits: CSTF1-3, that are involved in selection of PAS. Additional factors required for cleavage/polyadenylation are poly(A) polymerase, simplekin (the scaffold protein) and cleavage factor CFIm (tetramer) and CFIIm, as well as poly(A) binding proteins (PAB) (Davis & Shi, 2014).

Cleavage of RNA precursor starts with CPSF binding to polyadenylation signal, while CSTF recognises downstream element. CSTF than binds CPSF and forms a loop of RNA, that is stabilized with binding of CFI and CFII. Binding of poly(A) polymerase (PAP) to the complex promotes cleavage between polyadenylation signal and downstream element. Downstream RNA fragment is rapidly degraded, while the upstream fragment is polyadenylated by PAP. First 12 residues are added in a slow rate to the 3'end of the upstream fragment, and the process gets accelerated as PABP start binding to the initial poly(A) tail. When the length of

[3] Figure modified from (Elkon et al., 2013)

200-250 AMPs is reached, PABPs interacts with PAP to stop polymerysation (Zhao, Hyman, & Moore, 1999).

Poly(A) tail participates in regulation of many functions, such as mRNA export to cytoplasm, termination of transcription and stimulation of translation. Moreover, disintegration of the poly(A) tail is correlated with mRNA degradation (Wickens, Anderson, & Jackson, 1997). The reduction in the size of poly(A) tail is performed by poly(A) ribonucleases. After the mature mRNA is exported to cytoplasm, poly(A) tail is reduced to less than 50 AMP residues (Scorilas, 2002).

There are many pathologies caused by defects in 3' end processing. For example, defects in gene encoding poly(A) binding protein II (PABII) can result in oculopharyngeal muscular dystrophy. Depletion of CstF-64 factor leads to cell cycle arrest and apoptosis (Takagaki & Manley, 1998). Repression of PAP activity is normally occurring during M-phase and contributes to reduction of mRNA and protein anabolism, however, in pathological downregulation it can result with low growth rate and accumulation of cells in G0-G1 phase of cell cycle (Colgan, Murthy, Zhao, Prives, & Manley, 1998). On the other hand, increased activity of PAP is connected with several types of cancer, including acute lymphoblastic leukaemia (ALL) and chronic lymphocytic leukaemia (CLL), as well as breast cancer (Scorilas, 2002).

1.1.2.1 Alternative polyadenylation

It has been recently discovered that most human genes contain more than one polyadenylation site, suggesting that alternative polyadenylation (APA) is widespread mechanism of generating different 3'end isoforms. There four main classes of APA (Figure 1-5): tandem 3'UTR APA, alternative terminal exon APA, intronic APA and internal exon APA, of which the first two classes are the most frequent (Elkon et al., 2013). APA altering the 3'UTRs can potentially regulate stability, cellular localisation and translation of target RNA, as 3'UTR serve as docking platform for factors such as microRNA and RNA-binding proteins (Andreassi & Riccio, 2009). Through alterations in coding sequence, APA contributes to complexity of transcriptome by creating isoforms that vary in their coding sequence, and thus alter the function of target genes (Licatalosi & Darnell, 2010).

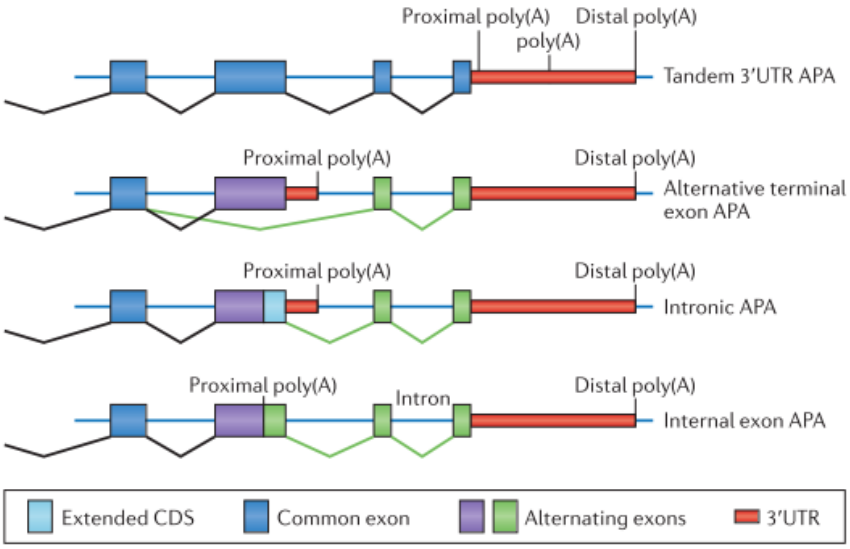

Figure 1-5. Four different types of alternative polyadenylation. Alternative polyadenylation types are: alternative terminal exon APA, in which alternative splicing generates isoforms that differ in their last exon; intronic APA, which involves cleaving at the cryptic intronic poly(A) signal (PAS), extending an internal exon and making it the terminal one; and internal exon APA, which involves premature polyadenylation within the coding region.[4]

Recognition of alternative polyadenylation signal can be modulated in several different ways. The overall rate of reaction depends on the concentration of all reactants, but also on how efficiently is PAS recognised by the processing machinery. According to the model "first come –first served", proximal PAS has an advantage over the distal one, as it is transcribed earlier and has more time to get recognised. As an outcome, both the sequence strength and the relative position of the alternative PAS are important in determining the final outcome of APA (Shi, 2012). A recent paper by Pinto and colleagues shows in *Drosophila* polo mRNA model that the slow mutant of RNAPII kinetically favors the recognition of the proximal PAS. In cells that express the wild-type form of RNAPII this advantage is less significant, and the usage of the distal PAS is increased (Pinto et al., 2011). A similar mechanism was previously described to regulate the alternative splicing (De La Mata et al., 2003).

The concentration of processing factors is an additional layer of APA regulation that can be altered on transcription level (Elkon et al., 2013). It can also be altered post-transcriptionally, by signalling pathways (Danckwardt et al., 2011), or via post-translational modifications and interactions with other factors (Millevoi & Vagner, 2009). For example, brain-specific RNA-binding protein Nova2 can affect alternative polyadenylation, in addition to its role in alternative

[4]Figure modified from (Elkon et al., 2013)

splicing. As recently revealed by high throughput analyses, Nova2 can bind to the RNA in PAS regions and suppress their usage (Licatalosi et al., 2008).

Involvement of transcription machinery in regulation of alternative polyadenylation is getting evident, with accumulating observations that transcription factors can regulate the recruitment of 3'end processing factors. Transcription elongation factor ELL2 has been suggested to promote the loading of CSTF onto transcription machinery while the transcription is ongoing, which is then more likely to promote the usage of proximal PAS, as well as skipping the last exon (Martincic, Alkan, Cheatle, Borghesi, & Milcarek, 2009).

1.2 Pre-mRNA splicing

Majority of eukaryotic genes are discontinuous: small pieces of coding DNA sequence within a gene (exons) are interspersed with much longer non-coding sequences (introns). The excision of the non-coding and merging coding parts in precursor RNAs is a well-conserved and precisely regulated process called splicing. Splicing is an important step of RNA processing contributing to diversity of functional products descending from the same gene (Pan, Shai, Lee, Frey, & Blencowe, 2008).

There are 4 different type of introns, on the basis of their mechanisms of splicing and/or characteristic sequences: group I are self-spliced introns removed by RNA catalysis, found in mitochondria, plastids and *Tetrahymena* pre-ribosomal RNA (Burke, 1988), group II are also self-spliced found in fungal mitochondria and plastids (Michel, Kazuhiko, & Haruo, 1989), introns in nuclear and archaeal transfer RNA (tRNA introns) removed by proteins, and finally, spliceosomal introns of pre-mRNA (Alberts, 2008). There is another group of catalytic introns, proposed to be group III, found in *Euglena gracilis* chloroplast genome (Copertino & Hallick, 1993).

Nuclear pre-mRNA introns are interspersed with short stretches of coding sequences (exons), and they are excised by multi-component machinery of spliceosome. Spliceosomal introns can be of two types: canonical introns, that start with GT and end with AT, and non-canonical ones, that usually have AT-AC at their termini. Canonical introns are the vast majority of spliceosomal introns (>99.5%), and they are spliced by the U2-dependent spliceosome. The minority of non-canonical introns are spliced by U12-dependent spliceosome, and they are found in plants, insects and vertebrates (Rodríguez-Trelles, Tarrío, & Ayala, 2006).

1.2.1 **Biochemistry of splicing reaction**

Splicing of precursor mRNA is performed in two sequential transesterification reactions (Figure 1-6) that takes place in a multi-component complex named spliceosome. In first reaction intron is cleaved at the intron-exon junction on its 5'end, and result is an intermedier composed of upstream exon with exposed 3' end and intron in the form of lariat. This catalytic step involves nucleophilic attack performed by 2'-OH group of adeonosine in the branch point, attacking the 5'splice site. The result is formed 2'-5' phosphodiester bond in intron and free 3'-OH group of the upstream exon. In the second reaction two exons are joined and intron is displaced. During the second transesterification the 3'-OH group from the cleaved 5'splice site is attacking phosphodiester bond within the 3' splice site, resulting in joining of two exons, and releasement of intron still in lariat form.

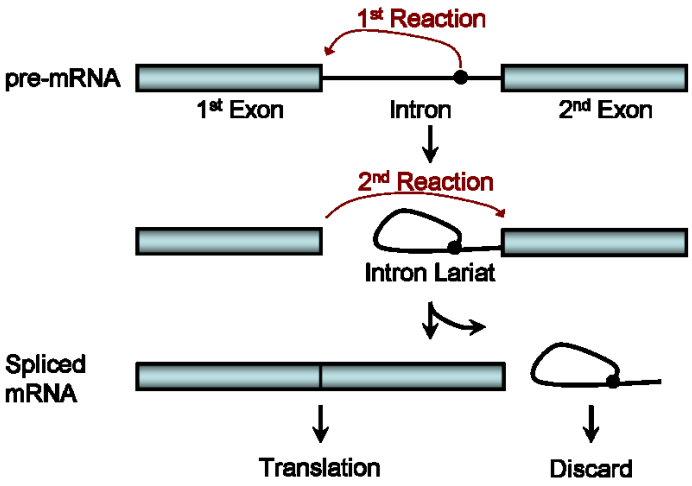

Figure 1-6. Catalytic steps of pre-mRNA splicing. In first transesterification, the 2' ribose hydroxyl of the branch site adenosine residue acts as a nucleophile and attacks the phosphodiester bond linking the 5' terminus of the intron to the 5' exon. This produces the lariat form of the intron (still joined to the 3' exon) and a free 5' exon, with a 3' ribose hydroxyl group, as reaction intermediates. In the second transesterification, the 3' ribose hydroxyl of the 5' exon acts as a nucleophile and attacks the phosphodiester bond linking the 3' terminus of the intron to the 3' exon.[5]

1.2.2 **Spliceosome**

The core of each spliceosome is composed of five RNA-protein complexes called small nuclear ribonucleoproteins (snRNP). Those RNA-protein complexes contain five different small

[5]Figure modified from (Lamond, 1993)

nuclear RNA molecules (snRNA): U1, U2, U4, U5 and U6. Prior to the catalytic steps of splicing, snRNPs together with additional splicing factors assemble on the pre-mRNA and form a spliceosome on each intron (Nilsen, 2003).

Spliceosome assembly starts with U1 snRNP binding to the pre-mRNA, forming a “commitment complex”, or E complex (early complex) – a point in which the committed pre-mRNA cannot be competed out by addition of an excess of competitor pre-mRNA (Lamond, 1993). Binding of U1 snRNP is performed through base pairing between the 5’ intron-exon junction and complementary sequence of 5’ end of U1 snRNA, without requirement for ATP (Michaud & Reed, 1991).

After formation of the commitment complex, U2 snRNP binds to the intron in the polypyrimidin region, through base-pairing, to form complex E. The stable binding of U2 snRNP is dependent upon ATP, and in mammalian cells it also requires a non-snRNP splicing factors: U2 snRNP auxiliary factor (U2AF), SF1 and SF3 (Krämer & Utans, 1991). The stable binding of U2 snRNP to the branchpoint converts E complex into a pre-spliceosomal complex A.

Spliceosomal B complex is formed when pre-assembled tri-snRNP consisted of U4, U5 and U6 is recruited to the A complex. This complex now undergoes some rearrangements to form the catalytically activated B* complex. Those rearrangements include displacing of U1 by U6 snRNP at the 5’ splice site, the U4/U6 binding is disrupted, an interaction between U2 and U6 is formed, and as a consequence U1 and U4 leave spliceosomal complex (Rino & Carmo-Fonseca, 2009).

The first transesterification reaction takes place in B* complex when adenosine from the branch point performs a nucleophilic attack immediately upstream of the 5’ splice site. This results in transforming of B complex into C complex, which now undergoes additional rearrangements to enable the second transesterification. The second catalytic step involves cleavage of the 3’ splice site and ligation of two exons, after which the spliceosome disassembles and mRNA is released.

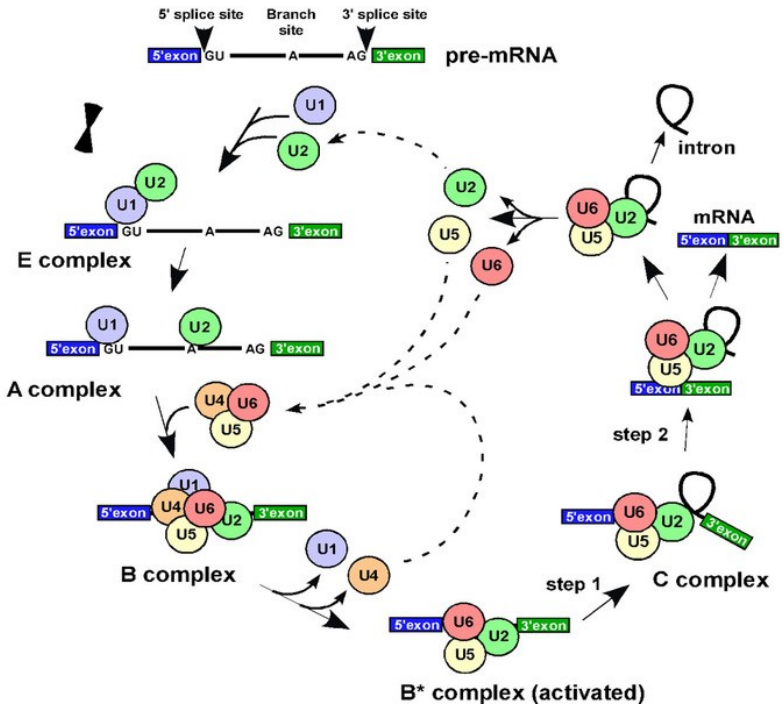

Figure 1-7. Pre-mRNA splicing by the major spliceosome[6]

1.2.3 **U snRNPs**

Each of the snRNP subunits of the spliceosome is comprised of a uridine-rich small nuclear RNA (U snRNA) and a set of proteins. A U snRNP can contain up to 40 proteins, divided in two groups: the Sm core proteins and specific proteins for each U snRNP.

The core of every snRNP is made of seven small polypeptides, from 9-29kDa in size. Designated as B/B', D1, D2, D3, E, F and G, these proteins form a ring shaped heptamer that interacts with a conserved sequence present in U1, U2, U4 and U5. In U6 snRNP the Sm proteins are replaced by seven Sm-like proteins (LSm). Protein-protein interactions are mediated through a Sm sequence motif present in each Sm protein (Hermann et al., 1995). The core of U snRNPs is important for the metabolic stability of these particles (Grimm, Stefanovic, & Schümperli, 1993).

U1, U2, U4 and U5 are transcribed by RNAPII, and in cytoplasm they assemble with the core snRNP proteins. After this assembly U1, U2, U4 and U5 acquire a modified 2, 2, 7 – trimethyl guanosine cap (m_3G-cap) on their 5' end, which is formed through hypermethylation

[6]Figure modified from (Wahl, Will, & Lührmann, 2009)

of the standard 7-methyl guanosine cap. Following the protein assembly and acquiring m_3G-cap, the mature snRNPs are reimported to nucleus (Lamond, 1993). The other spliceosomal snRNA, U6, is transcribed by RNAPIII and remains in nucleus. On its 5'terminus U6 has a γ-monomethyl triphosphate. U4 and U6 snRNAs interact with each other through extensive intermolecular base pairing to form U4/U6 particle, that afterwards interacts with U5 to form tri-snRNP particle (Lührmann, Kastner, & Bach, 1990).

1.2.3.1 U1snRNP

The U1 snRNP is a complex that recognizes the 5' splicing junction of pre-mRNA. It consists of U1 snRNA and 10 proteins: 7 Sm proteins and 3 U1 snRNP-specific proteins U1-A, U1-C and U1-70K. The recognition of pre-mRNA mediated by 5'-region of U1snRNA is critical for defining exon/intron junctions and is needed for a correct assembly of other components of spliceosome. In humans, small U1 RNA (U1 snRNA) is comprised in 164 nt, and it is transcribed from genes that occur in several copies within the human genome. The U1 snRNA molecules have a stem-loop 3d structure, and in 5'-region they contain a single-stranded sequence (9 nt in length) capable of binding to complementary sequence in splicing donor site on the pre-mRNA (C. L. Will & Lührmann, 2001).

70K is composed of unstructured N-terminus, one RNA binding domain through which it binds to the stem-loop 1 of U1 snRNA, and C-terminus rich with serine and arginine (RS domain). This protein is important for the recruitment of U1 snRNP to the 5' splice site (Kohtz et al., 1994). Its C-terminal domain is shown to be necessary and sufficient for interaction with splicing factor ASF/SF2 (Cao & Garcia-Blanco, 1998).

U1-A protein (32kDa) contains two RNA recognition motifs (RRMs), one in N-terminus and one in C-terminus, bridged by a linker. The N-terminal RRM binds to the stem-loop 2 (SL2) sequence AUUGCAC of U1 snRNA. It was previously demonstrated that this protein is not necessary for proper splicing *in vitro* (C. Will, Rümpler, Gunnewiek, Venrooij, & Lührmann, 1996).

U1-C protein (17kDa) is composed of C-terminal domain rich in RG residues, and a zinc-finger domain in N-terminus. Unlike U1-A and 70K, this protein cannot be directly bound to U1 snRNA, and requires binding of 70K and Sm-D3 (Nelissen, Will, van Venrooij, & Lührmann, 1994). Genome wide analyses show that U1-C depletion is not generally essential for splicing, but rather has a narrow group of dependent targets (Rösel-Hillgärtner et al., 2013).

However, the SRSF1 is able to interact with 70K only through interaction with U1-C, thus depletion of U1-C shows a significant effect on splicing (Kohtz et al., 1994).

Although it does not bind any proteins within U1 snRNP, stem-loop 4 has recently been shown important for splicing. It seems that this stem-loop interacts with U2 snRNP, more precisely with SF3A1 protein, during spliceosome assembly (Sharma, Wongpalee, Vashisht, Wohlschlegel, & Black, 2014).

At the first splicing step U1snRNA binds to the 5' splice site of pre-mRNA, by base-pairing between 5'-terminal 8 nucleotides of U1snRNA and complementary sequence in pre-mRNA. Most of the human natural 5'ss sequences have a modest complementarity with U1 snRNA (mean is 6/10nt), and many factors have been described that regulate U1 snRNP binding to the 5'ss (nuclear binding complex, CBC; T-cell intracellular antigen-1, TIA-1; SR proteins that bind splicing enhancers or silencers) (Black, 2003).

1.2.4 **Alternative splicing**

Alternative splicing (AS) is a regulated process that results in generating different forms of mature mRNAs from the same gene, by including or excluding different exons of the same gene (Figure 1-8). The outcome of alternative splicing can give a rise to a family of related proteins originated from the same gene. On protein level, changes in splicing can result in changes of peptide sequence, therefore affecting protein properties such as ligand binding, allosteric regulation, protein binding or enzymatic activity (Blencowe, 2006).

Most exons are always included or spliced from the final mRNA, and they are called constitutive. On the other hand, some exons are regulated in a manner that sometimes results with an inclusion and sometimes with exclusion from mRNA, and they are marked as cassette exons. Exons can also be mutually exclusive, meaning that inclusion of one means splicing of other cassette exons from mRNA (Blencowe, 2006). Coding sequences can also be lengthened or shortened by altering the position of their splice sites, both from 5' and 3' end, and in some cases the whole intron may be retained. The 5' terminal exons can be included/skipped through choice of alternative promoters, while 3' terminal exons can be regulated through alternative polyadenylation sites (Zavolan et al., 2003).

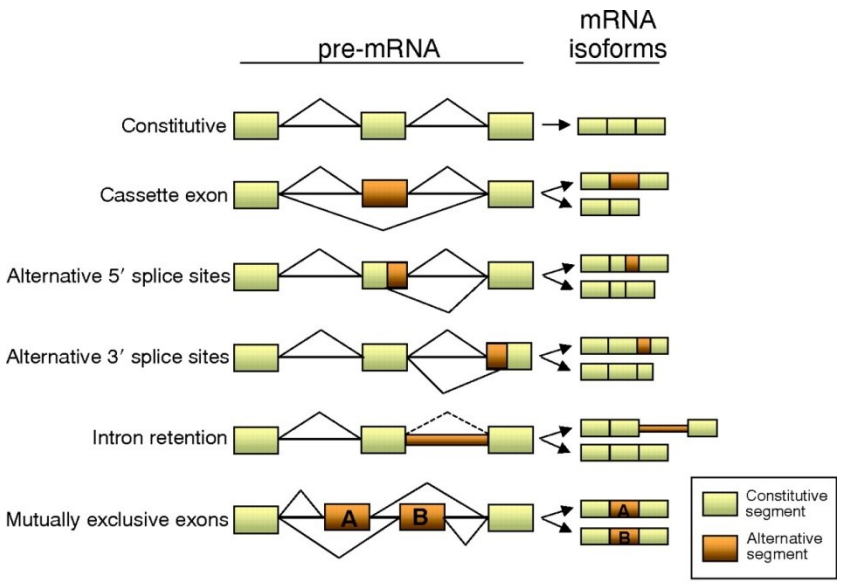

Figure 1-8. Possible outcomes of alternative splicing. Alternative splicing is generating functionally distinct transcripts[7]

Some studies show that transcription factors involved in initiation and elongation can impact the selection of splice sites (de la Mata & Kornblihtt, 2006). In particular, factors that reduce the rate of RNAPII elongation can increase the inclusion of alternative exons. This is explained by kinetic model in which the slow RNAPII favours the recognition of relatively weak splicing signals. Recent results indicate that the exons of nascent transcripts can be "tethered" to RNAPII during their processing within pre-mRNA (Dye, Gromak, & Proudfoot, 2006). This phenomenon can both affect the splice site choice and ensure order in the process of splicing so that the correct pairs of splice sites are merged together.

Alternative splicing can as well be coupled to post-transcriptional control of gene expression, by introducing a premature stop codon (PTC), which targets mRNA for nonsense mediated decay (NMD) (Lejeune & Maquat, 2005). Moreover, AS-NMD regulates the expression of SR proteins (Lareau, Inada, Green, Wengrod, & Brenner, 2007), hnRNPs (Wollerton, Gooding, Wagner, Garcia-Blanco, & Smith, 2004) and core spliceosomal proteins (Saltzman et al., 2008). As shown by Lareau and colleagues, some of the exons that are included or skipped to create premature stop codon are associated with perfectly conserved elements. This suggests that AS-NMD could be evolutionary important mechanism for

[7]Figure modified from (Blencowe, 2006)

establishing the correct balance of RNA binding proteins for different cell types and developmental stages (Boutz et al., 2007)

Since alternative splicing plays an important role in gene expression, irregularity of this process can cause a number of diseases (Faustino & Cooper, 2003). Diseases can arise from mutations that activate cryptic splice sites or disrupt sequences necessary for RNA processing, which results in transcript instability or specific protein isoforms. Irregularities in RNA binding proteins are linked to several neurodegenerative diseases: mutations of FUS and TDP43 proteins are discovered in familial amyotrophic lateral sclerosis (Sun et al., 2015), mutations in survival motor neuron protein 1 (SMN1) result in spinal muscular atrophy (Cooper, Wan, & Dreyfuss, 2009). In the same manner, mutations, deletions or inappropriate expression of miRNAs can lead to aberrant RNA regulation, present in neurological diseases, autoimmunity and cancer (Licatalosi & Darnell, 2010).

1.2.5 **Regulatory elements and factors in pre-mRNA splicing**

The key steps in RNA splicing are performed by these RNA molecules through base-pairing between the snRNAs and the consensus RNA sequences in the pre-mRNA. There are three main sequences of pre-mRNA to be recognized by spliceosome: the 5’ splice site (typically, a GU nucleotide sequence), the branch-point site (the most important is A nucleotide), and the 3’ splice site (typically, an AG sequence). In most cases, there is also polymypyrimidine tract of variable length just upstream from the 3’ splice site. The three aforementioned consensus sequences are generally not sufficient to determine whether the splice site will assemble the spliceosome or not, and other elements and factors are involved in that decision.

RNA elements that positively stimulate spliceosome assembly are called enhancers, and depending on their position they can be exonic (exon splicing enhancer, ESE), or intronic (intron splicing enhancer, ISE). Conversely, other RNA sequences may act on silencing of splicing, both in their exonic and intronic positions (ESS and ISS, respectively) (Franco Pagani & Baralle, 2004). Some regulatory sequences can create secondary structure that can affect the recognition of splice site (Buratti & Baralle, 2004; Muro et al., 1999). These sequences are important for distinguishing the correct splice sites from surrounding cryptic splice sites surrounding them.

Trans-acting factors interact with RNA regulatory elements, and they are divided in two main groups: members of the serine arginine (SR) family of proteins and heterogeneous nuclear RNPs (hnRNPs). In general, the SR proteins bind enhancers, while hnRNPs bind silencers (Franco Pagani & Baralle, 2004).

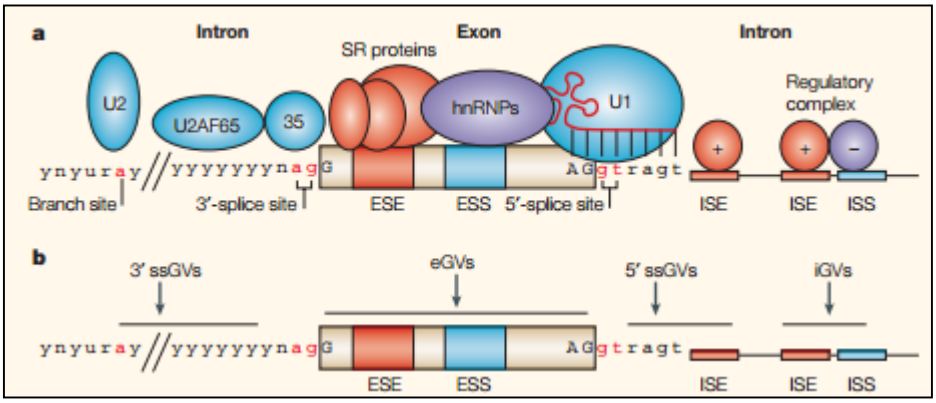

Figure 1-9. Regulatory cis-elements and trans-factors involved in regulation of splicing. ESE – exon splicing enhancer, ESS – exonic splicing silencer, ISE – intronic splicing enhancer, ISS – intronic splicing silencer, eGVs – exonic genomic variants [8]

1.3 Role of U1 snRNP in transcription and polyadenylation

The C-terminal domain (CTD) is an essential part of RNAPII, and not polymerases I and III, and it holds an important role in co-transcriptional pre-mRNA processing. Experiments done by Bentley group show that deletion of this domain or mutations within it prevent efficient pre-mRNA processing (McCracken et al., 1997). Proteomic studies later showed that phosphorylated CTD binds large number of proteins with role in transcription, RNA processing, chromatin structure and spliceosome (Phatnani, Jones, & Greenleaf, 2004). However, this interaction is not only one way, but rather a two-way process, meaning that proteins bound to CTD can also modulate the activity of RNAPII.

One of the first evidence that splicing can affect RNAPII processivity is represented by the fact that genes that contain introns are more actively transcribed. Work by Brinster et al was examining the expression of mRNA produced from gene constructs identical in everything except in presence/absence of introns. These constructs were introduced in murine germ line

[8]Figure modified from (Franco Pagani & Baralle, 2004)

by microinjection of fertilized eggs. In each of four gene construct pairs they analysed, intron containing construct produced 10 to 100 fold more mRNA in compare to their intronless pair (Brinster, Allen, Behringer, Gelinas, & Palmiter, 1988). Opposite from that, when introns were removed from two genes of *Saccharomyces cerevisiae*, it resulted with reduced transcriptional output (Furger, O'Sullivan, Binnie, Lee, & Proudfoot, 2002).

High-resolution kinetic analyses of transcription and splicing in budding yeast also revealed a strong positive correlation between splicing efficiency and transcription activity. Experiments showed transient accumulation of phosphorylated RNAPII around the 3' ends of introns. This pausing coincides with recruitment of spliceosomal components, and it is lost upon mutation of intron (Alexander, Innocente, Barrass, & Beggs, 2010). The similar output was observed in cells stably expressing HIV-1 or b-globin mRNAs (Damgaard et al., 2008).

Another confirmation comes from an *in vitro* study on HIV-1 template that shows spliceosomal snRNPs interacting with human transcription elongation factor TAT-SF1. This interaction was strongly stimulating polymerase II elongation, possibly through TAT-SF1 binding to elongation factor P-TEFb, a component of transcription elongation complex (Fong & Zhou, 2001). These results, all together with U1 snRNP being an important component of spliceosome, necessary for the very first step of its assembly, have triggered several research projects with focus of the role of U1 in modulating transcription. In one of them, authors demonstrated that a general transcription factor TFIIH is specifically associating with U1 snRNA, stimulating the rate of initiation by RNAPII. In an *in vitro* studies they showed that U1 stimulates the rate of formation of the first phosphodiester bond by polymerase. Additionally, after the first round of transcription, the presence of the 5' splice site increased efficiency of reinitiation, in the presence of U1 and TFIIH (Kwek et al., 2002).

Confirming the previously mentioned study, Proudfoot group demonstrated that inactivation of the 5' splice site reduces the level of nascent transcription in HeLa cells. Moreover, moving the 5' splice site away from the site of transcription initiation by inserting a spacer also negatively affects the steady-state RNA levels, as shown in a minigene system within the same study (Furger et al., 2002).

What is especially interesting, a 5' splice site can stimulate transcription even in the absence of splicing. Chromatin immunoprecipitation experiments done by Damgaard et al show that a gene that contains a functional 5' splice site has an enhanced promoter docking of transcription initiation factors TFIIB, TFIID and TFIIH, of which first two are specifically recruited to the 5' splice site (Damgaard et al., 2008).

In addition to its roles in splicing and transcription, U1 snRNP has also been implicated in regulation of polyadenylation in many genes (Spraggon & Cartegni, 2013). Adjacent to the

study done by Furger et al, the same group demonstrated in HIV-1 model that poly(A) site located in 5' long terminal repeat (LTR) of HIV is inhibited by U1 snRNP binding to the downstream major splice donor site (Ashe, Pearson, & Proudfoot, 1997). The occlusion of this poly(A) signal does not require the homologous HIV-1 promoter nor the close proximity of this signal to the transcription initiation site. Instead, the major splice donor site ~200bp downstream from the poly(A) signal keeps it in inactive state. If this donor site is mutated, this poly(A) signal starts being used, and additionally, if the U1 snRNA is loaded by base pairing near to the mutated donor splice site, signal is again suppressed. Described effect was shown to be dependent on the stem-loop 1 of U1, that binds the 70K protein (Ashe, Furger, & Proudfoot, 2000). Another *in vitro* study on repression of papillomavirus late gene expression showed that 70K, but not U1A, was responsible for interesting with and inhibiting poly(A) polymerase (Gunderson, Polycarpou-Schwarz, & Mattaj, 1998).

However, not all studies agree on which component of U1snRNP is responsible for the observed effects on APA regulation. It has been reported that the pre-mRNA encoding for human U1A protein contains a similar or identical sequence to the one that U1A binds in mature U1snRNA. In excess of the protein, U1A binds the stem-loop elements on its own pre-mRNA and stops the RNA processing through interaction with poly(A) polymerase. In this way U1A is down-regulating the amount of its own U1 mRNA and protein levels (Boelens et al., 1993; Gunderson, Vagner, Polycarpou-Schwarz, & Mattaj, 1997). Moreover, human poly(A) polymerase gamma (hsPAPγ) has been shown to contain a U1A protein-interaction site within its C terminus, through which U1A inhibits the activity of hsPAPγ (Kyriakopoulou, Nordvarg, & Virtanen, 2001). One study performed in cells describes a form of U1A protein which is not associated with U1snRNP, but it is rather a part of a protein complex called SF-A (snRNP-free U1A), which also includes PSF, p54 and p68 proteins (Liang & Lutz, 2006). When this complex is inhibited by antibodies, the output is *in vitro* inhibition of splicing and polyadenylation, suggesting that U1A itself might be important for these processes.

Development of new high throughput methods gave researchers a possibility to access the global effects of U1 on polyadenylation. One such study showed that U1 depletion results in termination of most nascent gene transcripts within ~1kb, whereas the moderate decrease that is insufficient to inhibit splicing, had a dose-dependent effect on shifting cleavage and polyadenylation. These changes are characteristic for neural cell differentiation, and were recapitulated by U1 decrease and antagonized by U1 overexpression in experiments done by Dreyfuss group (Berg et al., 2012). As they demonstrate in this study, the global activation of transcription can cause a transient U1 shortage (U1 to transcripts ratio), which in turn results in changes of PAS selection, frequently in introns near (~5kb) the start of the transcript. This

effect does not arise from U1's effect on splicing, according to their further experiments, as it is not appearing when the splicing is blocked through inhibition of U2 snRNP. They also show that U1 snRNAs base-pairing with mRNA far away from the 5' splice site, in a position where they do not have an effect on splicing, still have a protective role against activation of nearby intronic (cryptic) polyadenylation signals (Kaida et al., 2010). These observations were confirmed in a study by Cartegni group, working on receptor tyrosine kinases (RTK). Blocking of U1 snRNA can activate intronic polyadenylation sites, resulting in production of at least 31 alternative RTK isoforms (Vorlova et al., 2011). In these experiments, functional depletion of U1 with antisense oligonucleotides complementary to its 5'tail, induced the activation of intronic polyadenylation sites with changes in the regulation of alternative polyadenylation (Kaida et al., 2010; Vorlova et al., 2011).

1.4 Modified U1 as a tool for gene therapy

Modified U1 snRNA has potential to be used as gene therapy tool in diseases associated with exon skipping. One of the first improvements of the U1 snRNA–mediated rescue of donor splice sites at protein level was obtained on coagulation factor VII. The mutated FVII gene splicing was partially rescued by using an engineered small nuclear U1snRNA, with 5'end sequence changed to complementary fit the mutant 5'splice site (Pinotti et al., 2008).

Exon-specific U1 snRNAs (ExSpeU1) are specially modified human U1 snRNAs that are complementary to non-conserved sequences downstream of mutant donor splice sites in target mRNA. A single ExSpeU1 can rescue multiple splicing defects that affect a single exon (Franco Pagani & Pinotti, 2011). The reason for such a design is reducing off-targets: if modified U1 snRNAs differs for only one (mutated) base from the natural one, it could potentially interfere with the maturation of transcripts generated from other functional wild-type genes. Mechanism of these ExSpeU1 molecules is not strictly antisense, because they are active even when the endogenous U1 is not loaded on the 5'splice site (Alanis et al., 2012).

The target nucleotide sequence of the modified U1 snRNA molecule is located in a region of the pre-mRNA between 2 and 50 base pairs downstream of an exon/intron junction site (5'ss), so the junction site is not compromised. Preferably, the target nucleotide sequence is 5 to 50 nucleotides (more preferably 9-30 nt) in length (Franco Pagani & Pinotti, 2011).

Alanis and co-workers demonstrated the therapeutic splicing rescue by using ExSpeU1 snRNA on 3 different models: F9 exon 5 (involved in haemophilia B), CFTR exon 12 (involved

in cystic fibrosis) and SMN2 exon 7 (inclusion of this exon can be a new therapeutic strategy in patients with no functional SMN1 gene) (Alanis et al., 2012). ExSpeU1 was active on several 5'ss mutations in CFTR, on two transversions at the polypyrimidine tract in F9, on two exonic substitutions in CFTR and on defective SMN2 exon 7 splicing. In the model of haemophilia B, 16 disease causing mutations were completely rescued by ExSpeU1, through an SRSF2-dependent enhancement mechanism, resulting in a complete rescue and coagulation activity (Tajnik et al., 2016). Another good example of splicing rescue by ExSpe U1 is performed on mutations in exon 11 of SPINK gene known to be responsible for Netherton syndrom (Dal Mas et al., 2015).

The same strategy was brought to *in vivo* experiments, showing a striking result that a single chromosomal integrated copy of ExSpeU1 is able to significantly correct endogenous SMN2 exon7 splicing, thus restoring the SMN protein levels in SMA mice (Dal Mas, Rogalska, Bussani, & Pagani, 2015). The same group developed mouse model for severe spinal muscular atrophy and introduced ExSpeU1 by germline transgenesis, to consequently observe increased SMN2 exon7 inclusion and production of SMN protein. The lifespan of transgenic mice was significantly extended: affected animals died within 4-6 days from birth, while ~60% of rescued mice were alive on the last day of recording (250 days), without significant differences between hemizygote and homozygote for transgenic animals. *In vitro*, they show that most of the splicing rescue was mediated through 70K and stem-loop 4 elements of U1 (Rogalska et al., 2016).

1.5 Coagulation factor 7 deficiency

Inherited coagulation factor 7 (F7) deficiency is characterised by autosomal recessive inheritance, and it is not associated with complete absence of functional F7, as it would be incompatible with life (Rosen et al., 1997). This haemorrhagic disorder is clinically heterogeneous, and it can vary from almost asymptomatic to mild or lethal forms. Majority of the cases are mild, with symptoms that mimic a platelet disorder (Mariani & Bernardi, 2009).

The gene that encodes coagulation F7 is located on chromosome 13 (13q34) upstream from the F10, consists of 9 exons and spans around 12kb. F7 protein is a serine protease, produced in liver and delivered to blood, where it circulates in concentration of approximately 0.5µg/mL (10nM) (McVey, Boswell, Mumford, Kemball-Cook, & Tuddenham, 2001). This zymogen form is a single chain polypeptide that gets activated by proteolytic cleavage at Arg152-Ile153, separating the light chain (contains the gamma-carboxyglutamic acid domain

and two epidermal growth factor-like domains), from the heavy chain which contains the catalytic domain. F7 synthesis is dependent on vitamin K, and can be significantly decreased with usage of warfarin or similar anticoagulant (Leissinger, Blatt, Hoots, & Ewenstein, 2008).

Upon vessel injury, tissue factor (TF) that is normally not in touch with bloodstream now gets exposed and binds the circulating F7. Different proteases of coagulation factors (10a, 9a, 12a and F7a-TF) convert F7 to its active form, F7a. The active complex F7a-TF catalyses the conversion of factors 9 and 10 into active proteases, 9a and 10a (Wajima, Isbister, & Duffull, 2009).

Mutations of the F7 gene are most frequently (70-80%) missense changes, and some of them result in changes of splicing sites (Bernardi et al., 1993), especially in intron 7. Nevertheless, there are several described polymorphisms that are known affect plasma level of F7 in general population (Bernardi et al., 1996). In unpublished study, Pinotti et al have identified 4 patients with reduced plasma level of functional F7 as a consequence of mutation within the first donor splice site. The homozygous mutation -1G>A, affecting F7 pro-peptide, was associated with low F7 coagulant activity F7c and antigen F7ag levels of 4%. Two related heterozygous patients carrying +5G>A mutation showed F7 levels of 20% and one individual with +4C>T substitution had reduced F7 levels to 43%. All subjects showed a mild to asymptomatic haemorrhagic phenotype.

2. THE AIMS OF THE THESIS

The aim of my thesis is to understand how promoter proximal 5'ss mutations affect pre mRNA processing and evaluate the potential therapeutic effect of modified U1 snRNAs. To be more specific, I address the following:

- Characterise the effect of promoter proximal 5'ss mutations on spliced transcript and polyadenylation selection using F7 minigenes
- Identify the modified U1 snRNAs that rescue the pre-mRNA processing defects caused by promoter proximal 5'ss mutations
- Characterise the structural elements of modified U1 snRNAPs involved in the pre-mRNA rescue
- Evaluate the potential therapeutic effect of modified U1 snRNAPs in other gene systems

3. MATERIAL AND METHODS

2.1 Chemical reagents

Commercial reagents purchased are listed in table below:

Purpose	Reagent	Manufacturer
Standard solutions	NaCl	Riedel-de Haën™
	Na_2HPO_4	Riedel-de Haën™
	KCl	Riedel-de Haën™
	KH_2PO_4	Riedel-de Haën™
	HEPES	Sigma
	EDTA	Sigma
	Tris	Invitrogen
	Boric acid	Sigma-Aldrich
	KOH	Riedel-de Haën™
	Agarose	Euroclone
Bacterial cells	DMSO	Riedel-de Haën™
	PEG 4000	Serva
	$MgCl_2$ x 6 H_2O	Riedel-de Haën™
Eukaryotic cells	DMEM	Gibco, Life technologies
	Antibiotic Antimycotic	Sigma
	FBS	Gibco, Life technologies
	DMSO	Sigma-Aldrich
	Zeocin	Invitrogen
	Blasticidine S hydrochloride	Sigma-Aldrich
	Hygromycin B	Invitrogen
	Anhidrotetracycline hydrochloride	Vetranal
	Effectene	QIAGEN
	Lipofectamine 2000	Thermo Fisher Scientific
RNA	TRI Reagent®	Ambion
	Chloroform	Sigma-Aldrich
	2-Propanol	Riedel-de Haën™
	Ethanol	Merck
	DNase I	Promega
	dNTPs	Rovalab

Purpose	Reagent	Manufacturer
	Reverse transcription kit	Invitrogen
	Random primer	Invitrogen
	dT oligo	Clontech
DNA	Restriction endonucleases	New England Biolabs
	CIP	New England Biolabs
	T4 DNA Ligase	Promega
	NucleoSpin® Plasmid kit	Machery-Nagel
	NucleoBond® Xtra midi kit	Machery-Nagel
	QIAquick® Gel Extraction Kit	QIAGEN
PCR	*Taq* DNA polymerase	New England Biolabs
	Taq DNA polymerase	Roche
	dNTPs	Rovalab
	SYBR Green	Bio-Rad
	Customized oligo	Sigma
Cellular fractionation	Urea enzyme grade	Invitrogen
	Sucrose	Sigma
	Glycerol	Sigma-Aldrich
	NP-40	Calbiochem
	cOmplete protease inhibitor cocktail	Roche
	DTT	Thermo Fisher Scientific
Western blot	Methanol	Merck
	Tween 20	Sigma
	SDS	Sigma-Aldrich
	Glycine	Sigma-Aldrich
	ProtoGel (30%)	National Diagnostic
	Bromophenol blue	Sigma-Aldrich
	Skimmed milk	Gramm SpA
	ECL Luminata Crescendo	Milipore
	Protran Nitrocellulose Membranes	Amersham
	Hyperfilm™ ECL	Amersham

2.2 Standard solutions

Solutions that were prepared for various protocols are listed below:

Standard solutions

- PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4
- Luria-Bertani medium (LB): 1% w/v Difco Bactotyptone, 0.5% w/v Oxoid yeas extract, 1% w/v NaCl in dH_2O, pH 7.5
- TSS: 10% w/v PEG 4000, 5% v/v DMSO, 35 mM $MgCl_2$ in sterile LB medium, pH 6.5
- 10x TBE: 108 g/L Tris, 55 g/L boric acid, 9.5 g/L EDTA
- PBST: 0.1% v/v Tween 20 in 1x PBS

Solutions for cellular fractionation

- HLB+N: 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% v/v NP-40, 1x cOmplete protease inhibitor cocktail (Roche), 0.1 mM DTT in dH_2O
- HLB+NS: 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% v/v NP-40, 10% w/v sucrose, 1x cOmplete protease inhibitor cocktail, 0.1 mM DTT in dH_2O
- NUN1: 20 mM HEPES-KOH pH 7.6, 75 mM NaCl, 0.5 mM EDTA, 50% v/v glycerol, 1x cOmplete protease inhibitor cocktail, 0.1 mM DTT in dH_2O
- NUN2: 20 mM HEPES-KOH pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 7.5 mM $MgCl_2$, 1M urea, 1% v/v NP-40, 1x cOmplete protease inhibitor cocktail, 0.1 mM DTT in dH_2O

Solutions for western blot

- 10x running buffer: 250 mM Tris, 1.92 M glycine, 1% w/v SDS in dH_20, pH ~8.3
- 10x transfer buffer: 250 mM Tris, 1.92 M glycine in dH_2O, pH ~8.3. When diluted to 1x, methanol was added to 20% v/v
- Upper Tris buffer: 0.5 M Tris-HCl pH 6.8
- Lower Tris buffer: 1.5 mM Tris pH 8.8
- 2x lysis buffer: 30 mM HEPES pH 7.5, 0.5 mM NaCl, 1% v/v NP-40, 20% v/v glycerol in dH_2O
- 5x SDS sample loading buffer: 250 mM Tris-HCl pH 6.8, 10% w/v SDS, 30% v/v glycerol, 5% β-mercaptoethanol, 0.02% w/v bromophenol blue in dH_2O

2.3 Enzymes

Restriction enzymes were purchased from New England Biolabs Inc. DNA modifying enzymes, including DNase I and CIP were purchased from Promega, and T4 DNA ligase from New England Biolabs (NEB). For 3'RACE PCR *Taq* DNA polymerase from NEB was used, and for cloning purposes *Taq* DNA polymerase by Roche. All enzymes were used following manufacturer's instructions.

2.4 Synthetic oligonucleotides

Synthetic DNA oligonucleotides for PCR and cloning were obtained from Sigma-Aldrich:

Purpose	Name	5'-3' sequence
Reverse	dT (Clontech)	AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTT
Sequencing	CMV for	CGCAAATGGGCGGTAGGCGTG
	bGH rev	TAGAAGGCACAGTCGAGG
	U1 215 for	CGTGCTTCACCACGAACCAGTTCC
	SP6 rev	ATTTAGGTGACACTATAG
Cloning	U1 -1A+4C for	GATCTCATACAAGTGCGTGCAGGGGAGATACCAT
	U1 -1A+4C rev	GATCATGGTATCTCCCCTGCACGCACTTGTATGA
	U1F7 shift+9 for	TCATAACCTCCCCGCAGGGGAGATACCATGATC
	U1F7 shift+9 rev	GATCATGGTATCTCCCCTGCGGGGAGGTTATGA
	U1wt U1Amut for	CCCAGGGCGAGGCTTATCCCTTACGCTCCGGATGTGCTG
	U1wt U1Amut rev	CAGCACATCCGGAGCGTAAGGGATAAGCCTCGCCCTGGG
	U1wt 70Kmut for	CACAAGATCTCATACTTACCTGGCAGGGGAGATACCATTGACACGA
	U1wt 70Kmut rev	CCACATTCGTGTCAATGGTATCTCCCCTGCCAAGTACGTATGAGAT
	U1-1A+4C 70Kmut	CACAAGATCTCATACGTACTTGGCAGGGGAGATACCATTGACACGA
	U1-1A+4C 70Kmut	TTCGTGTCAATGGTATCTCCCCTGCCAAGTACGTATGAGATCTTGT
	L4mut for	CATAATTTGTGGTAGTGATATACTTATTTCGATATTTATA
	L4mut rev	CTCCAGAAAGTCAATATAAATATCGAAATAAGTATATCA
PCR	CFex1 284 dir	AGACCATGCAGAGGTCGCCTCTGG
	Cfex3 292 rev	CATCGCCGAAGGGCATTAATGAGTTTAGG
	F7ex1 22 for	CTCCTCTGCCTTCTGCTTGG
	F7int1 1408 rev	CCAGAAAACCCTCCTGGTGGATG
	ex2 glob for	GATGTTCCTGTCCTTCCCCAC
	ex2 glob rev	GTTGGGCATGTCGTCCAC
	glo 800 rev	GCTCACAGAAGCCAGGAACTTGTCCAGG
	qGAPDH fw	GACAGTCAGCCGCATCTTCT
	qGAPDH rev	TTAAAAGCAGCCCTGGTGAC
	FNBP4 fw	GAGATGCTGGCTGGCTGATA
	FNBP4 rev	GCCTTGGAGGAAGGAGATGG
	ACTIN B fw	CACACTGTGCCCATCTACGA
	ACTIN B rev	CCATCTCTTGCTCGAAGTCC
	XIST fw	CCACCACACGTCAAGCTCTTC
	XIST rev	CAGGTGGGAAGGCTGACTTC
	Anchor (Clontech)	AAGCAGTGGTATCAACGCAGAGT

2.5 Bacterial competent cells

The *E.coli* K12 strain DH5α was used for transformation of plasmid DNA following the heat shock protocol.

2.6 Preparation of bacterial competent cells

Bacterial competent cells were prepared using the protocol developed by Chung and Nimela (1989) (Chung, Niemela, & Miller, 1989). *E. coli* DH5α strain was grown in 10 mL of Luria-Bertani (LB) medium overnight at 37°C with vigorous shaking. On the following day, 140 mL of fresh medium were added, and cells were grown in the shaker at 37°C until the optical density measured at 600 nm (OD_{600}) was between 0.3 and 0.4. Cells were then placed in pre-cooled 50 mL tubes, and centrifuged at 180 xg (Eppendorf centrifuge 5804/5804 R) for 10 min at 4°C. Pellet was re-suspended in ice-cold TSS solution in 10% of the initial volume, aliquoted and rapidly frozen in liquid nitrogen. Competent were stored at -80°C, and their competence was determined by transforming with 100 pg of pUC19 in 100 µL of freshly thawed cells. If obtained number of colonies was exciding 10^2, the competency was considered to be satisfactory.

2.6.1 **Transformation of bacteria**

Briefly, for re-transformation 30 µL of freshly thawed DH5α were supplied with 1ng of plasmid DNA and incubated in ice for 30min. The heat shock was carried out at 42°C for 30 sec. In order to cool down the sample, the tube was immediately placed on ice for 5 min, and then plated on LB medium containing 1% agar, with addition of 100 µg/mL ampicillin as a selection marker. Plates were placed in 37°C incubator overnight, and next day selected colony was expanded in 10mL of liquid LB medium with ampicillin for small scale plasmid DNA purification, or 100mL for medium scale purification, and samples were incubated overnight in shaker at 37°C. Next day overnight cultures were centrifuged at 4500 xg (Eppendorf centrifuge 5804/5804 R) and pellet was used as input for plasmid purification kits.

For cloning purposes, 70µL of freshly thawed competent cells were transformed with inactivated ligation reaction using previously described heat shock protocol, and cells were plated on LB agar containing 100 µg/mL ampicillin.

2.7 Small and medium scale plasmid DNA purification

Pellets obtained by centrifuging the overnight culture of transformed and re-transformed DH5α competent cells were processed using kits according to the manufacturer's instructions. For small scale plasmid DNA purification was used NucleoSpin® Plasmid kit (Macherey-Nagel, Germany), yielding 20-30 μg of product starting from 10mL of overnight culture. For medium scale plasmid DNA purification was used NucleoBond® Xtra midi (Machery-Nagel, Germany), yielding ~250-350 μg of product starting from 100 mL of overnight culture.

After purification, plasmid DNA was stored in aqueous solution at -20°C. Concentrations were measured on NanoDrop 1000 (Thermo Scientific) spectrophotometer, and quality was checked on 1% agarose gel containing 0.5 μg/mL of ethidium bromide in 1x TBE buffer, and ran at 70-80 mA until desired separation level. Gels were visualized by UV transillumination and the result was recorded using Gel Logic 100 imaging system (Kodak).

2.8 Enzymatic modification of DNA

2.8.1 **Restriction endonucleases**

For the purpose of construction and analysis of recombinant plasmids, restriction endonucleases were used, purchased from NEB and supplied with their corresponding buffers for their optimal function.

For analytical purposes, 300ng of DNA were digested by 3U of appropriate enzyme in a 20 μL reaction, during 3 hours at 37°C, and bands were resolved on 1% agarose gel containing ethidium bromide and visualized under the UV light.

Preparative digestions for construction of recombinant plasmids were made using 3 μg of plasmid DNA and 30U of restriction enzyme in final volume of 50 μL.

2.8.2 **Calf intestinal phosphatase (CIP)**

Calf intestinal phosphatase (CIP) provided from NEB was used for removal of 5' phosphate groups from DNA. Fragments lacking the 5' phosphoryl termini required by ligases cannot self-ligate, and this property was used in preparing the vector for cloning strategies to reduce the background. The standard reaction was carried out in 50 µL of final volume using 2U of enzyme per 800 ng of linearized vector, for 30 minutes at 37°C, followed by inactivation 10 min at 65°C.

2.8.3 **T4 DNA ligase**

T4 DNA ligase provided from NEB was used for to join double stranded DNA fragments with compatible sticky ends during generation of recombinant plasmids. This enzyme catalyzes the formation of phosphodiester bonds between adjacent 3' hydroxyl and 5' phosphoryl termini in DNA, and 1U of it was used in reaction containing 20-40 ng of linearized vector and 3-fold molar excess of insert, in final volume of 20 µL. Reactions was carried out for 3 hours at room temperature for inserts up to 500 bp in length, or at 16°C overnight for larger fragments. Ligation reaction was deactivated at 75°C for 20 min prior to transformation in DH5α competent cells.

2.9 Elution and purification of DNA fragments from agarose gels

For the purpose of recombinant plasmid construction, linearized vectors or digested inserts were purified after separation on low percentage agarose gel. In brief, DNA samples obtained through PCR or enzymatic digestion in total volume of 50 µL, were separated on 0.8% agarose gel containing 0.5 µg/mL of ethidium bromide in 1x TBE buffer using 70mA current. DNA was visualized under the UV light, and the required fragment was excised from the gel and processed using QIAquick® Gel Extraction Kit (QIAGEN, Venlo, Netherlands) using manufacturer's directions. Finally, samples were eluted in 20-50 µL of dH_2O, measured on

NanoDrop 1000 (Thermo Scientific) spectrophotometer to determine concentration and small portion was re-loaded on gel to check the quality of extraction.

2.10 Amplification of DNA fragments for cloning purposes

The polymerase chain reaction (PCR) was performed on plasmid DNA following the basic protocols of the Roche Diagnostic *Taq* DNA polymerases in final volume of 50 µL. The content of reaction mix was following: 1x *Taq* buffer, 200 µM of each dNTP, 20 µM of each of two primers, 2.5U of *Taq* polymerase and 100 pg of plasmid DNA. The amplifications were performed on a Gene Amp PCR System (Applied Biosystems) using standard conditions listed further in text.

2.11 Plasmids used in this study

2.11.1 **Construction of minigene systems**

Minigene reporters are widely used and well characterized systems for studying alternative splicing. They usually contain few exons and introns of the gene of interest, and after being transfected in human cell lines, they express RNAs which can be tracked by semi quantitative (such as RT-PCR pattern of splicing) or quantitative (qPCR, NB) methods.

For this study, several minigenes were constructed in pcDNA3 vector containing exon 1 of coagulation factor 7 gene followed by exons 2 and 3 of α-globin (Figure 3-1), under control of CMV promoter and with bGH polyA signal at the end of the gene.

First, the piece containing the 5'UTR, exon 1 and 5' portion of intron 1 of human coagulation factor 7 was cloned in pcDNA3 vector using restriction enzymes *HindIII* and *BamHI*, giving an intermediary construct pcDNA3-F7ex1 (Figure 3-1 A). Then, the piece containing α-globin sequence (3' portion of intron 1, exon 2, intron 2 and exon 3) was cloned in pcDNA3-F7ex1 using endonucleases BamHI and NotI, giving the second intermediary plasmid (Figure 3-1 B).

This plasmid was further extended at the position of *BamHI* restriction enzyme with a piece of sequence coming from intron 24 of human fibronectin gene, expanding the intron 1 to the total length of 480 bp (short intron) or 1.2 kb (long intron) (Figure 3-1 C). In both short and long constructs this extension was cloned in two possible orientations to distinguish the possible effect of sequence from the effect of intron 1 length.

Each of aforementioned constructs was made in its wildtype (WT) form of exon 1, as well as the several mutant forms, corresponding to patient mutations: -1G/A, +4C/T and +5G/A (Figure 3-1 D). These mutations were introduced by PCR-mediated site directed mutagenesis.

For a separate study of the effect of 5'ss mutations on polyadenylation, in each construct was added an artificial AATAA polyadenylation site present in glucocorticoid receptor isoform β (Kaida et al., 2010) (Figure 3-1 E).

For creating cell lines stably expressing the construct of interest, WT and -1A mutant minigenes were cloned into pcDNA5-FRT-TO plasmid, keeping the same promoter and polyA signal at the end of the gene (Figure 3-1 F).

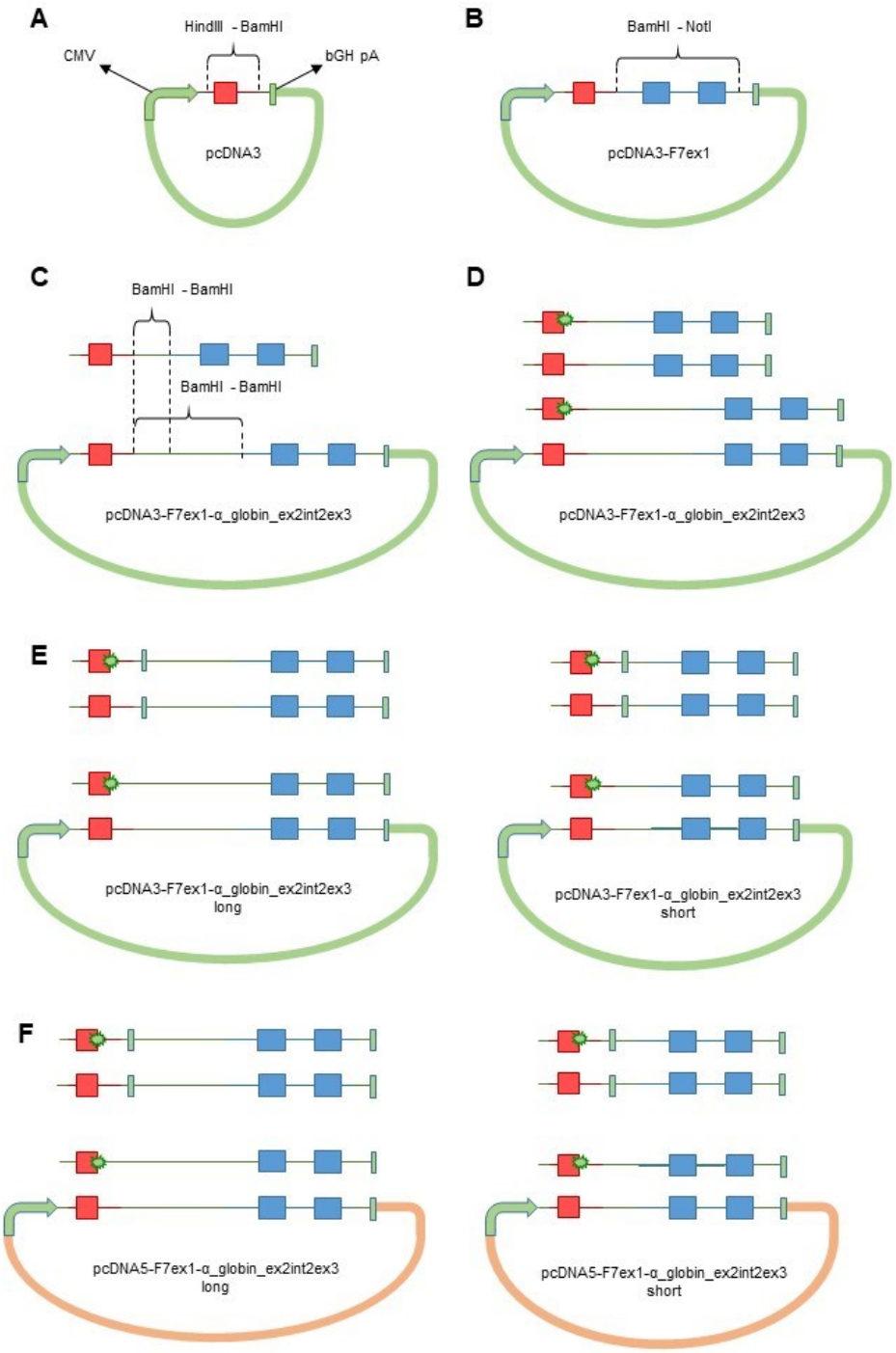

Figure 3-1. Construction of minigene systems.

pcDNA3was cut *HindIII-BamHI* to insert F7ex1 in WT form. This intermediate construct was then cut *BamHI-NotI* to insert α-globin ex2-int2-ex3. *BamHI* restriction site in first intron was used to expand the length up to 0.48 or 1.2 kb. For the purpose of studying polyadenylation mechanisms connected to 5'ss, several mutants of F7ex1 were created, as well as artificial polyA signal in first intron.

2.11.2 **pSV-CFTR-ex123**

Plasmid pSV-CFTR-ex123 has been previously described (F. Pagani, Raponi, & Baralle, 2005; Franco Pagani, Buratti, Stuani, & Baralle, 2003), and in this study it has been used in minigene transfection experiments as transfection control. In particular, 100 ng of this plasmid is transfected along with minigenes in each sample, the level of RNA transcribed from it was measured by qPCR and 3'RACE PCR methods and used for normalization of samples.

The equal amount of plasmid per each well was ensured by making dilution of this plasmid in transfection buffer, then 300 µL was aliquoted per each technical duplicate in separate tubes, minigenes were added and content of each tube was well mixed, and separated in single samples.

2.11.3 **pGEM-U1**

The U1 snRNA WT form of gene is cloned as BamHI-BamHI cassette into pGEM3 vector. For creating modified U1s, the 5' tail complementary to 5'ss between BglII and BclI restriction sites was replaced with specific annealed oligo as previously described (Alanis et al., 2012). The identity was ultimately confirmed through sequencing analysis (Sanger sequencing by GATC LightRun).

U1 snRNA mutants unable to bind U1A or 70K protein, and stem-loop4 mutant were created through overlapping PCR as previously described (Rogalska et al., 2016), with primers listed in Table below. Constructs were confirmed by sequencing.

Purpose	Name	5'-3' sequence
Cloning	U1F7 -1A+4C for	GATCTCATACAAGTGCGTGCAGGGGAGATACCAT
	U1F7 -1A+4C rev	GATCATGGTATCTCCCCTGCACGCACTTGTATGA
	U1F7 shift+9 for	
	U1F7 shift+9 rev	GATCATGGTATCTCCCCTGCGGGGAGGTTATGA
	U1wt U1Amut for	CCCAGGGCGAGGCTTATCCCTTACGCTCCGGATGTGCTG
	U1wt U1Amut rev	CAGCACATCCGGAGCGTAAGGGATAAGCCTCGCCCTGGG
	U1wt 70Kmut for	CACAAGATCTCATACTTACCTGGCAGGGGAGATACCATTGACAC
	U1wt 70Kmut rev	CCACATTCGTGTCAATGGTATCTCCCCTGCCAAGTACGTATGAG
	U1-1A+4C 70Kmut for	CACAAGATCTCATACGTACTTGGCAGGGGAGATACCATTGACA
	U1-1A+4C 70Kmut rev	TTCGTGTCAATGGTATCTCCCCTGCCAAGTACGTATGAGATCTT
	L4mut for	CATAATTTGTGGTAGTGATATACTTATTTCGATATTTATA
	L4mut rev	CTCCAGAAAGTCAATATAAATATCGAAATAAGTATATCA
Sequencing	U1 215 for	CGTGCTTCACCACGAACCAGTTCC
	SP6 rev	ATTTAGGTGACACTATAG

2.12 Eukaryotic cell lines

The cell line used for transfection and co-transfection experiments was HeLa. For creating stable clones Flp-In™ 293 T-REx cell line was used.

2.13 Maintenance of HeLa cells and transient transfection

HeLa cells were grown as adherent monolayer in Dulbecco's modified Eagle's medium with Glutamax I (Gibco, Life Technologies) (DMEM with glutamine, sodium pyruvate, pyridoxine and 4.5 g/l glucose) with addition of 10% fetal bovine serum (FBS) (Gibco, Life Technologies) and Antibiotic Antimycotic (Sigma). Cells were kept in humidified incubator at 37°C and 5% of CO_2.

24 hr transient transfection was performed on HeLa grown on six well plates using Effectene reagents (QIAGEN) according to manufacturer's protocol; cells were transfected with F7 minigene construct (0.03 pmol per well), and U1 construct (0.3 pmol per well for 5'ss and complementary U1, 0.6 pmol for L1, L2 and L4 mutants).

Total RNA was extracted using TRI Reagent (Ambion).

2.14 Working with Flp-In™ 293 T-REx cell line

2.14.1 **Creating and maintaining clones with stable integration of minigene system**

To study the role of 5'ss mutants in chromatin context, I used the advantage of Flp-In system (Thermo Fisher Scientific) to create a stable cell line expressing the construct of interest. In brief, laboratory for Human Molecular Genetic, in which I performed experiments for this thesis, has purchased Flp-In™ 293 T-REx cell line (Cat.no. R78007) that contains a single stably integrated FRT site at a transcriptionally active genomic locus, that is maintained in culture by supplementing the growth medium with antibiotic Zeocin™ (Invitrogen). Besides the integration site, these cells as well express Tet (Tetracycline resistance-encoding transposon) repressor protein that blocks the activity of CMV promoter, and selection of clones that contain high expression of Tet (and thus prevent leakage of promoter) is performed by keeping them in growth medium in the presence of 5 µg/mL of Blasticidine.

Next, I have cloned the mini gene of interest in pcDNA5™-FRT-TO plasmid containing CMV promoter, and co-transfected it with the Flp recombinase vector, pOG44. Co-transfection was performed using Lipofectamine 2000 (Thermo Fisher Scientific), in ratio pcDNA5:pOG44 1:9, and 24hrs post-transfection passed into p100. Another 24hrs later I started selection of clones containing integrated copy of mini gene by introducing 100 µg/mL of Hygromycin B into medium. Selection lasted ~10 days, until all non-transfected Flp-In™ 293 T-REx died. Since the integration site is cut and used for inserting a copy of gene of interest, these stable clones lose the resistance to Zeocin™. In fact, my stable clones are containing a single copy of mini gene, that is under control of Tetracycline-inducible CMV promoter. These clones are sensitive to Zeocin™, and resistant to Blasticidine and Hygromycin B.

These cells were frozen in FBS (fetal bovine serum) (Gibco, Life Technologies) with 5% of DMSO (Sigma-Aldrich), and when needed, thawed fast in DMEM with 10% FBS and Antibiotic Antimycotic (Sigma), in further text $DMEM^{++}$, in humidified incubator at 37°C and 5% of CO_2. After cells attached to the surface, medium was changed for the one containing 5 µg/mL Blasticidine (Sigma-Aldrich) and 100 µg/mL Hygromycin B (Invitrogen). In order to avoid

overgrowing, when cells would reach ~90% confluence, they are passed into new 1p100 in dilution from 1:6 to 1:8. In particular, medium is removed, and cells washed 2x PBS, and treated with 1 mL of 2x trypsin in EDTA for 5min in incubator at 37°. Trypsin is neutralized with 5mL of DMEM^{++} and centrifuged at 2800 rpm 7min. Cellular pellet is then re-suspended in 3 or 4 mL, and 0.5 mL is plated in new 1p100 culture dish in final volume of 10 mL and placed back to incubator.

2.14.2 **Induction of expression and transient transfection in stable clones**

All experiments in stable clones were performed on passages 2-4, to ensure reliable biological replicates and comparability of results obtained in different experiments.

For induction experiments, cells were plated in 6 well plates in technical replicates at 40% confluence in DMEM^{++} with Blasticidine and Hygromycin B. Next day, they are gently washed with PBS and selection medium was added, composed of DMEM^{++} with 5 µg/mL Blasticidine and 100 µg/mL Hygromycin B. At various time points, 0, 4, 8, 12, 16 and 24hrs, certain wells were washed 2x with PBS and cells were collected in TRI Reagent (Ambion) for further RNA extraction.

For transfection experiments, the day before cells were plated into 6 well plates in technical replicates in DMEM^{++} with Blasticidine and Hygromycin B at ~80% confluence. On the day of transfection medium is changed to DMEM^{++}, and transfection was performed using Effectene reagent (QIAGEN). U1 wt, U1 5'ss, U1 shift+9 and U1 sm25 were transfected with 1 µg per 6 well, while U1A, 70K and L4 mutants of corresponding U1s were transfected with 2 µg per 6 well, as it is known that mutants have lower expression from their corresponding fully assembled U1s. 1 and 2 µg per 6 well are chosen to ensure high transfection efficiency (>80%) and maximally saturated expression of U1s. 4-6 hours after transfection cells are gently washed with PBS once, and medium is again changed to DMEM^{++} with Blasticidine + Hygromycin B. 24 hours' post-transfection, cells were induced in the next 24hrs with 1 µg/mL of Tetracycline (Vetranal) in DMEM^{++}, and thereafter collected in TRI reagent (Ambion) for RNA extraction.

2.15 Cellular fractionation

Cellular fractionation is performed according to modified protocol of Proudfoot. In brief, Flp-In™ 293 T-REx cells and stable clones plated in 1p100 were washed with cold PBS, scraped and counted, and $1x10^7$ cells was used as input. Cells were lysed on ice with cold HLB+N buffer for 5min, and then gently underlaid with HLB+NS buffer that contains sucrose and serves as a cushion. Samples were centrifuged and supernatant was collected as cytoplasmic fraction. Pellets were checked under microscope to confirm the integrity of nuclei, and then re-suspended in buffer NUN1 containing glycerol. Buffer NUN2, containing strong detergent, was added and samples were incubated in ice 15min with occasional vortexing, with visible presence of chromatin strings. Samples were centrifuged on low speed, and supernatant was taken as nucleoplasmic fraction, while pellets (chromatin fraction) were washed once with PBS and re-suspended in 100 µL of 1x DNase buffer (Invitrogen) and treated with 10U of DNase. RNA from all three fractions, cytoplasmic (CP), nucleoplasmic (NP) and chromatin (CHR), were extracted using TRI reagent.

Reliability of each fraction was confirmed on RNA and protein level. cDNA from fractions was tested in qPCR with oligo pairs probing transcripts previously published: cytoplasm was tested for enrichment in spliced RNA of GAPDH (Rogalska et al., 2016), nucleoplasm was tested for enrichment of Actb (Bhatt et al., 2012; Werner & Ruthenburg, 2015), and chromatin was tested for enrichment with Xist transcript (Bhatt et al., 2012; Werner & Ruthenburg, 2015). On protein level, Western blot analysis was performed on equal portions of each fraction (method described in (Mayer & Churchman, 2016)). Cytoplasmic fraction was checked for enrichment with GAPDH, nucleoplasmic with 70K protein of U1 snRNP, and chromatin fraction was checked for enrichment with Histone 2B.

2.16 RNA extraction and DNase I treatment

RNA extraction was performed using TRIzol reagent (Ambion) according to manufacturer's instructions. Cells were washed with 1xPBS and resuspended in 750 µL of TRIzol reagent. After adding 150 µL of chloroform, samples were mixed by vortex, incubated 10 minutes at RT and centrifuged at 14000 rpm for 25 minutes at 4°C to separate the phases. 300 µL of the

aqueous phase containing RNA was collected in a new tube. For the RNA precipitation, 400 µL of isopropyl alcohol were added and the solution was incubated in dry ice for 30 minutes to facilitate precipitation. Then, samples were centrifuged at 14000 rpm for 30 minutes at 4°C. Supernatant was eliminated and pellet was washed in 150 µL 75% ethanol, then centrifuged for 10 minutes. At the end, pellet was air-dried and resuspended in 23 µL of autoclaved dH2O. Quality of RNA was checked by electrophoresis on a 0.8% agarose gel and quantified using NanoDrop 1000 (Thermo Scientific) spectrophotometer.

Extracted RNA was treated with DNase I (Thermo Fisher Scientific) to remove any traces of genomic and plasmid DNA. In final reaction volume of 10 µL 1x DNase I buffer, 2 U of RNase-free DNase I and 800 ng of RNA. Reaction was incubated at 37°C for 30 min, and inactivated by adding Stop Solution followed by incubation at 65°C for 10 min.

RNA was stored at -80°C, and prior to experiment was thawed in ice.

2.17 cDNA preparation

The first-strand cDNA synthesis for quantitative PCR was performed with M-MLV Reverse Transcriptase Kit (Invitrogen) according to manufacturer's instructions. Following components were combined: 200 ng of DNase I treated RNA, 200 ng of random primer (Pharmacia) and sterile water to reach the final volume of 12 µL. The mixture was then denatured at 95°C for 2 min, and let to cool down at room temperature. After this step another set of solutions was added, composed of 1x first strand buffer, 10 mM DTT, 1 mM dNTPs and 200 U of reverse transcriptase enzyme called M-MLV (Moloney Murine Leukemia Virus). Reaction was incubated at 37° for 90 minutes.

The first-strand cDNA synthesis for rapid amplification of 3' cDNA ends was performed in the same way, just with oligo dT (Clontech) replacing the random primer.

21.6% of the first strand reaction was used in triplicate mixture as a template for qPCR (~7% per sample), while 8% of the first strand reaction per sample was used as a template for 3'RACE PCR.

2.18 Real time Quantitative Polymerase Chain Reaction (qPCR)

Real time PCR was performed using SYBR® Green Master Mix (Bio-Rad) on CFX384 Touch™ Real-Time PCR Detection System (Bio-Rad). Composition of the qPCR mix per reaction was 7.5 µL of SYBR Green Master Mix, 5.1 µL of sterile water and 0.3 µL of forward and 0.3 µL of reverse primer. First, mix per each primer set was made, and 39.6 µL was aliquoted in sterile 0.5 mL tubes. Then, into each tube was added 5.4 0.3 µL of cDNA, mixture was well pipetted and aliquoted 13 µL per well. Melt curve for primer pairs was assessed in each reaction to ensure the specificity of aplification for every sample.

Protocol used is following:

1: 98.0°C for 0:30
2: 95.0°C for 0:05
3: 60.0°C for 0:25
Plate Read
4: GOTO 2, 39 more times
5: 95.0°C for 0:10
6: Melt Curve 66.0°C to 95.0°C: Increment 0.5°C 0:01
Plate Read

Results were analyzed in MC Office Excel and graphically represented, showing the mean of minimum 3 biological replicates with 2 technical replicates each, and standard deviation.

Samples comprised of total RNA were normalized to GAPDH, for cellular fractionation CP, NP and CHR fractions were normalized to FNBP4 (Werner & Ruthenburg, 2015). Transcripts in chromatin fraction were normalized to Xist.

In the table below are given primers used for real time quantitative PCR:

Target	Name	Sequence 5'-3'
pSV-CFTR	CFex1 284 dir	AGACCATGCAGAGGTCGCCTCTGG
	Cfex3 292 rev	CATCGCCGAAGGGCATTAATGAGTTTAGG
F7	F7ex1 22 for	CTCCTCTGCCTTCTGCTTGG
	F7int1 1408 rev	CCAGAAAACCCTCCTGGTGGATG
α globin	ex2 glob for	GATGTTCCTGTCCTTCCCCAC
	ex2 glob rev	GTTGGGCATGTCGTCCAC
	glo 800 rev	GCTCACAGAAGCCAGGAACTTGTCCAGG
GAPDH	qGAPDH fw	GACAGTCAGCCGCATCTTCT
	qGAPDH rev	TTAAAAGCAGCCCTGGTGAC
FNBP4	FNBP4 fw	GAGATGCTGGCTGGCTGATA
	FNBP4 rev	GCCTTGGAGGAAGGAGATGG
ActB	ACTIN B fw	CACACTGTGCCCATCTACGA
	ACTIN B rev	CCATCTCTTGCTCGAAGTCC
Xist	XIST fw	CCACCACACGTCAAGCTCTTC
	XIST rev	CAGGTGGGAAGGCTGACTTC

2.19 Rapid Amplification of cDNA 3' Ends (3' RACE)

To amplify only the polyadenylated transcripts, F7ex1 22 for was used as a forward primer, and Anchor (Clontech) as a reverse. Anchor specifically anneals with cDNA synthetized with oligo dT.

PCR reaction was performed in final volume of 25 µL, with final concentration of 1x standard *Taq* buffer (NEB), 1 mM dNTPs, 3 ng/mL of forward and reverse primers, and 0.625 U of *Taq* DNA polymerase (NEB), containing 2 µL of cDNA.

Program was set to include 5 min at 94°C, and 35 cycles composed of 45 seconds at 94°C, 45 seconds at 65°C and 45 seconds at 72°C, final elongation was 10 min at 72°C.

Samples were resolved at 1.5% agarose gel containing ethidium bromide, and result was visualized under UV light and recorded with Gel Logic 100 imaging system (Kodak). Strength of bands was quantified using ImageJ software (https://imagej.net).

Oligos used for 3' RACE are given in the table below:

Target	Name	Sequence 5'-3'
PolyA tail	Anchor (Clontech)	AAGCAGTGGTATCAACGCAGAGT
CFTRex1	CFex1 284 dir	AGACCATGCAGAGGTCGCCTCTGG
F7ex1	F7ex1 22 for	CTCCTCTGCCTTCTGCTTGG
α globin exon2	ex2 glob for	GATGTTCCTGTCCTTCCCCAC

2.20 Protein extraction and quantification

2.20.1 **Protein extraction from cellular fractions**

Cellular fractions CP, NP and CHR were obtained starting from 1p100 material, and diluted to approximately same concentrations. Cytoplasmic fraction was collected in volume of 1.5 mL, which represents 1/3 of total volume of this fraction. Nucleoplasm was collected in volume of 0.5 mL and was diluted with 1 mL of PBS. Chromatin fraction had a total volume of 100 µL and was diluted 9x with PBS. Estimated concentration of each fraction was ~220 ng/µL.

2.20.2 **Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)**

~10 µg of protein were added to the correct amount of 2xProtein loading buffer in a final volume of 56 µL, the mixture was then incubated at 94°C for 5 minute. SDS-PAGE was performed in discontinuous Laemli buffer system using 2xRunning buffer. The SDS polyacrylamide gel was composed of 5% stacking gel (pH 6.8) and 10% running gel (pH 8.8). The gels were run at 35 mA until the bands were well separated.

Three gels were loaded and run under same conditions for incubating with different primary antibodies.

2.20.3 **Western blot**

After the proteins were separated on SDS-PAGE gel, they were electroblotted onto PVDF membrane (Amersham Biosciences) according to standard protocol. Briefly, a sandwich was prepared with 3 pieces of paper, gel, membrane (previously activated with methanol) and 3 other pieces of paper. The sandwich was wet in 1xRunning buffer and after elimination of air bubbles, it was closed and put in the blotting support filled with 1xRunning buffer. The blotting was carried out at 200 mA for 120 minutes. The membrane was then blocked by incubation in 5% w/v non-fat dry milk in 1xPBS. Membrane was probed overnight at 4°C with different primary antibodies, washed 3x 10 min with PBST, and then incubated at 1h at room temperature with corresponding secondary antibodies conjugated with horseradish peroxidase (HRP). Membranes were washed 3x 10 min with PBST, and upon adding of its substrate contained in Luminata Crescendo reagent (Millipore), chemiluminescence was captured on X-ray film (Hyperfilm^{TM} ECL, Amersham).

Primary antibody	**Producer**	**Organism and type**	**Dilution**
Anti-GAPDH	Abcam 8245	Mouse monoclonal	1:800 in 5% milk/PBS
Anti-70K	Abcam 83306	Rabbit polyclonal	1:400 in 5% milk/PBS
Anti-Histone 2B	Santa Cruz 515808	Mouse monoclonal	1:250 in 5% milk/PBS

Secondary antibody	**Producer**	**Organism and type**	**Dilution**
Anti-mouse	Dako P 0447	Goat polyclonal Ig/HRP	1:2000 in 5% milk/PBS
Anti-rabbit	Dako P 0448	Goat polyclonal Ig/HRP	1:2000 in 5% milk/PBS

4. RESULTS

3.1 Patient mutations of coagulation factor 7 cause alternative exon skipping in hybrid mini gene context

Mutations in the promoter-proximal 5' splice site in human F7 selected for this thesis were chosen from F7 mutation database and the International Registry of F7 deficiency. The homozygous mutation -1G>A (c.64G/A), present in two patients, was associated with low F7 coagulant activity and antigen levels of 4%. One individual with +4C>T substitution had F7 antigen level 43%. And finally, two related heterozygous patients were found to carry +5G>A mutation resulting in 20% of F7 antigen level. All subjects showed mild to asymptomatic haemorrhagic phenotype.

Since mutations are located within the 5'splice site consensus, the first step was to evaluate their effect on pre-mRNA processing. To study the effect of -1A, +4T and +5A patient mutations on F7 exon 1a recognition, I have created hybrid alternatively spliced exon in the well characterized minigene reporter system of pTB.

3.1.1 **-1A, +4T and +5A mutations of first exon in factor 7 cause alternative exon skipping in pTB mini gene**

For the purpose of this thesis, hybrid exon was designed by fusing 3' part of CFTR exon12 with 5' end of F7 exon 1a (in further text F7ex1a), along with their corresponding flanking intronic sequences, and cloned in *NdeI* restriction site of pTB (Figure 4-1 A). Minigenes were transfected in HeLa cells, total RNA was extracted and reverse-transcribed into cDNA. I have evaluated their pattern on splicing by semi-quantitative PCR using primers α2.3 and Bra2.

Mutations that affect 5' splice site are known to reduce or even abolish the interaction with U1snRNP and thus cause the skipping of alternative exons. The WT sequence of F7ex1a has C instead of A in position +4, compared to the canonical 5'ss. The same situation is observed in +4T mutant, where instead of A is present T. The other two mutants have an additional base of mismatch with U1wt, -1 and +4 in case of -1A, and +4 and +5 in case of +5A mutant (Figure

4-1 B). Even though the predicted strength of 5'ss (fruitfly.org) was moderate to high in -1A (0.63), +4T (0.98) and +5A (0.68) mutants, all of them have caused skipping of alternative hybrid exon, compared to the WT hybrid minigene (5'ss strength 0.99), that showed ~35 % of exon inclusion (Figure 4-1 C and D).

This result indicates that the three different F7 mutations inserted in a central region of a minigene affect splicing by inducing exon skipping.

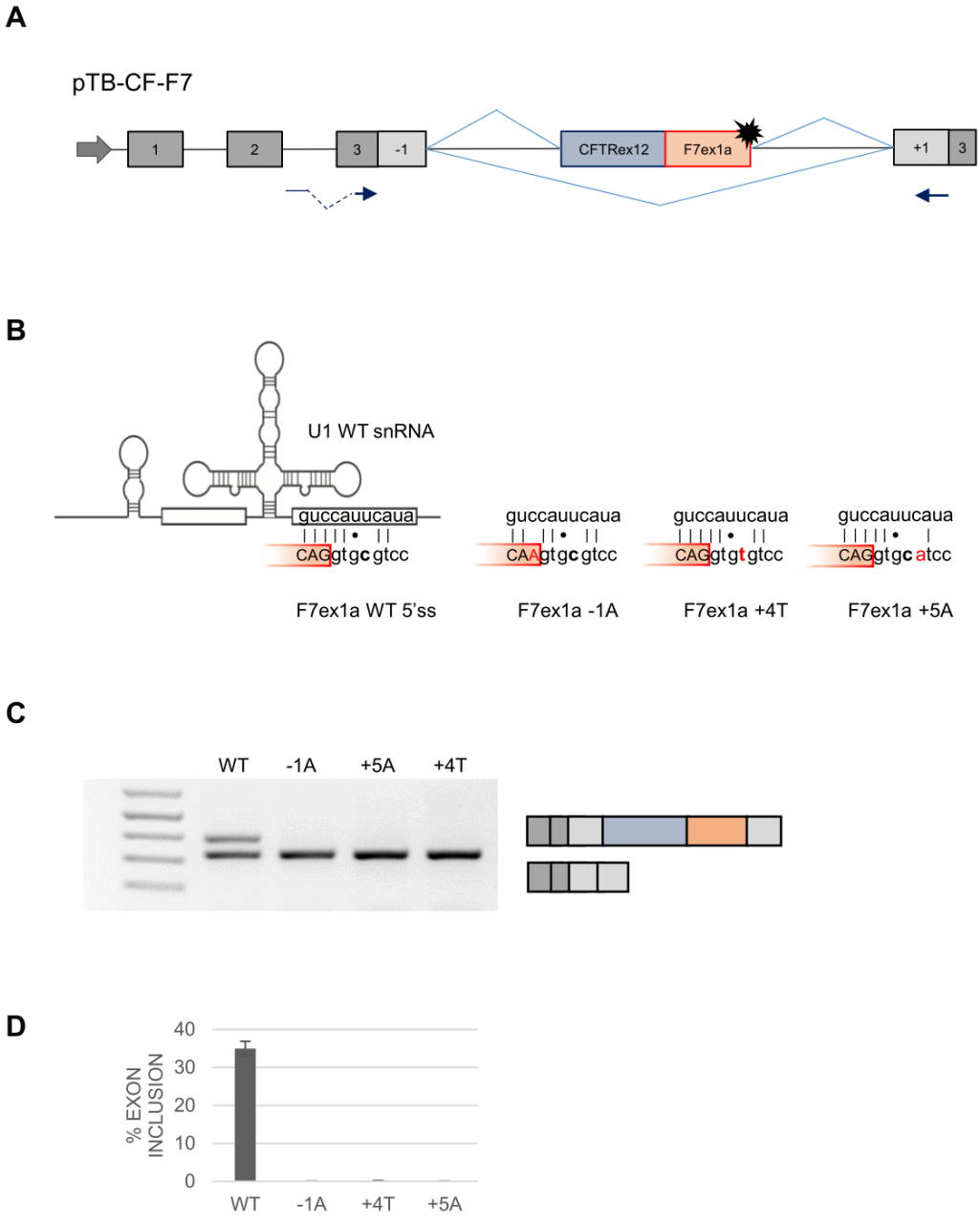

Figure 4-1 Mutations in 5'splice site of F7 exon1 cause skipping of alternative hybrid exon. A. Schematic representation of pTB-CF-F7 minigene. Dark grey boxes represent α globin exons, light grey are exons of fibronectin. In blue and red are given the parts of hybrid exon coming from CFTRex12 and F7ex1a, respectively.

Lines represent introns, arrows primers used for PCR amplification. Position of 5'ss mutations is indicated. B. Schematic representation of endogenous U1 snRNA base pairing with WT and three mutant 5'ss. Vertical lines represent base pairing, while black dots represent wobbles, mismatches are not marked. C. Agarose electrophoresis for analysis of spliced transcripts in WT and mutants, with the identity of bands indicated on the right side of the panel. D. Quantification of gel estimated by ImageJ software. Data are expressed as mean ±SD of three independent experiments.

3.1.2 **Exon-specific U1 shift+9 rescue the inclusion of hybrid exon in all three mutants of 5'ss**

To understand the role of U1 snRNP in the recognition of aforementioned hybrid exons and identify potential therapeutic splicing correctors, I have evaluated a set of modified U1 snRNAs, that specifically bind either at the 5'ss or to downstream intronic sequences. Two modified U1 snRNAs were created that bind by complementarity to the F7ex1a WT 5'ss or to the -1A mutant, respectively (Figure 4-2 A). Another set of 4 U1 snRNAs, named exon-specific U1s (ExSpeU1s), was designed to bind to downstream intronic sequences (Figure 4-2 A). The screening of modified U1s was performed by co-transfection of modified U1s with the -1A minigene in HeLa cells, followed by RNA extraction, cDNA synthesis and semi-quantitative PCR. All modified U1s were tested in their ability to increase exon inclusion, compared to WT. Co-transfection of the modified U1 that is complementary to the 5'ss of F7ex1a WT improved exon inclusion from ~35% to ~95% indicating that in the context of this hybrid minigene the 5'ss is defective. Co-transfection of the modified U1 that is complementary to the 5'ss of F7ex1a -1A increased the percentage of exon inclusion to ~24%, a level comparable to the non-treated WT (Figure 4-2 B and C). The ExSpeU1s were tested in the hybrid exon containing the -1A mutation and showed different rescue efficiency ranging from ~0.2% for U1 shift+20 and ~0.5% for U1 shift+12, to ~12.4% for U1 shift+6 and ~15.6% for U1 shift+9 (Figure 4-2 C). Among the ExSpeU1s, the shift+9 was the most effective in the -1A context and accordingly it was tested on the other mutants.

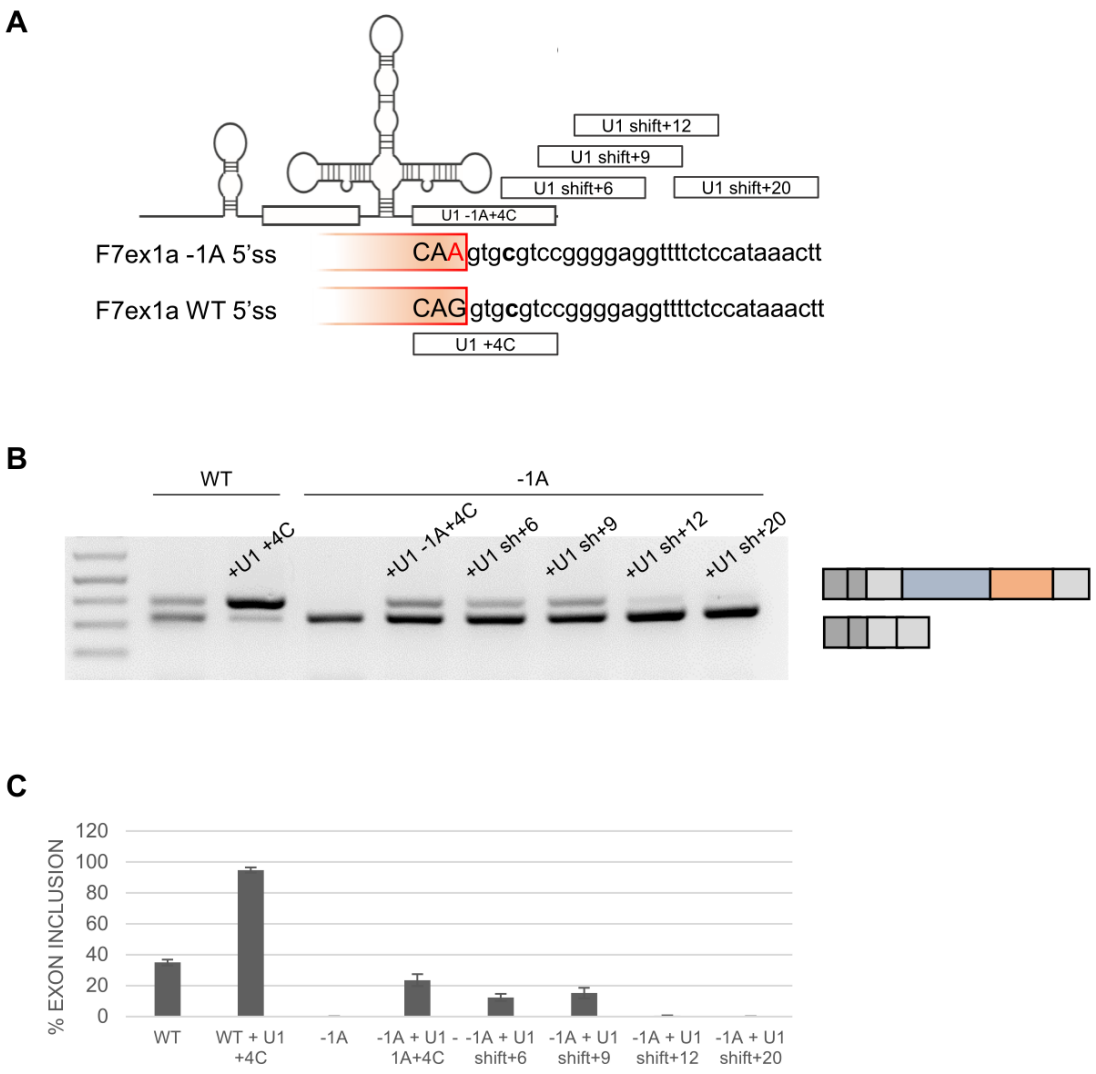

Figure 4-2 Modified U1 approach in rescue of alternative exon containing F7ex1a WT and mut 5'splice site. A. Schematic representation showing base pairing of modified U1s with WT and -1A 5'ss. Binding position of each U1 is indicated. B. Agarose electrophoresis for analysis of spliced transcripts in WT and mutants, and their rescue by modified U1s. The identity of bands is indicated on the right side of the panel. C. Quantification of gel estimated by ImageJ software. Data are expressed as mean ±SD of three independent experiments.

As -1A, +4T and +5A mutations are located upstream from U1 shift+9 binding sequence (Figure 4-3 A), and in accordance to literature (Alanis et al., 2012; Tajnik et al., 2016) one exon-specific U1 is predicted to rescue exon inclusion caused by different types of exon skipping mutations. In order to test this, I have co-transfected U1 shift+9 with these three mutant minigenes, and evaluated the effect on pattern of splicing. All three mutants were partially rescued: -1A was improved to ~15.6%, +4T to 10.6% and +5A to 12.5%, estimated by ImageJ software analysis.

This result indicates that the modified U1 snRNA that either bind to the 5'ss or to downstream intronic sequence can improve definition of the defective exons in the hybrid minigenes.

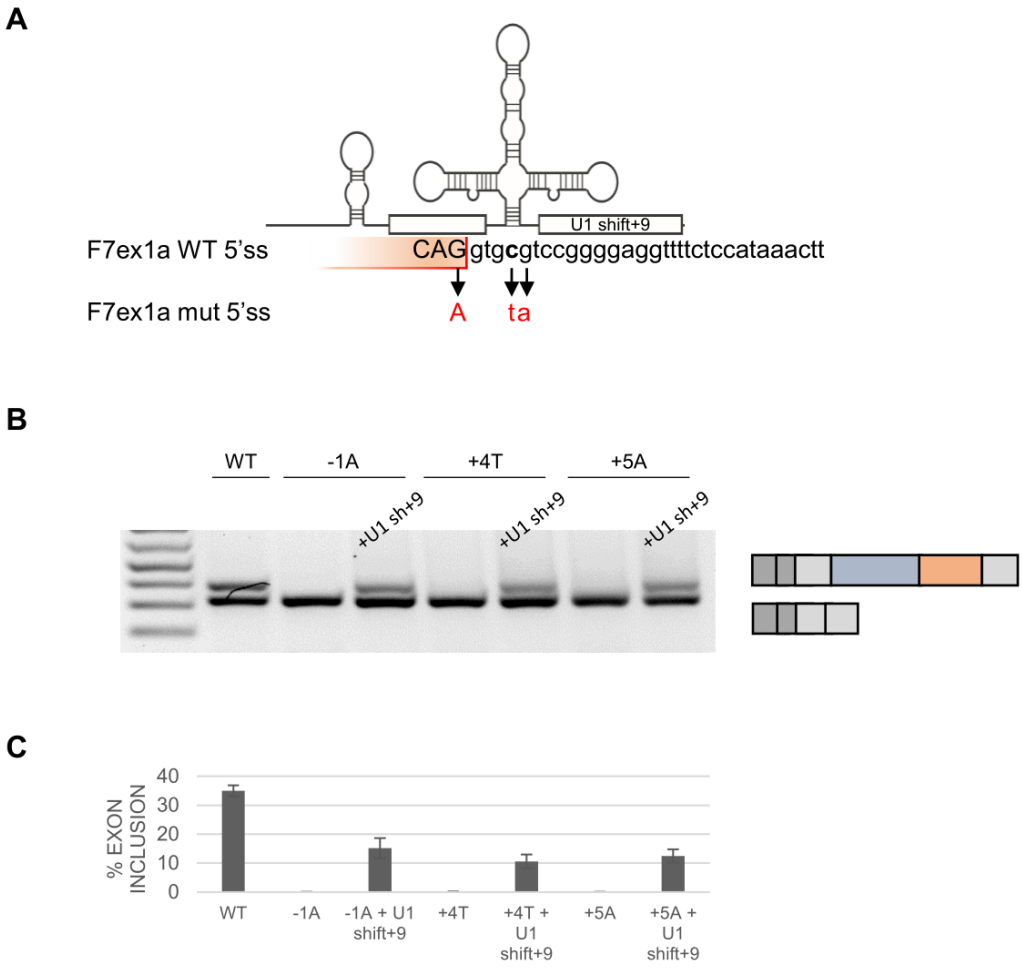

Figure 4-3 Exon-specific U1 shift+9 is able to improve exon inclusion in -1A, +4T and +5A mutants. A. Schematic representation showing base pairing of U1 shift+9 with 5'splice site of F7ex1a WT, with position of mutations indicated with arrows, and nucleotide substitutions given in red color. B. Agarose electrophoresis for analysis of spliced transcripts in WT, mutants and their rescue with U1 shift+9. The identity of bands is indicated on the right side of the panel. C. Quantification of gel estimated by ImageJ software. Data are expressed as mean ±SD of three independent experiments.

3.2 F7 promoter-proximal 5’ splice site mutations reduce the amount of mRNA in an intron length-dependent manner in transient transfection experiments

3.2.1 Effect of the promoter-proximal 5’ss mutation on splicing in transient transfection assay

In order to study 5’ splice site mutations of F7ex1a in the promoter proximal context, I have created a set of minigenes containing F7ex1a as first exon. F7ex1a was included together with ~50bp of its 5’UTR as well as ~150bp of flanking intron 1. This fragment was cloned *at HindIII-BamHI* sites, replacing the corresponding piece of sequence in α globin gene. Intron 2 of α globin is known to be efficiently spliced, and thus this gene and pcDNA backbone were chosen as well studied platform. The length of first intron was increased by cloning spacers in *BamHI* restriction site. These constructs with WT or mutated 5’ss in promoter-proximal position (pp 5’ss) were co-transfected in HeLa cells together with pSV-CFTRex1-2-3 plasmid as transfection control, and after cDNA synthesis, amount of spliced transcript was measured by real time PCR, normalizing samples against CFTRex1-2-3 transcript. First I analysed the minigene with the short intron (480bp) (Figure 4-4 A) and the results show that -1A and +5A mutants, in comparison to the WT significantly reduce the mRNA levels to ~58% and ~76%, respectively (Figure 4-4 B). In contrast, the +4T mutation in the short intron context does not significantly affect the amount of mRNA (Figure 4-4 B).

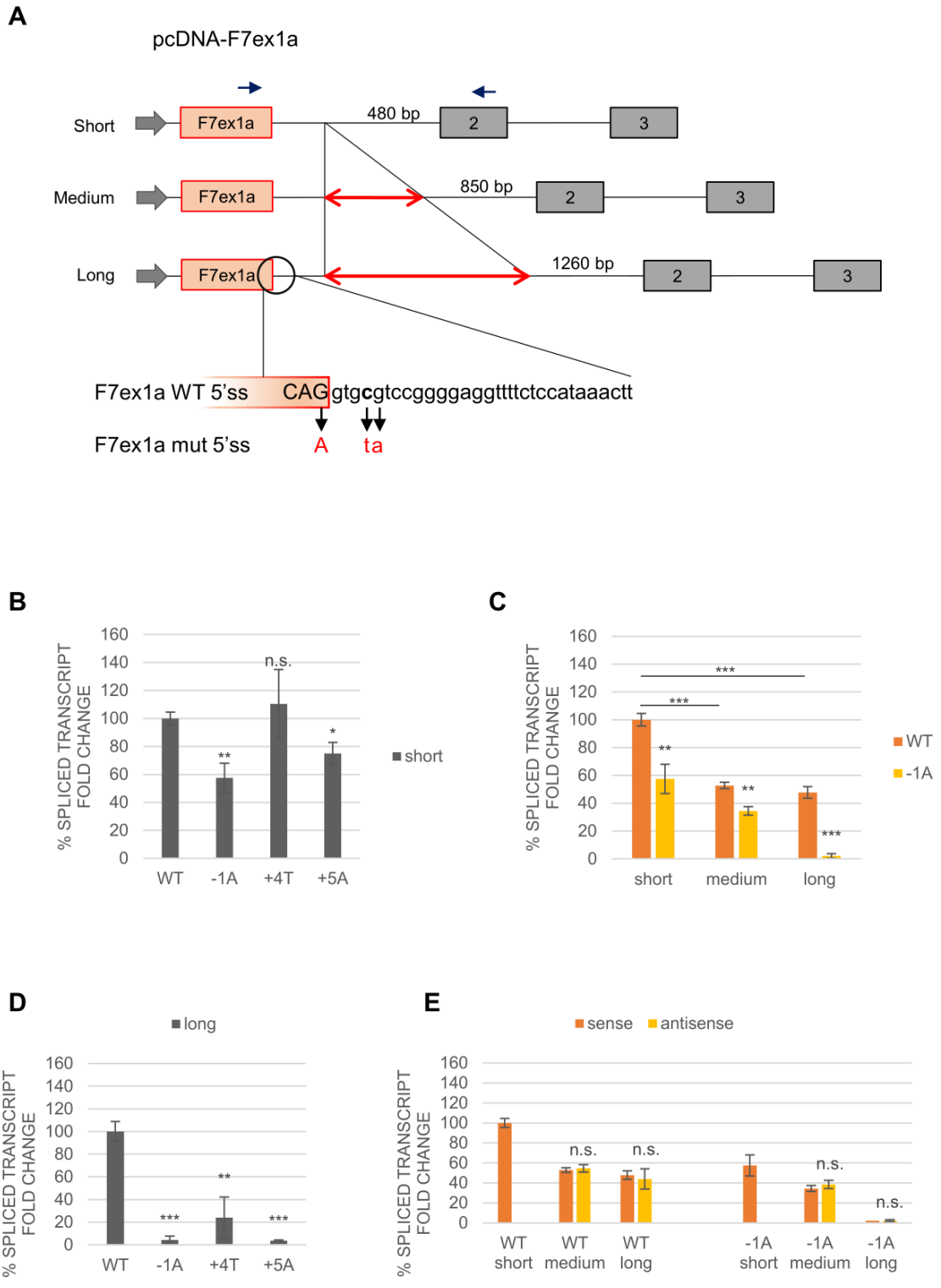

Figure 4-4 Promoter-proximal 5'ss mutations reduce the amount of spliced transcript in an intron-length-dependent manner in transient transfection assay.
A. Schematic representation showing minigene reporter, with F7ex1a in promoter-proximal context. Position of mutations are indicated with arrows and red letters. Intron 1 is expanded with spacer (indicated in red). Primers used for quantitative analysis of spliced transcript are indicated. B.C.D.E. Quantitative real time PCR analysis. The percent fold change in spliced transcript is expressed as ratio between WT and other samples. Data are expressed as mean

±SD of three independent experiments. Two-tailed Student T-test, * p<0.05, ** p<0.01, *** p<0.001, n.s. not significant. Mutants/orientation of insert is compared with WT of the same intron length, if not indicated different.

Next, I have evaluated the potential contribution of the length of the first intron (Figure 4-4 A). Interestingly, progressive increase of the length of the first intron from 480bp (referred to as “short”) to 850bp (“medium”) and 1200bp (“long”), proportionally reduced the amount of the transcript. Compared to WT short, the -1A mutant resulted in a decrease in transcript from ~58% (short) to ~34% (medium) and ~2% (long) (Figure 4-4 C). This effect is also evident for the WT pp 5’ss where, compared to WT short, the WT medium and WT long reduced the amount of transcript to ~53% and ~48%, respectively.

The negative effect of the intron length is also present for the other pp 5’ss mutants. In the long intron context, the levels were reduced to ~3.2% , ~23.8% and ~3.4% for the -1A, +4T and +5T mutants, respectively (Figure 4-4 D). This is also evident when comparing the mutants with the short intron with their corresponding long intron versions: -1A is reduced by 27 fold, +4T by 12 fold and +5A by 47 fold. To exclude that the inserted sequence by itself might affect splicing I also tested the spacer elements inserted in the antisense orientation. As expected, I did not observe any difference (Figure 4-4 E). Thus, in transient transfection assay, the negative effect of the pp 5’ss mutation on processed transcript is affected by the length of the intron.

3.2.2 **Rescue of promoter proximal 5’ss mutations by modified U1 snRNPs in transient transfection experiments**

3.2.2.1 U1 -1A+4C and U1 shift+9 rescue the -1A mutant defect (in long and short construct)

The analysis I have performed previously has identified different modified U1s that are able to rescue exon skipping cause by the 5’ss mutations. I have chosen three of them, two that are complementary to the 5’ss (+4C for WT and -1A+4C for -1A mutant) and one ExSpeU1 that bind downstream (shift+9) for further testing in the promoter-proximal context. In order to test the ability of U1s to rescue spliced transcript in constructs containing pp 5’ss mutation, I co-transfected minigenes with U1s and measured splicing of first intron by quantitative real time PCR (Figure 4-5 A). All U1s that bind to the 5’ss or to downstream intronic sequences significantly improved pre-mRNA processing. The most striking effect was observed for the U1s that bind to the 5’ss. Compared to the untreated WT short construct, co-transfection of the U1s

that bind to the 5'ss increased the amount of transcript to 384.4%, 423.6%, 235.6% and 267.4% in WT short, -1A short, WT long and -1A long, respectively. On the other hand, compared to the untreated WT short construct, co-transfection of the U1 shift+9 increased the amount of transcript to 224.4%, 289.4%, 111.7% and 62.4% in WT short, -1A short, WT long and -1A long, respectively Figure 4-5 B). Considering the -1A long constructs, the minigene context where -1A mutation has the most deleterious effect, the U1 that bind to the 5'ss and the shift+9 ExSpeU1 induced a dramatic increase in the transcript. In fact, the U1-5'ss and the shift+9 induced ~126.8 and ~29.6 -fold increase, respectively. U1 wt did not affect the amount of spliced transcript in all tested minigenes.

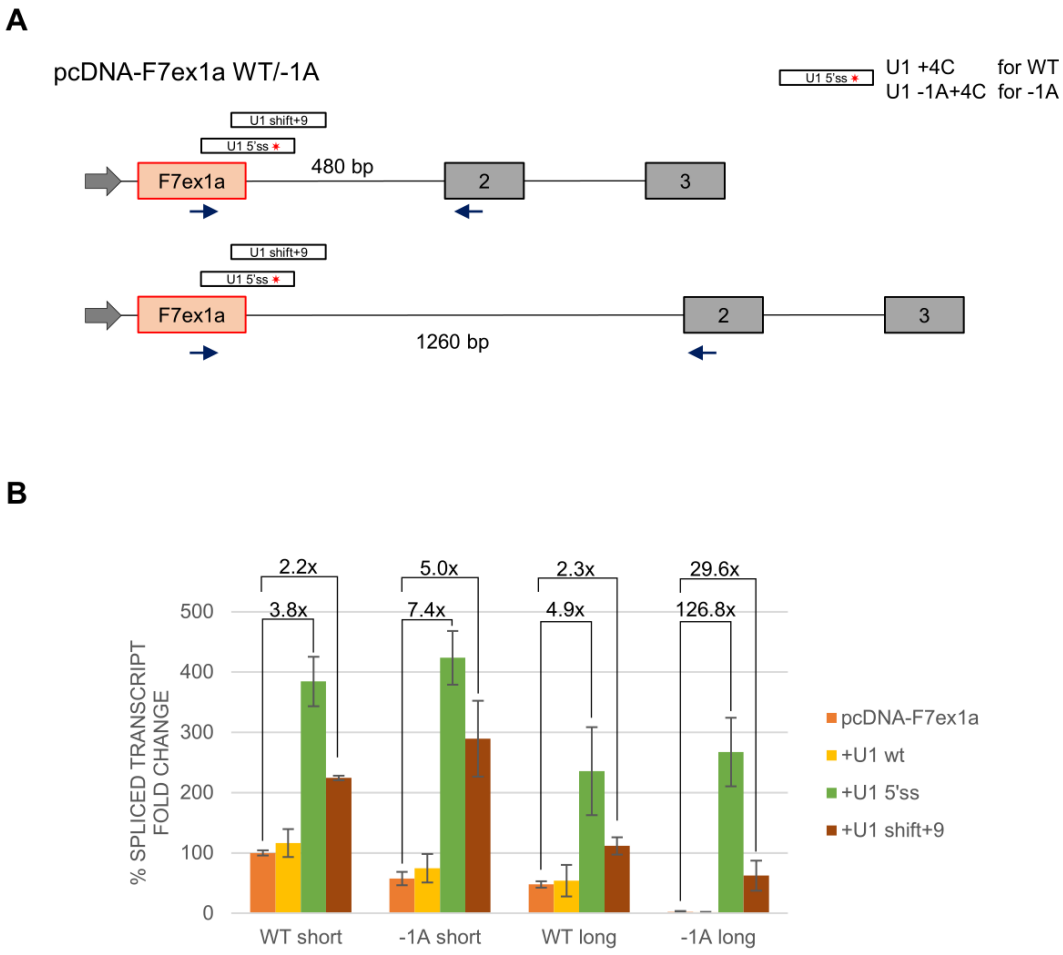

Figure 4-5 Modified U1s improve splicing in WT and -1A mutants. A. Schematic representation showing pcDNA-F7ex1a construct with binding sites of U1s indicated. For WT U1+4C was used as 5'ss binding one, while for -1A it was U1-1A+4C. Primers for real time PCR are shown as arrows. B. Quantitative real time PCR analysis. The percentage fold change in spliced transcript is expressed as ratio between WT short and other samples. Data are expressed as mean ±SD of three independent experiments.

3.2.2.2 *U1 mutants U1A, 70K and L4 reduce the rescue of splicing observed with modified U1s*

For better understanding of mechanism underlying observed rescue in spliced transcripts, I have created mutants of U1snRNA (Figure 4-6 A). By mutating 3 nucleotides in stem loop 2 I created an U1 snRNA that does not bind to U1-A protein, as shown previously in *in vitro* studies (Gunderson et al., 1998; Rogalska et al., 2016). To create the mutant of U1 snRNA unable to bind 70K, I replaced stem loop 1 with the binding sequence for the MS2 protein, as previously published (Rogalska et al., 2016). The mutant of stem loop 4 was created as previously described by (Sharma et al., 2014). Stem loop 4 does not bind to structural proteins within U1 snRNP, but it has been shown to interact with U2 snRNPs-associated proteins (Sharma et al., 2014). These three mutants were created for U1 snRNAs -1A+4C and shift+9, and their rescue activity was tested in co-transfection experiments (Figure 4-6 B). Co-transfection experiments show that, in comparison with corresponding non-mutated modified U1s, U1-A mutants reduced the rescue activity by only 50%. The 70K mutant shows activity reduced to 10% and 38%, for the U1 -1A+4C and U1 shift+9, respectively. The loop 4 mutants were the most disruptive, reducing the rescue to 7% and 10 %, for the U1 -1A+4C and U1 shift+9, respectively. (Figure 4-6 C). These data indicate that U1-A is partially involved in the rescue activity, whereas both 70K and stem loop 4 seem to be very important structural elements affecting the rescue activity of modified U1s. These results are similar to the ones previously reported for ExSpeU1s that recover exon skipping (Rogalska et al., 2016).

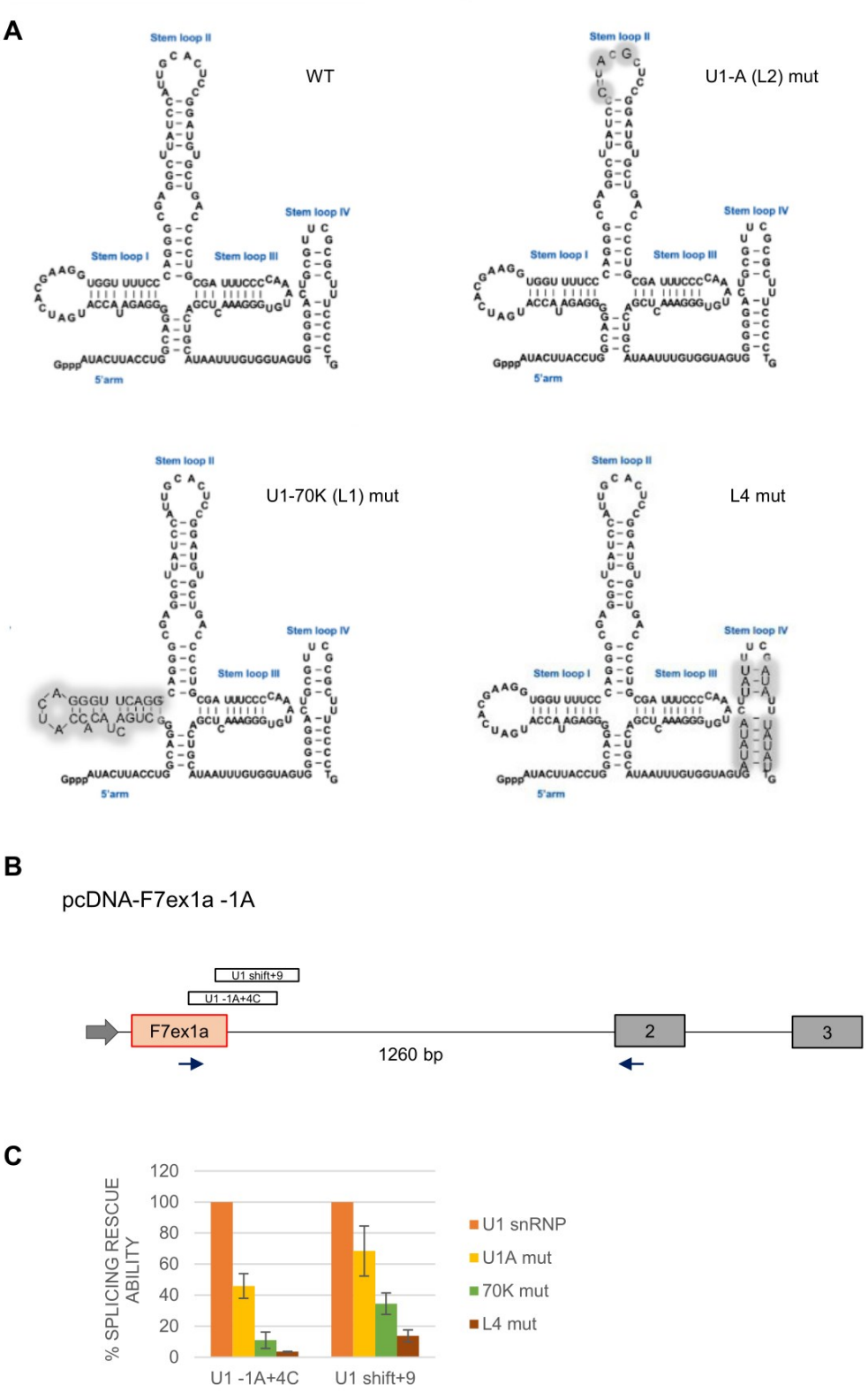

Figure 4-6 The role of U1 snRNPs components in modified U1 snRNPs. A. Schematic representation showing U1 snRNA WT and mutant sequences, each one indicating (dis)ability to bind certain U1snRNP-specific proteins. B. Schematic representation of F7 minigene with -1A mutation with long flanking intron. Binding positions of different modified U1s are indicated. C. Quantitative real time PCR analysis. The

percentage of splicing rescue is expressed as ratio between each fully functional modified U1 snRNP and their mutants, respectively. Data are expressed as mean ±SD of three independent experiments.

3.3 Effect of the promoter proximal 5'ss mutations in the chromosomal context

3.3.1 Flp-In T-REx 293 clones containing stably integrated minigene construct are inducible

In order to study the effect of the 5'ss mutations in more physiological context, I have created Flp-In T-REx 293 clones with a single copy of minigene stably integrated in chromatin (Figure 4-7 A). Stable clones containing a single integrated copy of the F7 minigenes were tested for inducible expression by adding tetracycline (tet) to growth medium, followed by total RNA extraction, cDNA synthesis and measurement of the spliced transcript by qRT-PCR. As normalization, I used the GAPDH housekeeping gene. I tested four clones corresponding to the WT 5'ss with the short and long intron and the -1A mutant with the short and long intron. The transcript analysis was performed at time 0, 4, 8 and 24 hours after the induction. As can be seen in Figure 4-7 at time 0 no expression can be detected in the different minigenes. After induction, the WT constructs (short and long) showed a significant rise in the expression: the percentage of spliced transcript at 8h reached 99% and 79% whereas at 24h reached 100% and 90% for WT short and WT long, respectively. In contrast, the -1A mutants showed only a minimal increase in response to the transcriptional activation: it was 5.6% for the -1A short and 2.0% for the -1A long at 8h, and 8.2% and 3.4% at 24 h, respectively (Figure 4-7). These results indicate the -1A pp 5'ss mutation in the Flp-In T-REx 293 cells significantly affects the biosynthesis of the spliced transcripts. In addition, the length of the first intron has a small effect on the transcript: in the WT minigenes the percentage of splicing of the long construct was lower at the different time points (compare WT S with WT L curves). In the WT constructs, I could observe only a small difference in the pattern of expression between them: the minigene with the short intron showed ~10% higher levels (1.1-fold change) in comparison to the minigene with the longer intron at the different time points after induction. Similar results but with lower percentages were observed for the -1A minigene, where difference between short and long intron was ~5% or 2.4-fold (Figure 4-7 C).

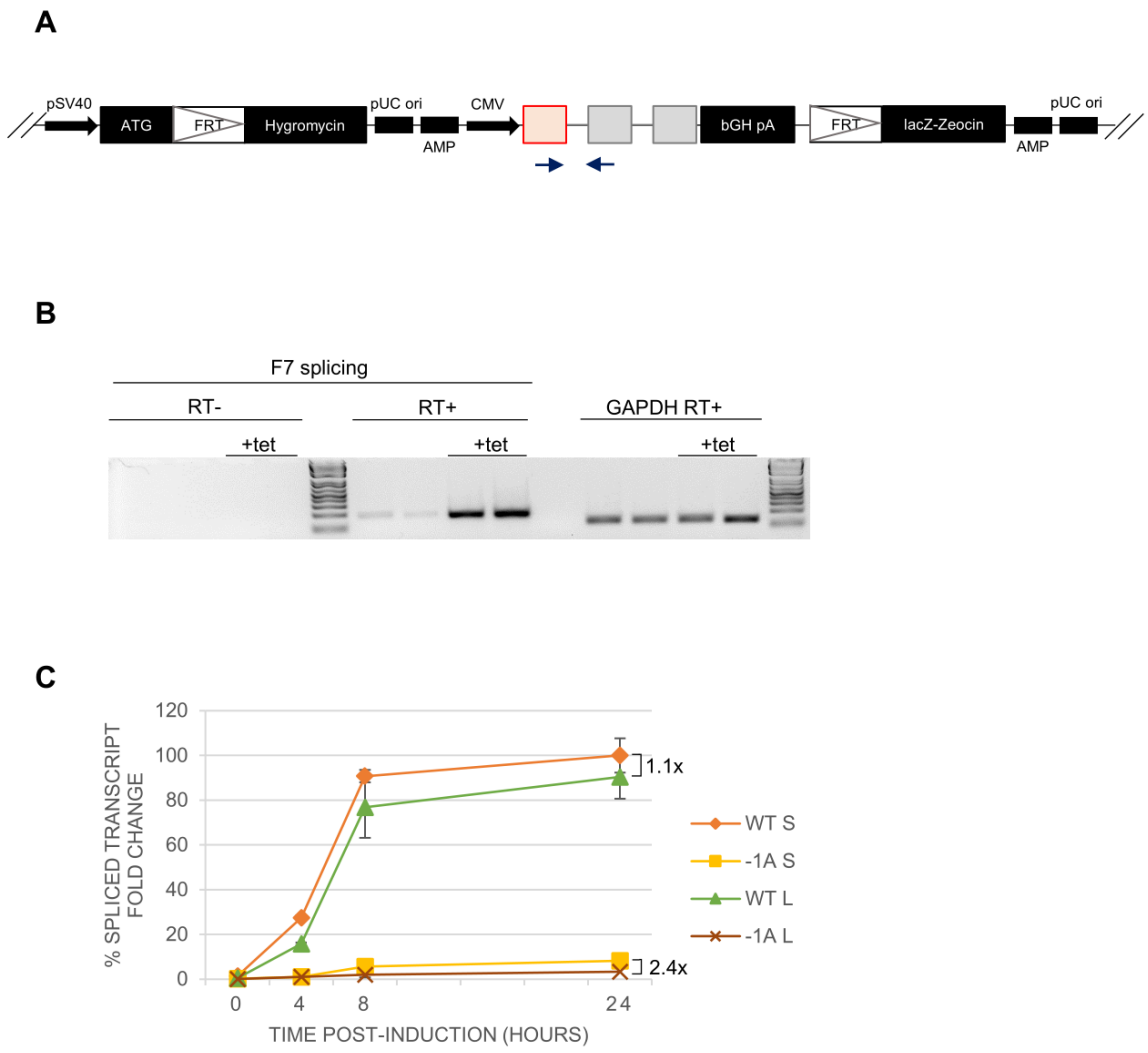

Figure 4-7 Flp-In T-REx 293 stable clones show inducible expression of F7 minigene. A. Schematic representation showing stably integrated copy of minigene, created using Flp-In system in HEK 293 T-REx cells. Primers used for PCR analysis are indicated. B. Agarose gel with RT-PCR samples obtained on stable clones. RT+/- refer to cDNA synthesis in presence/absence of reverse transcriptase. +tet refers to presence of tetracycline in medium for 24hrs. C. Quantitative real time PCR analysis. The percent fold change in spliced transcript is expressed as percentage between WT short at time point 24hrs set to 100%. Values were normalized to GAPDH expression. Data are expressed as mean ±SD of three independent experiments.

3.3.1.1 Modified U1s rescue the proximal 5' splice site mutation defects

In the following experiment I have evaluated the rescue activity of the modified U1 snRNPs on the proximal splice sites mutation defects in the four inducible stable clones. Cells were plated, and then transfected next day with plasmids expressing modified U1s, and 24h post-transfection cells were induced with tetracycline. Samples were collected at time points 0, 8h

and 24h after induction, and analysed using qRT-PCR (Figure 4-8). Co-transfection of the WT short minigene with U1 +4C induced ~178% and ~315% (2-3 fold) increase in the transcript at 8h and 24h, respectively, consistent with the fact that the WT 5'ss sequence is suboptimal, being not a perfect donor site consensus (Figure 4-1 B). In fact, the 5'ss of F7ex1a WT has one nucleotide difference at position +4 compared to canonical sequence. Strikingly, co-transfection of U1 -1A+4C or U1 shift+9 in the -1A mutant completely rescued the defect. Compared to untreated conditions (-1A S line), U1 -1A+4C increases mRNA levels ~22 fold at 8h and ~106 fold at 24h, respectively. U1 shift+9 improved mRNA levels ~18 fold at 8h and ~67 fold at 24h (Figure 4-8 B).

Co-transfection experiments using the long intron minigenes showed similar results. WT is increased ~5 fold at 8h and ~7 fold at 24h, and -1A ~132 fold at 8h and ~116 fold at 24h when U1 5'ss was used, or ~96 fold at 8h and ~78 fold at 24h when U1 shift +9 was used (Figure 4-8 C). The transfection of modified U1 in the absence of tetracycline did not change measured mRNA, as expected.

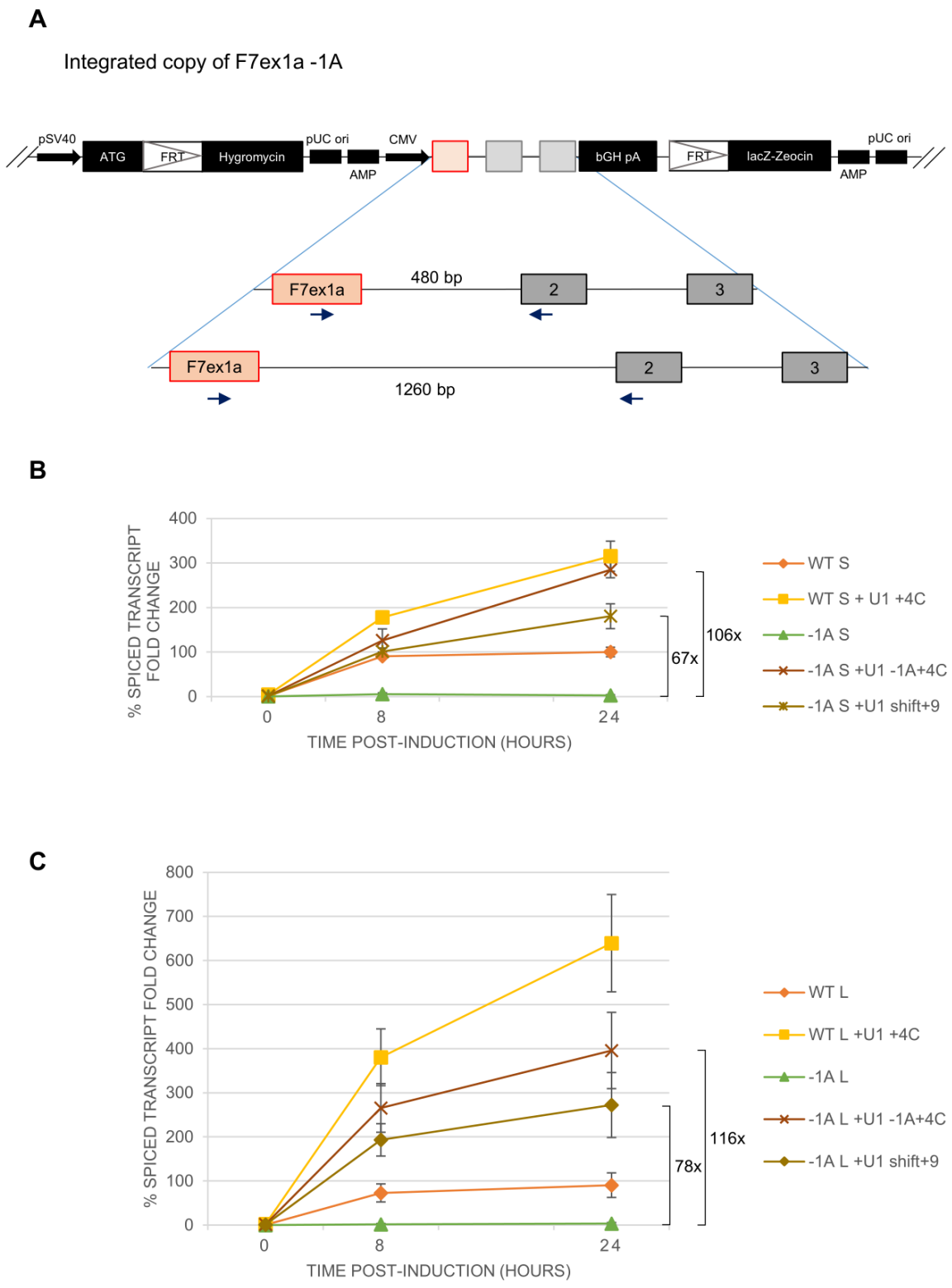

Figure 4-8 Modified U1s that bind to the pp 5'ss are able to rescue pre-mRNA processing defects in chromosomal context.

A. Schematic representation showing inducible stably integrated F7 minigene in short and long intron context. B. Quantitative real time PCR analysis showing rescue of WT and -1A mutant in short intron context. The percent fold change in spliced transcript is expressed as ratio between WT short at time point 24h and other samples. C. Quantitative real time PCR analysis showing rescue of WT and -1A long with modified U1s. The fold change in

spliced transcript is expressed as ratio between WT short at time point 24h (not shown on the graph) and other samples. Data are expressed as mean ±SD of three independent experiments.

3.3.2 **U1 mutants U1A, 70K and L4 reduce the rescue of splicing observed with modified U1s in chromosomal context**

In order to evaluate the effect of modified U1s on levels of mRNA, I transfected U1s expression plasmids in stable clones with long intron, and 24h later induced with tetracycline. Samples were collected 24h post-induction and analysed.

Results show that U1 -1A+4C and U1 shift+9 were able to increase the level of spliced transcript, ~12 and ~9 fold, respectively (Figure 4-9 B). Mutants of these two modified U1s, that are not able to bind U1A protein reduced the rescue to 60% of the initial, while 70K mutant was more important in the U1 -1A+4C than in U1 shift+9. Stem loop 4 was equally important for both U1s, reducing the ability to rescue to only 20-25% from the initial (Figure 4-9 C).

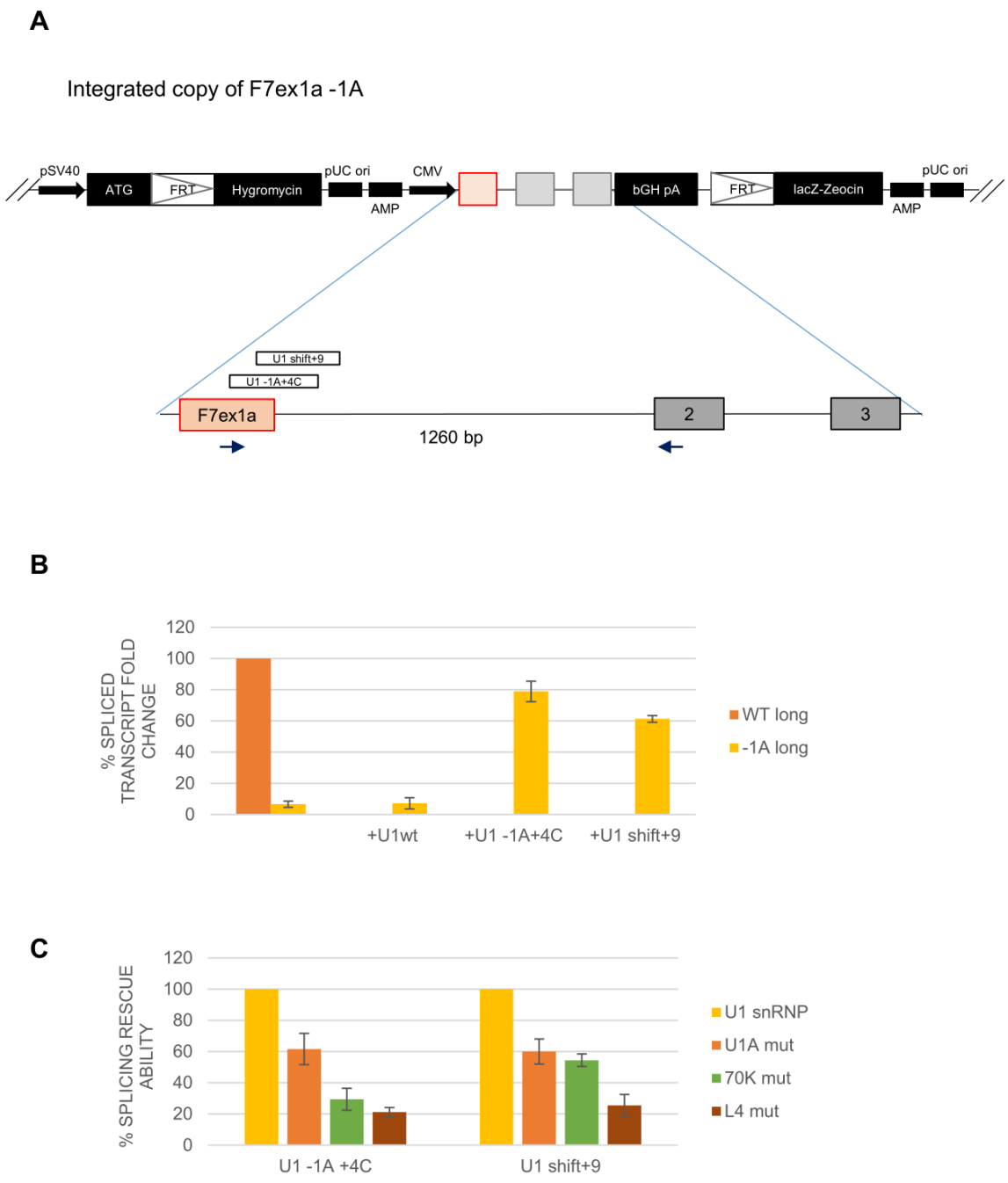

Figure 4-9 The role of U1 snRNPs components in ability of modified U1s to rescue pre-mRNA processing defects.

A. Schematic representation showing inducible stably integrated F7 minigene in long intron context. B. Quantitative real time PCR analysis showing rescue of F7 -1A long with modified U1s. The percent fold change in spliced transcript is expressed as ratio between WT and each sample. C. Quantitative real time PCR analysis. The percent of splicing rescue ability is expressed as ratio between each fully functional modified U1 snRNP and their mutants, respectively. Data are expressed as mean ±SD of three independent experiments.

3.4 Analysis of pre-mRNA processing defects in chromatin fraction

The effect of U1 snRNP on level of total transcript is not well understood, thus I used the advantage of minigenes in chromosomal context to look more closely to levels of RNA bound to chromatin. The method I used consists of cellular fractionation (Mayer & Churchman, 2016; Werner & Ruthenburg, 2015), and in-depth analysis of chromatin fraction by qRT-PCR. In this manner I intended to study RNA bound to chromatin, and examine the effect of modified U1s on nascent RNA.

3.4.1 Purity of extracted chromatin fraction is confirmed by qPCR and Western blot analysis

To better understand the effect of the pp 5'ss mutations I have performed cellular fractionation in the stable Flp-In T-REx 293 clones. The method I used consists of cellular fractionation (Mayer & Churchman, 2016) and analysis of chromatin, nucleoplasmic and cytoplasmic transcripts. As first step I checked the purity of the fractions on control RNAs, using qRT-PCR, and on control proteins by Western blot (Figure 4-10). For the purpose of inter-fractional comparisons all PCRs were normalized to Formin binding protein 4 (FNBP4), as it was shown to be equally abundant in fractions (Werner & Ruthenburg, 2015) . On the other hand, all western blots were normalized according to dilution: all fractions coming from one sample were resuspended in the same volume, as suggested by Werner & Ruthenburg.

Spliced GAPDH transcript was enriched 13-26 fold in cytoplasm compared to nucleoplasm, and 147-261 fold in cytoplasm compared to chromatin (Figure 4-10 A) (Mayer & Churchman, 2016). ACTB transcript was enriched 5.5-9 fold in nucleoplasm compared to chromatin (Figure 4-10 B) (Bhatt et al., 2012; Mondal, Rasmussen, Pandey, Isaksson, & Kanduri, 2010; Werner & Ruthenburg, 2015), while long noncoding RNA XIST was enriched in chromatin in comparison with the other two fractions (Bhatt et al., 2012; Engreitz et al., 2013; Mondal et al., 2010; Werner & Ruthenburg, 2015), 3-4 fold in chromatin compared to nucleoplasm, and 14-17 fold in chromatin compared to cytoplasm (Figure 4-10 C). Cytoplasmic, nucleoplasmic and chromatin fractions were also used as input for protein extraction and Western blotting. Analysis was performed using antibodies against GAPDH (enriched in cytoplasm), 70K protein of U1snRNP (enriched in nucleoplasm) and an antibody against histone 2B (enriched in

chromatin) (Figure 4-10 D) and the results show that the proteins are present in the corresponding fractions (Mayer & Churchman, 2016; Vaz-Drago et al., 2015). These results indicate the purity of the cellular fractions obtained.

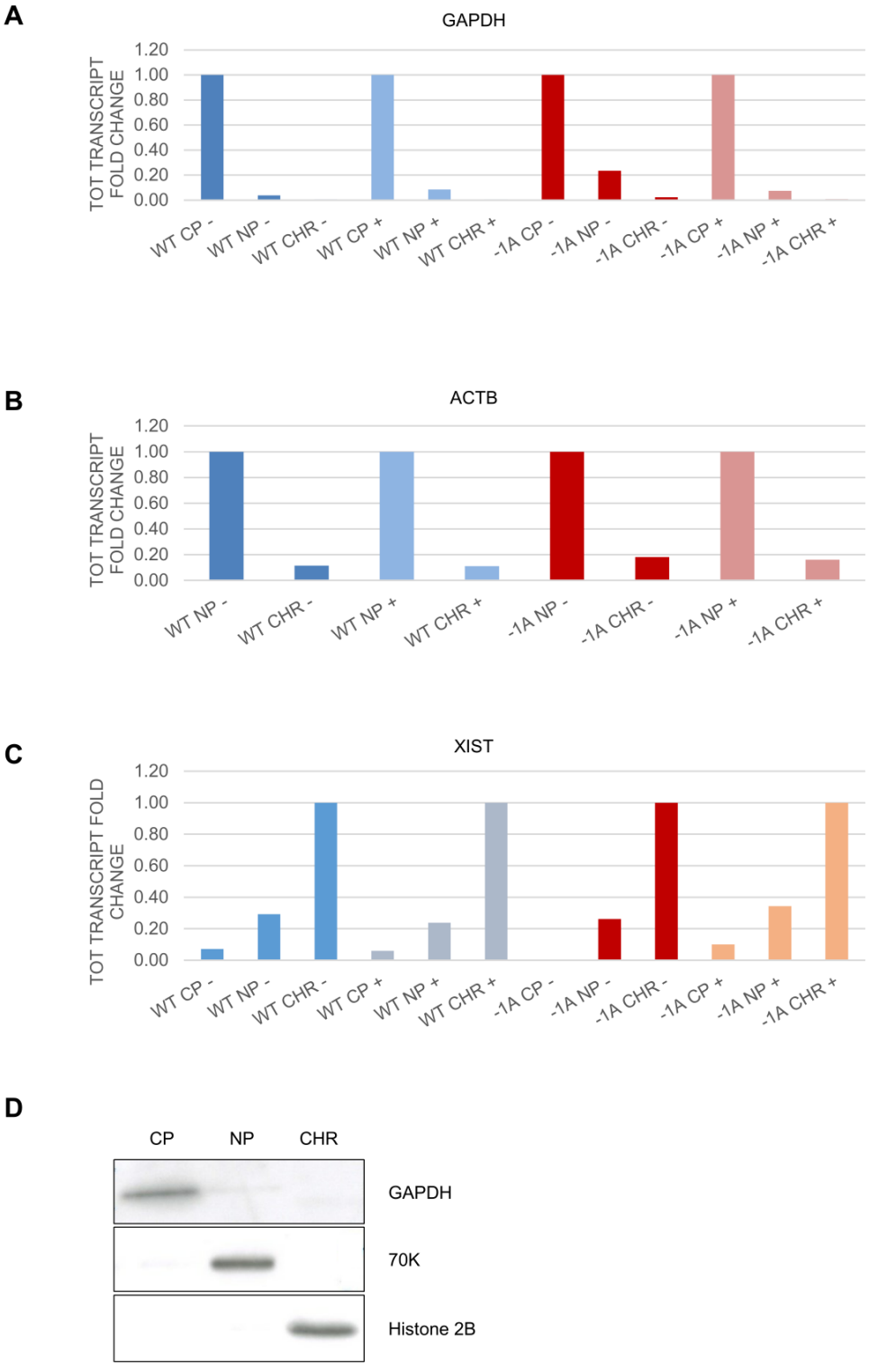

Figure 4-10 Analysis of the markers specific for the chromatin, nucleoplasmic and cytoplasmic cellular fractions.
A. Quantitative real time PCR analysis. The total transcript fold change in GAPDH is expressed as ratio between cytoplasmic fraction, and nucleoplasm or chromatin, for each sample separately. Samples were normalized using FNBP4. Fractions: CP – cytoplasm, NP – nucleoplasm, CHR - chromatin. B. Quantitative real time PCR analysis. The total transcript fold change in ACTB is expressed as ratio between nucleoplasmic fraction and chromatin, for each sample separately. Samples were normalized using FNBP4. Fractions: NP – nucleoplasm, CHR - chromatin. C. Quantitative real time PCR analysis. The total transcript fold change in XIST is expressed as ratio between chromatin fraction, and cytoplasm or nucleoplasm, for each sample separately. Samples were normalized using FNBP4. Fractions: CP – cytoplasm, NP – nucleoplasm, CHR - chromatin. D. Western blot analysis of fractions, using antibodies against GAPDH, 70K and histone 2B. Fractions: CP – cytoplasm, NP – nucleoplasm, CHR - chromatin.

3.4.2 Abundance of spliced transcript in the different cellular fractions in WT and -1A mutant Flp-In T-REx 293 cells.

After confirming the integrity of cellular fractions, this method of extraction was applied on stable clones containing F7 minigene in the long intron context. Cells were plated and next day induced with tetracycline. 24h post-induction cells were collected and fractionation was performed. Cytoplasm (CP), nucleoplasm (NP) and chromatin (CHR) fractions were treated with RNase free DNase I, RNA was extracted using standard TRIzol protocol, and cDNA was synthetized. Expression levels were normalized to FNBP4 which expression levels has been previously shown to be constant among the different fractions (Werner & Ruthenburg 2015 Cell Reports). qRT-PCR analysis revealed an 8.3 and 11.7 -fold enrichment of spliced transcript in WT and -1A clones, respectively in CP fraction relative to the NP fraction after tet induction. This difference was even more pronounced when comparing the CP fraction with CHR fraction: 486 and 522-fold enrichment in WT and -1A clones, respectively (Figure 4-11 B). These ratios between fractions are expected, since spliced transcripts are transported from the chromatin to nucleoplasm and then accumulate in the cytoplasm. Analysis of the CHR fraction showed that whereas the addiction of tet induced a significant increase in the amount of transcript (Fig. 2.9C compare WT CHR tet- with WT CHR tet+) in the WT minigene, this effect was completely absent in the -1A mutant (Fig. 2.9C compare -1A CHR tet- with -1A CHR tet+).

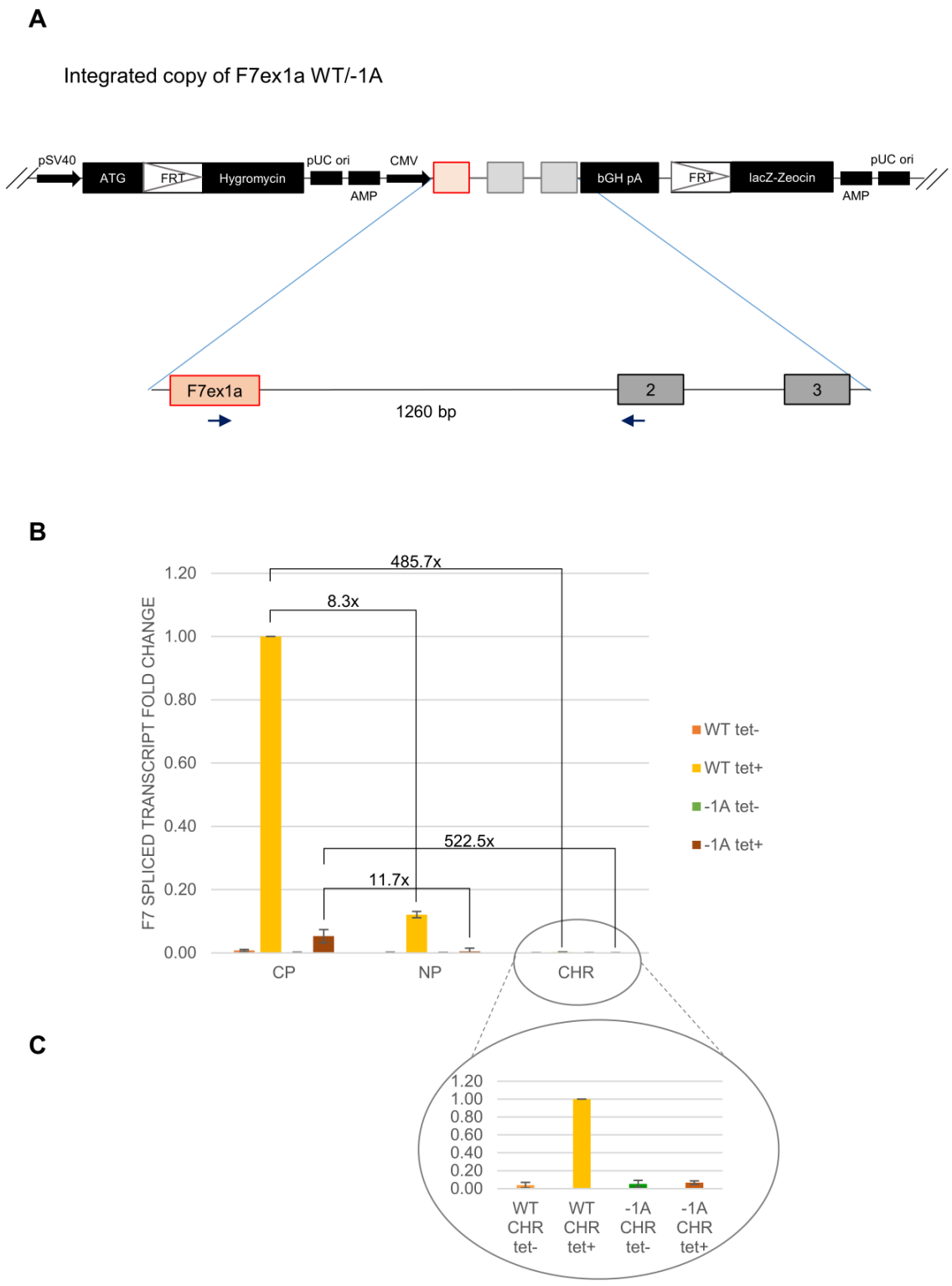

Figure 4-11 Abundance of spliced transcript in the different cellular fractions in WT and -1A mutant Flp-In T-REx 293 cells.
A. Schematic representation showing inducible stably integrated F7ex1a minigene in long intron context. B. Quantitative real time PCR analysis. The fold change in F7 spliced transcript is expressed as ratio between WT cytoplasmic fraction and all other samples. C. Quantitative real time PCR analysis. The fold change in F7 spliced transcript is expressed as ratio between WT chromatin fraction and all other samples. Fractions: CP – cytoplasm, NP – nucleoplasm, CHR - chromatin. Data are expressed as mean ±SD of three independent experiments.

3.5 Promoter-proximal 5'ss mutations affect the choice of polyadenylation signal

3.5.1 Promoter-proximal 5'ss mutations promote the usage of a proximal polyadenylation signal in an intron length-dependent manner in transfection experiments

To study the effect of pp 5'ss mutations on the choice of polyadenylation signal, I used a minigene system that include F7ex1a and exons 2 and 3 of α globin, cloned into pcDNA vector (Figure 4-12 A). These minigenes are similar to those used for splicing study, with the only difference being an artificial polyadenylation site inserted in first intron. Minigenes were transfected in HeLa cells, and for some experiments co-transfected with modified U1s, total RNA was extracted and reversely transcribed to cDNA using oligo dT to pick up the polyadenylated transcripts. Polyadenylation was analyzed using 3'RACE, a PCR modification that uses Anchor oligonucleotide as reverse primer. Products were resolved on 1.5% agarose gel with ethidium bromide, visualized under UV light, and acquired image was analyzed using ImageJ software.

To evaluate if the presence of intronic polyA signal affects the amount of spliced transcript I have compared F7ex1a minigenes with and without the polyA signal in first intron. Results show that the insertion of artificial polyA in first intron slightly reduced the spliced transcript levels in the different minigenes. (Figure 4-12 B). Next, WT and three mutant minigenes containing intronic polyA in the short intron context were transfected in HeLa cells, total RNA was extracted and retro-transcribed using oligo dT, and finally analysed using 3'RACE PCR. Products were resolved on agarose gel (Figure 4-12 C) and quantified by ImageJ software (Figure 4-12 D). Results showed that in WT ~96.6% of transcript use the distal polyA signal. The -1A and +5A mutations slightly promote the usage of the intronic polyadenylation site: the full length mRNA is reduced to ~75% and ~87%, respectively. +4T has no effect on polyadenylation pattern compared to WT (Figure 4-12 C and D).

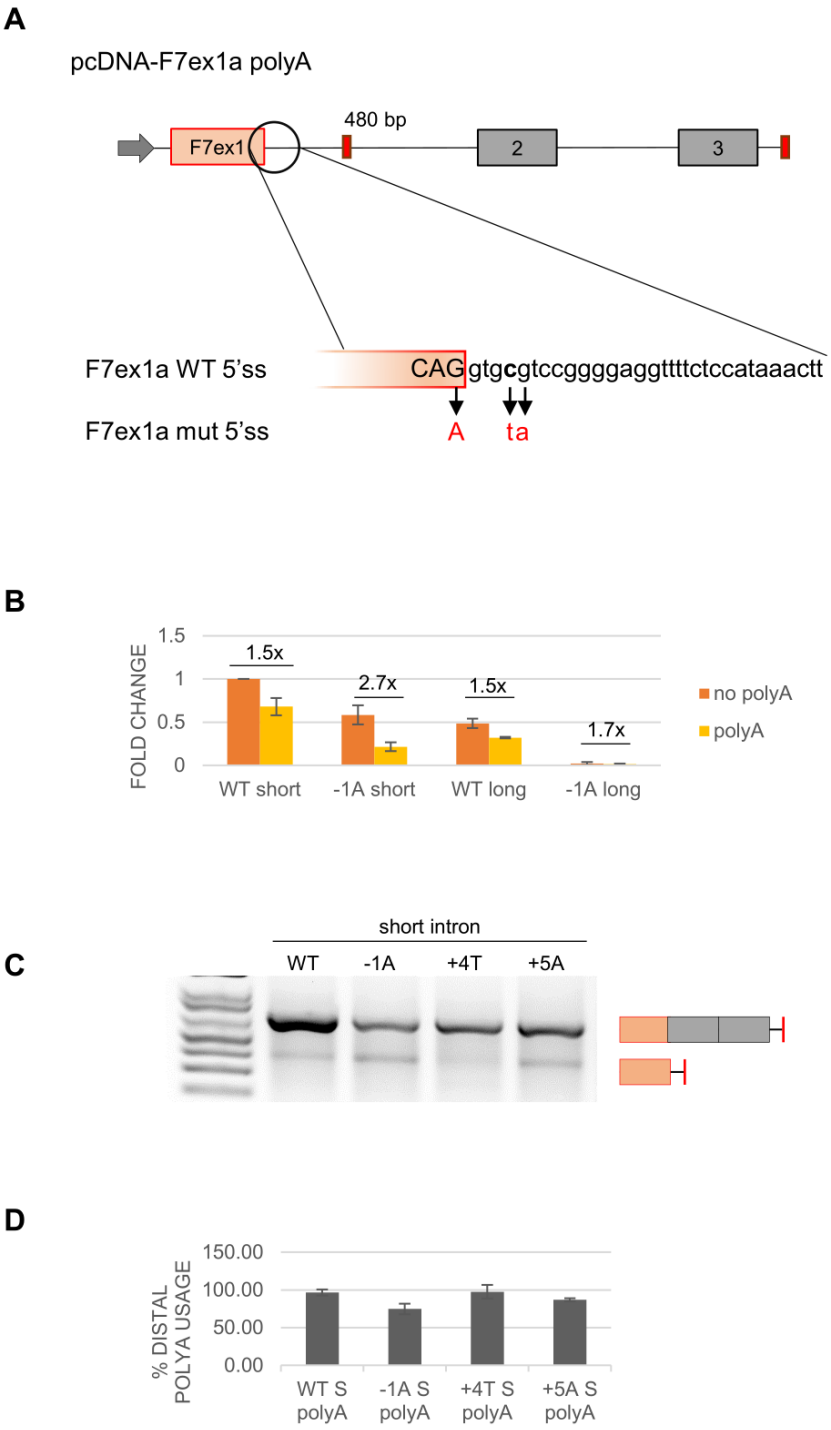

Figure 4-12 Pp 5'ss mutations mildly promote the usage of intronic polyadenylation signal in short intron context.

A. Schematic representation of F7ex1a polyA minigene. Positions of mutations are indicated. Red rectangle represents polyadenylation signal. B. Quantitative real time PCR. The fold change in spliced transcript is expressed as ratio between WT short and all other samples. Data are expressed as mean ±SD of three independent experiments. C. 3'RACE PCR resolved on 1.5% agarose gel. Identity of the bands is indicated. D. ImageJ quantification of agarose gel. Data are expressed as mean ±SD of three independent experiments.

To evaluate the role of intron length on polyadenylation selection I used minigenes containing the intronic polyA in a long intron context of 1260bp (Figure 4-13 A). Results show that in the WT long intron context ~50.4% of transcripts use the distal polyA site, compared to ~96.6% in WT with the short intron (Figure 4-13 B and C). This suggests that intron length by itself promotes usage of proximal polyA signal present in intron. Interestingly, the pp 5'ss mutations in the long context promote the usage of the proximal polyA site. In -1A and +5A the proximal polyA site it was the only variant detected in this experimental conditions, whereas in +4T ~40% of the transcripts use the distal one (Figure 4-13 B and C).

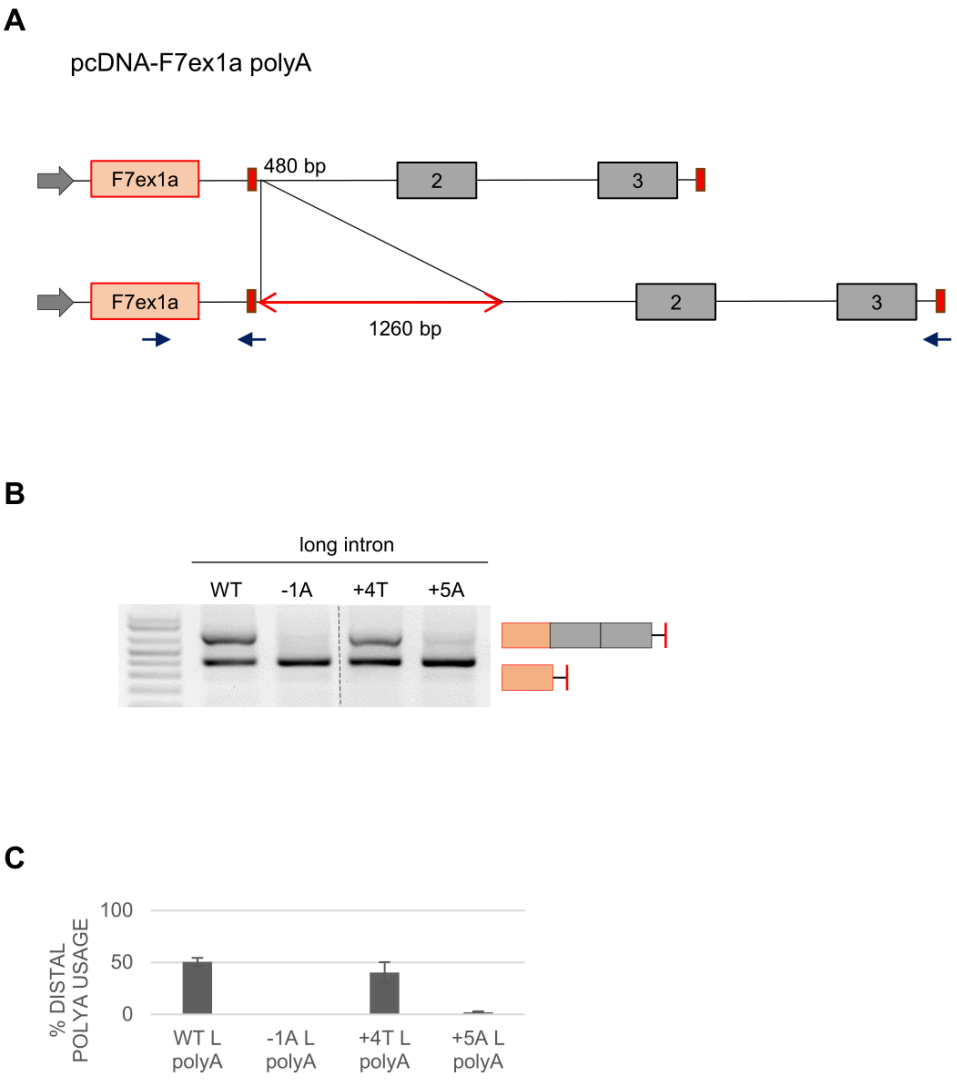

Figure 4-13 -1A and +5A Pp 5'ss mutations strongly reduce the full length mRNA in long intron context A. Schematic representation of F7ex1a polyA minigene in short and long intron context. Red rectangle represents polyadenylation signal. B. 3'RACE PCR resolved on 1.5% agarose gel. Identity of the bands is indicated. C. ImageJ quantification of agarose gel. Data are expressed as mean ±SD of three independent experiments.

3.5.2 **Modified U1 snRNPs inhibit utilization of the proximal polyadenylation site in minigene reporters**

*3.5.2.1 Polyadenylation pattern in -1A mutant is rescued with U1 -1A+4C and shift**+9***

To test the effect of modified U1s on usage of polyA sites in short and long intron context, I co-transfected these minigenes with U1 -1A+4C and U1 shift+9, and evaluated their effect by analysis of 3'RACE PCR products (Figure 4-14 A). Results show that co-transfection of -1A mutant with U1 -1A+4C increases the usage of distal polyA from ~75% to ~82% in short, while U1 shift+9 did not show an increase. In the long context, -1A was improved to ~38% and 11% when U1 -1A+4C and shift+9 were used, respectively (Figure 4-14 B and C).

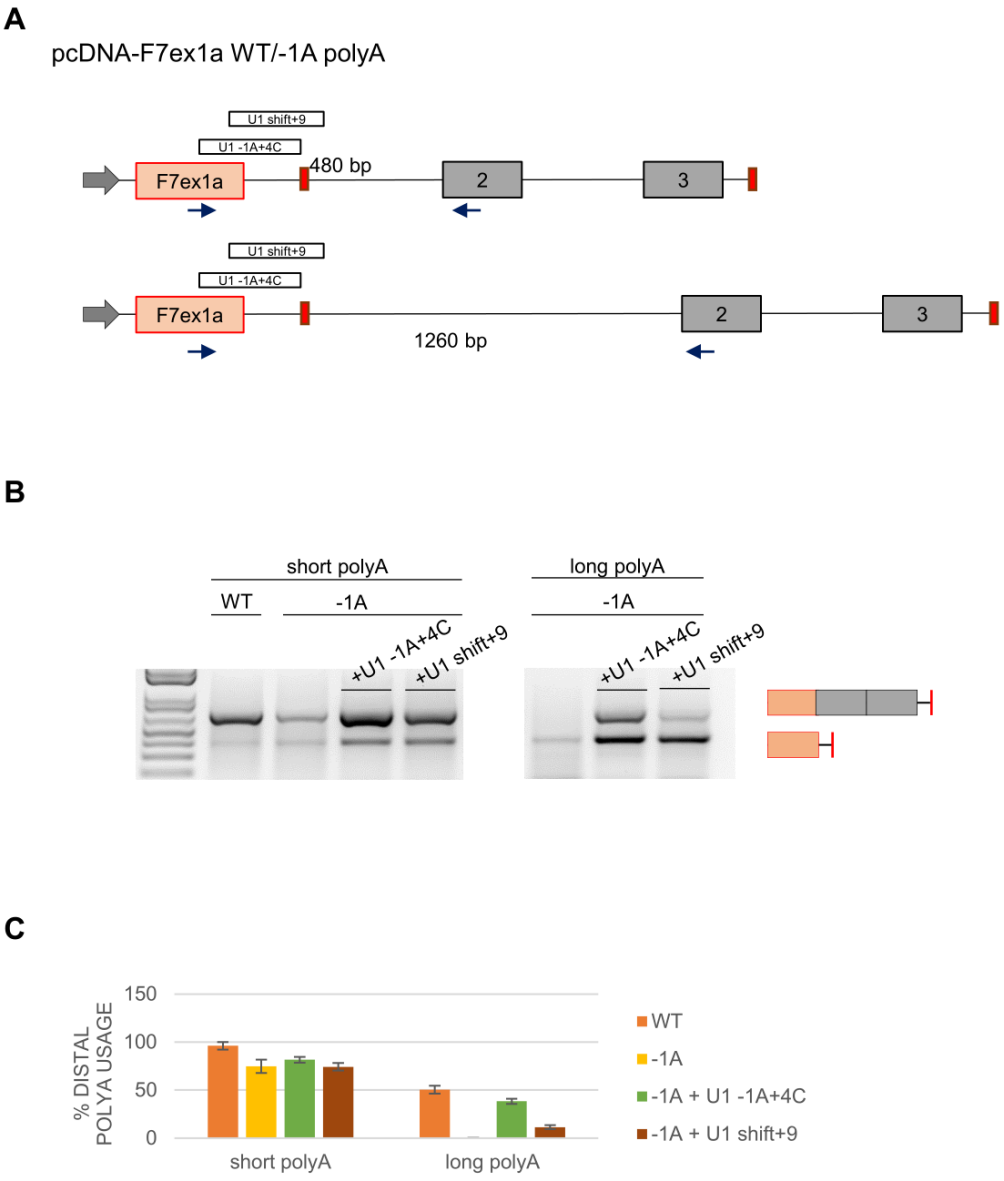

Figure 4-14 Modified U1s -1A+4C and shift+9 restore the usage of distal polyA signal in -1A mutant A. Schematic representation of F7ex1a polyA minigene in short and long intron context, with indicated binding sites of modified U1s. Red rectangle represents polyadenylation signal. B. 3'RACE PCR resolved on 1.5% agarose gel. Identity of the bands is indicated. C. ImageJ quantification of agarose gel. Data are expressed as mean ±SD of three independent experiments.

3.5.2.2 U1 snRNA mutants that do not bind U1A, 70K and L4 are defective in rescue of polyadenylation pattern in -1A long

For addressing the effect of different U1snRNP components in observed rescue of polyadenylation pattern, I have used previously described mutants of U1 -1A+4C and shift+9

that are unable to bind U1A protein (stem loop 2 mutant), 70K protein (stem loop 1 mutant) and the mutant of stem loop 4, which was the most important one for splicing rescue.

I have co-transfected these modified U1s and their mutants with -1A long polyA minigene (Figure 4-15 A), and evaluated their effect on ability to rescue the polyadenylation pattern. The U1A mutant of U1 -1A+4C reduced the rescue to ~74%, while in shift+9 it reduced the rescue to 40%. 70K mutant was again shown to be more delirious than U1-A mutant in U1 -1A+4C, reducing polyadenylation rescue of this particle to ~26%. In U1 shift+9 the same mutant caused reduction to ~37%. Finally, the mutant of stem loop 4 in both U1 snRNPs severely reduced rescue, leaving them only ~4-5% active (Figure 4-15 B and C).

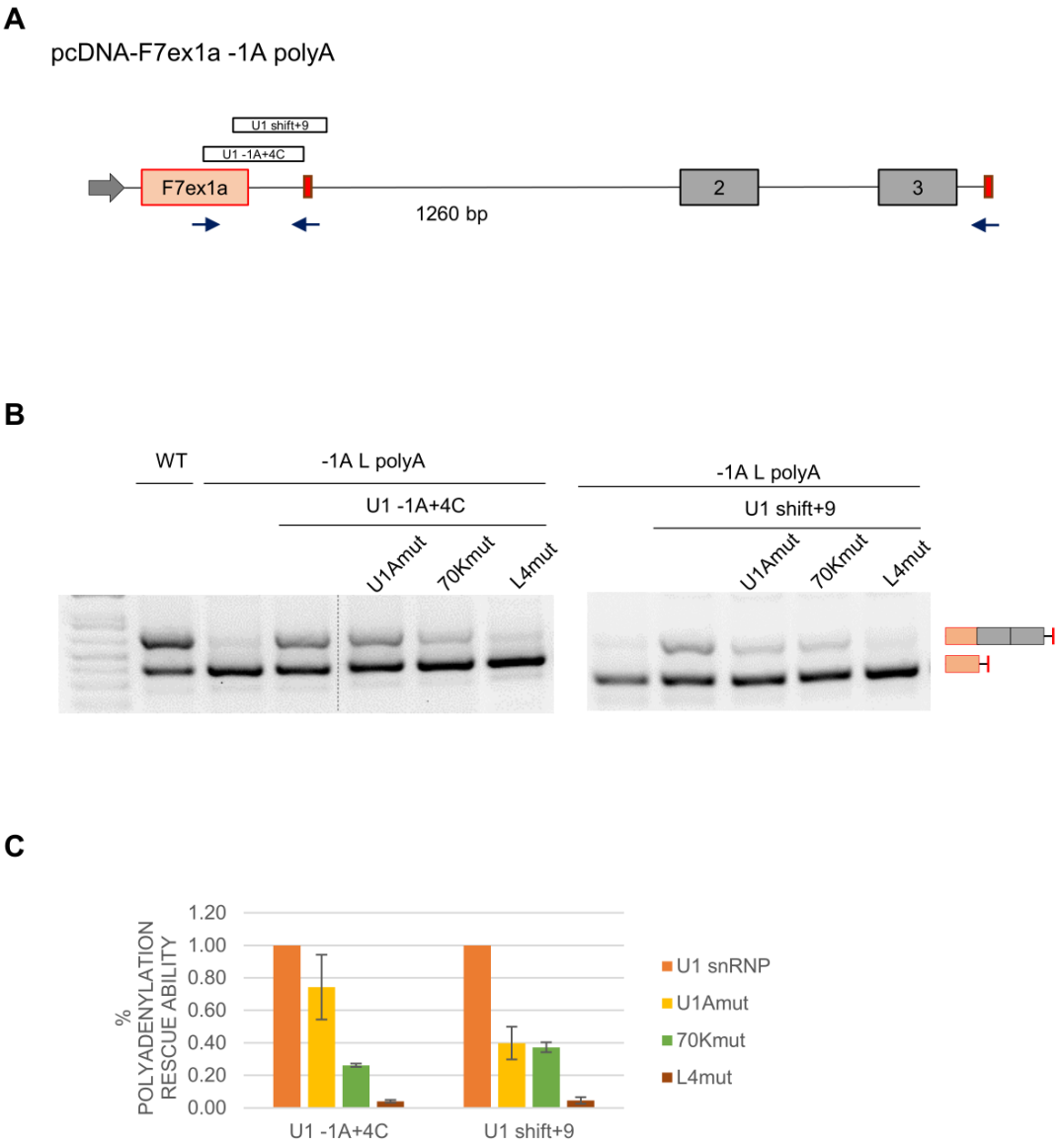

Figure 4-15 Stem loop 4 is the most important component of modified U1snRNP for the rescue of polyadenylation pattern.

A. Schematic representation of F7ex1a -1A polyA minigene in long intron context, with indicated binding sites of modified U1s. Red rectangle represents polyadenylation signal. Primers for 3'RACE are indicated. B. 3'RACE PCR resolved on 1.5% agarose gel. Identity of the bands is indicated. C. ImageJ quantification of agarose gel. Data are expressed as mean ±SD of three independent experiments.

3.6 Analysis of the WT and –1A mutant on polyadenylation selection in the Flp-In T-REx 293 inducible cells

3.6.1 **Pp 5'ss mutation and intron length promote intronic polyA usage**

For testing the effect of pp 5'ss mutation on selection of polyA signal in chromosomal context, I have created cell line with stably integrated copy of F7ex1a polyA minigene that is inducible upon the addition of tetracycline. These stable clones were created with WT and -1A versions of pp 5'ss in both short and long intron contexts (Figure 4-16 A). Experiments were performed as for splicing stable clones, and analysed by 3'RACE PCR.

Samples were collected at time points 0, 8h and 24h. At 8h post-induction WT short shows 84% of distal polyA usage, -1A short uses it in ~20%, while in the long context usage of distal signal is ~61% and >1% in WT and -1A, respectively. At the 24h time point values do not show big deviation from those previously measured, displaying ~90% and ~29% in WT and -1A short, and ~73% and ~1% in WT and -1A long (Figure 4-16 B and C). At time point 0 transcripts were barely detectable and unreliable for quantification.

These results all together confirm the observation from transient transfection experiments, showing that mutation in the pp 5'ss changes the ratio between proximal and distal polyA signal usage, reducing it with an intron length-dependent strength.

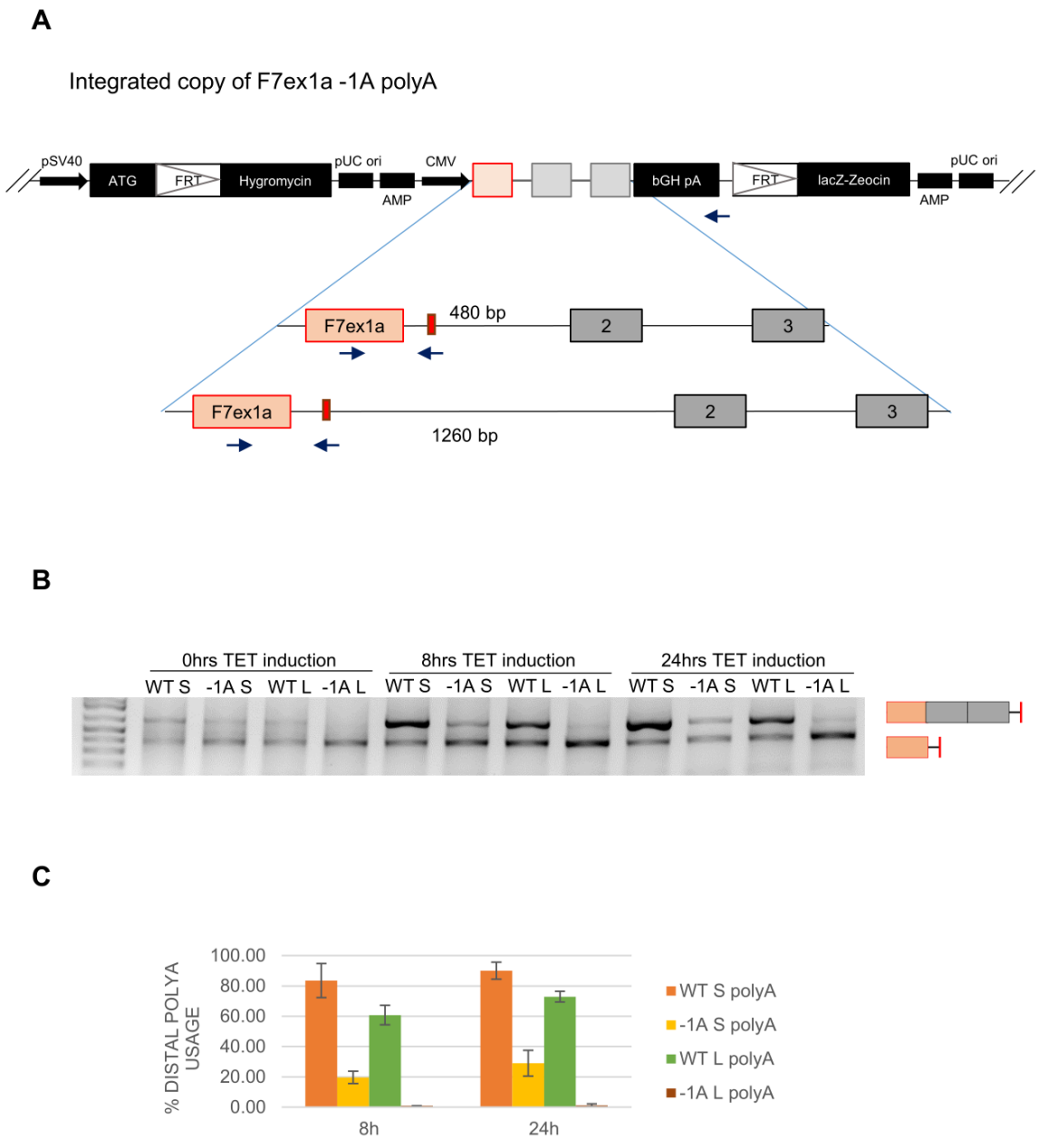

Figure 4-16 Pp 5'ss mutation promotes the usage of proximal polyadenylation signal in chromatin context. A. Schematic representation of F7ex1a polyA minigene in short and long intron context. Red rectangle represents intronic polyadenylation signal, while the regular one is presented in black in the upper part of scheme. Primers for 3'RACE are indicated. B. 3'RACE PCR resolved on 1.5% agarose gel. Identity of the bands is indicated. TET – tetracycline. C. ImageJ quantification of agarose gel. Data are expressed as mean ±SD of three independent experiments.

3.6.2 **Modified U1 snRNP 5'ss rescues the polyadenylation pattern in -1A mutant in chromosomal context**

My previous results suggest that pp 5'ss reduces the usage of distal polyA signal, in favor of the proximal one. This effect is a probable consequence of reduced binding of U1 snRNP, and to address this question I have performed rescue experiments in chromatin context using modified U1 that binds to the WT or -1A mutant 5'ss (Figure 4-17 A).

WT and -1A polyA Flp-In T-REx 293 cell lines with short and long first intron context were plated, transfected with modified U1 constructs, induced with tetracycline for 24h and collected for total RNA extraction.

U1 +4C did not significantly increase the usage of distal polyA signal in WT constructs, neither in short intron (from ~93% to ~94.7%) nor in the long intron context (from ~70% to ~72%). Short intron -1A mutant was rescued from ~34% to ~58% when U1 -1A+4C was used. More significantly, usage of distal polyA signal in -1A mutant of long intron was improved from ~3% to ~34% (Figure 4-17 A). These results are similar to those observed in minigene systems (Figure 3-3).

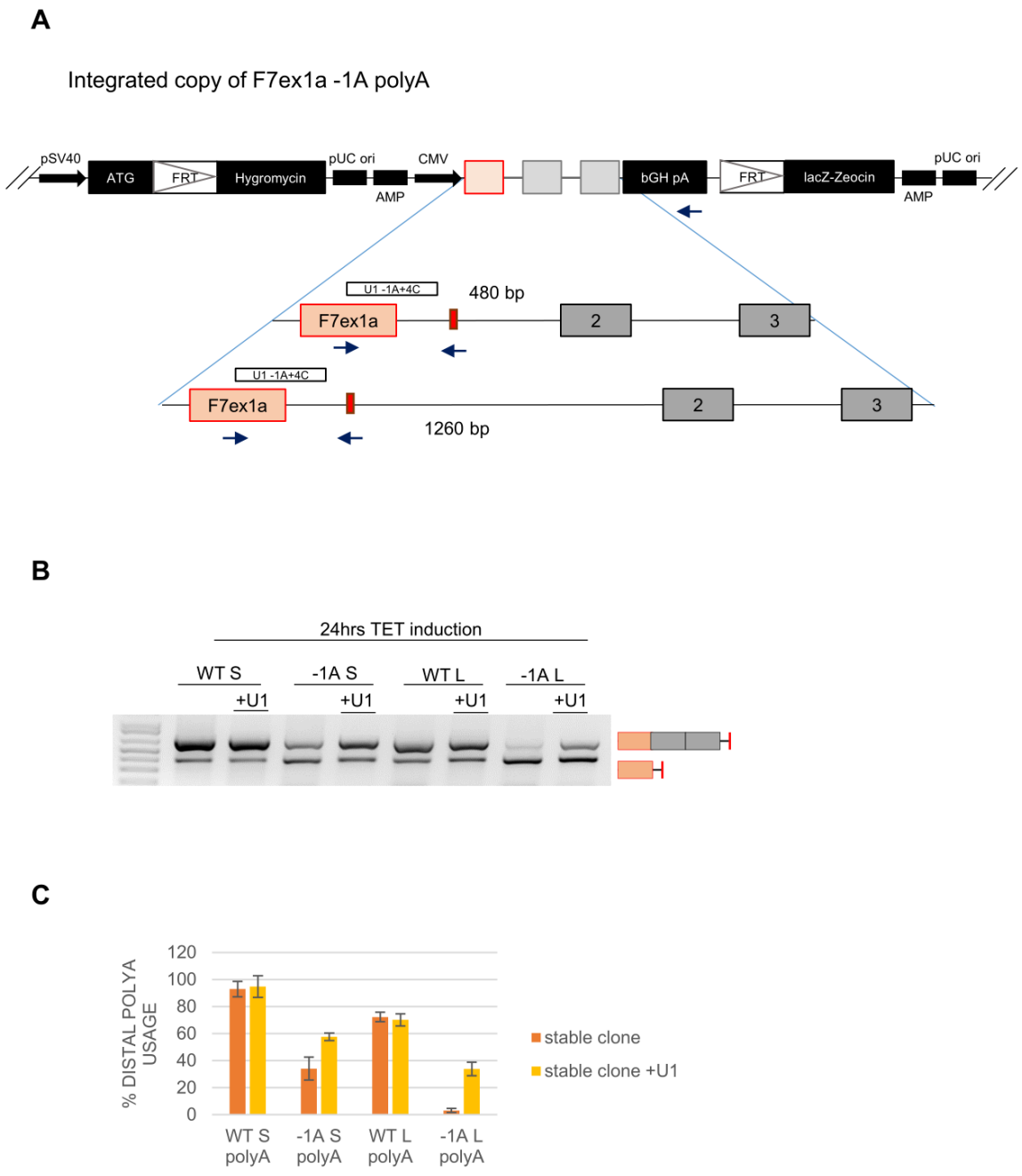

Figure 4-17 Modified U1s -1A+4C is able to restore the distal polyA signal usage in -1A mutant. A. Schematic representation of F7ex1a -1A polyA minigene in long intron context, with indicated binding sites U1 -1A+4C. Red rectangle represents polyadenylation signal, while the regular one is presented in black in the upper part of scheme. B. 3'RACE PCR resolved on 1.5% agarose gel. Identity of the bands is indicated. C. ImageJ quantification of agarose gel. Data are expressed as mean ±SD of three independent experiments.

3.7 Modified U1 rescues F7 protein and activity in a splicing competent minigene

Previously described and tested U1 -1A+4C was shown potent to rescue splicing in F7 mutant hybrid minigene reporter, promoter-proximal system, as well as in chromosomal context. In order to evaluate the potential therapeutic effect of splicing rescue, cells were co-transfected with minigene that include of all F7 exons and U1 -1A+4C (Figure 4-18 A). Growth medium was collected after 48h, followed by ELISA measurement of secreted F7 protein, as well as coagulant activity. This experiment was done in collaboration with prof M.Pinotti (UniFe). Results show that modified U1 induce a significant rescue of the F7 protein levels, as shown by ELISA measurement (Figure 4-18 B) and Western blot analysis (Figure 4-18 D). Compared to WT, upon addition of U1 -1A+4C the protein is rescued to ~87.7%. The overall coagulant activity of conditioned medium is increased ~60-fold upon rescue indicating that the protein is functional (Figure 4-18 C). Overall, the -1A mutant has reduced protein level, but the protein itself retains normal specific activity, which could explain the coagulation phenotype of patients.

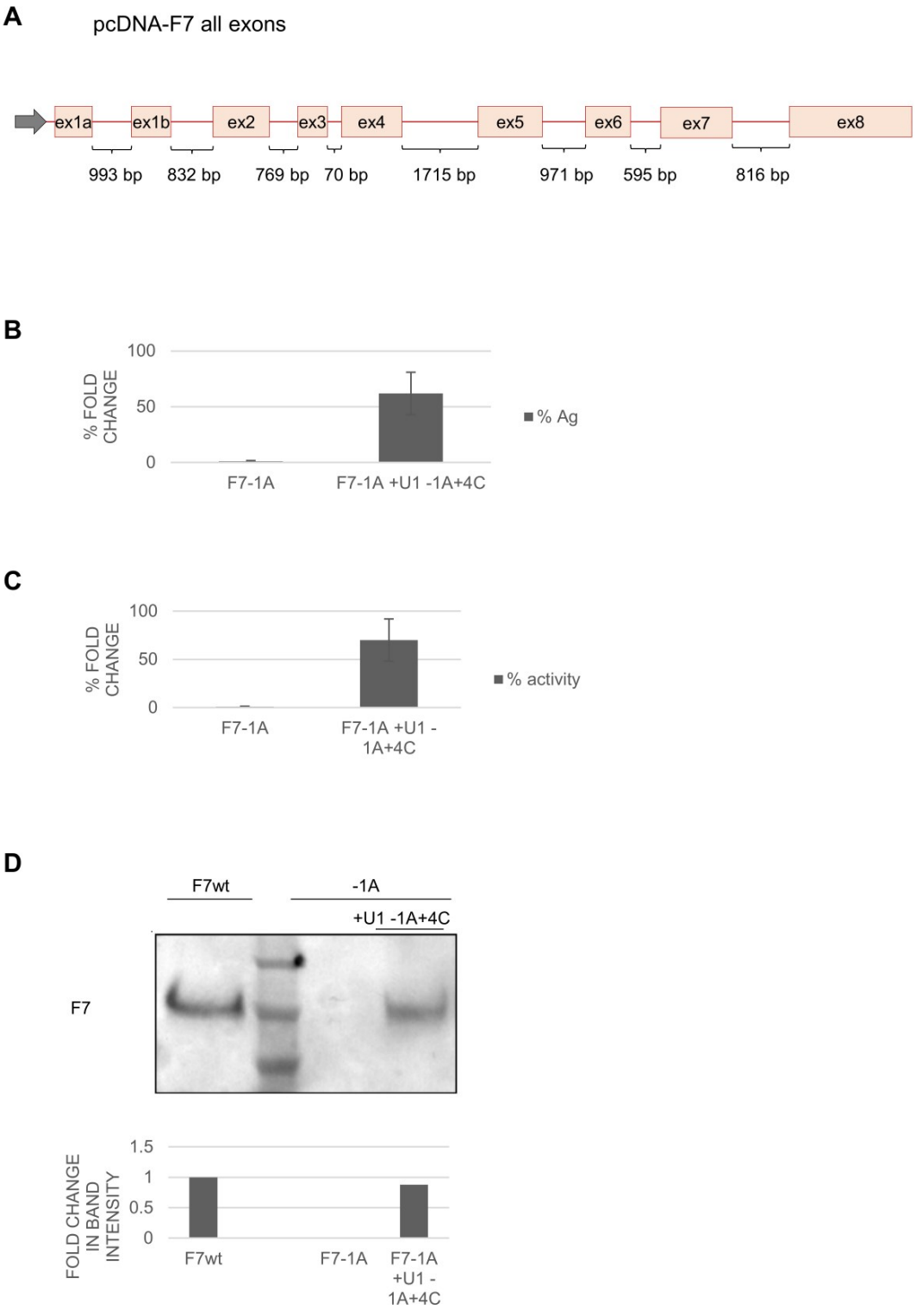

Figure 4-18 Modified U1 rescues protein level and coagulant activity of F7. A. Schematic representation of F7 minigene comprising all exons. Introns are shortened compared to hF7, with lengths indicated on the scheme. B. F7 antigen levels measured by ELISA in culture medium derived from cells transfected with full length F7 minigene. Values are shown as fold change with -1A mutant level set to 1. Data are expressed as mean ±SD of three independent experiments. C. Coagulant activity of F7 derived from growth medium

from cells transfected with full length F7 minigene. The activity of protein derived from -1A mutant is set to 1, and rescue is expressed in fold increase. Data are expressed as mean ±SD of three independent experiments. D. Western blot for detection of F7 in culture medium derived from cells transfected with full length F7 minigene. Bands are quantified using ImageJ and represented on the graph bellow.

3.8 Analysis of promoter proximal mutations in *SLC26A2* gene associated to diastrophic dysplasia and evaluation of modified U1s

Mutations in the DiasTrophic Dysplasia Sulfate Transporter (DTDST or *SLC26A2*) are associated with recessive inherited chondrodysplasias. The severity of the disease range from lethality shortly after birth (achondrogenesis 1B and atelosteogenesis type 2) (Ballhausen et al., 2003; J Hästbacka et al., 1996; Andrea Superti-Furga et al., 1996), major physical impairments (diastrophic dysplasia) (Johanna Hästbacka et al., 1994), and mild clinical phenotype (recessive form of multiple epiphyseal dysplasia) (A Superti-Furga et al., 1999). The SLC26A2 +2T>C Finnish founder mutation to affect the pp 5'ss and severely reduce the mRNA levels to 3-6% (Johanna Hästbacka et al., 1999). Roughly 90% of Finnish DTD chromosomes are considered to carry this mutation, located within the 5'UTR. In homozygous patient fibroblasts sulphate uptake was seriously impaired, and DTDST mRNA was undetectable on northern blots (Johanna Hästbacka et al., 1999). To study the molecular basis of this pp 5'ss mutation and evaluate the potential therapeutic role of U1s I created a minigene reporter system that includes SLC26A2 exon 1, part of intron 1 and exon 2 (Figure 4-19 A). I also designed two modified U1 that bind to the 5'ss of WT and +2C mutant. I co-transfected WT and mutant minigenes with corresponding U1s in HeLa cells, together with pSV-CFTR plasmid, here used for normalization. After total RNA extraction and cDNA synthesis, I performed semi-quantitative RT-PCR analysis (Figure 4-19 B). Samples were separated on 1.5% agarose gel with ethidium bromide, and products were visualized under UV light. In WT construct only one band was present, corresponding to properly spliced product. However, in +2C mutant another higher band was visible, corresponding to intron retention (Figure 4-19 B). My results show that modified U1 is not significantly affecting levels of WT transcript, but is increasing transcript coming from +2C mutant ~2.3-fold (Figure 4-19 C). Interestingly, modified U1 did not rescue splicing defect in +2C, suggesting that additional mechanism might be involved.

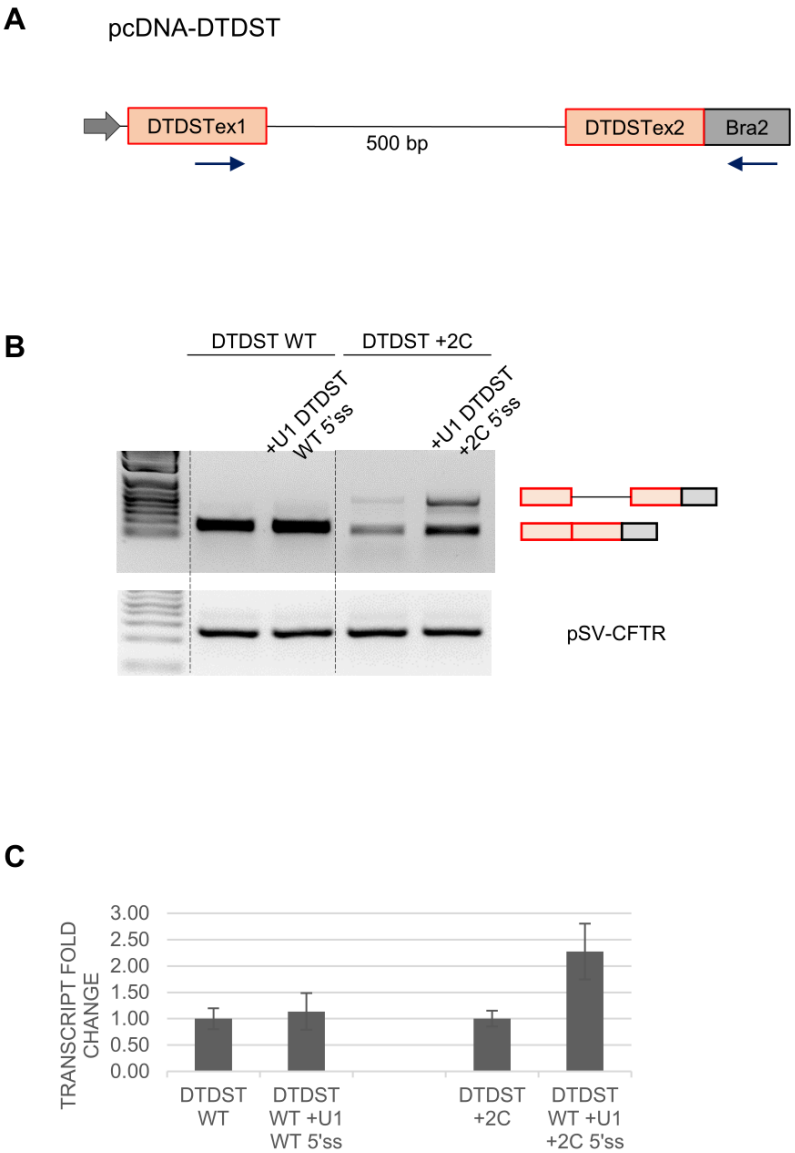

Figure 4-19. Modified U1 increases amount of transcript in DTDST minigene. A. Schematic representation of DTD minigene. PCR primers are indicated. B. Semi-quantitative RT-PCR products loaded on 1.5% agarose gel with ethidium bromide. Identity of the bands is indicated. pSV-CFTR is co-transfected together with DTD plasmid, and is used for normalization of samples. C. ImageJ quantification of agarose gel. Data are expressed as mean ±SD of three independent experiments.

3.9 Modified U1 rescues amount of spliced transcript in SMN2 minigenes

To test the effect of modified U1 on pp 5'ss I have also evaluated the SMN2 gene. Increase in SMN transcription is considered a potential therapeutic strategy for treatment of Spinal Muscular Atrophy (reviewed in (Sendtner, 2010)). Previous strategies included nonselective histone deacetylase (HDAC) inhibitors, that are known to upregulate transcription of 2% of all genes (Marks, Richon, Miller, & Kelly, 2004) including SMN. The major drawback with this approach is the fact that the drug treatment is not specific for SMN promoter, thus the modified U1 strategy could be beneficial for obtaining specific increase in SMN transcription while limiting off targets at the same time.

To test this strategy, I created a minigene that include SMN2 exon 1a and part of its intron 1 under control of CMV promoter (Figure 4-20 A). I designed two ExSpeU1s that bind the intron 1 sequence 3 and 13 nucleotides downstream from the pp 5'ss, called U1 shift+3 and U1 shift+13, respectively. After transfection in HeLa, total RNA extraction and cDNA synthesis, I performed quantitative real time PCR using primer pair positioned and the first and second exon of the minigene. My results show that ExSpeU1 shift+3 and shift+13 increase level of spliced transcript 1.4-fold and ~4 fold, respectively (Figure 4-20 B). Since CMV is not natural promoter for SMN gene, I also replaced CMV with the endogenous SMN promoter (Figure 4-20 C). In this case, co-transfection experiments showed that shift+3 and shift+13 induced a ~1.4 fold and ~1.9-fol increase in the transcript, respectively (Figure 4-20 D).

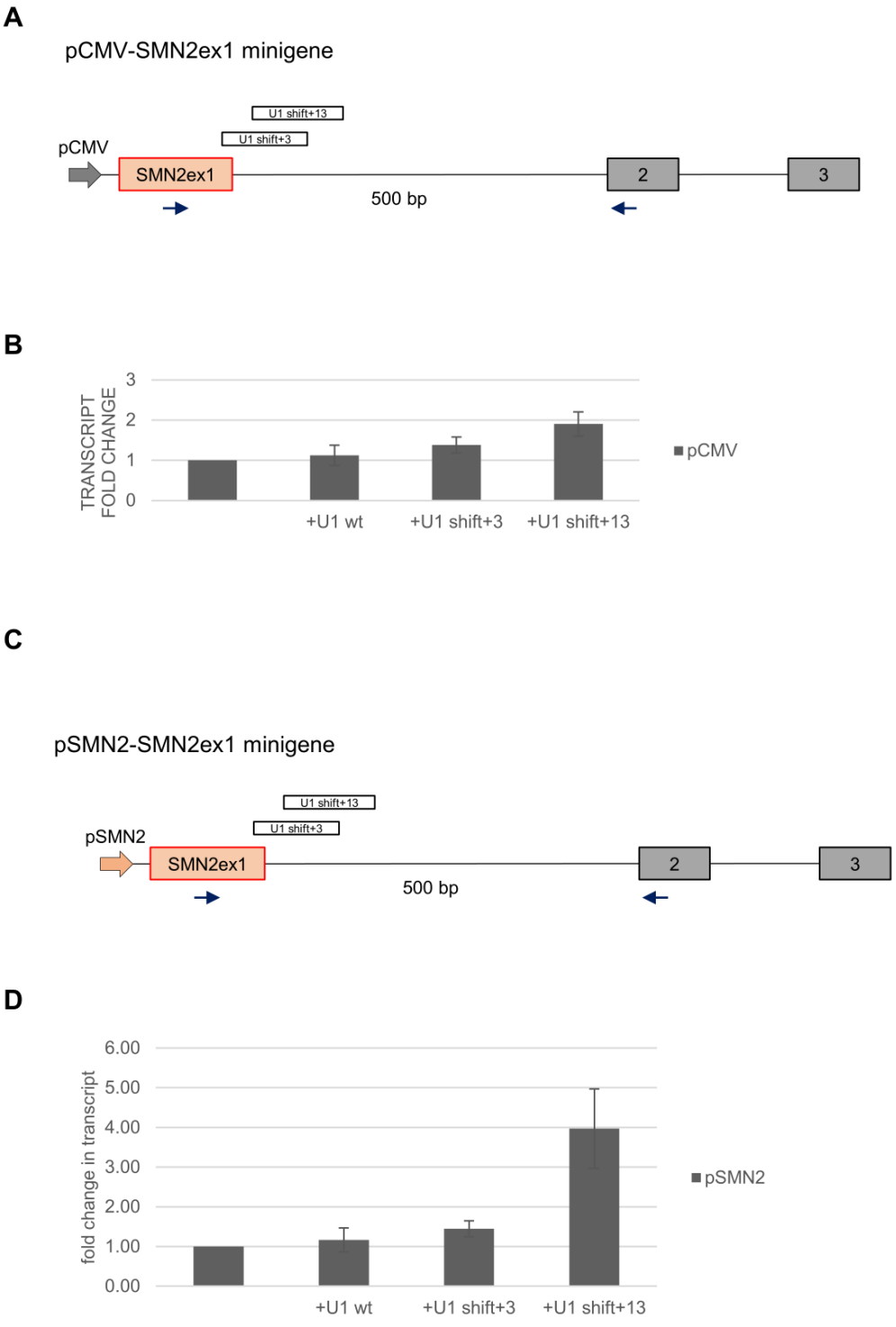

Figure 4-20. Spliced transcript is increased in SMN minigenes upon the ExSpeU1 rescue. A. Schematic representation of pcDNA SMNex1 minigene under the regulation of CMV promoter. Primers for quantitative real time PCR are indicated. B. Quantitative real time RT-PCR. Fold change in spliced transcript is expressed in comparison to expression in non-treated minigene. Data are expressed as mean ±SD of three independent experiments. C. Schematic representation of pcDNA SMN2ex1 minigene under the regulation of SMN promoter. Primers for quantitative real time PCR are indicated. D Quantitative real time RT-PCR. Fold change in spliced transcript is expressed in comparison to expression in non-treated minigene. Data are expressed as mean ±SD of three independent experiments.

5. DISCUSSION

The major aim of my studies was to understand the effect on pre-mRNA processing of a peculiar class of mutations located at the promoter proximal 5' splice site (pp 5'ss) and identify potential strategies to correct the underlying splicing defect. Donor splice site mutations at internal exons typically can induce activation of cryptic splice site or, more frequently, cause skipping of the entire exon. Skipping of the internal exon is mainly due to fact that the 5'ss mutations, reducing or abolishing binding of the U1 snRNP, interfere with the exon definition during the first step of the splicing process (Alanis et al., 2012; Pinotti et al., 2008; Susani et al., 2004; Tanner et al., 2009). In this case, modified U1s that bind to 5'ss or to downstream intronic sequences have been shown to rescue exon skipping defects (Alanis et al., 2012; Franco Pagani et al., 2002; Tajnik et al., 2016). However, when a pp 5'ss mutation affect the first exon, it obviously cannot induce exon skipping and the potential therapeutic effect of modified U1s RNA is not known. In this thesis, using different minigene model systems I have specifically addressed the effect of pp 5'ss mutations on two pre-mRNA processing steps related to the U1 snRNP function, splicing and alternative polyadenylation. My results clearly show that pp 5'ss mutations can indeed have a negative effect of these two events and thus explaining some of the molecular events at the basis of the diseases. In addition, and interestingly from the therapeutic point of view, I show that the pp 5'ss mutation-induced defects can be rescued by modified U1 snRNAs that bind either at the donor site or at downstream intronic sequences. These results further extend the potential usage of modified U1 snRNA for the therapeutic correction of different pre-mRNA splicing defects.

3.10 Promoter proximal 5'ss mutations affect splicing

To understand the role of pp 5'ss in pre-mRNA processing, I focused on three pp 5'ss mutations (-1A, +4T and +5A) that are located in the first exon of Factor 7 (F7) and associated to coagulation defects with different severity. These mutations do not affect the invariant GT dinucleotide of the splice site but reduce the complementarity of the donor site with the U1 snRNA (Figure 3-1 B). In fact, the normal WT pp 5'ss has one mismatch in position +4 (C instead of A), the -1A mutant two mismatches (at +1 and +4), +5A mutant two mismatches (at position +4 and +5), whereas the +4T mutant maintain the mismatch present in the WT at position +4 (T instead of A). The predicted effect of the mutations according to bioinformatics

programs indicate that, even though no mutation completely abolishes the score, -1A and +5A were the most deleterious reducing the score to 0.63 and 0.68, respectively. Interestingly, these three mutations tested in an artificial minigene as part of an internal exon induce exon skipping (Figure 3-1 C, D), strongly suggesting that they reduce to some degree the recognition of the 5'ss. However, it is important to note that this artificial system is probably very sensitive to subtle changes in the 5'ss strength as the WT hybrid exon is only partially included in the final processed transcript (~35%, Figure 3-1 C, D). Nevertheless, this system proved useful for a rapid screening of the therapeutic modified U1s (see below). In order to study the effect of the mutations in the promoter proximal context, I created another set of minigenes with the F7ex1a as the first exon. These minigenes with a different length of the first intron were initially evaluated in simple transient transfection experiments. My results show that the pp 5'ss mutations in this context reduce the spliced transcripts and thus provide a useful model that can be used to understand the pathological effect of these type of mutations. In general, depending on the type of pp 5'ss mutation, I observed a gradient of severity related to the residual base-pair complementary of the pp 5'ss sequence with the endogenous U1 (Figure 3-1 B). In transfection experiments, even if the magnitude of effect was modulated by the intron length, (see below for the role of the intron length in splicing and polyadenylation), the -1A was the most severe followed by +5A mutant and by the +4T (Figure 3-4). The pathological effect of the most severe pp 5'ss mutation was also tested in the stable Flp-In T-REx 293 cells in order to assess the kinetic of transcriptional activation. My results clearly show that in the -1A pp 5'ss variant the transcriptional activation induced by the addition of tetracycline resulted in extremely low amount of spliced transcript. To determine at which stage this effect was occurring, I also performed cellular fractionation and my results indicate that the defect start to be evident in the chromatin RNAs (Figure 3-11). The lack of spliced transcript after transcriptional induction suggest that the lack of the U1 snRNP at the pp 5'ss due to the mutations affect also transcription. Indeed, previous studies using artificial mutations at the pp 5'ss reported a reduction in transcription. This phenomenon was observed *in vitro* (Fong & Zhou, 2001), in transient transfection (Furger et al., 2002) and also stably integrated minigenes (Damgaard et al., 2008). This is also in line with the report that U1 is involved in TFIIH-facilitated transcription (Damgaard et al., 2008; Kwek et al., 2002), since this transcription factor is involved in promoter clearance (Kugel & Goodrich, 1998; Kumar, Akoulitchev, & Reinberg, 1998), but the exact mechanism through which U1 snRNP promotes transcription remains unclear.

3.11 Promoter proximal 5'ss mutations affect alternative polyA selection.

In eukaryotes, first introns are typically long (Bradnam & Korf, 2008), and stochastically contain more cryptic polyadenylation signals compared to internal introns. At the same time, first exons of human genes are often relatively short, so that donor splice site is placed near the promoter (International Human Genome Sequencing Consortium 2001). In this setting, U1 binding to the pp 5'ss can also be involved in protecting premature cleavage and polyadenylation (PCPA). Indeed, there are studies suggesting general role of U1 in inhibition of polyadenylation (Berg et al., 2012; Kaida et al., 2010; Oh et al., 2017). In order to test the effect of pp 5'ss mutations on polyadenylation, I inserted an artificial polyadenylation signal in the first intron of the F7 minigenes. My results show that the pp 5'ss mutations promote usage of proximal polyA signal in both transient transfection (Figure 3-12 and 3-13) and chromosomal context (Figure 3-16). As observed with splicing, a gradient of severity was observed that is related to the strength of pp 5'ss sequence analuzed (Figure 3-1 B), and also modulated by the intron length (See below for discussion). This observation expands the knowledge of pathological effects caused by pp 5'ss mutations, suggesting that this type of mutations might also induce cryptic polyA selection in the first intron. However, even if the artificial minigene system I used for these experiments indicates a possible relation with polyadenylation, future studies should be performed in more physiological contexts to fully appreciate its relevance. Considering F7, several potential cryptic 5'ss are located in the first intron downstream the sequences investigated in our experimental conditions and accordingly they might contribute *in vivo* to the pre-mRNA defect.

3.12 Role of the first intron length on pre mRNA processing defects associated to the promoter proximal 5'ss mutations.

Interestingly, I observed that the length of the first intron affects pre-mRNA processing by modulating the severity of the pp 5' mutation defects. This effect was evident both for splicing and polyadenylation even if some differences in the extent of the effect can be observed between the experimental procedures particularly for splicing. In transient transfection assay, the pp 5'ss mutation showed a modest effect on splicing in the short intron context (Figure 2-1B) (for example, the -1A mutant was 60% of WT) but the same mutation nearly completely abolish the biosynthesis of the transcript in the long intron context (Figure 2-1C and D). This

intron length mediated effect was clearly evident both in the case of the pp 5' mutations and for the WT 5'ss. This could be due to the fact that also the WT pp 5'ss of F7 is suboptimal as it does not perfectly match to the canonical donor site consensus (Figure 3-1 B). The length of the first intron affected also the amount of transcript in the experiments performed with the integrated inducible minigenes in Flp-In™ 293 T-REx cells (Figure 2-5). However, in this case the effect was less pronounced, suggesting that the integration status of the minigene and the resulting chromosomal context might modulate the RNAPII pre-mRNA processivity (see below). The intron length resulted also in changes on the selection of alternative polyadenylation sites. A longer intron resulted in the preferential selection of the proximal intronic polyA site and this effect was evident both in transfection experiments and in the Flp-In™ 293 T-REx cells (Figures 3-13 and 3-16). It is important to note that the position of intronic polyA signal was kept on constant distance from the promoter in all constructs, and spacers were inserted downstream to increase the intron length. Thus, the observed effect is due to the increase of intron length, rather than moving the polyA signal further away from the 5'ss.

This intron length-dependent effect on splicing and polyA selection could be due to a reduced processivity of the RNAPII related to the amount of U1 snRNP available at the pp 5'ss. It is possible that the lack, or the severely reduced amount of U1 snRNP present on the defective pp 5'ss reduces the elongation properties or the processivity of the RNAPII. U1 snRNP have been indeed reported to be associated with RNAPII (Kwek et al 2002 Nat Struc Biol, Tian et al 2001 FEBS Letters, Das et al 2007 Mol Cell, Spiluttini et al 2010 Journal of Cell Science). In the presence of a long intron, a poorly processive RNAPII not correctly primed by the pp 55'ss U1 ssRNP will detach from transcription before reaching the 3'ss. However, in the case of a relatively short intron the RNAPII will be quickly engaged in splicing thus recovering its fully processivity. Nevertheless, the effect on splicing is reduced when the minigenes are inserted in the chromosomal context as unique copy in the Flp-In™ 293 T-REx model cells (Figure 3-7). The reduced chromatinisation of transfected plasmid DNA compared to integrated copy potentially might explain the differences in the effect of intron length. In fact, in transient transfection experiments, plasmids are not well chromatinised as the endogenous genes. In my experiments, total plasmid RNA was extracted from transformed *E.coli* DH5α using NucleoSpin® Plasmid kit (Macherey-Nagel, Germany), which permits extraction of both supercoiled (ccc) and open circular form. Ratio between these two forms was assumed to be similar between different constructs. Chromatinisation of plasmids is shown to depend on the level of condensation. For example, negatively supercoiled plasmids have rapid and more efficient core histone transfer compared to open circular ones (Nakagawa, Bulger, Muramatsu, & Ito, 2001; Pfaffle & Jackson, 1990). Furthermore, although transfected DNA is arranged in

nucleosome-like particles, its structure is anomalous (Jeong & Stein, 1994). In fact, using micrococcal nuclease digestion and Southern hybridization techniques, the plasmid forms a ladder of DNA fragments in length ~300, 500, 700, 900bp whereas in the case of endogenous genes typical 180-190bp multiples are detected. Thus is possible that the different chromatin status might influence the pp 5'ss U1-mediated effect on the RNAPII processivity and in particular affect differentially splicing versus polyA processivity.

3.13 Modified U1 snRNAs rescue the pre-mRNA processing defects induced by the promoter proximal 5'ss mutations.

Several in vitro and in vivo evidence indicate that modified U1 snRNAs that either bind to the 5'ss or to downstream intronic sequences rescue exon skipping defects at internal exons (Dal Mas et al., 2015; Rogalska et al., 2016; Tajnik et al., 2016; van der Woerd et al., 2015). In these cases, the modified U1 snRNA correct exon skipping caused by mutations that reduce the exon definition. These mutations can be located in different regulatory elements: the 5'ss consensus, the exonic regulatory elements or the polypyrimidine tract (Alanis et al., 2012; Dal Mas et al., 2015; Tajnik et al., 2016). In this thesis, I provide evidence that modified U1s can also correct the pre mRNA processing defects caused by pp5'ss mutations. Both in transient transfection and chromosomal context I show that these molecules rescue the splicing (Figure 3-5 and 3-8) and the polyadenylation defects (Figure 3-14 and 3-17) caused by the pp 5'ss mutations. In general, I observed that modified U1s that bind to the pp 5'ss are more potent in comparison to those that bind downstream, the shifted ones. This can be due either to the position of binding or to the fact that the shifted U1s, containing a more consistent change in their tail are expressed at lower levels. The shifted U1s (also named ExSpeU1s when acting on internal exons) were conceived to reduce potential off targets effect and thus should be preferable in comparison the modified U1s that bind at the 5'ss consensus. In fact, their binding into intronic sequences is expected to have less off target sites on the transcriptome in comparison to the modified U1s that bind directly to the 5'ss consensus. However, in the most relevant context represented by the inducible system in Flp-In T-REx 293 cells the shifted U1s maintained a significant rescue activity: shift+9 U1 induced a 67-fold increase, compared to 106-fold for U1 5'ss in the short context (Figure 3-8), and a similar trend was observed in the long construct, where rescue was 116-fold for U1 5'ss and 78-fold for U1 shift+9 (Figure 3-8). To have the first evidence that the modified U1s correct also the protein I have created a

splicing-competent minigene that code for the entire F7 and the experiments showed that U1 -1A+4C completely recovered F7 expression to the level obtained with the WT minigene (Figure 2-12). To have a further evidence of the potential therapeutic effect of modified U1 snRNA I focused on diastrophic dysplasia (DD) and spinal muscular atrophy (SMA). In DD, I evaluated in minigene assay a frequent mutation that affect the pp 5'ss of the causative gene encoding a sulphate transporter (DTDST). This mutation reduces *in vivo* the DTDST mRNA levels to only ~5% (Johanna Hästbacka et al., 1999) and any treatment that will increase the amount of transcript is considered beneficial. Using minigene assay, I showed that pp 5'ss mutation reduces the amounts of normal transcript and that modified U1 significantly increased the total transcript levels. Unexpectedly, in this case I also observed that the mutations induce some degree of intron retention (Figure 2-13), that was not affected by the modified U1s, suggesting that the pp 5'ss might also induce, in some cases, additional pre-mRNA processing defects. In addition, I applied the ExSpeU1 approach on exon1 of SMN2 gene, which is involved in SMA. SMN2 transcriptional activation has been considered beneficial. For example, treatment with trichostatin A, sodium butyrate and valproic acid have been done with this purpose and showed some phenotypic improvement in mouse models (Avila et al., 2007; Chang et al., 2001; Tsai, Tsai, Lin, Hwu, & Li, 2006) and in patients (Swoboda et al., 2009). To evaluate the effect of modified U1 on SMN2 transcriptional activation, I designed a minigene that contains the exon 1 of SMN2 gene under its constitutive core promoter or the CMV promoter, along with modified U1 snRNA that bind in the first intron just downstream the pp 5'ss. The results showing that these modified U1snRNAs increase the level of SMN2 transcript ~1.5-2 fold (Figure 3-20) suggest a broader activity of these molecules also on non-mutated pp 5'ss.

3.14 Role of structural elements of modified U1 snRNAs in rescue of the pre-mRNA processing defects

In order to understand how the modified U1 snRNAs rescue the pp 5'ss defects, I have evaluated U1 snRNA mutants that contain specific mutations in the stem loops. Previous studies showed that the U1 particles with mutations in stem loop 2 and stem loop 1 lack U1-A and 70K binding, respectively (Rogalska et al., 2016), whereas mutants of stem loop 4 affect the interaction of particle with components recognising 3'ss (Sharma et al., 2014). Comparing the rescue activity of the different mutants with the activity of the corresponding modified U1 snRNA I was able to assess the relative importance of the different U1 structures on rescuing

the pre-mRNA processing defects induced by the pp 5'ss variants. My result performed in the factor 7 minigenes show that the two most important elements involved in both splicing and polyadenylation rescue are represented by the 70K and stem loop 4 (Figure 3-6). Modified U1 snRNA unable to bind 70K had severely reduced rescuing capabilities dropping to only 10-38% in splicing (Figures 3-6) and 26% in polyadenylation rescue (Figure 3-15). Stem loop 4 showed similar or even stronger effect, as in its absence rescue was reduced to less than 10% (Figure 3-6) in transient transfection and 20-25% in chromosomal context of splicing (Figure 3-9), and only 4-5% in polyadenylation rescue experiments (Figure 3-15). On the other hand, the U1-A seems dispensable and not strictly required for splicing or polyadenylation as shown by co-transfection experiments with the corresponding mutants (Figures 3-6 and 3-15). Interestingly, the analysis performed with mutants of modified U1 snRNAs that target defects at internal exons showed similar results (Rogalska et al., 2016) suggesting that 70K and loop 4, but not U1-A, are important in different pre mRNA processing events including polyadenylation. To my knowledge, the role of structural components of U1 snRNP in protection from premature cleavage and polyadenylation has been assessed only in one study so far. U1 binding to the major splice donor site ~200bp downstream from promoter has been shown to suppress the usage of polyA site and stem loop 1 binding U1-70K, but not the stem loop 2 binding U1-A, has been shown to be important for this function (Ashe et al., 2000).

6. CONCLUSIONS

In this thesis, using as a model the coagulation factor 7, I studied the mechanisms through which patients' mutations located in promoter proximal 5'ss affect pre-mRNA processing. Using minigene systems in transient transfection or in the chromosomal context, I show that this type of mutations has a complex effect on pre-mRNA processing inducing both alterations in splicing and polyadenylation. These events, caused by the reduced binding of the U1 snRNP at the first donor site, result in a low level of mature transcript that will then reduce the amount of the corresponding protein. The intriguing effect mediated by the length of the intron that modulates the severity of the splicing and polyadenylation defects suggests the involvement of RNAPII. As changes in the RNAPII processivity might affect both events, additional studies can now be performed to better examine the relationship between RNAPII status and the observed pre-mRNA processing defects. Modified U1s that bind to the defective promoter proximal 5'ss are able to efficiently rescue the underlying defects and these molecules represent a novel therapeutic strategy for the correction of these type of processing errors. These results represent the basis for a future development of modified U1s in the correction of promoter proximal 5'ss mutations. In general, approximately 2-3 % of reported mutations that affect the splice sites are located in the promoter proximal 5'ss (UMD central database) and in some diseases, like the DTDST, a consistent number of patients has this type of defect and consequently might benefit from the approach. In addition, my results on the SMN2 minigene suggest also the possibility that recruitment of modified U1 snRNA on a non-mutated promoter proximal 5'ss can also increase the level of transcript. Future studies should test the effect of modified U1 snRNAs in more physiological systems, like primary cells derived from patients and appropriate animal models to establish the efficacy and safety profile of this novel strategy. This approach, already successfully tested for the correction of splicing mutations at internal exons (Dal Mas et al., 2015; Dal Mas et al., 2015; Glaus, Schmid, Da Costa, Berger, & Neidhardt, 2011; Rogalska et al., 2016), will further extend the applicability of modified U1s in the therapy of human diseases.

REFERENCES

Alanis, E. F., Pinotti, M., Mas, A. D., Balestra, D., Cavallari, N., Rogalska, M. E., … Pagani, F. (2012). An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Human Molecular Genetics*, *21*(11), 2389–2398. https://doi.org/10.1093/hmg/dds045

Alberts, B. (2008). Molecular Biology of the Cell. *Figure*, 1601. https://doi.org/10.1024/0301-1526.32.1.54

Alexander, R. D., Innocente, S. A., Barrass, J. D., & Beggs, J. D. (2010). Splicing-Dependent RNA polymerase pausing in yeast. *Molecular Cell*, *40*(4), 582–593. https://doi.org/10.1016/j.molcel.2010.11.005

Andreassi, C., & Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. *Trends in Cell Biology*, *19*(9), 465–474. https://doi.org/10.1016/j.tcb.2009.06.001

Ashe, M. P., Furger, A., & Proudfoot, N. J. (2000). Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. *RNA (New York, N.Y.)*, *6*(2), 170–7. https://doi.org/10.1017/S1355838200991957

Ashe, M. P., Pearson, L. H., & Proudfoot, N. J. (1997). The HIV-1 5' LTR poly (A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO Journal*, *16*(18), 5752–5763. https://doi.org/10.1093/emboj/16.18.5752

Avila, A. M., Burnett, B. G., Taye, A. A., Gabanella, F., Knight, M. A., Hartenstein, P., … Sumner, C. J. (2007). Trichostatin A increases SMN expression and survival in a mouse model of spinal muscular atrophy. *The Journal of Clinical Investigation*, *117*(3), 659–71. https://doi.org/10.1172/JCI29562

Ballhausen, D., Bonafé, L., Terhal, P., Unger, S. L., Bellus, G., Classen, M., … Superti-Furga, A. (2003). Recessive multiple epiphyseal dysplasia (rMED): phenotype delineation in eighteen homozygotes for DTDST mutation R279W. *Journal of Medical Genetics*, *40*(1), 65 LP-71. Retrieved from http://jmg.bmj.com/content/40/1/65.abstract

Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M., & Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Research*, *10*(7), 1001–1010. https://doi.org/10.1101/gr.10.7.1001

Benne, R., Van Den Burg, J., Brakenhoff, J. P. J., Sloof, P., Van Boom, J. H., & Tromp, M. C. (1986). Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*, *46*(6), 819–826. https://doi.org/10.1016/0092-8674(86)90063-2

Bentley, D. L. (2005). Rules of engagement: Co-transcriptional recruitment of pre-mRNA processing factors. *Current Opinion in Cell Biology*, *17*(3), 251–256. https://doi.org/10.1016/j.ceb.2005.04.006

Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., … Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, *150*(1), 53–64. https://doi.org/10.1016/j.cell.2012.05.029

Bernardi, F., Marchetti, G., Pinotti, M., Arcieri, P., Baroncini, C., Papacchini, M., … Mariani, G. (1996). Factor VII gene polymorphisms contribute about one third of the factor VII level variation in plasma. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *16*(1), 72–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8548429

Bernardi, F., Patracchini, P., Gemmati, D., Ferrati, M., Arcieri, P., Papacchini, M., … Marchetti, G. (1993). Molecular analysis of factor VII deficiency in Italy: a frequent mutation (FVII Lazio) in a repeated intronic region. *Human Genetics*, *92*(5), 446–450. https://doi.org/10.1007/BF00216448

Bhatt, D. M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M. M., Natoli, G., … Smale, S. T. (2012). Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. *Cell*, *150*(2), 279–290. https://doi.org/10.1016/j.cell.2012.05.043

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, *72*(1), 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720

Blencowe, B. J. (2006). Alternative Splicing: New Insights from Global Analyses. *Cell*, *126*(1), 37–47. https://doi.org/10.1016/j.cell.2006.06.023

Boelens, W. C., Jansen, E. J. R., van Venrooij, W. J., Stripecke, R., Mattaj, I. W., & Gunderson, S. I. (1993). The human U1 snRNP-Specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, *72*(6), 881–892. https://doi.org/10.1016/0092-8674(93)90577-D

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

Bradnam, K. R., & Korf, I. (2008). Longer First Introns Are a General Property of Eukaryotic Gene Structure. *PLoS ONE*, *3*(8), e3093. https://doi.org/10.1371/journal.pone.0003093

Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E., & Palmiter, R. D. (1988). Introns increase transcriptional efficiency in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(3), 836–40. https://doi.org/10.1073/pnas.85.3.836

Buratti, E., & Baralle, F. E. (2004). Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Molecular and Cellular Biology*, *24*(24), 10505–10514. https://doi.org/10.1128/MCB.24.24.10505-10514.2004

Burke, J. M. (1988). Molecular genetics of group I introns: RNA structures and protein factors required for splicing — a review. *Gene*, *73*(2), 273–294. https://doi.org/10.1016/0378-1119(88)90493-3

Cao, W., & Garcia-Blanco, M. A. (1998). A Serine/Arginine-rich Domain in the Human U1 70k Protein Is Necessary and Sufficient for ASF/SF2 Binding. *Journal of Biological Chemistry*, *273*(32), 20629–20635. https://doi.org/10.1074/jbc.273.32.20629

Chang, J. G., Hsieh-Li, H. M., Jong, Y. J., Wang, N. M., Tsai, C. H., & Li, H. (2001). Treatment of spinal muscular atrophy by sodium butyrate. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(17), 9808–13. https://doi.org/10.1073/pnas.171105098

Chou, Z. F., Chen, F., & Wilusz, J. (1994). Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals. *Nucleic Acids Research*, *22*(13), 2525–31. https://doi.org/10.1093/nar/22.13.2525

Chung, C. T., Niemela, S. L., & Miller, R. H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences* , *86*(7), 2172–2175. Retrieved from http://www.pnas.org/content/86/7/2172.abstract

Colgan, D. F., Murthy, K. G. K., Zhao, W., Prives, C., & Manley, J. L. (1998). Inhibition of

poly(A) polymerase requires p34(cdc2)/cyclin B phosphorylation of multiple consensus and non-consensus sites. *EMBO Journal*, *17*(4), 1053–1062. https://doi.org/10.1093/emboj/17.4.1053

Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and Disease. *Cell*, *136*(4), 777–793. https://doi.org/10.1016/j.cell.2009.02.011

Copertino, D. W., & Hallick, R. B. (1993). Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends in Biochemical Sciences*, *18*(12), 467–471. https://doi.org/10.1016/0968-0004(93)90008-B

Cruz-Reyes, J., & Sollner-Webb, B. (1996). Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(17), 8901–6. https://doi.org/10.1073/pnas.93.17.8901

Dal Mas, A., Rogalska, M. E., Bussani, E., & Pagani, F. (2015). Improvement of SMN2 Pre-mRNA Processing Mediated by Exon-Specific U1 Small Nuclear RNA. *The American Journal of Human Genetics*, *96*(1), 93–103. https://doi.org/10.1016/j.ajhg.2014.12.009

Dal Mas, A., Fortugno, P., Donadon, I., Levati, L., Castiglia, D., & Pagani, F. (2015). Exon-Specific U1s Correct SPINK5 Exon 11 Skipping Caused by a Synonymous Substitution that Affects a Bifunctional Splicing Regulatory Element. *Human Mutation*, *36*(5), 504–12. https://doi.org/10.1002/humu.22762

Damgaard, C. K., Kahns, S., Lykke-Andersen, S., Nielsen, A. L., Jensen, T. H., & Kjems, J. (2008). A 5′ Splice Site Enhances the Recruitment of Basal Transcription Initiation Factors In Vivo. *Molecular Cell*, *29*(2), 271–278. https://doi.org/10.1016/j.molcel.2007.11.035

Danckwardt, S., Gantzert, A. S., Macher-Goeppinger, S., Probst, H. C., Gentzel, M., Wilm, M., … Kulozik, A. E. (2011). P38 MAPK Controls Prothrombin Expression by Regulated RNA 3' End Processing. *Molecular Cell*, *41*(3), 298–310. https://doi.org/10.1016/j.molcel.2010.12.032

Davis, R., & Shi, Y. (2014). The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation. *Journal of Zhejiang University. Science. B*, *15*(5), 429–37. https://doi.org/10.1631/jzus.B1400076

De La Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., …

Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Molecular Cell*, *12*(2), 525–532. https://doi.org/10.1016/j.molcel.2003.08.001

de la Mata, M., & Kornblihtt, A. R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nature Structural & Molecular Biology*, *13*(11), 973–980. https://doi.org/10.1038/nsmb1155

Dye, M. J., Gromak, N., & Proudfoot, N. J. (2006). Exon tethering in transcription by RNA polymerase II. *Molecular Cell*, *21*(6), 849–859. https://doi.org/10.1016/j.molcel.2006.01.032

Elkon, R., Ugalde, A. P., & Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics*, *14*(7), 496–506. https://doi.org/10.1038/nrg3482

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., … Guttman, M. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science (New York, N.Y.)*, *341*(6147), 1237973. https://doi.org/10.1126/science.1237973

Faustino, N. A., & Cooper, T. A. (2003, February 15). Pre-mRNA splicing and human disease. *Genes and Development*. Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gad.1048803

Fong, Y. W., & Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, *414*(6866), 929–933. https://doi.org/10.1038/414929a

Frech, G. C., & Simpson, L. (1996). Uridine insertion into preedited mRNA by a mitochondrial extract from Leishmania tarentolae: stereochemical evidence for the enzyme cascade model. *Molecular and Cellular Biology*, *16*(8), 4584–4589. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8754859

Furger, A., O'Sullivan, J. M., Binnie, A., Lee, B. A., & Proudfoot, N. J. (2002). Promoter proximal splice sites enhance transcription. *Genes and Development*, *16*(21), 2792–2799. https://doi.org/10.1101/gad.983602

Furuichi, Y., LaFiandra, A., & Shatkin, A. J. (1977). 5′-Terminal structure and mRNA stability. *Nature*, *266*(5599), 235–239. https://doi.org/10.1038/266235a0

Glaus, E., Schmid, F., Da Costa, R., Berger, W., & Neidhardt, J. (2011). Gene therapeutic

approach using mutation-adapted U1 snRNA to correct a RPGR splice defect in patient-derived cells. *Molecular Therapy : The Journal of the American Society of Gene Therapy*, *19*(5), 936–41. https://doi.org/10.1038/mt.2011.7

Gott, J. M., & Emeson, R. B. (2000). Functions and mechanisms of RNA editing. *Annual Review of Genetics*, *34*, 499–531. https://doi.org/34/1/499 [pii]\r10.1146/annurev.genet.34.1.499

Grimm, C., Stefanovic, B., & Schümperli, D. (1993). The low abundance of U7 snRNA is partly determined by its Sm binding site. *The EMBO Journal*, *12*(3), 1229–38. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8458335

Gross, J. D., Moerke, N. J., Von Der Haar, T., Lugovskoy, A. A., Sachs, A. B., McCarthy, J. E. G., & Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell*, *115*(6), 739–750. https://doi.org/10.1016/S0092-8674(03)00975-9

Grummt, I. (2003). Life on a planet of its own: Regulation of RNA polymerase I transcription in the nucleolus. *Genes and Development*, *17*(14), 1691–1702. https://doi.org/10.1101/gad.1098503R

Gunderson, S. I., Polycarpou-Schwarz, M., & Mattaj, I. W. (1998). U1 snRNP Inhibits Pre-mRNA Polyadenylation through a Direct Interaction between U1 70K and Poly(A) Polymerase. *Molecular Cell*, *1*(2), 255–264. https://doi.org/10.1016/S1097-2765(00)80026-X

Gunderson, S. I., Vagner, S., Polycarpou-Schwarz, M., & Mattaj, L. W. (1997). Involvement of the carboxyl terminus of vertebrate poly(A)polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes and Development*, *11*(6), 761–773. https://doi.org/10.1101/gad.11.6.761

Hästbacka, J., de la Chapelle, A., Mahtani, M. M., Clines, G., Reeve-Daly, M. P., Daly, M., … Lander, E. S. (1994). The diastrophic dysplasia gene encodes a novel sulfate transporter: Positional cloning by fine-structure linkage disequilibrium mapping. *Cell*, *78*(6), 1073–1087. https://doi.org/10.1016/0092-8674(94)90281-X

Hästbacka, J., Kerrebrock, A., Mokkala, K., Clines, G., Lovett, M., Kaitila, I., … Lander, E. S. (1999). Identification of the Finnish founder mutation for diastrophic dysplasia (DTD). *European Journal of Human Genetics*, *7*(6), 664–670.

https://doi.org/10.1038/sj.ejhg.5200361

Hästbacka, J., Superti-Furga, A., Wilcox, W. R., Rimoin, D. L., Cohn, D. H., & Lander, E. S. (1996). Atelosteogenesis type II is caused by mutations in the diastrophic dysplasia sulfate-transporter gene (DTDST): evidence for a phenotypic series involving three chondrodysplasias. *American Journal of Human Genetics*, *58*(2), 255–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8571951

Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H., & Lührmann, R. (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *The EMBO Journal*, *14*(9), 2076–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7744013

Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science (New York, N.Y.)*, *308*(5718), 118–120. https://doi.org/10.1126/science.1106910

Jeong, S., & Stein, A. (1994). Micrococcal nuclease digestion of nuclei reveals extended nucleosome ladders having anomalous DNA lengths for chromatin assembled on non-replicating plasmids in transfected cells. *Nucleic Acids Research*, *22*(3), 370–375. https://doi.org/10.1093/nar/22.3.370

Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., & Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, *468*(7324), 664–668. https://doi.org/10.1038/nature09479

Kim, H.-J., Jeong, S.-H., Heo, J.-H., Jeong, S.-J., Kim, S.-T., Youn, H.-D., … Cho, E.-J. (2004). mRNA capping enzyme activity is coupled to an early transcription elongation. *Molecular and Cellular Biology*, *24*(14), 6184–93. https://doi.org/10.1128/MCB.24.14.6184-6193.2004

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Lührmann, R., Garcia-Blanco, M. A., & Manley, J. L. (1994). Protein–protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, *368*(6467), 119–124. https://doi.org/10.1038/368119a0

Krämer, A., & Utans, U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *The EMBO Journal*, *10*(6), 1503–9. https://doi.org/10.1111/j.1471-4159.1991.tb06421.x

Kugel, J. F., & Goodrich, J. A. (1998). Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(16), 9232–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9689063

Kumar, K. P., Akoulitchev, S., & Reinberg, D. (1998). Promoter-proximal stalling results from the inability to recruit transcription factor IIH to the transcription complex and is a regulated event. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(17), 9767–72. https://doi.org/10.1073/PNAS.95.17.9767

Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., … Akoulitchev, A. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nature Structural Biology*, *9*(11), 800–805. https://doi.org/10.1038/nsb862

Kyriakopoulou, C. B., Nordvarg, H., & Virtanen, A. (2001). A Novel Nuclear Human Poly(A) Polymerase (PAP), PAP. *Journal of Biological Chemistry*, *276*(36), 33504–33511. https://doi.org/10.1074/jbc.M104599200

Lamond, A. I. (1993). The Spliceosome. *BioEssays*, *15*(9), 595–603. https://doi.org/10.1002/bies.950150905

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., & Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, *446*(7138), 926–929. https://doi.org/10.1038/nature05676

Leissinger, C. A., Blatt, P. M., Hoots, W. K., & Ewenstein, B. (2008). Role of prothrombin complex concentrates in reversing warfarin anticoagulation: A review of the literature. *American Journal of Hematology*, *83*(2), 137–143. https://doi.org/10.1002/ajh.21046

Lejeune, F., & Maquat, L. E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Current Opinion in Cell Biology*, *17*(3), 309–315. https://doi.org/10.1016/j.ceb.2005.03.002

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., & Mattaj, L. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes and Development*, *10*(13), 1683–1698. https://doi.org/10.1101/gad.10.13.1683

Liang, S., & Lutz, C. S. (2006). p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation. *RNA (New York,*

N.Y.), *12*(1), 111–121. https://doi.org/10.1261/rna.2213506

Licatalosi, D. D., & Darnell, R. B. (2010). RNA processing and its regulation: global insights into biological networks. *Nature Reviews. Genetics*, *11*(1), 75–87. https://doi.org/10.1038/nrg2673

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., … Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, *456*(7221), 464–469. https://doi.org/10.1038/nature07488

Lührmann, R., Kastner, B., & Bach, M. (1990). Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1087*(3), 265–292. https://doi.org/10.1016/0167-4781(90)90001-I

Maniatis, T., & Reed, R. (2002). Expression Machines. *Nature*, *416*(6880), 499–506. https://doi.org/10.1038/416499a

Mariani, G., & Bernardi, F. (2009). Factor VII deficiency. *SEMINARS IN THROMBOSIS AND HEMOSTASIS*, *35*(4), 400–406. https://doi.org/10.1055/s-0029-1225762

Marks, P. A., Richon, V. M., Miller, T., & Kelly, W. K. (2004). Histone Deacetylase Inhibitors. *Advances in Cancer Research*, *91*, 137–168. https://doi.org/10.1016/S0065-230X(04)91004-4

Martincic, K., Alkan, S. A., Cheatle, A., Borghesi, L., & Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nature Immunology*, *10*(10), 1102–1109. https://doi.org/10.1038/ni.1786

Mayer, A., & Churchman, L. S. (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nature Protocols*, *11*(4), 813–833. https://doi.org/10.1038/nprot.2016.047

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

McVey, J. H., Boswell, E., Mumford, A. D., Kemball-Cook, G., & Tuddenham, E. G. (2001). Factor VII deficiency and the FVII mutation database. *Human Mutation*, *17*(1), 3. Retrieved from

http://search.proquest.com/openview/293dadc1da9625d208e49a7e1e3bc59c/1?pq-origsite=gscholar

Michaud, S., & Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes & Development*, *5*(12b), 2534–2546. https://doi.org/10.1101/gad.5.12b.2534

Michel, F., Kazuhiko, U., & Haruo, O. (1989). Comparative and functional anatomy of group II catalytic introns — a review. *Gene*, *82*(1), 5–30. https://doi.org/10.1016/0378-1119(89)90026-7

Millevoi, S., & Vagner, S. (2009). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research*, *38*(9), 2757–2774. https://doi.org/10.1093/nar/gkp1176

Mondal, T., Rasmussen, M., Pandey, G. K., Isaksson, A., & Kanduri, C. (2010). Characterization of the RNA content of chromatin. *Genome Research*, *20*(7), 899–907. https://doi.org/10.1101/gr.103473.109

Muro, A. F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E., & Baralle, F. E. (1999). Regulation of Fibronectin EDA Exon Alternative Splicing: Possible Role of RNA Secondary Structure for Enhancer Display. *Molecular and Cellular Biology*, *19*(4), 2657–2671. https://doi.org/10.1128/MCB.19.4.2657

Nakagawa, T., Bulger, M., Muramatsu, M., & Ito, T. (2001). Multistep chromatin assembly on supercoiled plasmid DNA by nucleosome assembly protein-1 and ATP-utilizing chromatin assembly and remodeling factor. *The Journal of Biological Chemistry*, *276*(29), 27384–91. https://doi.org/10.1074/jbc.M101331200

Nelissen, R. L., Will, C. L., van Venrooij, W. J., & Lührmann, R. (1994). The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *The EMBO Journal*, *13*(17), 4113–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8076607

Nilsen, T. W. (2003). The spliceosome: The most complex macromolecular machine in the cell? *BioEssays*, *25*(12), 1147–1149. https://doi.org/10.1002/bies.10394

Oh, J.-M., Di, C., Venters, C. C., Guo, J., Arai, C., So, B. R., … Dreyfuss, G. (2017). U1 snRNP telescripting regulates a size–function-stratified human genome. *Nature Structural & Molecular Biology*, *24*(11), 993–999. https://doi.org/10.1038/nsmb.3473

Pagani, F., & Baralle, F. E. (2004). Opinion: Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Reviews Genetics*, *5*(5), 389–396. https://doi.org/10.1038/nrg1327

Pagani, F., Buratti, E., Stuani, C., & Baralle, F. E. (2003). Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *The Journal of Biological Chemistry*, *278*(29), 26580–8. https://doi.org/10.1074/jbc.M212813200

Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dörk, T., & Baralle, F. E. (2002). A new type of mutation causes a splicing defect in ATM. *Nature Genetics*, *30*(4), 426–429. https://doi.org/10.1038/ng858

Pagani, F., & Pinotti, M. (2011). Modified human U1snRNA molecule, a gene encoding for the modified human U1snRNA molecule, an expression vector including the gene, and the use thereof in gene therapy.

Pagani, F., Raponi, M., & Baralle, F. E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the National Academy of Sciences*, *102*(18), 6368–6372. https://doi.org/10.1073/pnas.0502288102

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415. https://doi.org/10.1038/ng.259

Pfaffle, P., & Jackson, V. (1990). Studies on rates of nucleosome formation with DNA under stress. *Journal of Biological Chemistry*, *265*(28), 16821–16829. Retrieved from http://www.jbc.org/content/265/28/16821.short

Phatnani, H. P., Jones, J. C., & Greenleaf, A. L. (2004). Expanding the Functional Repertoire of CTD Kinase I and RNA Polymerase II: Novel PhosphoCTD-Associating Proteins in the Yeast Proteome †. *Biochemistry*, *43*(50), 15702–15719. https://doi.org/10.1021/bi048364h

Pinotti, M., Rizzotto, L., Balestra, D., Lewandowska, M. A., Cavallari, N., Marchetti, G., … Pagani, F. (2008). U1-snRNA mediated rescue of mRNA processing in severe factor VII deficiency. *Blood*, *111*(5), 2681–2684. https://doi.org/10.1182/blood-2007-10-117440

Pinto, P. A. B., Henriques, T., Freitas, M. O., Martins, T., Domingues, R. G., Wyrzykowska, P. S., … Moreira, A. (2011). RNA polymerase II kinetics in polo polyadenylation signal

selection. *The EMBO Journal*, *30*(12), 2431–2444. https://doi.org/10.1038/emboj.2011.156

Rino, J., & Carmo-Fonseca, M. (2009). The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends in Cell Biology*, *19*(8), 375–384. https://doi.org/10.1016/j.tcb.2009.05.004

Rodríguez-Trelles, F., Tarrío, R., & Ayala, F. J. (2006). Origins and evolution of spliceosomal introns. *Annual Review of Genetics*, *40*, 47–76. https://doi.org/10.1146/annurev.genet.40.110405.090625

Rogalska, M. E., Tajnik, M., Licastro, D., Bussani, E., Camparini, L., Mattioli, C., & Pagani, F. (2016). Therapeutic activity of modified U1 core spliceosomal particles. *Nature Communications*, *7*, 11168. https://doi.org/10.1038/ncomms11168

Rösel-Hillgärtner, T. D., Hung, L. H., Khrameeva, E., Le Querrec, P., Gelfand, M. S., & Bindereif, A. (2013). A Novel Intra-U1 snRNP Cross-Regulation Mechanism: Alternative Splicing Switch Links U1C and U1-70K Expression. *PLoS Genetics*, *9*(10), e1003856. https://doi.org/10.1371/journal.pgen.1003856

Rosen, E. D., Chan, J. C., Idusogie, E., Clotman, F., Vlasuk, G., Luther, T., … Carmeliet, P. (1997). Mice lacking factor VII develop normally but suffer fatal perinatal bleeding. *Nature*, *390*(6657), 290–4. https://doi.org/10.1038/36862

Saltzman, A. L., Kim, Y. K., Pan, Q., Fagnani, M. M., Maquat, L. E., & Blencowe, B. J. (2008). Regulation of Multiple Core Spliceosomal Proteins by Alternative Splicing-Coupled Nonsense-Mediated mRNA Decay. *Molecular and Cellular Biology*, *28*(13), 4320–4330. https://doi.org/10.1128/MCB.00361-08

Schramm, L., & Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes and Development*, *16*(20), 2593–2620. https://doi.org/10.1101/gad.1018902

Schwer, B., Mao, X., & Shuman, S. (1998). Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme. *Nucleic Acids Research*, *26*(9), 2050–2057. https://doi.org/10.1093/nar/26.9.2050

Scorilas, A. (2002). Polyadenylate polymerase (PAP) and 3’ end pre-mRNA processing: function, assays, and association with disease. *Crit Rev Clin Lab Sci*, *39*(3), 193–224. https://doi.org/10.1080/10408360290795510

Sendtner, M. (2010). Therapy development in spinal muscular atrophy. *Nature Neuroscience*, *13*(7), 795–799. https://doi.org/10.1038/nn.2565

Sharma, S., Wongpalee, S. P., Vashisht, A., Wohlschlegel, J. A., & Black, D. L. (2014). Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes and Development*, *28*(22), 2518–2531. https://doi.org/10.1101/gad.248625.114

Sheets, M. D., Ogg, S. C., & Wickens, M. P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Research*, *18*(19), 5799–805. https://doi.org/10.1093/nar/18.19.5799

Shi, Y. (2012). Alternative polyadenylation: new insights from global analyses. *RNA (New York, N.Y.)*, *18*(12), 2105–17. https://doi.org/10.1261/rna.035899.112

Shuman, S. (2015). RNA capping: progress and prospects. *RNA (New York, N.Y.)*, *21*(4), 735–7. https://doi.org/10.1261/rna.049973.115

Spraggon, L., & Cartegni, L. (2013). U1 snRNP-dependent suppression of polyadenylation: Physiological role and therapeutic opportunities in cancer. *International Journal of Cell Biology*, *2013*, 1–10. https://doi.org/10.1155/2013/846510

Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., & Bartel, D. P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, *508*(7494), 66–71. https://doi.org/10.1038/nature13007

Sun, S., Ling, S.-C., Qiu, J., Albuquerque, C. P., Zhou, Y., Tokunaga, S., … Cleveland, D. W. (2015). ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. *Nature Communications*, *6*, 6171. https://doi.org/10.1038/ncomms7171

Superti-Furga, A., Hästbacka, J., Wilcox, W. R., Cohn, D. H., van der Harten, H. J., Rossi, A., … Gitzelmann, R. (1996). Achondrogenesis type IB is caused by mutations in the diastrophic dysplasia sulphate transporter gene. *Nature Genetics*, *12*(1), 100–102. https://doi.org/10.1038/ng0196-100

Superti-Furga, A., Neumann, L., Riebel, T., Eich, G., Steinmann, B., Spranger, J., … Superti-Furga, A. (1999). Recessively inherited multiple epiphyseal dysplasia with normal stature, club foot, and double layered patella caused by a DTDST mutation. *Journal of*

Medical Genetics, *36*(8), 621–4. https://doi.org/10.1136/jmg.40.1.65

Susani, L., Pangrazio, A., Sobacchi, C., Taranta, A., Mortier, G., Savarirayan, R., … Pagani, F. (2004). TCIRG1-dependent recessive osteopetrosis: Mutation analysis, functional identification of the splicing defects, andin vitro rescue by U1 snRNA. *Human Mutation*, *24*(3), 225–235. https://doi.org/10.1002/humu.20076

Swoboda, K. J., Scott, C. B., Reyna, S. P., Prior, T. W., LaSalle, B., Sorenson, S. L., … Simard, L. R. (2009). Phase II Open Label Study of Valproic Acid in Spinal Muscular Atrophy. *PLoS ONE*, *4*(5), e5268. https://doi.org/10.1371/journal.pone.0005268

Tajnik, M., Rogalska, M. E., Bussani, E., Barbon, E., Balestra, D., Pinotti, M., … Ravazzolo, R. (2016). Molecular Basis and Therapeutic Strategies to Rescue Factor IX Variants That Affect Splicing and Protein Function. *PLOS Genetics*, *12*(5), e1006082. https://doi.org/10.1371/journal.pgen.1006082

Takagaki, Y., & Manley, J. L. (1998). Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Molecular Cell*, *2*(6), 761–771. https://doi.org/10.1016/S1097-2765(00)80291-9

Tanner, G., Glaus, E., Barthelmes, D., Ader, M., Fleischhauer, J., Pagani, F., … Neidhardt, J. (2009). Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. *Human Mutation*, *30*(2), 255–263. https://doi.org/10.1002/humu.20861

Tsai, L.-K., Tsai, M.-S., Lin, T.-B., Hwu, W.-L., & Li, H. (2006). Establishing a standardized therapeutic testing protocol for spinal muscular atrophy. *Neurobiology of Disease*, *24*(2), 286–295. https://doi.org/10.1016/j.nbd.2006.07.004

van der Woerd, W. L., Mulder, J., Pagani, F., Beuers, U., Houwen, R. H. J., & van de Graaf, S. F. J. (2015). Analysis of aberrant pre-messenger RNA splicing resulting from mutations in ATP8B1 and efficient in vitro rescue by adapted U1 small nuclear RNA. *Hepatology*, *61*(4), 1382–1391. https://doi.org/10.1002/hep.27620

Varani, G., & Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.*, *27*(1), 407–445. https://doi.org/10.1146/annurev.biophys.27.1.407

Vaz-Drago, R., Pinheiro, M. T., Martins, S., Enguita, F. J., Carmo-Fonseca, M., Custodio, N., & Custódio, N. (2015). Transcription-coupled RNA surveillance in human genetic

diseases caused by splice site mutations. *Human Molecular Genetics*, *24*(10), 2784–2795. https://doi.org/10.1093/hmg/ddv039

Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B., & Mattaj, I. W. (1996). A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. *Journal of Cell Biology*, *133*(1), 5–14. https://doi.org/10.1083/jcb.133.1.5

Vorlova, S., Rocco, G., LeFave, C. V., Jodelka, F. M., Hess, K., Hastings, M. L., … Cartegni, L. (2011). Induction of Antagonistic Soluble Decoy Receptor Tyrosine Kinases by Intronic PolyA Activation. *Molecular Cell*, *43*(6), 927–939. https://doi.org/10.1016/j.molcel.2011.08.009

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*. https://doi.org/10.1016/j.cell.2009.02.009

Wahle, E., & Keller, W. (1992). The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem*, *61*, 419–40.

Wajima, T., Isbister, G. K., & Duffull, S. B. (2009). A comprehensive model for the humoral coagulation network in humans. *Clinical Pharmacology and Therapeutics*, *86*(3), 290–8. https://doi.org/10.1038/clpt.2009.87

Werner, M. S., & Ruthenburg, A. J. (2015). Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Reports*, *12*(7), 1089–1098. https://doi.org/10.1016/j.celrep.2015.07.033

Wickens, M., Anderson, P., & Jackson, R. J. (1997). Life and death in the cytoplasm: Messages from the 3' end. *Current Opinion in Genetics and Development*, *7*(2), 220–232. https://doi.org/10.1016/S0959-437X(97)80132-3

Will, C. L., & Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology*, *13*(3), 290–301. https://doi.org/10.1016/S0955-0674(00)00211-8

Will, C., Rümpler, S., Gunnewiek, J. K., Venrooij, W. J. van, & Lührmann, R. (1996). In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Research*, *24*(23), 4614–4623. https://doi.org/10.1093/nar/24.23.4614

Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A., & Smith, C. W. . (2004). Autoregulation of Polypyrimidine Tract Binding Protein by Alternative Splicing Leading to Nonsense-Mediated Decay. *Molecular Cell*, *13*(1), 91–100. https://doi.org/10.1016/S1097-2765(03)00502-1

Woychik, N., & Hampsey, M. (2002). The RNA Polymerase II MachineryStructure Illuminates Function. *Cell*, *108*(4), 453–463. https://doi.org/10.1016/S0092-8674(02)00646-3

Zavolan, M., Kondo, S., Schönbach, C., Adachi, J., Hume, D. A., Group, R. G., … Gaasterland, T. (2003). Impact of Alternative Initiation, Splicing, and Termination on the Diversity of the mRNA Transcripts Encoded by the Mouse Transcriptome. *Genome Research*, *13*(6), 1290–1300. https://doi.org/10.1101/gr.1017303

Zhao, J., Hyman, L., & Moore, C. (1999). Formation of mRNA 3’ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews : MMBR*, *63*(2), 405–445. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10357856