The impact of the U2AF35 zinc finger domains and cancer-associated mutations on protein function

Inaugural Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

OLGA HERDT

from Maiskoe, Kazakhstan

February 2018

This work was carried out in the period of January 2014 to February 2018 under the supervision of Prof. Dr. Florian Heyd at the Institute of Chemistry and Biochemistry, Freie Universität Berlin, Germany.

First Reviewer:	Prof. Dr. Florian Heyd Institute for Chemistry and Biochemistry Laboratory of RNA-Biochemistry Takustraße 6 14195 Berlin, Germany
Second Reviewer:	Prof. Dr. Markus C. Wahl Institute for Chemistry and Biochemistry Laboratory of Structural Biology Takustraße 6 14195 Berlin, Germany
Date of defense:	20.04.2018

Table of Contents

ABSTRACT

Splicing of pre-mRNA is a tightly regulated process that has to be performed very accurately to ensure the generation of functional mRNAs. The accuracy is accomplished by the recognition of conserved RNA sequences like the splice sites at the exon-intron boundaries. The U2 auxiliary factor 35 (U2AF35) is a splicing factor involved in the initial steps of spliceosome assembly. Together with U2AF65, U2AF35 builds the heterodimer U2AF that determines the 3' splice site (3' ss). The RNA contacting domain in U2AF35 was just recently characterized for the yeast homolog U2AF23. Assignment of the crystal structure and *in vitro* binding studies of yeast U2AF23 revealed that the two zinc finger (ZnF) domains cooperatively recognize a 4-bases long RNA sequence representing the 3' ss with the conserved AG dinucleotide in the center. The ZnFs of U2AF35 got further attention since mutations therein were implicated in hematopoietic disorders.

In this study, we have analyzed key functionalities of mammalian U2AF35 ZnFs in cell based approaches and could verify and extend the observations made for the yeast homolog U2AF23. Protein stability assays revealed that ZnF2 is required for U2AF35 protein stability. Furthermore, the ZnF2 is needed for a stable U2AF heterodimer complex formation. Knockdown and complementation assays showed that both ZnFs are indispensable for splicing regulation of U2AF35-dependent exons. The observations for U2AF35 could be confirmed with the highly conserved paralog U2AF26 and these results also demonstrate that U2AF26 can functionally substitute for U2AF35 *in vivo*. The importance of the ZnF2 was further verified with a naturally occurring splice variant of U2AF26 which lacks the largest part of the ZnF2 (U2AF26ΔE7). In contrast to U2AF26fl, U2AF26ΔE7 localizes into the cytoplasm and is upregulated in activated primary mouse T cells. An MS2-based reporter assay was used to elucidate the impact of U2AF26ΔE7 on cytoplasmic RNAs and revealed that U2AF26ΔE7 increases the translation of a reporter gene when tethered to its 5'UTR. To identify endogenous U2AF26ΔE7 targets, mouse T cells with a stable U2AF26ΔE7 expression were generated. A targeted approach revealed an elevated surface expression of the T cell marker Ly6a in the U2AF26ΔE7 expressing cells.

Further experiment revealed that the elevated Ly6a surface expression observed in U2AF26ΔE7 expressing cell might be regulated on the level of translation as well as mRNA stability.

We then focused our investigations on the U2AF35 ZnF2 and analyzed the consequences of the AML-associated Q157R and Q157P mutations therein. Interestingly, we discovered that the c.470A>G mutation that causes the Q157R substitution in addition generates an alternative 5' ss. Usage of the alternative 5' ss leads to the deletion of four amino acids within the ZnF2 creating the Q157Rdel variant. Genome wide examination of splicing targets and binding motifs of the Q157P, Q157R and Q157Rdel mutants resulted in considerable differences in targeted exon caused by different 3' ss binding preferences of the mutants. Differences in splicing regulation were also observed in RNA sequencing analysis of AML patients carrying the Q157R or Q157P mutation, suggesting that the mutations alter binding specificities of U2AF35. Furthermore, we identified Q157Rdel mRNA expression in the RNA sequencing data of c.470A>G patients. Comparison of Q157Rdel- and Q157R-specific splicing in cells and in patients demonstrated that the Q157Rdel variant likely contributes to the missplicing observed in the Q157R patients.

Splicing is an essential and highly controlled pre-mRNA processing step whose misregulation causes many diseases. U2AF35 plays a tremendous role in splicing regulation by the initial recognition of intron-exon boundaries in a sequence specific manner. With this study, we contributed to the functional characterization of the RNA-recognizing domains in U2AF35 and deciphered the consequences of disease-associated mutations therein.

ZUSAMMENFASSUNG

Das Spleißen von prä-mRNA ist ein stark regulierter Prozess und erfordert eine enorm hohe Präzision um die Herstellung von funktionalen mRNAs zu gewährleisten. Die Präzision wird durch das Erkennen von konservierten RNA-Sequenzen, wie beispielsweise den Spleißstellen, erreicht. Der Spleißfaktor U2AF35 (U2 auxiliary factor 35) übernimmt eine wesentliche Rolle in der Spleißinitiation. U2AF35 bildet zusammen mit dem Spleißfaktor U2AF65 das U2AF Heterodimer, welches die 3'-Spleißstelle erkennt und dadurch den Intron-Exon-Übergang markiert. Die RNA-bindende Domäne in U2AF35 wurde vor kurzem für das homologe Protein in Hefe (U2AF23) beschrieben. Die Bestimmung der Kristallstruktur und die Durchführung von RNA Bindestudien für U2AF23 ergaben, dass beide Zinkfingerdomänen gemeinsam an eine 4 Basen lange Sequenz binden. Die gebundene Sequenz stellt die 3'-Spleißstelle dar und enthält das konservierte AG-Dinukleotid in der Mitte. Die Zinkfingerdomänen von U2AF35 erlangten weitere Aufmerksamkeit durch die Identifizierung von darin vorkommenden Punktmutationen. Mutiertes U2AF35 führt zu veränderten Spleißmustern und trägt damit zur Entstehung von hämatopoetischen Erkrankungen bei.

Ein Ziel dieser Studie war die funktionelle Untersuchung der Zinkfingerdomänen von dem in Säugern vorkommenden U2AF35. Mit Hilfe von zellkulturbasierten Experimenten konnten die Erkenntnisse, die für das U2AF35 Homolog in Hefe erlangt wurden, für das murine U2AF35 verifiziert und erweitert werden. Die Untersuchung der Proteinstabilität von U2AF35 ergab, dass der zweite Zinkfinger (ZnF) für die Stabilität von U2AF35 benötigt wird. Darüber hinaus trägt der zweite ZnF zur Bildung eines stabilen U2AF-Heterodimers bei. Die Substitution des endogenen volle Länge U2AF35-Proteins mit ZnF-deletierten U2AF35-Proteinen ergab, dass beide ZnF für die Spleißregulation von U2AF35-abhändigen prä-mRNAs benötigt werden. Die beschriebenen Erkenntnisse wurden für das U2AF35-verwandte Protein U2AF26 bestätigt und zeigen auch, dass U2AF26 und U2AF35 äquivalente Funktionen in Zellen haben. Die Bedeutung des zweiten ZnFs wurde außerdem mit einer natürlich vorkommenden Spleißvariante von U2AF26 bestätigt, bei der durch die Exklusion von Exon 7 (U2AF26ΔE7) große Teile des zweiten ZnFs fehlen. Im Gegensatz zur Lokalisation des volle Länge U2AF26 in den Kern, lokalisiert U2AF26ΔE7 in das Zytoplasma und ist verstärkt exprimiert in aktivierten primären T-

Zellen aus der Maus. Um die Funktion des zytoplasmatischen U2AF26ΔE7 aufzuklären, wurde eine MS2-basierte Reportergenanalyse durchgeführt, die zeigt, dass U2AF26ΔE7 die Translation des Reportergens erhöht, wenn es zum 5'-Ende des untranslatierten Bereichs rekrutiert wird. Als nächstes wurde eine murine T-Zelllinie mit stabiler U2AF26ΔE7 Überexpression erzeugt, um mögliche endogene U2AF26ΔE7 Ziel-mRNAs zu identifizieren. Basierend auf den Vorergebnissen wurde die Proteinexpression von einigen T-Zelloberflächenmarkern untersucht. Diese Untersuchung ergab eine erhöhte Oberflächenexpression des Proteins Ly6a auf den U2AF26ΔE7-exprimierenden Zellen, welche durch eine erhöhte Ly6a mRNA Stabilität und eine leicht verstärkte Translation reguliert zu sein scheint.

Wir konzentrierten unsere Untersuchungen auf den zweiten ZnF von U2AF35 und untersuchten die Konsequenzen der darin vorkommenden AML-assoziierten Punktmutationen Q157R und Q157P. Interessanterweise, führt die c.470A>G Mutation, zusätzlich zur Q157R Substitution, zur Verstärkung einer sich direkt dahinter befindenden alternativen 5'-Spleißstelle. Zusätzlich zu der Q157R-Substitution führt die Nutzung der alternativen 5'-Spleißstelle zur Deletion von vier Aminosäuren und bildet dadurch die Q157Rdel Spleißvariante von U2AF35. Genomweite Untersuchungen von Ziel-prä-mRNAs und darin enthaltenen Spleißstellen ergaben, dass die Q157P-, Q157R- und Q157Rdel-Mutanten bestimmte 3'-Spleißstellen bevorzugt binden und dadurch das Spleißen unterschiedlicher prä-mRNAs regulieren. Unterschiedliches Spleißen von prä-mRNAs wurde auch in genomweiten Untersuchungen von AML-Patienten mit den Q157R- und Q157-Mutation gefunden. Dieses deutet auf eine veränderte Bindungsspezifität von U2AF35 hin, die durch die Mutationen verursacht wird. Des Weiteren ist von Bedeutung, dass die Q157Rdel-Spleißvariante auch in AML-Patienten mit der c.470A>G Mutation exprimiert ist. Der Vergleich von Spleißunterschieden, verursacht durch die Q157R- und die Q157Rdel-Mutante in Zellen und in Patienten, deutet darauf hin, dass die Q157Rdel-Spleißvariante zum abnormalen Spleißen in Q157R-Patienten beiträgt.

Spleißen ist ein essentieller und stark regulierter Schritt während der prä-mRNA-Prozessierung dessen Deregulierung viele Krankheiten verursacht. Zu Beginn der Spleißreaktion markiert U2AF35 sequenzspezifisch den Intron-Exon-Übergang und ist deswegen enorm wichtig für die Spleißregulation. Mit dieser Studie tragen wir zur funktionellen Charakterisierung der RNA-bindenden Domänen in U2AF35 bei und helfen, die Auswirkungen der darin vorkommenden krankheitsassoziierten Mutationen, zu verstehen.

LIST OF PUBLICATIONS

1. Herdt O*, Neumann A*, Timmermann B, Heyd F (2017). The cancer-associated U2AF35 470A>G (Q157R) mutation creates an in-frame alternative 5′ splice site that impacts splicing regulation in Q157R patients. RNA *23, 1796-1806.*

   * These authors contributed equally to this work.

OH designed and cloned all minigene and expression constructs; performed all minigene transfections, knock down and rescue assays and the subsequent splicing sensitive RT-PCRs, quantitative RT-PCRs and Western blots; prepared samples for sequencing; analyzed data of the wet lab experiments; and co-wrote the manuscript.

2. Herdt O, Preußner M, Heyd F (2018). The zinc finger domains in U2AF35 and U2AF26 control stability, localization, interaction with U2AF65 and play a role in translation. NAR, *in preparation for re-submission.*

OH designed and cloned U2AF35 and U2AF26 expression constructs; performed stability assays, co-immunoprecipitations and, knock down and rescue assays and splicing sensitive RT-PCRs; purified primary mouse T cells; generated stable mouse T cells and performed flow cytometry and quantitative RT-PCRs; verified the results of the MS2-assays performed by MP; analyzed the data; and co-wrote the manuscript.

1. INTRODUCTION

1.1 Pre-mRNA splicing

Most eukaryotic genes are transcribed into precursor messenger RNAs (pre-mRNAs) that undergo several processing steps prior to export to the cytoplasm and translation into a protein. These processing steps include the addition of a 7-methylguanosine cap at the 5' end and a poly-adenosine tail at the 3' end of the pre-mRNA. Furthermore, splicing occurs to remove non-coding introns from intron-containing pre-mRNAs and to join together the coding exons that flank the introns. Splicing is achieved by two transesterification reactions which are detailed shown in *Figure 1.1.* During the first transesterification step, the 2'OH group of the branch point (BP) adenosine attacks the phosphodiester bond at the 5' splice site (ss). In the second step, the free OH group at the 5' ss attacks the phosphodiester bond at the 3' ss resulting in the ligation of the exons and the release of the intron lariat.

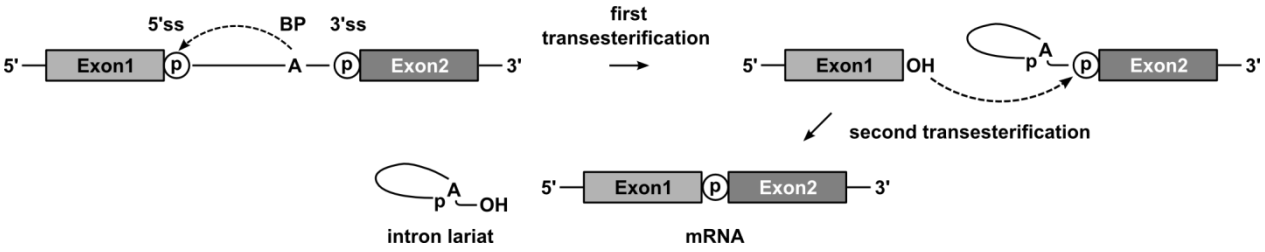


Figure 1.1: Pre-mRNA splicing reaction. The exons are depicted as grey boxes and the introns as solid lines. The splicing reaction is carried out in two transesterification reactions. Nucleophilic attacks are labeled as dashed lines.

Splicing is accomplished by the spliceosome which is a large dynamic RNA-protein complex. The spliceosome assembles in a stepwise manner on the pre-mRNA; this involves five small nuclear ribonucleoproteins (U snRNPs, e.g. U1) which are composed out of many snRNP-specific proteins and a specific uridine-rich small nuclear RNA *(Figure 1.2).* Additionally, many spliceosome-associated non-snRNP factors participate in splicing. Initiation of splicing requires

the recognition of conserved sequences within the pre-mRNA. The 5' splice site (ss) is bound by the U1 snRNP, the 3' ss by the U2 auxiliary factor (U2AF) and the branch point sequence (BPS) by the splicing factor 1 (SF1). Together, these interactions result in the spliceosomal E complex that defines the exon/intron boundaries. Subsequent to this initial step, SF1 is replaced from the BPS by the U2 snRNP building the A complex. This is followed by recruitment of the pre-assembled U4/U6.U5 tri-snRNP forming the pre-catalytic B complex. Rearrangements in the snRNPs release the U1 and the U4 snRNPs leading to the activation of the B complex. The active B complex carries out the branching in the first transesterification reaction resulting in the C complex. Further rearrangements activate the C complex that ligates the exons in the second transesterification step. Finally, the spliced mRNA, the intron-lariat and the U2, U5 and the U6 snRNP are released *(Figure 1.2*; Wahl et al., 2009).

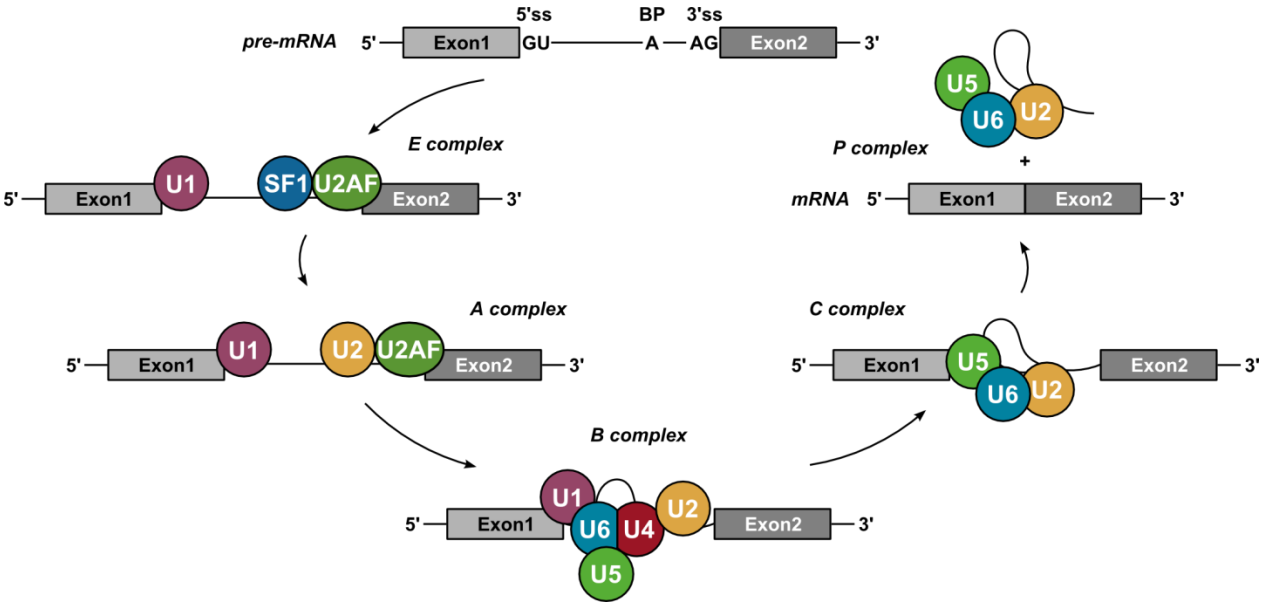


Figure 1.2: Simplified representation of the assembly of the major spliceosome. The exons are depicted as grey boxes. The introns are shown as solid lines with highlighted splice sites consensus sequences. The auxiliary factors and the five snRNPs are depicted in colored circles. Adopted from Wahl, Will, & Lührmann, 2009.

1.2 Alternative splicing

Most eukaryotic pre-mRNAs consist of several non-coding introns and coding exons. Additionally to the excision of introns during constitutive splicing, a vast majority of pre-mRNAs contain exons that can be alternatively included or excluded from the mature mRNA. Alternative splicing results in several different mRNAs encoded by one gene thereby dramatically increasing the proteome complexity (Nilsen and Graveley, 2010). One famous example is the *D. melanogaster* Dscam (Down syndrome cell adhesion molecule) gene whose 95 alternative exons are organized into four clusters and are extensively spliced allowing the possible expression of 38,016 different mRNAs (Celotto and Graveley, 2001). The development of RNA deep sequencing technologies enabled transcriptome wide analysis of alternative splicing and revealed that around 65% of the murine and 95% of human pre-mRNAs undergo alternative splicing (Barbosa-Morais et al., 2012; Merkin et al., 2012). Alternative splicing is regulated in many different cellular processes and in a development and tissue specific manner. Cell stage specific splicing was for example characterized during neuronal differentiation (Raj and Blencowe, 2015), heart development (Van Den Hoogenhof et al., 2016) or hematopoiesis (Inoue et al., 2016). Furthermore, alternative splicing occurs in differentiated cells as response to cellular signaling and cellular stresses like temperature changes, UV-light or infections (Biamonti and Caceres, 2009; Martinez and Lynch, 2013; Preußner et al., 2017; Sharma and Lou, 2011; Shin and Manley, 2004).

RNA sequencing technologies enabled the detection of a plethora of splice variants at the transcriptome level, however the extent to which the transcripts are translated is poorly known. Currently used mass spectrometry techniques are insufficient in sensitivity and coverage to detect adequate numbers of splice variant specific peptides. However, recent studies identified thousands of splice variants in polysomal fractions indicating that they are translated (Floor and Doudna, 2016; Sterne-Weiler et al., 2013; Weatheritt et al., 2016). Another challenge is the functional characterization of spliced variants which is currently determined individually for every event. It was shown that splicing variants of a protein can possess changed protein stability, cellular localization, ligand binding activity and enzymatic activity (Kelemen et al., 2013). Recently, a first systematic analysis revealed that alternative spliced variants possess different protein-protein interaction profiles and share less than 50% of their interactomes pointing to different functionalities of the splice variants (Yang et al., 2016).

1.3 Modes of alternative splicing

Alternative pre-mRNA splicing occurs in different forms as depicted in *Figure 1.3*. Exclusion or inclusion of an entire exon is the most common alternative splicing event. Those exons are termed cassette exons. Additionally to the splice sites at exon-intron boundaries, alternative 5’ ss and 3’ ss within exons or introns can be used leading to the inclusion of a lengthened or shortened exons. Another form of alternative splicing, which occurs rather rarely in eukaryotes, is the inclusion of an entire intron into the mature mRNA. Furthermore, two adjacent exons can be included in a mutually exclusive fashion (Keren et al., 2010). As a result of alternative splicing, the translated protein lacks or gains amino acid segments or has a completely new C-terminus through a shift in the reading frame. Furthermore, alternative splicing can result in a translation-dependent degradation of the mRNA by the non-sense mediated mRNA decay (NMD) pathway through introduction of premature termination codons (McGlincy and Smith, 2008).

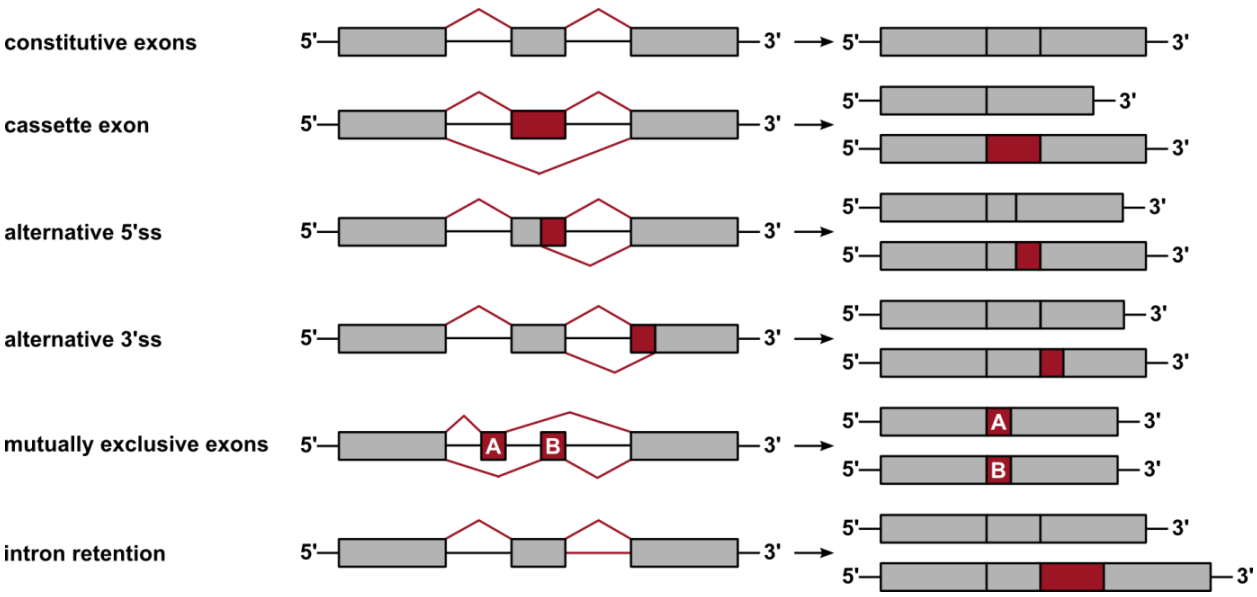


Figure 1.3: Alternative splicing of pre-mRNA. Depicted are the five basic types of alternative splicing. Constitutively included regions are shown in grey and alternative regions in red. The possible usage of splice sites is indicated with red lines on the left site and the spliced mRNA variants are shown on the right side.

1.4 Regulation of alternative splicing

Alternative splicing is a tightly regulated process involving *cis*-acting RNA elements that are bound by *trans*-acting protein factors that impact positively or negatively on the selection of a splice site. *Cis*-acting RNA elements involved in alternative splicing regulation are divided into four groups. Based on the position in the pre-mRNA and the direction of the regulation they are divided into exonic splicing silencer (ESS), exonic splicing enhancer (ESE), intronic splicing silencer (ISS) and intronic splicing enhancer (ISE) (Black, 2003). *Trans*-acting factors that bind to *cis*-acting RNA elements and are involved in alternative splicing regulation can be divided into two very well characterized protein families: the heterogeneous nuclear ribonucleoprotein particles (hnRNPs) and the serine/arginine- rich (SR proteins) splicing factors (*Figure 1.4*; Busch and Hertel, 2012).

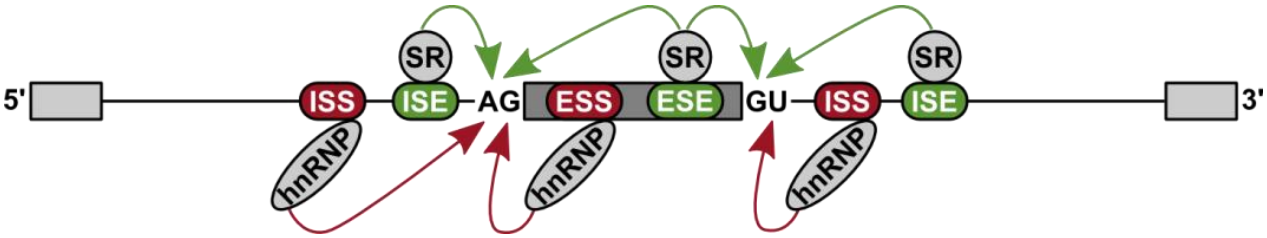


Figure 1.4: Assistance of RBPs in splice site recognition. RNA binding *trans*-acting factors recognize *cis*-acting elements and promote (green arrows) or prevent (red arrows) splice site recognition. *Cis*-acting elements that enhance splicing are depicted as green boxes and splicing silencer as red boxes. *Trans*-acting factors are represented as grey circles.

SR proteins contain at least one RNA recognition motif and a family-defining domain enriched for arginine/serine (RS) dipeptides. SR proteins are widely seen as positive regulators of splicing which is accomplished by direct protein-protein interactions through the RS domains of splicing factors. SR proteins enhance exon definition by the recruitment of spliceosomal components like the U1 snRNP to the 5' ss and the U2AF heterodimer to the 3' ss. In contrast, hnRNPs are widely seen as negative regulators of alternatives splicing. For hnRNPs several different regulatory mechanisms have been identified as for example looping out of exons,

inhibition of snRNP recruitment or multimerization of hnRNPs along exons (*Figure 1.4*; Busch and Hertel, 2012; Fu and Ares, 2014). However, this simple division of SR proteins as positive and hnRNPs as negative regulators is not always correct as there were exceptions described with opposite regulatory functions. This is because the position of the splicing regulatory element (SREs) relative to the 5' and 3' ss as well as surrounding sequences also have a strong impact on the regulatory potential of the bound *trans*-acting factor (Fu and Ares, 2014). In addition to SR proteins and hnRNPs, a variety of other RNA-binding proteins were shown to regulate alternative splicing, mostly in a tissue–specific manner (Fu and Ares, 2014; Vuong et al., 2016).

Regulation of alternative splicing through *cis*-elements and *trans*-factors is moreover substantially affected by RNA secondary structures and transcription. Through formation of stem loop structures in the pre-mRNA, large distances between regulatory elements can be bypassed and splice sites and SREs can be occluded in the double strand stems or exposed in the loops (Kornblihtt, 2015; McManus and Graveley, 2011; Taliaferro et al., 2016). During transcription, regulation of alternative splicing can additionally be affected through the RNA polymerase II (RNAPII) in a kinetic and recruiting coupled mode. In the kinetic coupled mechanism, varying transcriptional elongation rates determine the splice site choice. The weak splice site of the alternative exon competes with the strong splice site of the downstream constitutive exon resulting in inclusion during slow RNAPII elongation and exclusion during fast RNAPII elongation. Whereas in the recruiting coupled model, splicing factors are recruited through the C-terminal domain of RNAPII to the pre-mRNA and can prevent or favor alternative splicing (Naftelberg et al., 2015).

1.5 3' ss recognition through the U2AF heterodimer

During the initial steps of splicing, conserved sequences in the pre-mRNA have to be recognized to enable the selection of exon/intron boundaries. The 5' splice site in the majority of eukaryotic pre-mRNAs typically consists of an 8 nucleotides sequence stretch with an invariant intronic GU dinucleotide defining the exon/intron boundary. The intron/exon boundary at the 3' ss is determined by a pyrimidine rich segment followed by a conserved AG dinucleotide. The branch point sequence, characterized by a highly degenerate 7 nucleotide sequence with an

invariable branch point adenosine, is found 18-40 nucleotides upstream of the 3' ss (*Figure 1.5*; Wu and Fu, 2015). During splicing initiation, the 5' ss is bound by the U1 snRNP through base pairing of the complementary sequences found in the U1 snRNA. The 3' ss is initially recognized by the essential heterodimer U2-auxiliary factor (U2AF) in a temperature- and ATP-independent manner. Subsequently, U2AF recruits the U2 snRNP to the branch point A where it binds through U2 snRNA base pairing (*Figure 1.5*; Ruskin et al., 1988; Zamore and Green, 1989).

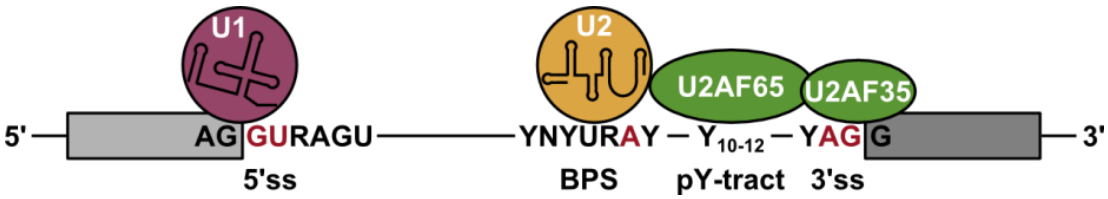


Figure 1.5: Recognition of conserved RNA elements during splicing initiation. The consensus sequences of the 5' ss, 3' ss and the BP are shown and invariant nucleotides are highlighted in red. The colored circles indicate snRNPs and the ovals auxiliary factors that bind the consensus sequences and thereby define exon/intron boundaries. Y, R and N indicate pyrimidine, purine and any nucleotide, respectively. Based on Wu & Fu, 2015.

Purification of the U2AF heterodimer showed that it is composed of the large 65 kDa subunit U2AF65 also termed U2AF2 and the small 35 kDa subunit U2AF35 also termed U2AF1 (Zamore and Green, 1989). Binding analysis revealed that U2AF65 binds to the poly-pyrimidine (pY) tract of the 3' ss and that its affinity correlates with the length of the pY-tract (Singh et al., 1995; Zamore et al., 1992). For U2AF35 alone, only weak binding to the 3' ss was observed however a significant contribution to a high-affinity binding of the U2AF heterodimer was detected for U2AF35. It was shown that U2AF35 especially increases the binding of U2AF65 to splice sites where several pyrimidines in the pY-tract are mutated (Rudner et al., 1998a). Site-specific crosslinking of U2AF35 showed that it directly binds to the AG dinucleotide following the pY-tract in the 3' ss and thereby assist U2AF65 in 3' ss recognition (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). *In vitro* splicing assays revealed that assistance of U2AF35 is especially needed for 3' ss with short or weak pY-tracts and led to the classification of exons into AG-dependent and - independent exons (Guth et al., 1999; Wu et al., 1999).

1.6 U2AF35 and U2AF35-related proteins

U2AF35 is an essential and highly conserved protein encoded in the genomes of fission yeast up to human. Deletion of U2AF35 orthologs in *C. elegans* (UAF-2), *S. pombe* (U2AF23) and *D. melanogaster* (dU2AF38) causes lethality and the human U2AF35 was shown to be essential for cell viability (Pacheco et al., 2006a; Rudner et al., 1996; Webb and Wise, 2004; Zorio and Blumenthal, 1999). Additionally, mammals encode several U2AF35-related proteins with distinct functions in the cell. In the human genome, three U2AF35-related genes are present (*Figure 1.6*; Mollet et al., 2006). The gene *U2AF1L4* encodes for a 26 kDa protein termed U2AF26 and shows the highest conservation to U2AF35 of 76%. The most divergent part of these two proteins is the C-terminal RS domain that is reduced to a few RS repeats in U2AF26 compared to U2AF35 (Shepard et al., 2002). The genes *ZRSR1* and *ZRSR2* encode for the 58 kDa proteins ZRSR1 and ZRSR2 that are 91% identical between themselves. Both proteins have an acidic 170 amino acid (aa) long N-terminus preceding the U2AF35-homology regions that have an average conservation of 45% (Mollet et al., 2006; Tronchère et al., 1997).

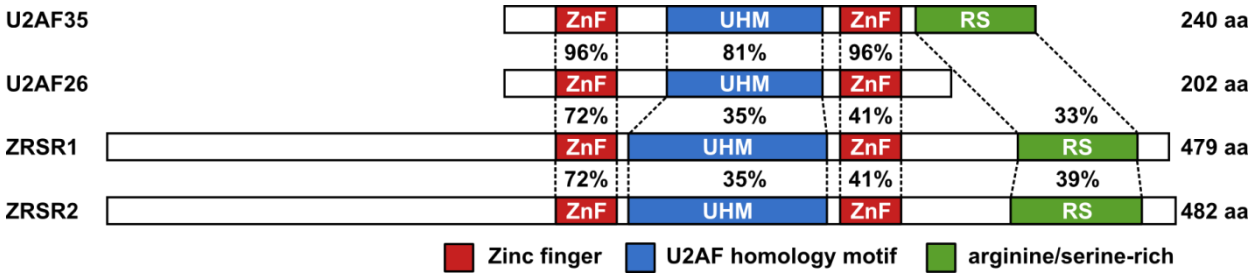


Figure 1.6: Domain organization and conservation of human U2AF35-related proteins. Protein domains of U2AF26, ZRSR1 and ZRSR2 were aligned to U2AF35 and between the dashed lines the identity to U2AF35 was indicated as percent. The amino acid (aa) length is notified on the right side. Based on Mollet, Barbosa-Morais, Andrade, & Carmo-Fonseca, 2006.

ZRSR2 (zinc finger RNA binding motif and serine/arginine rich 2) is ubiquitously expressed in all human tissues and has distinct functions in splicing regulation. It was shown that splicing of an U2-type intron is strongly reduced in ZRSR2-depleted extracts and can be restored through addition of recombinant ZRSR2 but not through U2AF35. Like U2AF35, ZRSR2

interacts with U2AF65 through U2AF35-homologous regions but the interaction is less efficient (Tronchère et al., 1997). Additionally, ZRSR2 is associated with the human U11/U12 snRNP and was shown to regulate splicing of U12-type introns through direct binding to their 3' ss (Shen et al., 2010; Will et al., 2004). U12-type introns are spliced by the minor spliceosome that has a similar protein composition as the major spliceosome presented in *Chapter 1.1* but contains specific snRNAs. Additionally to the GT-AG consensus sequence bound by the major spliceosome, the minor spliceosome recognizes introns flanked with AT-AC sequences and with shortened pY-tracts (Turunen et al., 2013). For U2-type introns, ZRSR2 was shown to be important for the execution of the second step of splicing, also through contacting the 3' ss (Shen et al., 2010).

ZRSR1 was created through retro transposition of the processed *ZRSR2* mRNA. The mouse *ZRSR1* gene is located in one intron of the *Murr1* gene and probably lacks its own promoter. Still, mRNA expression of mouse *ZRSR1* was detected in different tissues with the highest expression in the brain (Hatada et al., 1993). Furthermore, ZRSR1 expression is increased during erythroid differentiation and seems to be required for normal erythropoiesis. ZRSR1 was also shown to interact with U2AF65 and other spliceosomal components indicating a role in splicing. Along with that, splicing of the erythroid-specific exon 8 of muscleblind-like 2 (*Mbnl2*) was deferentially regulated in *ZRSR1* knockout mice (Andrade, 2011).

1.6.1 U2AF35

The human U2AF35 was initially cloned and characterized by Zhang *et al*. in 1992. Since then, many studies focused on the analysis of U2AF35 splicing regulation and the function of U2AF35 protein domains. Like many other factors involved in splicing regulation, U2AF35 has a C-terminal arginine/serine-rich (RS) domain which is interrupted by a glycine stretch (Zhang et al., 1992). In SR proteins, RS domains serve as protein-protein interaction surfaces for other RS-domain containing proteins and contain the nuclear localization signal (NLS). Both, interaction and localization were shown to be regulated by phosphorylation of the RS domain (Twyffels et al., 2011). Although the RS repeats in U2AF35 are more dispersed and less periodically arranged than in SR proteins, the RS domain possesses similar functions. U2AF35 but not U2AF65 was shown to interact with the SR proteins SRSF1 and SRSF2. Moreover, the RS domain of U2AF35

is crucial for enhancer dependent splicing regulation (Zuo and Maniatis, 1996). Together with further studies, the model was established that SR proteins bind to ESEs and recruit the U2AF heterodimer to weak 3' ss through binding of the RS domain of U2AF35 (Graveley et al., 2001; Zuo and Maniatis, 1996). SRSF1 and SRSF2 simultaneously interact with U2AF35 and the U1 snRNP and thereby connect the 3' ss with the 5' ss during exon definition (Wu and Maniatis, 1993).

As in other SR proteins, the RS domains of both U2AF subunits are important for their cellular localization. It was shown that the RS domains of U2AF35 and U2AF65 function as a NLS that brings the proteins into the nucleoplasm where they are further enriched in nuclear speckles. Both U2AF subunits can shuttle between the cytoplasm and the nucleus in an mRNA-independent manner through a carrier-mediated mechanism that is sufficiently enabled through one of the two RS domains (Gama-Carvalho et al., 2001). In agreement with that, *in vivo* studies showed that the presence of at least one RS domain of *D. melanogaster* U2AF is needed for fly viability (Rudner et al., 1998b).

Direct RNA-binding of U2AF35 to the 3' ss AG dinucleotide was first demonstrated in the late 90s by performing site-specific UV-crosslinking and systematic evolution of ligands by exponential enrichment (SELEX) analysis (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). Still, the domain that directly contacts the RNA was unknown for many years. RNA-recognition motifs (RRMs) are the most common domains of RNA-binding proteins and are known to mediate RNA-binding. An RRM is composed of two conserved ribonucleoprotein (RNP) motifs separated by a 40-residue stretch. In total, an RRM domain comprises 80-90 residues that form four antiparallel ß-sheets packed against two α-helices (Muto and Yokoyama, 2012). The N-terminus of U2AF35 encodes an atypical RRM that has a prolonged amino acid stretch between both RNPs that forms a 30-residue long α-helix instead of a 10-residue α-helix found in classical RRMs (Kielkopf et al., 2001). Still, *in vitro* splicing assays indicated that the RRM of U2AF35 together with full length U2AF65 is sufficient to activate the splicing of a weak 3' ss substrate suggesting that the RRM of U2AF35 is responsible for RNA binding (Guth et al., 2001). However, RNA binding of the U2AF35 RRM was found to be rather weak in a micromolar range (K_d=6 µM) compared to the strong affinity of U2AF65 (K_d=1-10 nM) (Kielkopf et al., 2001; Singh et al., 1995). Furthermore, binding of the U2AF35 RRM to RNA induced just small perturbations in nuclear magnetic resonance (NMR) studies arguing

against a direct role of the RRM in RNA binding. Additionally, the RRM alone failed to discriminate between the consensus 3' ss AG dinucleotide and mutations therein indicating that other domains are responsible for RNA binding specificity of U2AF35 (Soares et al., 2014).

Webb *et al*. provided first indications that the two zinc finger (ZnF) domains of U2AF35 are indispensable for RNA binding (Webb and Wise, 2004). ZnFs are abundant protein domains that obtain their structure by coordinating zinc ions with a combination of cysteine and histidine residues and were initially described to bind DNA in a sequence-specific manner (Klug, 2010). However, some classes of ZnFs, like CCCH and C_2H_2 ZnFs, are often found in RNA-binding proteins where they were shown to bind RNA (Fu and Blackshear, 2017; Hall, 2005). The ZnF domains of U2AF35 are highly conserved between species and in U2AF35-related proteins (*Figure 1.6*). U2AF35 has two ZnFs, the first is located in the N-terminal part of U2AF35 preceding the RRM and the second ZnF follows the RRM *(Figure 1.6)*. In the first ZnF, the zinc complexing residues are arranged in a C-X_8-C-X_5-C-X_3-H and in the second ZnF in a C-X_7-C-X_5-C-X_3-H motif (Webb and Wise, 2004; Yoshida et al., 2015). Webb and Wise performed comprehensive mutational analysis of the *S.pombe* U2AF35-ortholog U2AF23, which revealed that the RRM as well as both ZnFs of U2AF23 are indispensable for viability. Furthermore, utilizing a modified RNA three-hybrid system they showed that all three domains contribute to RNA-binding. These observations were clarified in the recent assignment of the crystal structure of full length U2AF23 (Yoshida et al., 2015). The U2AF23 structure demonstrates that the ß-sheet surface of the RRM serves as a scaffold for the ZnFs on which they interact with each other thereby covering the possible RNA-binding surface of the RRM (*Figure 1.7*). Furthermore, isothermal titration calorimetry (ITC) measurements revealed a cooperative binding of both ZnF to a 4-bases long RNA comprising the 3' ss sequence 5'-UAGG-3' with a K_d value of 3.3 µM, whereas they do not observe RNA-binding of the RRM alone. ITC measurements with mutated 3' ss as well as ZnF mutated U2AF23 further verified the ZnF as RNA-binding domain in U2AF23. In small-angle X-ray scattering (SAXS) analysis no major conformational changes were observed upon incubation with the RNA indicating that the RRM is not surface-exposed upon RNA binding and stays covered by the two ZnFs (Yoshida et al., 2015).

The initial characterization of the U2AF heterodimer revealed stable complex formation between U2AF35 and U2AF65 in a 1:1 stoichiometry (Zamore and Green, 1989). The interaction was mapped to a 125 residue long region in U2AF35 which mainly encodes the RRM (Zhang et

al., 1992). In *D. melanogaster* U2AF65 a proline-rich fragment of 28 residues in the N-terminus was mapped as the U2AF35-interacting region (Rudner et al., 1998c). The first crystal structure of the RRM was assigned using a minimal human U2AF heterodimer composed of the U2AF35 RRM and the 28 residue interacting region in U2AF65. The structure revealed a novel "tongue to groove" interaction mode involving one tryptophan in either of the proteins that reciprocally binds in a groove of the other protein (Kielkopf et al., 2001). In line with that, NMR studies showed that the U2AF35 RRM alone is unstructured in solution but becomes structured upon binding of the U2AF65 proline-rich motif (Kellenberger et al., 2002). The "tongue to groove" interaction mode was also confirmed in the structure of yeast U2AF23. Furthermore, the structure of full length yeast U2AF23 identified an additional interaction surface between U2AF23 and the large subunit U2AF59 in the very C-terminal helix of U2AF23 (*Figure 1.7*; Yoshida et al., 2015). The novel interaction mechanism in U2AF is also observed between the RRM of U2AF65 and SF1 (Selenko et al., 2003). Based on that, such RRM-like domains with distinct functions in protein recognition were termed U2AF homology motifs (UHMs) (Kielkopf et al., 2004).

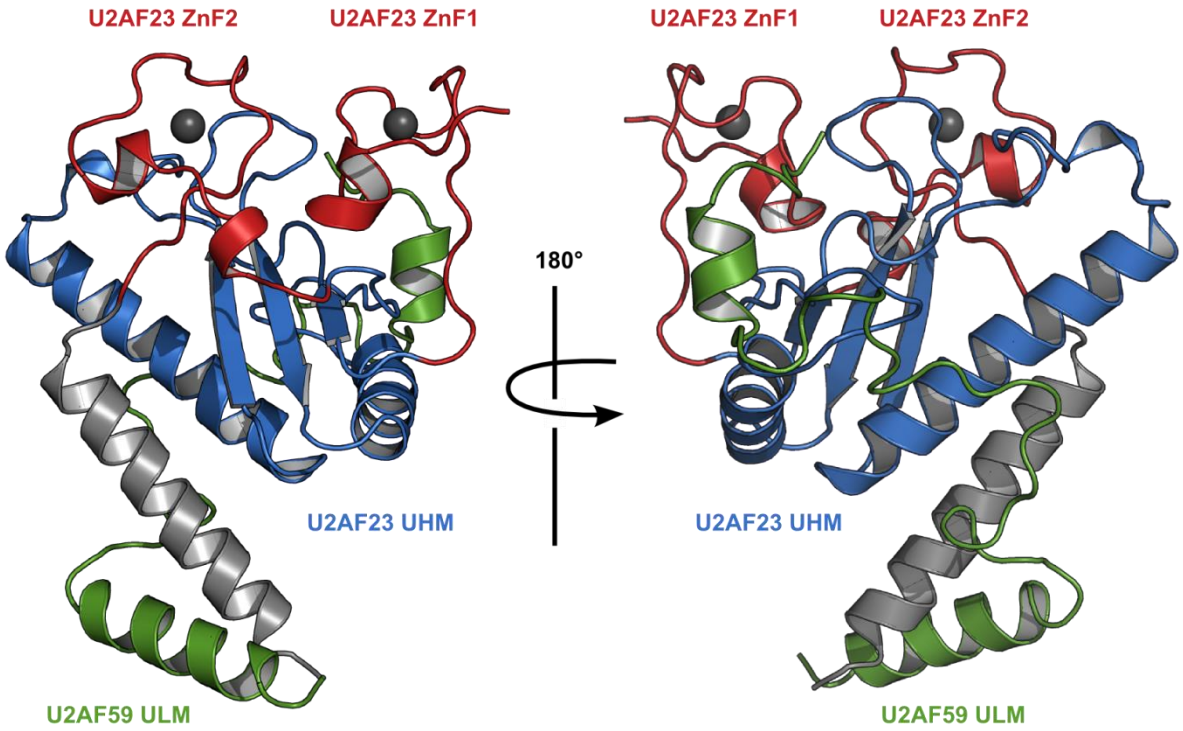


Figure 1.7: Crystal structure of the* S.pombe *U2AF heterodimer. The zinc finger domains (ZnF1 and ZnF2), the U2AF homology motif (UHM) and the C-terminal helix 6 of U2AF23 are colored in red, blue and grey, respectively. The U2AF ligand motif (ULM) of U2AF59 used for co-crystallization is depicted in green. PDB ID: 4YH8; Yoshida *et al*., 2015.

1.6.2 U2AF26

U2AF26 is encoded by the gene *U2AF1L4* and was initially cloned and described by Shepard *et al*. in 2002. U2AF26 possesses a similar domain organization as U2AF35 and is 76% identical to U2AF35. The N-terminus of U2AF26 is composed of the UHM and the U2AF65 interaction regions flanked by two ZnFs and has 89% identical amino acids to U2AF35 (*Figure 1.6 and Figure 1.8*). The C-terminal domain of U2AF26 is shortened and lacks most of the RS dipeptides and the glycine-rich stretch present in U2AF35. Like U2AF35, U2AF26 localizes into the nucleus and builds a functional heterodimer with U2AF65. Furthermore, *in vitro* splicing assays showed that a U2AF26-U2AF65 dimer can substitute for U2AF35-U2AF65 in constitutive and enhancer dependent splicing through to binding to the 3' ss (Jeremiah Brian Shepard, 2004; Shepard et al., 2002). Splicing regulation by U2AF26 was also observed *in vivo*. Knock-down and overexpression studies of U2AF26 in cells and in mice revealed a regulation of *CD45* alternative splicing towards a less active variant which negatively regulates T cell activation. This regulation is counteracted by an interaction of U2AF26 with the transcription factor GFI1. *CD45* splicing was specifically regulated by U2AF26 and not U2AF35 indicating that both proteins regulate different splicing events (Heyd et al., 2006).

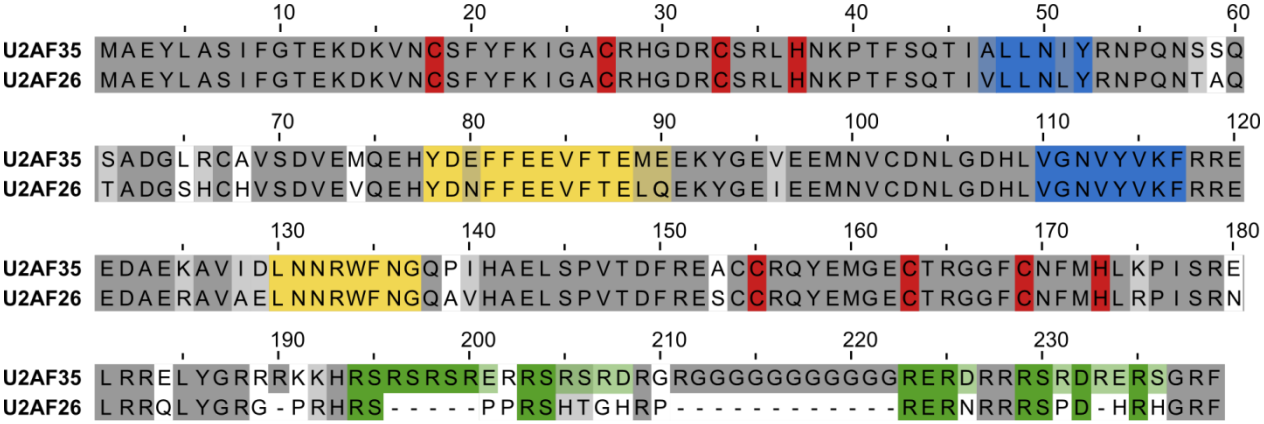


Figure 1.8: Sequence alignment of mouse U2AF35 and U2AF26. Sequence comparison of U2AF35 and U2AF26 shows identical residues boxed in dark grey, conservative residues in light grey and different residues in white. Zinc-complexing cysteines and histidines are boxed in red and dipeptides of the RS domain in green. The RRM, composed of RNP1 and RNP2, is labeled in blue and U2AF65-contacting regions are shown in yellow. Adopted from Shepard, Reick, Olson, & Graveley, 2002.

The *U2AF1L4* pre-mRNA itself is alternatively spliced giving rise to two additional protein variants apart from the full length (fl) protein. In the first U2AF26 variant, exon 7 is excluded from the mature mRNA (U2AF26ΔE7) which encodes 32 amino acids of the C-terminus of U2AF26. In contrast to U2AF26fl, U2AF26ΔE7 loses its ability to shuttle between cytoplasm and nucleus and localizes into the cytoplasm with a so far unknown function (Heyd et al., 2008). In the second U2AF26 variant, exon 6 and 7 are skipped from the mature mRNA. Splicing of exon 6 and 7 is regulated in a circadian manner and changes the reading frame of U2AF26 thereby driving translation far into the supposed 3'UTR. Like U2AF26ΔE7, U2AF26ΔE67 localizes into the cytoplasm where it impacts on the circadian clock regulation by destabilizing the clock protein PERIOD1 (Preußner et al., 2014).

1.7 Spliceosomal mutations in cancer

Besides altered mRNA expression, cancer cells can possess altered splicing patterns leading to the expression of protein variants that promote their growth and survival (David and Manley, 2010). Starting in 2011, many cancer-associated mutations in genes encoding splicing factors were discovered that might induce aberrant splicing patterns in cancer cells. Mutations of splicing factors appear in solid cancers like breast and pancreatic cancer and lung adenocarcinoma, but most frequently in different types of leukemia and in patients with myelodysplastic syndromes (MDS) which are precursor diseases of leukemia. Spliceosomal mutations predominantly occur in the proteins SF3B1, SRSF2, U2AF1 and ZRSR2, all involved in the initial steps of splicing and mostly as somatic, heterozygous missense mutations in a mutually exclusive manner (*Figure 1.9*). Except for ZRSR2, the mutations appear in hot spots

and rarely induce frame shifts or premature stops, indicating a gain-of-function of these mutations (reviewed in Yoshida and Ogawa, 2014).

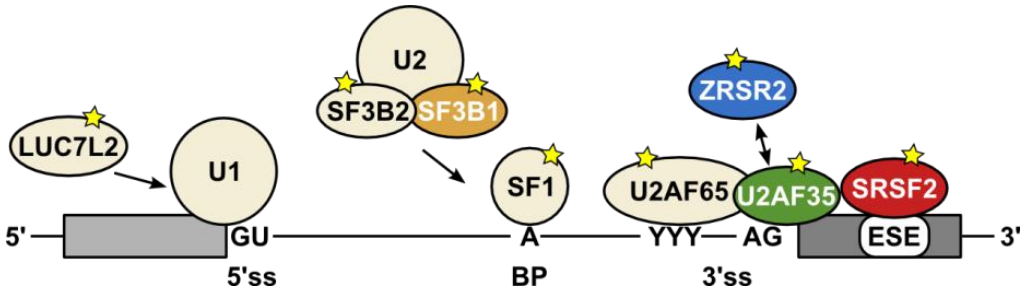


Figure 1.9: Splicing factor mutations in hematopoietic diseases. Splicing factors that are mutated in hematopoietic diseases are labeled with a yellow star. The factors highlighted by colored circles are most frequently mutated.

The most frequently mutated spliceosomal component is the splicing factor 3b subunit 1 (SF3B1) with mutation rates up to 80% in some MDS subclasses (Yoshida et al., 2011). SF3B1 is part of the SF3B complex which associates with SF3A to build the U2 snRNP and binds to U2AF65 during BPS selection (Wahl et al., 2009). Mutations in SF3B1 are enriched in a few residues located in three of the 22 HEAT (Huntington, Elongation factor 3, protein phosphatase 2A, Targets of rapamycin 1) repeats (Yoshida and Ogawa, 2014). Mutations in SF3B1 were shown to induce usage of cryptic 3' ss located closely upstream of the canonical 3' ss through utilization of an alternative branch point (*Figure 1.10*). Half of the alternative mRNAs generated through cryptic 3' ss usage induced by mutated SF3B1 undergo nonsense-mediated mRNA decay (NMD) thereby reducing the amount of canonical isoforms (Alsafadi et al., 2016; Darman et al., 2015). Recent structural information revealed that the mutations in SF3B1 might impact on the conformation of the bound pre-mRNA (Cretu et al., 2016; Yan et al., 2016; Jenkins and Kielkopf, 2017).

The second most mutated splicing factor is the serine/arginine-rich splicing factor 2 (SRSF2) that belongs to the SR protein family and promotes splice site recognition (*Chapter 1.4*). SRSF2 is most often mutated in chronic myelomonocytic leukemia (CMML; 50%) exclusively

affecting proline at position 95 that is found in the linker region between the N-terminal RRM and the C-terminal RS domain (Yoshida et al., 2011). SRSF2 mutations were shown to impair hematopoietic differentiation in mice by missplicing of key hematopoietic regulators besides many other misspliced mRNAs found in this study. This is due to a conformational change of the nearby RRM induced by the P95 mutation. As a consequence, the sequence preference of mutated SRSF2 is shifted towards CCNG whereas wt SRSF2 binds the motifs CCNG and GGNG equally well (*Figure 1.10*; Kim et al., 2015; Zhang et al., 2015).

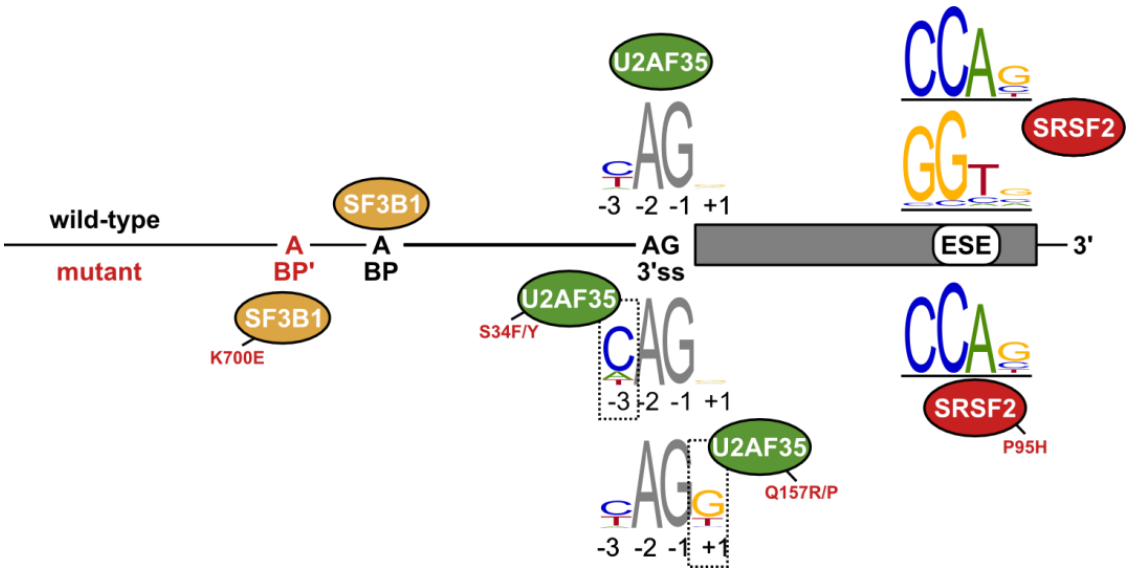


Figure 1.10: Mechanistic effects of SF mutations on RNA splicing. Shown are the binding specificities of wild-type (top) and mutant (bottom) SF3B1, U2AF35 and SRSF2. Adapted from Inoue *et al*., 2016.

U2AF35 mutations were detected in patients with MDS (5-12%), CMML (5-17%) and AML (2-9%) and predominantly affect the residue S34 in the first ZnF and, less frequently, the residue Q157 in the second ZnF (Yoshida and Ogawa, 2014). The serine 34 to phenylalanine substitution (S34F) is by far the most frequent mutation in U2AF35 representing 60-80% of all U2AF35 mutations; serine 34 to tyrosine (S34F) represents around 10-15%. Mutations of glutamine 157 to either proline (Q157P) or arginine (Q157R) add up to 20% of all U2AF35 mutations with a very diverse prevalence for the Q157P or Q157R mutation in different studies (Graubert et al., 2011; Makishima et al., 2012; Yoshida et al., 2011). In patients and in cell lines

with U2AF35 mutations no widespread splicing failure was observed, but a changed splicing program for a few hundred pre-mRNAs with increased skipping of cassette exons as most frequent event type (Brooks et al., 2014; Ilagan et al., 2014; Okeyo-Owuor et al., 2015; Przychodzen et al., 2013). The occurring splicing changes were different from those observed in knock down analysis of U2AF35 wt indicating an altered function of the mutants rather than a simple loss-of-function character of the mutations (Przychodzen et al., 2013). Furthermore, U2AF35 S34F was shown to favor the selection of the distal over the proximal cleavage and polyadenylation site in many target mRNAs (Park et al., 2016).

Motif analysis of differentially spliced cassette exons revealed variations in the bound 3' ss sequences for the S34 and Q157 mutations (*Figure 1.10*). S34 mutants preferentially include exons with a C or an A at the -3 position directly preceding the conserved AG dinucleotide and preferentially exclude exons with a T at this position (Fei et al., 2016; Ilagan et al., 2014; Okeyo-Owuor et al., 2015; Shirai et al., 2015). Whereas for the Q157 mutants a preferential inclusion of exons with a G and exclusion of exons with an A at the +1 position directly following the conserved AG dinucleotide was observed (Ilagan et al., 2014). Mutational analysis of minigenes verified the sequence preferences for S34 and Q157R (Ilagan et al., 2014; Okeyo-Owuor et al., 2015).

The fourth most common mutated splicing factor is the U2AF35-related protein ZRSR2. ZRSR2 is mutated in 3-11% of MDS and 8% of CMML patients (Yoshida and Ogawa, 2014). In contrast to SF3B1, SRSF2 and U2AF35, mutations in ZRSR2 are distributed over its complete coding region. Most of the mutations are nonsense or frame shift inducing mutations that cause large structural changes or truncations of the ZRSR2 protein indicating a loss-of-function character of the mutations (Yoshida et al., 2011). Consistent with its known function in regulation of U12-dependent splicing, RNA sequencing analysis of ZRSR2-depleted cells and patients with ZRSR2 mutations revealed aberrant splicing of U12-type introns (Madan et al., 2015).

1.8 Aims of this study

Splicing of pre-mRNA is a tightly regulated process that has to be performed very accurately to ensure the generation of functional mRNAs. The accuracy is accomplished by the recognition of conserved RNA sequences like the splice sites at the exon-intron boundaries. Almost 20 years ago, human U2AF35 was identified as 3' ss binding protein, yet the domain that directly contacts the RNA stayed enigmatic for a long time. Three years ago, a structural approach revealed that both ZnFs cooperatively mediate sequence-specific binding of yeast U2AF35 to RNA. Furthermore, disease-associated mutations within the ZnF domains of U2AF35 were shown to be associated with missplicing through changed binding specificity. To extend the insights into the function of mammalian U2AF35 ZnF we addressed the key properties of the ZnF as a whole and focused on the functional consequences of mutations within the second ZnF of U2AF35 that so far got little attention.

The protein U2AF26 is highly conserved to U2AF35 and *in vitro*-based studies suggests an involvement of U2AF26 in splicing regulation. Furthermore, U2AF26 is itself alternatively spliced giving rise to the cytoplasmic U2AF26ΔE7 variant with yet unknown function. In this study, we aimed to identify functional similarities and differences between U2AF26 and U2AF35 in splicing regulation *in vivo*. Furthermore, we aimed to decipher the potential role of the U2AF26ΔE7 variant on cytoplasmic RNAs.

2. RESULTS

(This chapter presents and summarizes the most important results obtained during this study. Further results and controls are shown in the appended manuscripts.)

2.1 The zinc finger domains in U2AF35 and U2AF26 control protein stability, splicing regulation and interaction with U2AF65.

Refers to**:**

Herdt, O., Preußner, M., Heyd, F. (2018). The zinc finger domains in U2AF35 and U2AF26 control stability, localization, interaction with U2AF65 and play a role in translation. NAR, *in preparation for re-submission.*

Lately, many studies identified cancer-associated mutations in both ZnFs of U2AF35 and subsequently focused on functional implications of the mutations in splicing regulation as the ZnFs were shown to be responsible for 3' ss binding in U2AF35 (Yoshida and Ogawa, 2014; Yoshida et al., 2015). So far, functional analysis of the ZnF domains in general were limited to the yeast ortholog U2AF23 (Webb and Wise, 2004; Yoshida et al., 2015). Hence, the first focus of this study was the analysis of key functionalities of the ZnFs of mammalian U2AF35 concerning expression, stability, protein-protein-interaction and splicing regulation. The ZnFs of U2AF35 and U2AF26 are highly conserved, but previous studies indicate slightly different binding specificity for both proteins (Shepard et al., 2002). Hence, we included U2AF26 in our study to clarify functional similarities and differences to U2AF35 to address the second aim of this study. Additionally, we asked whether the cancer-associated mutations in the ZnFs might impact on U2AF35 properties besides RNA binding. The term U2AF1 is used whenever both U2AF35 (encoded by U2AF1) and U2AF26 (encoded by U2AF1L4) are meant together.

To elucidate the key functionalities of the ZnF domains of mammalian U2AF1-proteins, we cloned full length and ZnF-deleted variants of the mouse U2AF35 and U2AF26. Additionally, we introduced the cancer-associated substitutions S34F in the first and the Q157R and Q157P in the second ZnF of U2AF35 *(Figure 2.1.1 A)*. The U2AF1-mutants were transiently transfected in HEK293T cells, treated for different time points with the translation inhibitor cycloheximide

(CHX) and subjected to Western blot analysis. Investigation of the protein expression and stability revealed that deletion of ZnF1 slightly reduces the basal expression of U2AF35 and the protein half-life to less than 8 hours, whereas introduction of the point mutations in both ZnFs just lead to a slight reduction of the basal expression. Deletion of the second ZnF had a more severe impact on the basal expression and protein half-life of U2AF35 reducing it below 4 hours *(Figure 2.1.1 B)*. Similar results were obtained for U2AF26 indicating that ZnF2 stabilizes U2AF1-proteins (appended data). For all mutants the negative effect on expression and stability was restored through expression of the dimerization partner U2AF65 *(Figure 2.1.1 B)*. Next, we asked whether the ZnF domains affect the interaction of the U2AF heterodimer in co-immunoprecipitation assays. Under low salt conditions, the interaction of U2AF35 with U2AF65 was unaffected by the deletion of either the ZnF1 or ZnF2, however increase of the stringency (1200 mM salt) diminished the interaction with U2AF35 lacking the ZnF2 indicating that ZnF2 is required for stable U2AF heterodimer formation. In contrast, neither of the point mutations in U2AF35 showed an impact on heterodimer formation *(Figure 2.1.1 C)*.

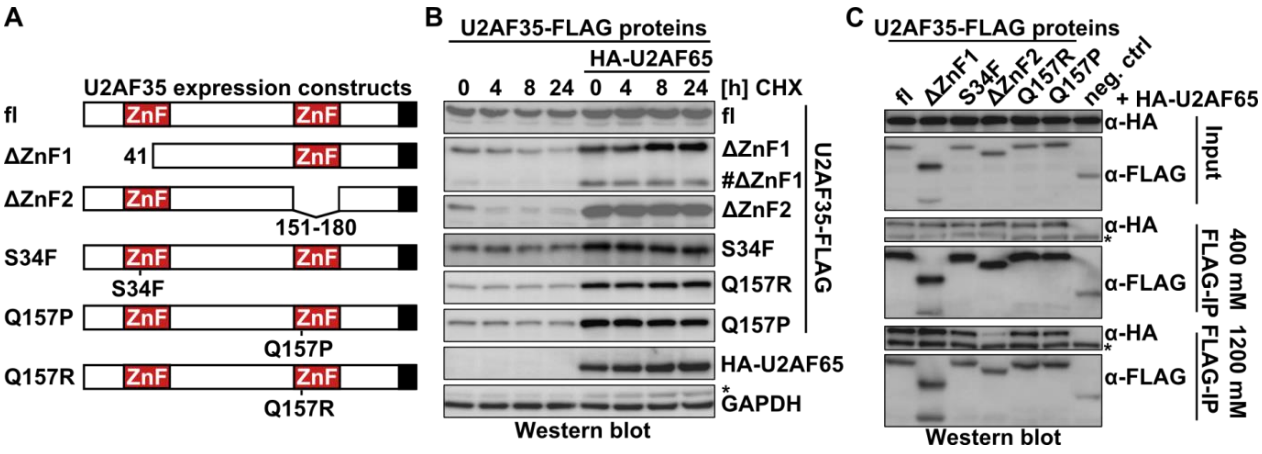


Figure 2.1.1: U2AF35 ZnF2 is important for protein stability and stable U2AF heterodimer formation. **(A)** Schematic depiction of U2AF35 expression constructs with deletion of either of the two ZnFs or disease-associated mutations therein. The C-terminal FLAG-tag is shown in black. **(B)** Western blot analysis to investigate the protein stability of U2AF35-mutants shown in **(A)** in the presence or absence of U2AF65. Cells were treated with cycloheximide for the indicated times prior lysis. GAPDH was used as loading control. **(C)** Co-immunoprecipitation of U2AF65 with U2AF35-mutants in the presence of 400 mM or 1200 mM NaCl.

To analyze the impact of the ZnF domains on splicing regulation, we knocked down endogenous U2AF35, substituted the expression with ZnF-deleted U2AF35 and analyzed alternative splicing of known U2AF35-dependent cassette exons. For all targets investigated, rescue of splicing could not be achieved through U2AF35 lacking either of the two ZnF domains, indicating that both ZnFs are indispensable for splicing regulation *(Figure 2.1.2)*. Full length U2AF26 exhibited similar rescue abilities as full length U2AF35 and no rescue was seen for ZnF-deleted mutants indicating that U2AF26 can functionally substitute for U2AF35 *in vivo* and that splicing regulation is also mediated through its ZnF domains.

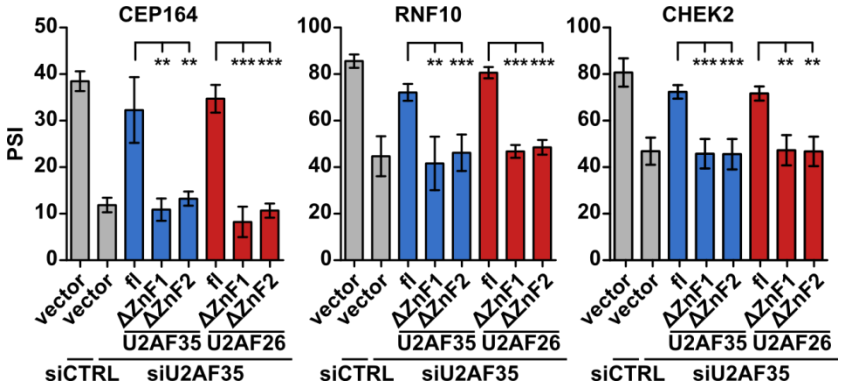


Figure 2.1.2: U2AF35 and U2AF26 ZnFs are indispensable for splicing regulation. Quantifications of splicing sensitive RT-PCRs for U2AF35-dependent cassette exons. U2AF35-specific siRNA was used to deplete endogenous U2AF35 in HEK293T cells followed by overexpression of siRNA-resistant U2AF35 or U2AF26 full length or ZnF-deleted variants. Shown is the mean percent spliced in (PSI) value ±SD. n=4; **p<0.01; ***p<0.001.

To verify the significance of the second ZnF on U2AF35 and U2AF26 function, we repeated the analysis with a naturally occurring splice variant of U2AF26 that lacks the largest part of the second ZnF through skipping of exon 7 (U2AF26ΔE7). Basal expression and protein stability, as well as the interaction with U2AF65 under stringent conditions is reduced for the U2AF26ΔE7 variant and it has also lost its ability to regulate splicing of U2AF35-dependent exons *(Figure 2.1.3)*.

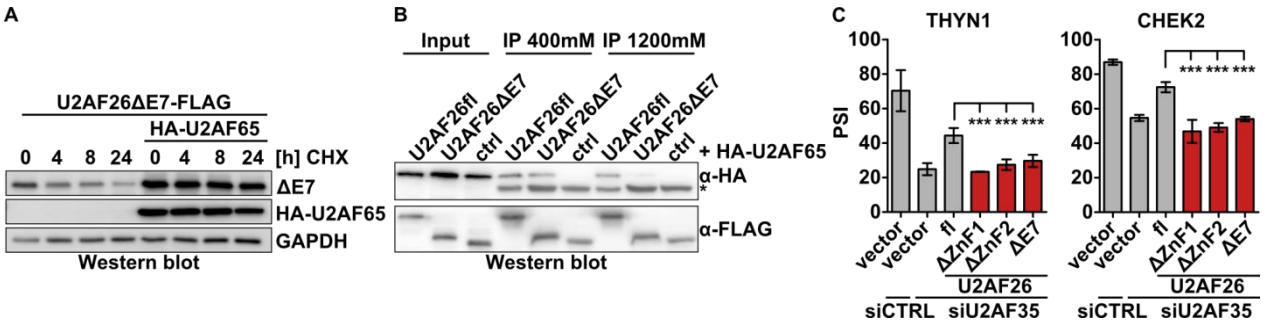


Figure 2.1.3: The naturally occurring U2AF26ΔE7 behaves similar as the ZnF2 deletion mutant. **(A)** Western blot analysis to investigate the protein stability of U2AF26ΔE7. **(B)** Co-immunoprecipitation of U2AF65 with U2AF26 and U2AF26ΔE7 in the presence of 400 mM or 1200 mM NaCl. **(C)** Knockdown and complementation assay as described in Figure 2.1.2. (mean PSI ± SD). n=4 ; ***p<0,001.

2.2 A variant of splicing factor U2AF26 impacts on translation and is induced during T cell activation

Refers to**:**

Herdt, O., Preußner, M., Heyd, F. (2018). The zinc finger domains in U2AF35 and U2AF26 control stability, localization, interaction with U2AF65 and play a role in translation. NAR, *in preparation for re-submission.*

For the U2AF35-related splicing factor U2AF26 two protein variants deviating from the full length U2AF26 were identified in different cell types. The first variant lacks exon 6 and 7 (U2AF26ΔE67) which leads to a shifted reading frame allowing translation far into the supposed 3'UTR of U2AF26. U2AF26ΔE67 thereby possesses a new C-terminus and, in contrast to the nuclear U2AF26fl, localizes into the cytoplasm where it functions in the regulation of the circadian clock (Preußner et al., 2014). In the second U2AF26 variant, skipping of exon 7 induces the exclusion of the largest part of the second ZnF and the downstream nuclear localization signal leading as well to a cytoplasmic localization of the U2AF26ΔE7 variant (Heyd et al., 2008). So far, the function of cytoplasmic U2AF26ΔE7 remained unidentified. Interestingly, the expression of different U2AF26 proteins seems to be regulated during mouse T cell activation. These protein variants could reflect alternative spliced variants of U2AF26 which were uncharacterized so far (Heyd et al., 2006).

In this project we aimed to identify the function of cytoplasmic U2AF26ΔE7. As splicing occurs in the nucleus and as U2AF26ΔE7 is incapable of rescuing U2AF35-dependent splicing through disruption of the ZnF2 seen in the chapter above, U2AF26ΔE7 is presumably not implicated in splicing regulation. Since U2AF26ΔE7-associated mRNAs in the cytoplasm are not known yet, we recruited U2AF26ΔE7 to a model RNA to investigate the potential regulatory function of cytoplasmic U2AF26ΔE7. We used a firefly luciferase reporter gene with MS2 loops in its 5' and 3' untranslated region (UTR) in forward and reverse orientation. Tethering MS2-tagged U2AF26ΔE7 to MS2-loops forward orientated in the 5'UTR led to increased luciferase expression *(Figure 2.2.1 A, B)* whereas binding to the 3'UTR had no effect. Quantitative RT-PCR revealed that the increased luciferase protein expression is not due to increased mRNA stability or increased export as the total and cytoplasmic amount of luciferase mRNA was unaffected by MS2-U2AF26ΔE7 expression *(Figure 2.2.1 C)*. These results indicate that

U2AF26ΔE7 increases the translation of a model RNA when recruited to the 5'UTR. Using the luciferase reporter assay and different truncations of MS2-U2AF26ΔE7 we assessed that the N-terminus of U2AF26ΔE7 containing the ZnF1 and the RNP2 regulates the translation of the luciferase reporter (appended data).

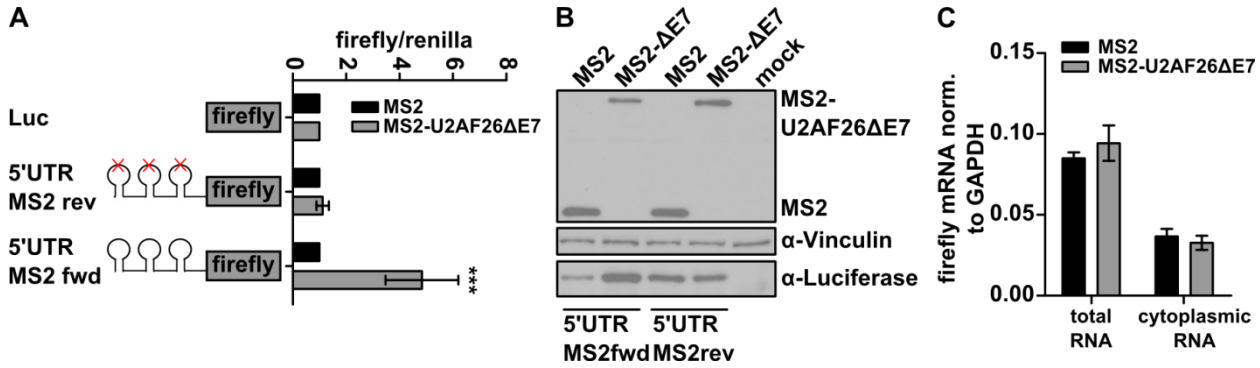


Figure 2.2.1: U2AF26ΔE7 regulates translation in a MS2 tethering assay. (A) Scheme of firefly luciferase reporter genes with MS2-protein binding loops in its 5'UTR in forward (fwd) and reverse (rev) orientation (left side). Expression of firefly luciferase in the presence of MS2-protein or MS2-tagged U2AF26ΔE7 normalized to a cotransfected renilla control and to MS2-protein alone signal (right side). Shown is the mean ± SD. n=9, one sample t-test, ***p<0.001. **(B)** Western blot analysis of the luciferase assay shown in **(A)**. MS2-protein and MS2-tagged U2AF26ΔE7 were detected with a α-FLAG antibody and vinculin was used as loading control. **(C)** Quantitative RT-PCR for firefly luciferase mRNA expression normalized to GAPDH of total and cytoplasmic RNA. Prior RNA extraction, cells were transfected with MS2-protein or MS2-tagged U2AF26ΔE7 and 5'UTR MS2fwd luciferase reporter.

In a previous study, several U2AF26 protein species were detected in lysates of activated mouse T cells (Heyd et al., 2006). As T cells undergo many alternative splicing changes upon activation (Martinez et al., 2012), we asked whether the different U2AF26 protein species might originate from alternative splicing of U2AF26 during T cell activation. For that, we purified primary mouse T cells, activated them for 48 hours with different stimuli and analyzed the expression of U2AF26ΔE67 and ΔE7 splicing variants by splicing-sensitive RT-PCR. We observed strongly increased expression of the U2AF26ΔE7 variant upon T cell activation

whereas the expression of U2AF26ΔE67 was just mildly affected, indeed indicating a regulation of U2AF26 splicing during T cell activation *(Figure 2.2.2 A,B).*

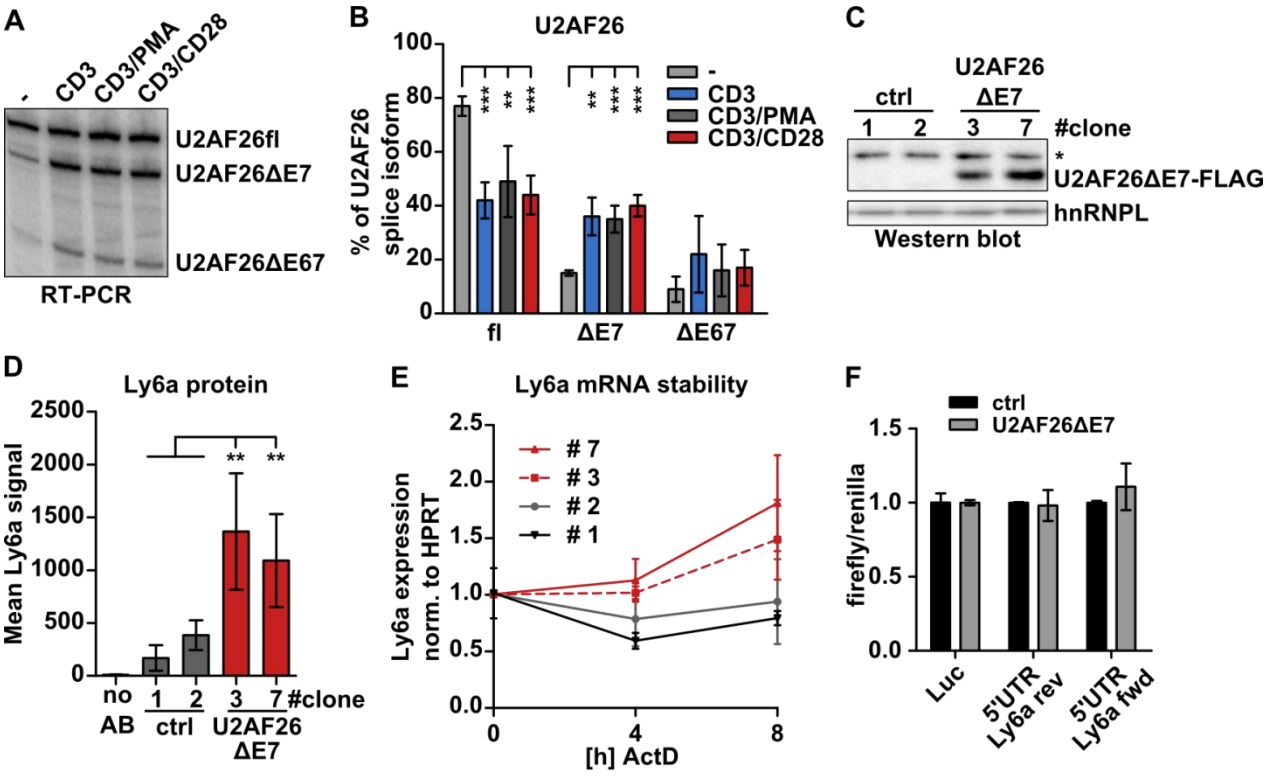


Figure 2.2.2: U2AF26ΔE7 increases Ly6a expression in mouse T cells. **(A)** Splicing sensitive RT-PCR for U2AF26 exons 4-8 in primary mouse T cells treated for 48h with CD3 alone, CD3/CD28 or CD3/PMA. **(B)** Quantification of U2AF26 splice isoforms shown in **(A)**. n=4; **p<0.01; ***p<0.001. **(C)** Western blot analysis of mouse EL4 T cells stably transduced with FLAG-tagged U2AF26ΔE7. hnRNPL was used as loading control. *Marks an unspecific band. **(D)** Surface expression of Ly6a on stable cells shown in **(B)**. Shown is the mean Ly6a signal intensity ± SD. n≥3; **p<0.01. **(E)** Ly6a mRNA stability in EL4 stable clones treated with Actinomycin D for 4 and 8 hours. Expression was normalized to HPRT (mean ±SD). #1: n=2; #2/3/7: n=4. **(F)** Expression of firefly reporter genes with the Ly6a 5'UTRs in forward and reverse orientation in HEK293T cells in the presence of U2AF26ΔE7 or a control protein. n=2.

Based on the results of the MS2-tethering assay, we aimed to identify endogenous targets of U2AF26ΔE7 translational regulation. To uncouple the process of T cell activation, which is

accompanied by a plethora of changes from the increased expression of the U2AF26ΔE7 variant, we generated mouse T cells stably expressing FLAG-tagged U2AF26ΔE7 *(Figure 2.2.2 C)*. In a targeted approach, we assessed the expression of seven known T cell surface markers on the surface of U2AF26ΔE7 positive T cells by flow cytometry. For the surface marker Ly6a we observed a significant increase in protein abundance on the surface of U2AF26ΔE7 expressing cells *(Figure 2.2.2 D)*. Quantitative RT-PCR revealed a slight increase in Ly6a mRNA abundance and stability *(Figure 2.2.2 E)*. Based on the observations for the model RNA these results hint to a, at least partially, translational regulation of the Ly6a protein expression through U2AF26ΔE7. To analyze whether the regulation is due to binding of U2AF26ΔE7 to the 5'UTR of the Ly6a mRNA we cloned reporter gene constructs with the 5'UTR of Ly6a upstream of the luciferase gene. We did not observe increased luciferase expression in the presence of U2AF26ΔE7 through placing the 5'UTR upstream of the luciferase gene *(Figure 2.2.2 F)* indicating that the Ly6a expression regulation through U2AF26ΔE7 might be accomplished by recognition of other elements in the Ly6a mRNA. Furthermore, the Ly6a surface expression could be regulated on the level of mRNA stability as we see a slightly increased Ly6a mRNA stability in the presence of U2AF26ΔE7 (*Figure 2.2.2 E*).

2.3 The cancer-associated U2AF35 470A>G (Q157R) mutation creates an in-frame alternative 5′ splice site that impacts splicing regulation in Q157R patients

Refers to:

Herdt, O.*, Neumann, A.*, Timmermann, B., Heyd, F. (2017). The cancer-associated U2AF35 470A>G (Q157R) mutation creates an in-frame alternative 5′ splice site that impacts splicing regulation in Q157R patients. ***RNA*** 23, 1796-1806.

* These authors contributed equally to this work.

Several recent studies described mutations in both ZnFs of U2AF35 to be associated with diseases of the blood system like myelodysplastic syndromes (MDS), different types of leukemia and solid tumors like lung adenocarcinoma (Imielinski et al., 2012; Yoshida et al., 2011). In U2AF35, the amino acid substitutions S34F and S34Y in the ZnF1 and Q157R and Q157P in the ZnF2 are associated with disease formation. Functional studies predominantly focused on the most frequent mutation S34F and identified aberrant splicing through a changed 3’ ss binding specificity of U2AF35S34F (Brooks et al., 2014; Ilagan et al., 2014; Okeyo-Owuor et al., 2015; Przychodzen et al., 2013). As mutations in the second ZnF are poorer understood and the results above show that especially ZnF2 impacts on U2AF35 function we concentrated this study on the disease-associated mutations in the second ZnF.

The mutations in the second ZnF both affect the glutamine 157 which is substituted to either an arginine (Q157R) or a proline (Q157P). On DNA level, the adenine at position 470 is mutated either to a guanine (Q157R) or to a cytosine (Q157P). Interestingly, the mutated nucleotide is followed by a GT-nucleotide that might act as a potential alternative 5’ ss whose strength could be affected by the mutations. Calculation of the splice site scores for the alternative 5’ ss revealed an increased score when the Q157R (c.470A>G) mutation is introduced whereas the score is only mildly affected by the Q157P (c.470A>C) mutation *(Figure 2.3.1 A)*. Splicing sensitive RT-PCRs of mutated U2AF35 minigenes and sequencing analysis confirmed the use of the alternative 5’ ss only when the Q157R mutation is introduced *(Figure 2.3.1 B)*. Usage of the alternative 5’ ss leads to the skipping of 12 nucleotides encoding four amino acids that are missing in the Q157R mutant additionally to the Q157R substitution *(Figure 2.3.1 C)*. We therefore termed this splice variant Q157Rdel.

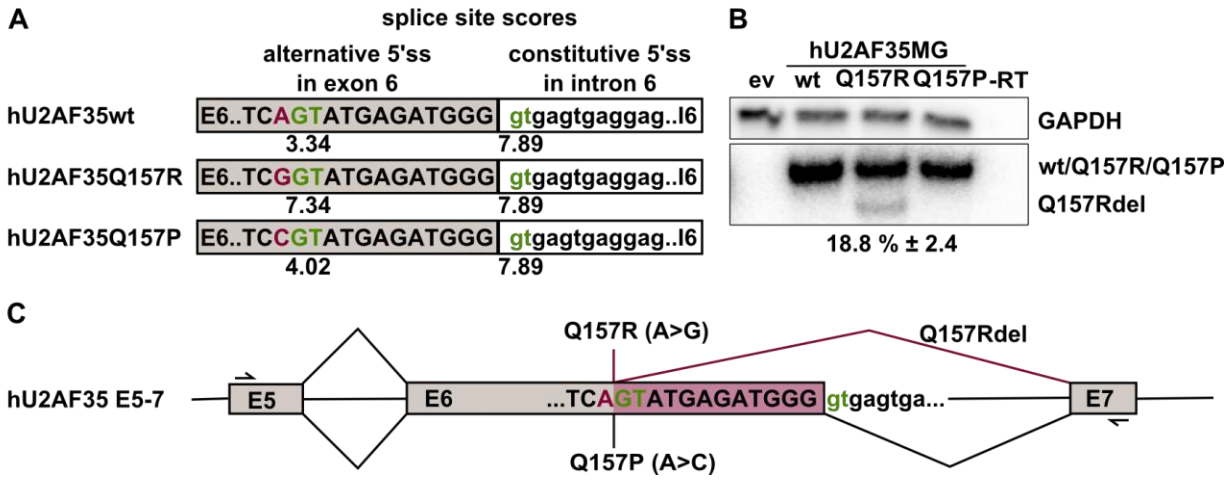


***Figure 2.3.1: The U2AF35 Q157R mutation creates an alternative 5' ss. (*A)** Nucleotide sequence of the human U2AF35 exon6-intron6 junction. The nucleotide position mutated in MDS-patients is shown in red. The alternative and constitutive 5' ss are labeled in green and their splicing scores are indicated below. **(B)** Splicing sensitive RT-PCR and quantification of U2AF35 minigenes with MDS-associated mutations. GAPDH was used as loading control. **(C)** Schematic representation of the splicing patterns of the U2AF35 E5-7 minigenes analyzed in **(B)**.

To assess target exons of the Q157R and the Q157P mutants and to analyze a potential role of the Q157Rdel mutant on splicing regulation, we overexpressed the three Q157 mutants and U2AF35wt in U2AF35-depleted cells and performed deep sequencing analysis of total RNA *(Figure 2.3.2 A, B*; bioinformatic analysis performed by Alexander Neumann). This approach revealed that each Q157 mutant rescues the splicing of a specific subset of U2AF35-dependent cassette exons. Q157R rescued 66%, Q157P 53% and Q157Rdel 44% of all U2AF35-dependent exons. Interestingly, the Q157Rdel has the largest subset of uniquely rescued target exons and a rather small overlap with Q157R or Q157P responsive exons, whereas the largest overlap is obtained for Q157R and Q157P *(Figure 2.3.2 C)*. We went on with defining the 3' ss signatures of the three Q157 mutants by analyzing the 3' ss of the uniquely rescued exons. Interestingly, as suggested before, the sequence specificity for the mutants in the second ZnF is accomplished by the +1 nucleotide directly downstream of the invariant AG dinucleotide (Ilagan et al., 2014). For the Q157R mutant the +1 position is strongly enriched for cytosine and for the Q157P a strong enrichment for guanine was observed. The most remarkable enrichment was observed for Q157Rdel where 95% of the unique targets have an adenine at the +1 position *(Figure 2.3.2 C)*.

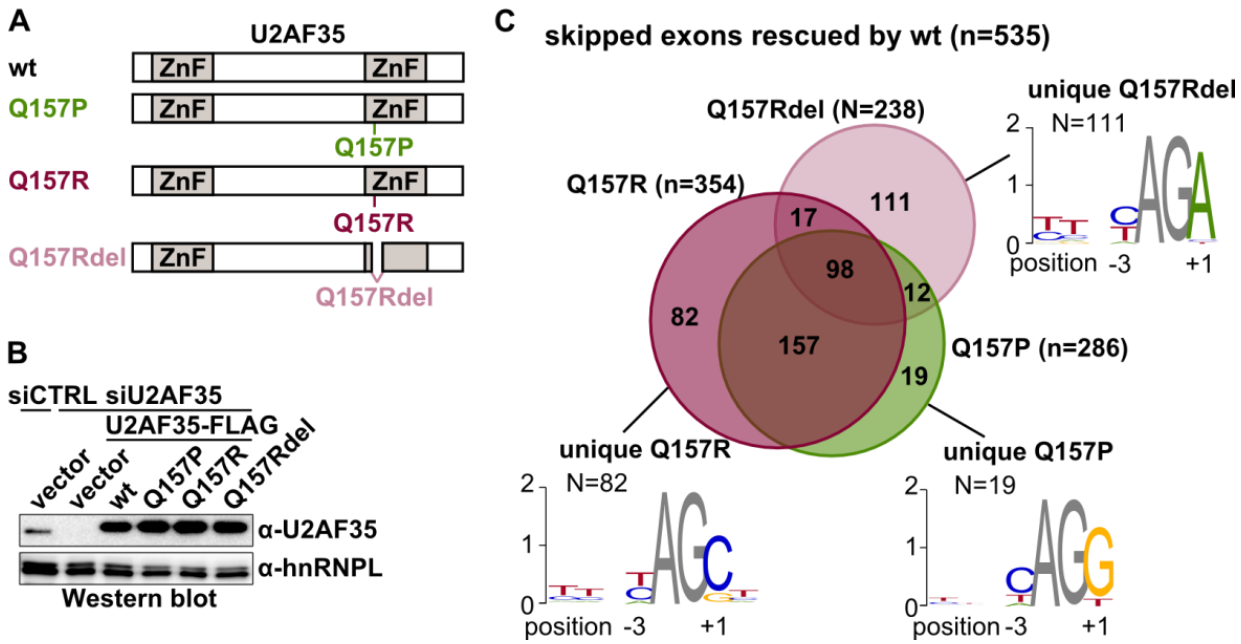


Figure 2.3.2: Q157P, Q157R and Q157Rdel regulate different splicing events through distinct 3' ss preferences. **(A)** Scheme of U2AF35 expression constructs with disease-associated mutations in the ZnF2. **(B)** Western blot analysis of HEK293T cells siRNA-depleted for endogenous U2AF35 and substituted with expression constructs of U2AF35 shown in **(A)**. **(C)** Identification of Q157P, Q157R and Q157Rdel responsive U2AF35-dependent exons in RNA deep sequencing analysis. Consensus motifs of uniquely targeted exons were used to analyze binding preferences of the U2AF35 Q157 mutants.

So far, one study addressed the impact of Q157 mutations on splicing in patients. However, the study combined S34 and Q157 mutants due to insufficient patient number and thereby missed Q157R- and Q157P-specific targets (Qiu et al., 2016). We therefore aimed to find Q157R and Q157P patients and subsequently to analyze splicing defects observed *in vivo* caused by either of the two mutations. We exploited publicly available RNA sequencing data and identified three Q157R and two Q157P patients in a cohort of 533 AML patients (bioinformatic analysis performed by Alexander Neumann). We compared the splicing patterns of those U2AF35 mutant patients to ten AML patients without U2AF35 mutations. This analysis identified 912 cassette exons that are alternatively spliced in the Q157R or Q157P patients. Interestingly, only 20% of those exons are similarly affected by Q157R and Q157P indicating that the two mutants are not equivalent but rather cause missplicing of a distinct set of exons. Comparison of the differentially spliced cassette exons in Q157R/Q157P patients to U2AF35-

dependent exons in cell culture revealed 70 overlapping exons *(Figure 2.3.3 A).* Plotting of the differences in percent spliced in values (dPSI) between U2AF35 wt and U2AF35 mutants in cells against the dPSIs in patients revealed reasonable correlation coefficients for the Q157R and Q157P mutants *(Figure 2.3.3 B).* Using splicing-sensitive RT-PCR, we could further validate the splicing targets identified in the analysis of Q157R and Q157P patients.

Lastly, we wanted to see whether the Q157Rdel variant is present in Q157R patients and might contribute to splicing patterns observed in the patients. To see whether the c.470A>G mutations also induces alternative splicing of U2AF35 in patients, we searched for U2AF35 transcripts carrying the Q157Rdel, Q157R or Q157P sequence in the wt and Q157 mutated AML-patients described above. As a consequence of the heterozygous type of the U2AF35 mutations, slightly below 50% of the U2AF35 transcripts carried the Q157R or the Q157P mutation in the respective Q157 mutated patients. Furthermore, we indeed found expression of the Q157Rdel variant in all three Q157R patients indicating that the alternative 5' ss is also used *in vivo* although it is used less frequent than in the minigene context *(Figure 2.3.1 A and 2.3.3 C).* As expected from the minigene experiments, no mRNA coding for the Q157Rdel variant was detected in wt and Q157P patients. To investigate a potential contribution of the Q157Rdel variant on splicing in Q157R patients, we had a closer look on the 70 targets overlapping between cell culture and patients *(Figure 2.3.3 A).* Interestingly, of the 70 overlapping events splicing of sixteen target exons in Q157R patients is more closely simulated by the Q157Rdel variant than by the Q157R point mutant in cell culture suggesting a role of the Q157Rdel variant in splicing regulation in Q157R patients *(Figure 2.3.3D).*

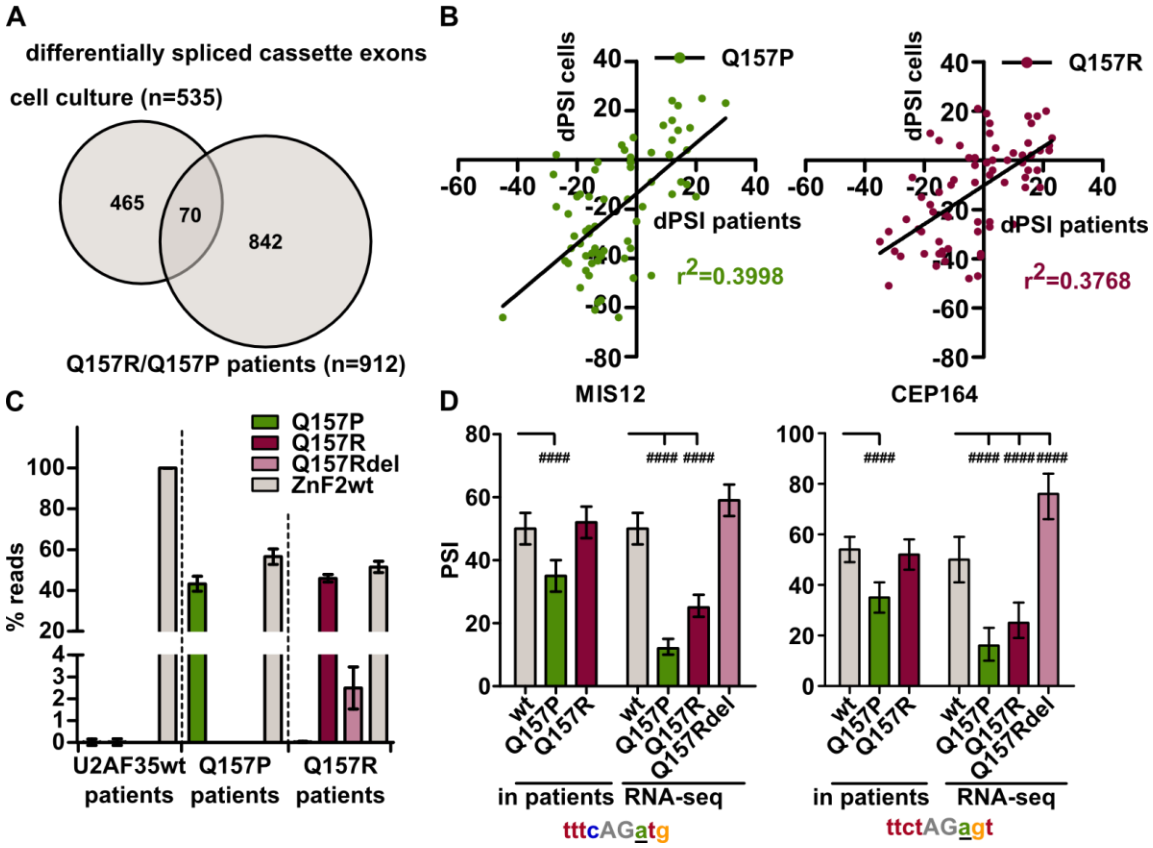


Figure 2.3.3: Splicing regulation in Q157R and Q157P patients. **(A)** Overlap of differentially spliced cassette exons in cell culture and Q157R/Q157P patients. **(B)** Correlation of percent spliced in (PSI) values of Q157P or Q157R in cell culture and in patients. Plotted are the dPSI values reflecting the difference of Q157P or Q157R to U2AF35 wt. **(C)** Identification of the Q157Rdel variant in AML patients. Depicted are the number of reads obtained for the three Q157 mutants and the U2AF35 wt sequence in Q157P, Q157R and U2AF35 wt patients. **(D)** Examples of differentially spliced exons in RNA sequencing data of the cell culture approach and the AML-patients. The corresponding 3' ss are shown below the graphs. (###) Bayes factor > 100.

3. DISSCUSSION

3.1 Functional implications of mouse U2AF35 ZnFs

Although the function of U2AF35 in binding the AG di-nucleotide in the 3' ss is known for more than 15 years, the ZnFs were just recently characterized as the domains responsible for RNA binding. Yoshida *et al*. solved the structure of the *S. pombe* U2AF35 homolog U2AF23 and performed *in vitro* binding studies to conclude that the ZnFs, rather than the UHM domain, is responsible for RNA binding (Yoshida et al., 2015). In this study, we aimed to verify the role of the ZnFs for mammalian U2AF35. For that purpose, we deleted either the first or the second ZnF of U2AF35 and analyzed their impact on protein stability, protein-protein interactions and splicing regulation. Deletion of the second ZnF of mouse U2AF35 led to a dramatic decrease in the basal expression of U2AF35 and to a strongly reduced protein-half life, when compared to the wt and the ZnF1-lacking U2AF35. Co-expression of the hetero-dimerization partner U2AF65 rescued this phenotype which is consistent with *in vitro* data showing correct folding of the UHM of U2AF35 only upon interaction with U2AF65 (Kellenberger et al., 2002). The results indicate that in the absence of U2AF65 the second ZnF contributes to the proper folding of U2AF35. In the crystal structure of yeast U2AF35 both ZnF domains interact with each other on top of the UHM domain (Yoshida et al., 2015). Deletion of either of the ZnFs would disrupt this intramolecular interaction and could therefore lead to destabilization of the protein, however only the deletion of ZnF2 has a strong impact on protein stability indicating that the ZnFs are not equivalent.

In co-immunoprecipitation assays, we further show that the ZnF2 is required for a stable interaction with U2AF65, as the interaction is disrupted under high salt conditions. Interestingly, the crystal structure of full length yeast U2AF23 identified an additional interaction surface between both U2AF subunits besides the known "tongue to groove" interaction between the UHM domain of U2AF35 and the ULM domain of U2AF65 of the minimal human U2AF heterodimer (Kielkopf et al., 2001). In U2AF23, the additional interaction surface to U2AF65 is mediated through a helix in the C-terminus of U2AF23 directly following the ZnF2. The C-terminus is poorly conserved between yeast and mouse U2AF1 on amino acid level (13% identity; 23% similarity). Still, the sequence following the ZnF2 in mouse U2AF35 is predicted

to form a helix, however, the helix is slightly shortened compared to yeast U2AF23 *(Figure 3.1)*. Removal of the ZnF2 in U2AF35 probably changes the orientation, location or folds of this helix and thereby reduces the interaction to U2AF65 under stringent conditions. These observations indicate that, also for the mouse U2AF heterodimer, more than one protein region mediates the interaction. Our results are further supported by the observation that the expression of recombinant yeast U2AF23 that lacks this C-terminal helix is reduced whereas deletion of the ZnF1 had no impact on expression (Yoshida et al., 2015). Certainly, the salt conditions used in the immunoprecipitation experiments do not reflect physiological salt conditions. However, any *in vitro* approach used to examine protein-protein interactions is artificial. The intracellular environment represents an extremely crowded milieu which was shown to affect protein folding, structure and also protein-protein and protein-nucleic acid interactions that cannot be recapitulated *in vitro* (reviewed in (Kuznetsova et al., 2014)). We therefore assume that the helix following the ZnF2 of U2AF35 might also be important for stable complex formation of mouse U2AF in cells, however further evidences are needed.

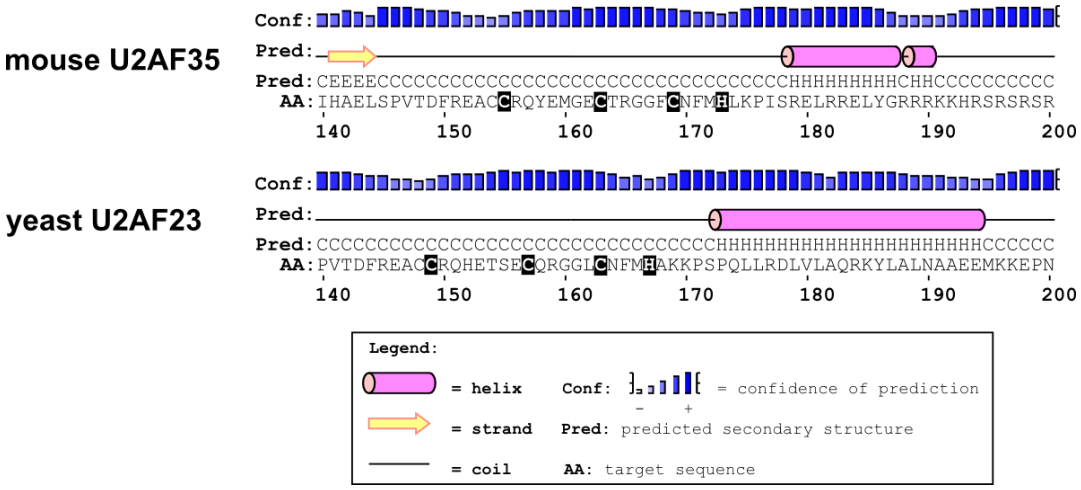


Figure 3.1: Secondary structure prediction of mouse U2AF35 and yeast U2AF23. Secondary structure prediction is shown for the ZnF2 and the downstream region of U2AF35 and U2AF23. Zinc-coordinating residues are highlighted with black boxes.

To investigate the importance of the ZnF domains of yeast U2AF35 on RNA binding, Yoshida *et al*. performed *in vitro* RNA binding assays with recombinant U2AF35 mutated in either of the two ZnFs. Deletion of the first ZnF abolished RNA binding, whereas point mutations had diverse effects ranging from improvement to loss of binding and to changed sequence selectivity (Yoshida et al., 2015). We investigated the importance of the ZnF domains of mouse U2AF35 on splicing in a cell-based assay. We reconstituted U2AF35-depleted cells with ZnF-deleted U2AF35 and analyzed the alternative splicing of known U2AF35-dependent exons. Deletion of either of the two ZnFs led to a complete loss of rescue of U2AF35-dependent splicing defects showing that both ZnFs are indispensable for splicing regulation. We obtained similar results in the presence of co-expressed U2AF65 showing that the rescue-failure is not due to reduced protein levels through deletion of the ZnFs. Based on the results of yeast U2AF23, the loss in splicing regulation is likely to be caused by a loss of RNA binding. Thus, with this cell culture based study we could confirm purely structural and *in vitro* based conclusions made from yeast U2AF23 for the mouse U2AF35. The correlation of the yeast *in vitro* and our mouse cell culture-based data is supported by a high degree of conservation between the U2AF35-homologs of the respective organisms. The N-terminal region of yeast and mouse U2AF1 is 60% conserved, whereas the ZnF1 and ZnF2 have an even higher conservation of 83% and 76%, respectively. The observations made for mouse U2AF35 can most likely be applied for human U2AF35 as the sequence similarity is 98%.

3.2 U2AF26 can functionally substitute for U2AF35

U2AF26 is a highly conserved homolog of U2AF35 for which previous *in vitro* data indicate functional similarities to U2AF35 (Shepard et al., 2002). We included U2AF26 in our cell-based assays discussed in *Chapter 3.1* to verify and to extend the previous observations. For all conducted experiments, we obtained similar results for U2AF26 as for U2AF35 indicating that U2AF26 can functionally substitute U2AF35 *in vivo*. U2AF26 interacts with U2AF65 which was observed before in pull down assays with purified recombinant proteins (Shepard et al., 2002). Furthermore, the second ZnF of U2AF26 has similar functions in protein and heterodimer stability as in U2AF35, validating our observations for U2AF35. Replacement of endogenous U2AF35 with U2AF26 in HEK293T cells revealed that full length U2AF26 can restore splicing

defects induced by U2AF35-depletion. And, as for U2AF35, deletion of either the first or second ZnF of U2AF26 abolished the rescue abilities. This is consistent with *in vitro* splicing assays performed with one model pre-mRNA showing that U2AF26 can functionally substitute for U2AF35 in constitutive and enhancer-dependent splicing (Shepard et al., 2002).

As the ZnF-deleted variants of U2AF35 and U2AF26 reflect an artificial situation, we extended our analysis to a naturally occurring splice variant of U2AF26 in which exon 7 (U2AF26ΔE7) is excluded from the mRNA. Exon 7 encodes the largest part of the ZnF2 and the nuclear localization signal of U2AF26 leading to a cytoplasmic localization of this variant (Heyd et al., 2008). U2AF26ΔE7 had similar properties as the ZnF2-deleted U2AF26 and U2AF35. Compared to the full ZnF2-deletion, the basal expression and the protein half-life of U2AF26ΔE7 is slightly reduced but again stabilized by the co-expression of U2AF65. Although U2AF26ΔE7 and U2AF65 are found in different cellular compartments, we see interaction of both proteins which is lost under stringent salt conditions as for the full ZnF2 deletion. U2AF65 is known to shuttle between nucleus and cytoplasm (Gama-Carvalho et al., 2001) and thus has the chance to interact with cytoplasmic U2AF26ΔE7. Another explanation is that the proteins of the cytoplasmic and nuclear compartment are mixed together during cell lysis and interact during incubation for the co-immunoprecipitation. Similar as the ZnF2-deleted U2AF26, the U2AF26ΔE7 variant was not able to rescue U2AF35-dependent splicing defects. However, we cannot distinguish whether this is due to its localization into the cytoplasm or due lack of the second ZnF.

SELEX-experiments with recombinant U2AF35 or U2AF26 together with U2AF65 resulted in the enrichment of the consensus 3’ ss although with slight differences at the +1 position (Jeremiah Brian Shepard, 2004). For U2AF35, the author observed a slight enrichment of a guanine and uracil at the +1 position and for U2AF26 of a cytosine. However in cross-linking experiments, aimed to validate the SELEX experiments, U2AF26 appeared to prefer uracil at the +1 position (Jeremiah Brian Shepard, 2004). In our knockdown and complementation assay, alternative splicing for seven out of eight analyzed U2AF35-dependent exons were rescued equally through U2AF26 and U2AF35. For one target (RIPK2), we observed stronger rescue abilities of U2AF26 compared to U2AF35. Interestingly, this is the only target included in our study with a uracil at the +1 position matching the *in vitro* crosslink-based determination of the binding motif identified by Shepard (Jeremiah Brian Shepard, 2004).

Analysis of the binding preferences of hematopoietic diseases-associated U2AF35 mutations revealed that the first ZnF recognizes the nucleotide preceding the AG dinucleotide in the 3' ss whereas the second ZnF binds to the AG-following nucleotide. Interestingly, the first ZnF of mouse U2AF35 and U2AF26 is fully conserved, whereas the second ZnF has one non-conservative (alanine 153 to serine) and one conservative (lysine 175 to arginine) amino acid exchange (*Figure 1.8*). Both positions are highly conserved in U2AF35 within different species and were shown to form hydrogen bonds with the UHM of U2AF35 (Yoshida et al., 2015). These results indicate that there might be slight differences in the binding specificity of U2AF26 and U2AF35 defined by the AG-following nucleotide and the second ZnF. Interestingly, it was shown that splicing of the transmembrane tyrosine phosphatase CD45 is specifically regulated through U2AF26 and not through U2AF35 suggesting a distinct function of both proteins in splicing regulation (Heyd et al., 2006). However, RNA sequencing of brains of U2AF26 wt and knock-out mice failed to identify further U2AF26-dependent splicing events (Preußner et al., 2014). Further genome-wide analyses are needed to provide convincing evidence for a differential binding of U2AF26 and U2AF35 and for the identification of further U2AF26-specific targets.

3.3 Function of cytoplasmic U2AF26ΔE7

Alternative splicing of U2AF26 was described in different mouse cell lines however the function of the cytoplasmic U2AF26ΔE7 variant was not addressed so far (Heyd et al., 2008; Preußner et al., 2014, 2017). We aimed to analyze a potential role of U2AF26ΔE7 on cytoplasmic RNAs by tethering U2AF26ΔE7 to the UTRs of a model mRNA. The analysis revealed that U2AF26ΔE7 enhances the translation of a model mRNA when recruited to the 5'UTR and that this regulation is accomplished by the N-terminus of U2AF26ΔE7 *(Figure 3.2)*. There are several examples known where splicing variants of RNA-binding proteins change localization from nucleus to the cytoplasm and affect processing of cytoplasmic mRNAs. One example is the alternative splicing regulator muscleblind-like 1 (MBNL1) that plays a crucial role in skeletal and heart muscles development and is itself alternatively spliced during development (Lin et al., 2006; Terenzi and Ladd, 2010). The MBNL1 variant lacking exon 5 localizes, in contrast to solely nuclear MBNL1 full length protein, to the cytoplasm as well as to the nucleus (Lin et al., 2006; Terenzi and Ladd, 2010). To date, the function of cytoplasmic MBNL1ΔE5 is

unknown. However, its paralog MBNL3 that lacks the region encoded by MBNL1 exon 5 is as well localized in the nucleus and the cytoplasm. The function of cytoplasmic MBNL3 was determined, showing that MBNL3 associates with translated mRNAs in polysomes and binds to 3'UTRs of genes involved in cell growth and proliferation (Poulos et al., 2013). Another example is the RNA-binding protein Quaking (Qk) in which alternative splicing of the C-terminus localizes the Qk6 and Qk7 splice variants in the cytoplasm and the nucleus, whereas the Qk5 splice variant localizes just into the nucleus (Fagg et al., 2017). The study further shows that cytoplasmic Qk6 promotes its own translation and represses Qk5 translation through binding to the 3'UTR; however the mechanistic details are still missing.

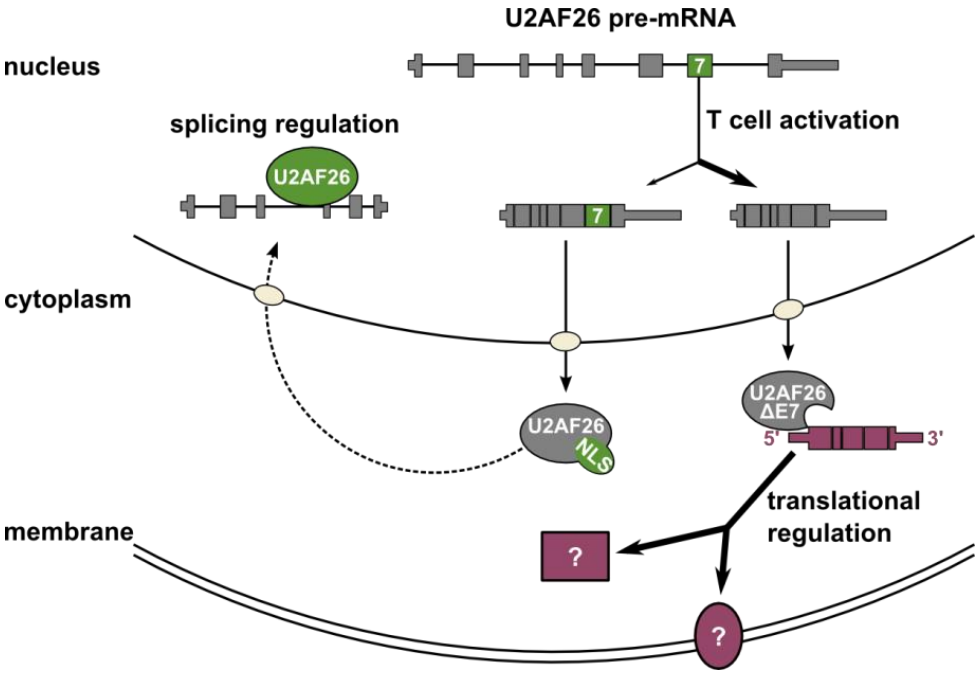


Figure 3.2: Regulation of mRNA processing through U2AF26 variants. U2AF26 pre-mRNA is alternatively spliced upon T cell activation. Full length U2AF26 is translated in the cytoplasm and imported into the nucleus where it binds to 3' ss of pre-mRNAs and regulates splicing. U2AF26ΔE7 remains in the cytoplasm due to absence of the NLS where it might bind to the 5'UTR of mRNAs and enhances their translation.

By using the MS2-tethering assay, we are not able to determine a possible RNA-binding mechanism for U2AF26ΔE7 to cytoplasmic RNAs. As U2AF26ΔE7 lacks the largest part of the ZnF2, RNA binding might be accomplished by the ZnF1 alone. RNA binding of a single ZnF was shown for one ZnF domain of the tristetraprolin (TTP) protein (Michel et al., 2003), which, interestingly, has two CCCH-type ZnFs similar to U2AF26. TTP protein family members are characterized by two tandemly arranged ZnFs which bind to an AU-rich element with each ZnF recognizing a 4-mer (Hudson et al., 2004). Furthermore, binding could be mediated through the atypical RRM of U2AF26 which is unaffected in the U2AF26ΔE7 variant. Another possibility is an indirect binding of U2AF26ΔE7 to target RNA mediated through recruitment by other RNA-binding proteins. To expand and deepen the role of U2AF26ΔE7 on endogenous cytoplasmic mRNAs, it would be of large interest to analyze the transcriptome and the polysome-associated mRNAs in our U2AF26ΔE7-expressing T cells. To understand the mechanism by which U2AF26ΔE7 functions on cytoplasmic mRNAs, the protein interactome of U2AF26ΔE7 could be determined through co-immunoprecipitation followed by mass spectrometry or through a yeast two hybrid screen. An initial yeast two hybrid screen identified the poly(rC)-binding protein 2 (PCBP2), an RNA-binding protein with functions in translational regulation, as potential interaction partner of U2AF26ΔE7, however co-immunoprecipitation experiments could not confirm the interaction so far. Interestingly, human and drosophila U2AF65 was shown to associate with spliced and intronless mRNAs in the cytoplasm (Blanchette et al., 2004; Gama-Carvalho et al., 2006; Zolotukhin et al., 2002). It might be possible that U2AF65 mediates recruitment of U2AF26ΔE7 to cytoplasmic mRNAs.

The best-known RBP family that fulfills nuclear as well as cytoplasmic functions is the SR-protein family. Cytoplasmic roles of SR-proteins were shown in the regulation of mRNA stability and translation (reviewed in Twyffels et al., 2011). For example, SRSF1 was shown to bind to the 3'UTR of protein-kinase-C-interacting protein mRNA and to induce its degradation (Lemaire et al., 2002). Furthermore, SRSF1 and SRSF7 associate with monosomes or polysomes, which seems to be regulated by the phosphorylation level of the RS domain (Sanford et al., 2005). As U2AF26 contains some RS-repeats in its C-terminal domain that are not removed by skipping of exon 7 it will also be very interesting to analyze whether these repeats are phosphorylated and impact on translational regulation by U2AF26ΔE7. In SRSF1, the RRM2 is implicated in translational control probably through protein-protein interactions mediated by this

domain (Sanford et al., 2005). We obtained similar results for U2AF26ΔE7 when we used the MS2-reporter assay and truncated versions of MS2-U2AF26ΔE7. This analysis revealed that deletion of the RNP2 and the ZnF1 leads to the loss of translational activation of the reporter gene. To unravel the mechanism by which U2AF26ΔE7 might regulate translation when recruited to the 5'UTR of target mRNAs, it would be of interest to identify RRM2/ZnF1 interacting proteins.

We further aimed to identify endogenous targets of cytoplasmic U2AF26ΔE7. Interestingly, we discovered that U2AF26ΔE7 variant expression is upregulated during activation of primary mouse T cells. We therefore used immortalized mouse T cells to identify endogenous targets of cytoplasmic U2AF26ΔE7. Utilizing mouse T cells stably overexpressing U2AF26ΔE7, we observed an upregulated surface expression of the Ly6a protein upon U2AF26ΔE7 expression. However, further examinations revealed that the regulation is not obtained through the 5'UTR of Ly6a, as seen in the MS2-based tethering assay. These results indicate that the increase in Ly6a expression through U2AF26ΔE7 is regulated through a distinct mechanism. One possibility might be an increased Ly6a mRNA stability which is weakly observed in the Ly6a positive EL4 cells. In regard to the MS2-based reporter assay, it is also possible that U2AF26ΔE7 promotes translation only in the context of a structured RNA, e.g. by recruiting a RNA helicase. Furthermore, it might be that U2AF26ΔE7 binds to the gene body rather than to the 5'UTR of Ly6a. Our MS2-based reporter gene assays did not address recruitment of U2AF26ΔE7 to the gene body which could also promote translation of an mRNA.

Nevertheless, Ly6a (lymphocyte antigen 6 a) also known as Sca-1 (stem cell antigen-1) is still an interesting U2AF26ΔE7-target. Ly6a is a cell surface marker expressed on hematopoietic stem cells and on primary CD4-positive T cells where its expression is strongly upregulated upon activation (Holmes and Stanford, 2007). It was suggested that Ly6a has both a positive and negative role in T cell response upon antigen stimulation probably mediated through TCR signaling (Flood et al., 1990; Hanson et al., 2003; Henderson et al., 2002; Stanford et al., 1997). U2AF26fl was shown to attenuate T cell activation by promoting the formation of a less active CD45 protein isoform through alternative splicing regulation (Heyd et al., 2006). U2AF26ΔE7 might further affect the T cell response by increasing the expression of Ly6a or by potentially regulating other cell surface markers. The comprehensive identification of endogenous mRNA targets of U2AF26ΔE7 is desirable to elucidate the mechanism by which U2AF26ΔE7 functions

on cytoplasmic mRNAs and to link U2AF26ΔE7 expression during T cell activation to T cell function.

3.4 The U2AF35 470A>G mutation creates two distinct protein variants

Another focus of our study was the analysis of the cancer-associated mutations Q157R and Q157P in the second ZnF2 of U2AF35. We realized that mutating the adenine at position 470 to a guanine (c.470A>G), a mutation that codes for the Q157R substitution, induces the recognition of the mutation following GT as a 5' ss. Use of the alternative 5' ss leads to the skipping of 12 nucleotides, which encode for four amino acids within the ZnF2 of U2AF35 creating the Q157Rdel variant *(Figure 3.3)*. The four amino acids do not comprise any of the zinc complexing residues of the CCCH-type ZnF2 of U2AF35. However, the distance between the first and the second zinc complexing cysteine is shortened from seven to three amino acids. In mammals, there are around 60 proteins known that contain CCCH-type ZnF domains which mostly bind to RNA and are implicated in regulation of different steps of RNA metabolism. The spacing between the zinc coordinating cysteine and histidine residues in CCCH-ZnFs is very diverse with the most common type having a C-X_{7-9}-C-X_{4-6}-C-X_3-H arrangement (Fu and Blackshear, 2017). Though, the distance between the two first cysteines can range from five to twelve residues in humans and from four to 15 residues in plants, whereas the spacing between the other residues is very rigid (Fu and Blackshear, 2017; Wang et al., 2008). This diversity in the arrangement of CCCH-type ZnFs suggests that the ZnF2 in the Q157Rdel variant might still be functional. A structure prediction of the U2AF35 Q157Rdel variant revealed the loss of a short helix within the ZnF2 and a slight change in the distance and orientation of the zinc complexing residues but the overall structure of the Q157Rdel ZnF2 stays intact. In knock down and complementation assays, the Q157Rdel mutant rescued splicing of many U2AF35-dependent exons that largely differentiate from those rescued by the Q157R point mutant indicating that the ZnF2 tolerates the deletion of four amino acids, but possesses an altered binding specificity *(Figure 3.3)*.

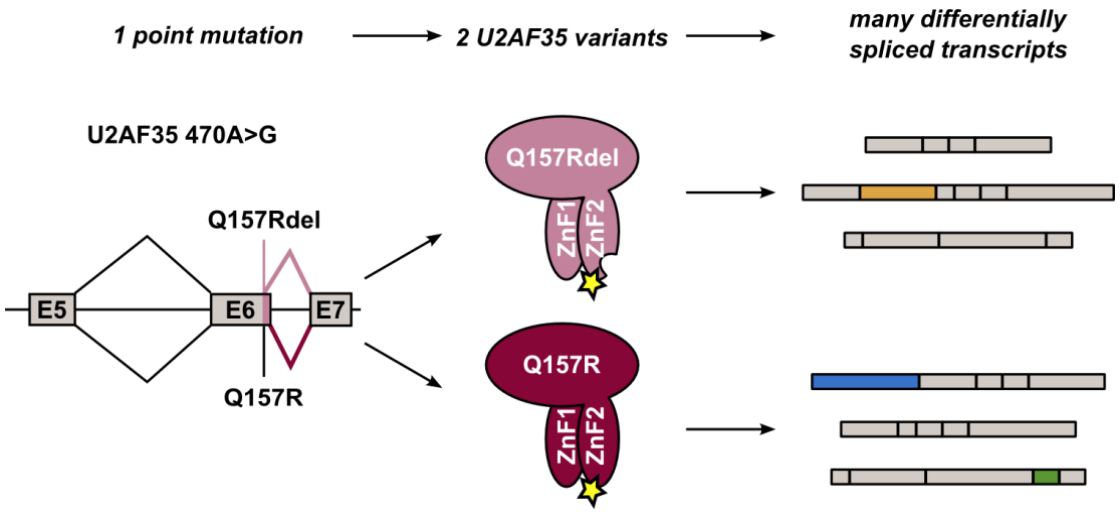


Figure 3.3: The c.470A>G mutation yields two U2AF35 proteins. Schematic depiction of the U2AF35 pre-mRNA splicing in the presence of the c.470A>G mutation in exon 6. Either the constitutive 5' ss is used resulting in the full length U2AF35 protein with the Q157R substitution (bottom) or the alternative 5' ss is used resulting in a deletion of four amino acids within the ZnF2 additionally to the Q157R substitution (top). Both U2AF35 mutants regulate splicing of different transcripts.

In our study, we show that the c.470A>G mutation yields two U2AF35 proteins. The first harbors the amino acid substitution Q157R and the second lacks, additionally to the substitution, four amino acids through the use of an alternative 5' ss (Q157Rdel). Often mutations are solely classified based on the genomic sequence in frameshift, missense or nonsense mutations *(Figure 3.4)*, although, already more than 25 years ago, it was estimated that up to 15% of disease-associated mutations cause RNA splicing defects of the mutated pre-mRNA (Krawczak et al., 1992). Since then, examples were identified where disease-associated mutations induce exon skipping by changing the consensus splicing sequences or lead to shortened or prolonged coding sequences through generation of cryptic splice sites (Cartegni et al., 2002). Mutations that create or destroy 5' ss and 3' ss can be rather easily identified using splice site strength predictions tools as implemented in this study (Yeo and Burge, 2004). More difficult is the prediction of mutations affecting splicing silencer and enhancer sequences as these are more diverse. Mutation-induced RNA splicing alterations can affect the protein sequence either by modifying or destroying the protein function, or by reducing the protein amount through frameshift induced premature stops and subsequent RNA degradation (*Figure 3.4*; Cartegni et al., 2002).

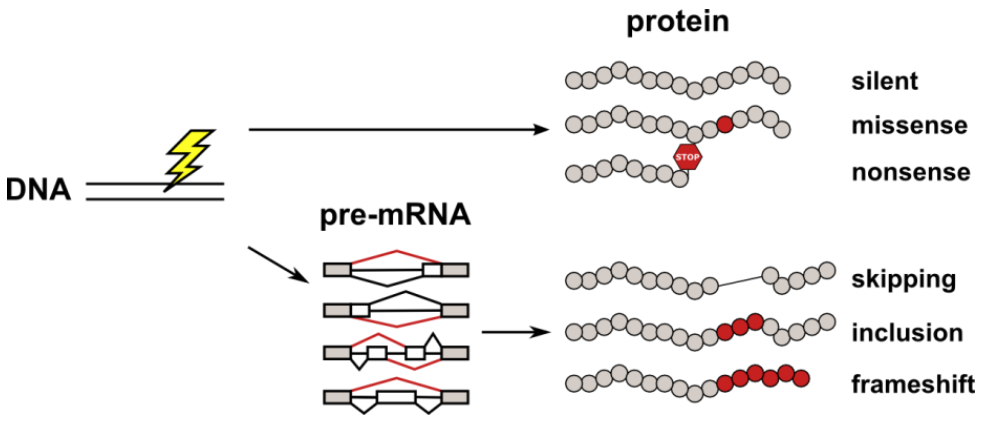


Figure 3.4: Translation of genomic point mutations. The yellow flash illustrates a point mutation in the genomic DNA. Exonic DNA mutations can be translated into protein sequence and classified in silent, missense or nonsense mutations. Exonic and intronic DNA mutation can be examined on pre-mRNA level to identify splicing affecting mutations that induce skipping or inclusion of sequences or induce a frameshift.

By classifying mutations simply based on the genomic sequence, many disease-causing sequence alterations might have been overlooked. This especially applies for silent exonic mutations that are considered to be neutral or intronic mutations that can affect splicing silencer and enhancer. Bioinformatic and experimental approaches just started to identify mutations that induce aberrant RNA splicing in a comprehensive manner. Minigene-based analysis of almost 5000 disease-causing exonic mutations showed that 10% of the mutations alter splicing (Soemedi et al., 2017). As this approach was limited to exons with a maximum length of 100 nucleotides, further experimental or bioinformatic approaches are needed to identify splicing-inducing mutations in introns and longer exons. Interestingly, a recent bioinformatic study that aimed to quantify the contribution of mutations on different stages of gene regulation revealed that genetic variations and diseases are mainly connected through RNA splicing (Li et al., 2016).

3.5 Splicing regulation through U2AF35 Q157 mutants in cell culture

Several studies examined the impact of U2AF35 ZnF mutations on splicing but they mainly focused on the most frequent mutation S34F in the first ZnF of U2AF35. These studies

observed aberrant splicing in cells and in transgenic mice, as well as in AML and MDS patients, with mutated U2AF35 (Brooks et al., 2014; Ilagan et al., 2014; Okeyo-Owuor et al., 2015; Przychodzen et al., 2013; Shirai et al., 2015; Yoshida et al., 2011). Consistently, the studies see that the S34F mutant has a changed specificity for the nucleotide preceding the invariant AG dinucleotide in the 3' ss where it strongly prefers a cytosine or an adenine over a thymine. This indicated a direct binding of the ZnF1 to the pre-mRNA which was later on shown for the yeast homolog of U2AF35 through crystallization and ITC measurements (Yoshida et al., 2015). The other mutations in ZnF1 (S34Y) and the mutations in ZnF2 (Q157R and Q157P) got less attention so far.

In this study, we aimed to analyze splicing changes caused by either the Q157R or the Q157P mutation in the second ZnF of U2AF35 in cells. RNA deep sequencing analysis of U2AF35-depleted cells rescued by U2AF35 wt identified over 500 cassette exons that were excluded upon knockdown and again included when the U2AF35-depleted cells were rescued with U2AF35 wt. Q157R and Q157P rescued splicing of 65% and 50% of these U2AF35-dependent exons, respectively. Although many of the target exons are rescued by both Q157 mutants, there were also target exons specifically rescued by either one of the mutations (30% for Q157R and 10% for Q157P) which was observed before and suggests a slightly different preference of the Q157 mutants for particular exons (Ilagan et al., 2014). Interestingly, substitution of the cells with the newly identified Q157Rdel variant revealed the lowest rescue ability of U2AF35-dependent exons of 45% and the highest number of uniquely rescued target exons. This probably results from the more severe interference of the Q157Rdel on the ZnF compared to the point mutations. We could validate our RNA sequencing analysis with splicing-sensitive RT-PCRs showing perfect accordance for nine out of 16 analyzed targets and the same tendency for two further targets.

To investigate whether splicing regulation of distinct sets of exons is due to distinct binding preferences of the Q157 mutants, as repeatedly seen for S34F, we analyzed the 3' ss sequences of Q157 mutant target exons. Exons uniquely targeted by either of the Q157 mutants revealed a binding preference based on the nucleotide following the invariant AG at the +1 position. Q157R preferentially binds exons with a cytosine, Q157P with a guanine and Q157Rdel with an adenine at this position *(Figure 3.5)*. Ilagan *et al*. also observed that the specificity of Q157R and Q157P is obtained though the +1 position of the 3' ss (Ilagan et al., 2014). However,

contrary to our results, the authors see for both Q157 point mutants an enrichment of guanine at the +1 position which might be due to consideration of all and not only uniquely targeted exons. When considering all Q157R or Q157P rescued exons, we receive similar results as Ilagan *et al*. meaning an enrichment of a G at the +1 position. Also consistent with Ilagan et al. is the enrichment of an A at the +1 position for exons that were not rescued by Q157R/P. By mutating the 3' ss of an U2AF35-target exon, we could validate the preferential binding specificity observed for the Q157 mutants. Ilagan *et al*. also provided minigene-based evidence for differential 3' ss binding of the Q157R and the S34Y mutant (Ilagan et al., 2014). These results demonstrate that the three Q157 U2AF35 mutants control distinct sets of target exons through a different binding specificity to the 3' ss. The results further suggest that U2AF35 wt has a high flexibility in recognizing different 3' ss whereas the Q157 U2AF35 variants are more restricted to specific nucleotides at the +1 position.

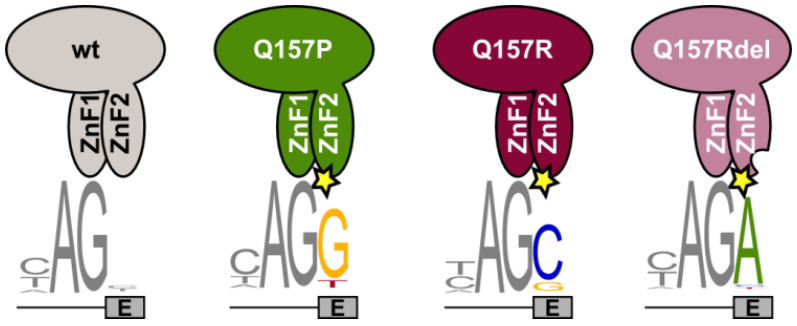


Figure 3.5: Binding specificities of U2AF35 Q157 mutants to the 3' ss. Depicted are 3' ss motifs preferentially bound by U2AF35 wt, Q157P, Q157R and Q157Rdel. The intron is represented as a line and the exon as grey box. The invariant AG dinucleotide is shown in grey and marks the intron-exon boundary. The nucleotides determining the binding-preferences at the +1 position in the exon are colored.

3.6 Splicing regulation through U2AF35 Q157 mutants in AML patients

So far, investigations of Q157R- or Q157P-specific splicing targets in patients were missing. We therefore aimed to confirm our results obtained in cell culture for human patients carrying Q157 mutations. First, we searched for RNA sequencing data of patients with Q157R

and Q157P mutations in the ZnF2 of U2AF35. In a publicly available RNA sequencing study of 533 AML patients we identified three Q157R and two Q157P patients representing 0.6% and 0.4% of the patients, respectively. These incidences just slightly differ from the incidences observed in the initial study identifying U2AF35 Q157R (1.2%) and Q157P (0.7%) mutations on genomic level (Yoshida et al., 2011). Previous studies that intended to analyze splicing patterns in Q157 patients lacked either patients with the Q157R or the Q157P mutation in their cohort and thus missed the identification of mutation-specific targets (Ilagan et al., 2014; Qiu et al., 2016). We therefore used the mRNA sequencing data of the identified Q157R and Q157P patients to analyze aberrant splicing.

Investigation of splicing alterations in the three Q157R and two Q157P patients revealed almost 2000 alternatively regulated splicing events affected in patients of both mutants when compared to AML-patients with wt U2AF35. Detailed analysis of the cassette exons, which represent the largest group of differentially regulated events (~900), showed that 22% are similarly deregulated in both mutants compared to wild-type, whereas 46% and 32% are only affected in Q157R or Q157P patients, respectively. Comparison of differently regulated exons in Q157 mutated patients to the exons regulated in HEK293T cells with U2AF35 Q157 mutant expression revealed a rather low overlap of 70 targets. The small overlap likely results from the divergent gene expression profiles of the hematopoietic primary cells compared to the immortalized HEK293T cells kept in culture. Also, as the U2AF35 mutations occur in a heterozygous manner, the patients express the mutated as well as the wt U2AF35 protein and therefore the effects observed in patients reflect a combinatorial effect of the wt and the mutated U2AF35. Contrary to that, we rescued U2AF35-depletion in HEK293T cells solely with the mutated variants of U2AF35. In a previous study, the overlap of splicing events in S34F/Y patients compared to cell cultures stably expressing those mutants also displayed a rather low overlap of 16% even though the stably transduced K562 cells line used in the study was derived from blood cells with a leukemic background (Ilagan et al., 2014). We therefore continued our analysis with these data sets. Despite the above discussed pitfalls, splicing patterns for seven out of 16 targets analyzed by RT-PCR were almost perfectly coincided in the patients validating our patient analysis. In addition, when comparing the splicing alterations of the 70 overlapping targets between the Q157 patients and the RNA sequencing of the HEK293T cells, we got fairly reasonable correlations, validating our observed differences in Q157R and Q157P patients.

Furthermore, hierarchical clustering analysis of splicing patterns grouped the three Q157R and the two Q157P patients together, indicating that the differences observed between Q157R and Q157P patients mainly result from the mutations and not through random tumor heterogeneity.

Having characterized Q157R- and Q157P-regulated target exons in patients, we asked whether we see similar nucleotide enrichments at the +1 position as in our cell culture system. Consensus motif analysis revealed an enrichment of a G at the +1 position in patients with either the Q157R or the Q157P mutation for exons that are more often included in the Q157 mutant compare to wt patients. This reflects our cell culture results although the enrichment in patients is less prominent than in cell culture. For exons that show less inclusion in Q157R compared to wt, we also see a tendency matching our cell culture results, which is the enrichment of an A at the +1 position. The weaker enrichment in patients might be due to the presence of 50% wt U2AF35 in the cells caused by the heterozygous character of the mutations.

3.7 Impact of the Q157Rdel variant on splicing in Q157R patients

In patients, splicing patterns between Q157R and Q157P are more different and the motif enrichment is less prominent compared to our cell culture based results, which might be a combinatorial effect of the expression of wt and the mutated U2AF35 in the patients. In Q157R patients however, this might be also due to the contribution of the newly identified Q157Rdel variant on splicing regulation. To test this hypothesis, we first analyzed whether the alternative 5’ ss generated by the Q157R mutation is used in Q157R patients. For this purpose, we searched for junction reads that encode for the Q157Rdel variant in the RNA sequencing data of the Q157R patients. Indeed, 2.5% of the U2AF35 transcripts in Q157R patients encoded for the Q157Rdel variant providing evidence that the alternative 5’ ss is used in Q157R patients. However, the observed usage was weaker than expected from our minigene experiments in cells. RNA stability assays excluded that the reduced presence of the Q157Rdel variant is due to reduced mRNA transcript stability. The reduced presences might be rather due to a differential regulation of splicing in the AML patients compared to the HEK293T cells. Possible differences might be the expression of cell type specific *trans*-acting regulators which are absent in HEK293T cells, different elongation rates of the PolII or the presence of *cis*-acting elements that are not included in our minigene sequence (Fu and Ares, 2014; Naftelberg et al., 2015).

Having shown that the Q157Rdel variant is expressed in Q157R patients and that it has a specific 3' ss binding sequence in cell culture analysis *(Figure 3.4),* which is even avoided by the Q157R mutant, we aimed to explore a potential role of the Q157Rdel variant on splicing patterns observed in Q157R patients. For that, we utilized the 70 target exons that were differentially spliced in patients and in cell culture and searched for exons where Q157R splicing regulation in patients is better mimicked by the Q157Rdel variant than by the Q157R point mutant in cell culture. The analysis revealed that 16 out of 70 targets show this pattern indicating a contribution of the Q157Rdel variant on splicing regulation in Q157R patients. In these 16 targets the nucleotide A is enriched at the +1 position of the 3' ss, reflecting the motif preferred by Q157Rdel. Compared to U2AF35 wt and Q157R just a small proportion of the U2AF35 mRNA encodes for the Q157Rdel variant, however as the Q157Rdel binds a specific 3' ss consensus motif it is likely that it contributes to splicing regulation in the Q157R patients.

3.8 Splicing-unrelated function of mutated U2AF35

It is obvious to assume that disease formation in patients with splicing factor (SF) mutations is accomplished though changed splicing regulation, however, for most of the mutated SFs, splicing-unrelated functions in the cell are known. RNA sequencing analysis of the pro-B cell line Ba/F3 transformed by U2AF35 S34F revealed an altered cleavage and polyadenylation (CP) site usage as the most frequently affected RNA processing event in those cells (Park et al., 2016). A connection between U2AF and polyadenylation regulation was already observed beforehand. It was shown that the U2AF heterodimer interacts with components of the cleavage factor I and II (CFIm and CFIIm) complexes and that alternative polyadenylation sites are used in U2AF35-depleted cells (Kralovicova et al., 2015; Millevoi et al., 2006; de Vries et al., 2000). Park et al. further show that U2AF35 S34F induces usage of the distal CP cite of the *Atg7* pre-mRNA by differential recruitment of enhancing and repressing CFIm complex components in an RNA-independent manner. The authors also detect usage of the distal CP site of *Atg7* in MDS patients with U2AF35 S34F mutation. As a consequence of the distal CP site usage, the *Atg7* mRNA has a prolonged 3'UTR that represses translation and results in reduced ATG7 protein expression. ATG7 is an essential autophagy factor whose depletion in hematopoietic cells in mice displayed an autophagy defect and developed an MDS-like syndrome, functionally linking

U2AF35 S34F mutation to disease transformation (Mortensen et al., 2011). Analysis of two U2AF35 S34Y patients and one Q157P patient revealed as well an increase in distal CP site usage in the *Atg7* pre-mRNA indicating that this might be a common path by which U2AF35 drives hematopoietic disorder formation. Recently it was shown that the activator function of CFIm is mediated through its RS-like domains and GST-pull-downs showed that U2AF65 directly binds to the RS-like domain in CFIm59 (Millevoi et al., 2006; Zhu et al., 2018), however direct binding of U2AF35 to CFIm59 was not yet assessed. Although these results are very interesting, it needs to be clarified how exactly the impaired recruitment of CFIm components through the S34F-mutant might work and whether the same mechanism is true for the Q157P mutant.

As U2AF35 is implicated in splicing regulation, our analysis discussed in *Chapter 3.6* and *Chapter 3.7* was designed to identify differentially spliced alternative events of the U2AF35 Q157 mutants in AML patients and in HEK293T cells. MISO, the software tool used for this investigation, analyzes the splicing of annotated alternative splicing events and is not able to discover *de novo* splicing. As the U2AF35 Q157 mutants possess distinct 3' ss binding specificities and probably also distinct binding affinities, they might also affect constitutive splicing or bind to cryptic 3' ss. RNA sequencing analysis of the U2AF35-depleted HEK293T cells revealed 200 skipped exons which could not be properly rescued by the U2AF35 wt but could be rescued by Q157R, Q157P or Q157Rdel indicating higher affinities of the Q157 mutants than of the wt U2AF35 for this exons. Further analysis of the RNA sequencing data in patients and in cells with an unbiased tool could identify *de novo* splicing events regulated by the Q157 mutants. Furthermore, the RNA sequencing analysis can be extended to identify altered CP site usage to in the HEK293T cells and in the Q157 patients.

3.9 Role of mutated U2AF35 in malignant hematopoiesis

Great progress was made in clarifying the mechanistic consequences of SF mutations on protein function and many differentially spliced mRNAs were identified. However, details of the direct contribution of these mutations on cell viability and development of blood cell disorders just started to be understood. In cell cycle analysis, we see an accumulation of HEK293T cells in the G2/M phase when substituted by either of the Q157 mutants following the U2AF35-

depletion. Furthermore, we observe a reduced proliferation of the Q157 mutant rescued HEK293T cells, which might be caused by the impaired cell cycle progression (Herdt et al., 2017). Proliferation suppression through the U2AF35 mutants was already described before (Shao et al. 2014; Yoshida et al. 2011) and an enrichment of cells in the G2/M phase is also seen in cells depleted for U2AF35 and in cells with ZnF-mutated U2AF35 (Chen et al., 2018; Pacheco et al., 2006b). These results suggest that U2AF35 is important for a normal cell cycle progression and cell proliferation and that the U2AF35 mutants are impaired in fulfilling this function. However, these initial attempts to analyze the impact of the U2AF35 mutations on cell function were based on overexpression experiments in already highly transformed tumor cells of non-hematopoietic origin. More sophisticated approaches are required to determine the contribution of the U2AF35 mutations on disease formation.

A first attempt to analyze the impact of U2AF35 mutations on normal and malignant hematopoiesis *in vivo* was made by using a mouse model in which transgenic mouse bone marrow with inducible mutant U2AF35 S34F expression was transplanted into immunodeficient mice. U2AF35 S34F-recipient mice had reduced numbers of leukocytes in the peripheral blood and an increased apoptosis and a reduced competitiveness of maturing hematopoietic cells in the bone marrow. Although these are known early characteristics of MDSs, the overall survival of mice with mutated U2AF35 was not affected and they did not developed MDS or AML (Shirai et al., 2015). Spliceosomal mutations occur in very early stages of MDS disorder development and are therefore thought to be driver mutations in MDS diseases (Papaemmanuil et al., 2013). Together with the cell culture results and the mouse model phenotypes, it was speculated that the reduction of mature blood cells observed in early phases of MDS development might be caused by U2AF35 mutations. In later stages, MDSs are characterized by aberrant clonal expansion of hematopoietic cells, which might be accomplished through the stepwise acquisition of further mutations in proteins that are especially involved in chromatin modification and cell signaling (Papaemmanuil et al., 2013; Sperling et al., 2017). Contrary to a proliferation inhibition through S34F-mutated U2AF35, another study observed increased proliferation and tumor formation when U2AF35 S34F-expressing cells were injected subcutaneously in immunocompromised mice. However, the authors performed the analysis with cells originating from lymphoid not myeloid progenitors and injected them subcutaneously, which might provide possible explanations for the different phenotype (Park et al., 2016). U2AF35 mutations co-occur with

mutations in the polycomb group protein ASXL1 (Haferlach et al., 2014; Papaemmanuil et al., 2013), thus more sophisticated analysis are required to address which genetic context U2AF35 mutations need to induces MDS or AML formation; and whether U2AF35 may especially contribute to ineffective hematopoiesis resulting in initial reduction of mature blood cells, whereas further mutations cause clonal expansion. In contrast to U2AF35 S34F mice, SRSF2 P95H and SF3B1 K700E knock-in mice and SF3B1 K700E mice with additional loss of Tet2 developed MDS indicating a more severe impact of the SF3B1 and SRSF2 mutations on disease development (Kim et al., 2015; Kon et al., 2017; Mupo et al., 2016; Obeng et al., 2016). In patients however, the so far published reports show conflicting observation on disease outcome and overall survival in patients with splicing factor mutations (Gangat et al., 2016).

The mutually exclusive presence of SF3B1, SRSF2, U2AF35 and ZRSR2 mutations in different hematopoietic malignancies might suggest similar effects of these mutations on disease formation, however several observations point to distinct downstream effects. At first, the mechanisms by which the mutations affect splicing are different as explained in the introduction. Also, the prevalence of the spliceosomal mutations is very diverse in various hematopoietic malignancies indicating a differential contribution to disease formation. And furthermore, the co-occurring genetic alterations differ between each spliceosomal mutation (reviewed in Inoue et al., 2016). In addition to that, considering their important role in splicing initiation, mutations in splicing factors induce rather minor changes in splicing, indicating that specific splicing events might drive disease pathogenesis rather than total splicing failure. Comparison of splicing patterns observed in patients with mutations in SF3B1, SRSF2 and U2AF35 revealed misregulation of largely non-overlapping splicing events. However, the affected genes for all three splicing factors belong to pathways that regulate cell cycle and DNA damage (Qiu et al., 2016). These results indicate that, although the mutated SFs regulate distinct splicing events by different mechanisms, the impact on disease formation is similar as proteins from similar functional pathways are affected.

Additionally to that, a recent study suggests involvement of an alternative common mechanism for SF-induced disease formation. The authors observe an enhanced R-loop formation in cells expressing mutated SRSF2 and U2AF35 (Chen et al., 2018). They suggest that R-loops induce DNA damage and replication stress, thereby leading to a reduced proliferation and a cell cycle arrest in the G2/M phase observed in cells with mutated SRSF2 and U2AF35.

Interestingly, R-loop formation occurred predominantly at the transcription start site and was underrepresented in the gene body and at splice sites, linking R-loop formation mediated by mutated SF to transcription rather than to splicing. For SRSF2, a splicing-independent role in transcription was described beforehand, however for U2AF35 no transcription modulating mechanism is known so far. Impaired cell cycle progression in SF-mutated cells could be partially rescued by expression of the RNA-DNA-duplex specific Ribonuclease H1 indicating that R-loops to some extend impact on SF-induced cellular defects (Chen et al., 2018).

3.10 Therapy opportunities for SF-induced hematopoietic disorders

The identification of SF mutations in MDS patients opened new opportunities to treat MDS patients. Two strategies might be feasible to treat patients with SF-induced hematopoietic disorders. The first is the inhibition of splicing factors with small molecules. As SF mutations occur in a mutually exclusive and heterozygous manner, it was hypothesized that the cells cannot tolerate full loss of one or partial loss of more than one splicing factor, thus those cells might be more sensitive to further splicing inhibition. Indeed, treatment of S34F-mutant cells with the SF3B-inhibitor sudemycin D6 was shown to affect the cell cycle progression and to reduce the survival of S34F-mutant cells more drastic compared to wt cells. In mice, this treatment slowed down the cell death of hematopoietic progenitor cells that constitute an initial phenotype during MDS (Shirai et al., 2017). The authors observe almost 20.000 misspliced events with a dPSI larger than 10% upon treatment of non-mutated cells with 1 µM sudemycin D6 for 6 hours, which reflects the half maximal inhibitory concentration (IC_{50}) of sudemycin D6. *In vivo*, the authors used lower concentrations of sudemycin D6, which resulted in fewer splicing changes and seemed not to affect the maturation of normal blood cells (Shirai et al., 2017). Increased selectivity of SF-mutated cell for further splicing inhibition was also observed for hematopoietic cells with SF3B1- or SRSF2-mutations, although again very strong splicing changes were induced in both wild-type and mutated cells (Lee et al., 2016; Obeng et al., 2016). For Myc-driven cancers the spliceosome was also shown to be a potential therapeutic entry point. Here, inhibition of the spliceosome led to increased intron retention in the tumor but not in healthy tissue, where the splicing machinery could cope with reduced spliceosome activity (Hsu et al., 2015; Koh et al., 2015)

The second approach to treat SF-induced hematopoietic disorders could be the specific modulation of splicing events that are targeted by the mutated SFs. Short synthetic antisense oligonucleotides (ASOs) can be used to specifically regulate splicing. ASOs bind in a sequence specific manner to the target pre-mRNA. As ASOs function through complementary base pairing, a high specificity is achieved and the off target effects are minimized. ASOs can be designed to either directly target splice site sequences or splicing regulatory elements like enhancers or silencers. Thereby ASOs prevent recognition of the *cis*-elements through the spliceosome or regulatory factors. In mouse models, ASOs were already used to treat various cancers, neurological disorders or viral, metabolic or inflammatory diseases (Havens and Hastings, 2016). Nusinersen is the first ASO treatment that passed the clinical trials and was recently approved by the US food and drug administration (FDA) for the treatment of spinal muscular atrophy (SMA). SMA patients have insufficient expression of the survival motor neuron (SMN) protein through deletions in the *SMN1* gene. Some full length SMN protein can be produced from the duplicated gene *SMN2,* however most of the transcripts arising from the *SMN2* gene lack exon 7 which leads to the degradation of the protein. The ASO used for SMA treatments binds to an ISS downstream of exon 7 in the *SMN2* gene, thereby blocking binding of the repressors hnRNPA1/ A2. This leads to increased exon 7 inclusion and to the expression of stable full length SMN (Chiriboga, 2017). To target splicing events induced by SF-mutations in MDS disorders, much effort must be put in the characterization of splicing events that drive the tumor transformation and maintenance.

.

REFERENCES

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. *7*, 1–12.

Andrade, J.M.L.M.C. (2011). Dissertation: Functional analysis of the U2AF35 family of splicing factors. Univ. Lisbon 126.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Çolak, R., et al. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science. *21*, 1587–1594.

Biamonti, G., and Caceres, J.F. (2009). Cellular stress and RNA splicing. Trends Biochem. Sci. *34*, 146–153.

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

Blanchette, M., Labourier, E., Green, R.E., Brenner, S.E., and Rio, D.C. (2004). Genome-Wide Analysis Reveals an Unexpected Function for the Drosophila Splicing Factor U2AF 50 in the Nuclear Export of Intronless mRNAs. Mol. Cell *14*, 775–786.

Brooks, A.N., Choi, P.S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., Pedamallu, C.S., Sivachenko, A., Rosenberg, M., Chmielecki, J., et al. (2014). A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events. PLoS One *9*, e87361.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein an hnRNP splicing regulatory factors. Wiley Interdiscip. Rev. RNA *3*, 1–12.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening To Silence and Understanding Nonsense: Exonic Mutations That Affect Splicing. Nat. Rev. Genet. *3*, 285–298.

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

Chen, L., Chen, J., Huang, Y., Abdel-wahab, O., Chen, L., Chen, J., Huang, Y., Gu, Y., Qiu, J., Qian, H., et al. (2018). The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Article The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. Mol. Cell *69*, 412–425.e6.

Chiriboga, C.A. (2017). Nusinersen for the treatment of spinal muscular atrophy. Expert Rev. Neurother. *17*, 955–962.

Cretu, C., Schmitzová, J., Ponce-Salvatierra, A., Dybkov, O., De Laurentiis, E.I., Sharma, K., Will, C.L., Urlaub, H., Lührmann, R., and Pena, V. (2016). Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. Mol. Cell *64*, 307–319.

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., et al. (2015). Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. Cell Rep. *13*, 1033–1045.

David, C.J., and Manley, J.L. (2010). Alternative pre-mRNA splicing regulation in cancer: Pathways and programs unhinged. Genes Dev. *24*, 2343–2364.

Fagg, W.S., Liu, N., Fair, J.H., Shiue, L., Katzman, S., Donohue, J.P., and Ares Jr, M. (2017). Autogenous cross-regulation of Quaking mRNA processing and translation balances Quaking functions in splicing and translation. Genes Dev. *31*, 1–16.

Fei, D.L., Motowski, H., Chatrikhi, R., Prasad, S., Yu, J., Gao, S., Kielkopf, C.L., Bradley, R.K., and Varmus, H. (2016). Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival. PLoS Genet. *12*, 1–26.

Flood, P.M., Dougherty, J.P., and Ron, Y. (1990). Inhibition of Ly-6A antigen expression prevents T cell activation. J. Exp. Med. *172*, 115–120.

Floor, S.N., and Doudna, J.A. (2016). Tunable protein synthesis by transcript isoforms in human cells. Elife *5*, 1–25.

Fu, M., and Blackshear, P.J. (2017). RNA-binding proteins in immune regulation: A focus on CCCH zinc finger proteins. Nat. Rev. Immunol. *17*, 130–143.

Fu, X.-D., and Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat. Rev. Genet. *15*, 689–701.

Gama-Carvalho, M., Carvalho, M.P., Kehlenbach, A., Valcarcel, J., and Carmo-Fonseca, M. (2001). Nucleocytoplasmic shuttling of heterodimeric splicing factor U2AF. J. Biol. Chem. *276*, 13104–13112.

Gama-Carvalho, M., Barbosa-Morais, N.L., Brodsky, A.S., Silver, P. a, and Carmo-Fonseca, M. (2006). Genome-wide identification of functionally distinct subsets of cellular mRNAs associated with two nucleocytoplasmic-shuttling mammalian splicing factors. Genome Biol. *7*, R113.

Gangat, N., Patnaik, M.M., and Tefferi, A. (2016). Myelodysplastic syndromes: Contemporary review and how we treat. Am. J. Hematol. *91*, 76–89.

Graubert, T. a, Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2011). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. *44*, 53–57.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA *7*, 806–818.

Guth, S., Martínez, C., Gaur, R.K., and Valcárcel, J. (1999). Evidence for substrate-specific requirement of the splicing factor U2AF35 and for its function after polypyrimidine tract

recognition by U2AF65. Mol. Cell. Biol. *19*, 8263–8271.

Guth, S., Tange, T.O., Kellenberger, E., and Thomas, O. (2001). Dual Function for U2AF 35 in AG-Dependent Pre-mRNA Splicing. Mol. Cell. Biol. *21*, 7673–7681.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., et al. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia *28*, 241–247.

Hall, T.M.T. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr. Opin. Struct. Biol. *15*, 367–373.

Hanson, P., Mathews, V., Marrus, S.H., and Graubert, T.A. (2003). Enhanced green fluorescent protein targeted to the Sca-1 (Ly-6A) locus in transgenic mice results in efficient marking of hematopoietic stem cells in vivo. Exp. Hematol. *31*, 159–167.

Hatada, L., Sugama, T., and Mukai, T. (1993). A new imprinted gene cloned by a methylation-sensitive genome scanning method. Nucleic Acids Res. *21*, 5577–5582.

Havens, M.A., and Hastings, M.L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. Nucleic Acids Res. *44*, 6549–6563.

Henderson, S.C., Kamdar, M.M., and Bamezai, A. (2002). Ly-6A.2 expression regulates antigen-specific CD4+ T cell proliferation and cytokine production. J. Immunol. *168*, 118–126.

Herdt, O., Neumann, A., Timmermann, B., and Heyd, F. (2017). The cancer-associated U2AF35 470A>G (Q157R) mutation creates an in-frame alternative 5' splice site that impacts splicing regulation in Q157R patients. RNA J. *23*, 1796–1806.

Heyd, F., ten Dam, G., and Möröy, T. (2006). Auxiliary splice factor U2AF26 and transcription factor Gfi1 cooperate directly in regulating CD45 alternative splicing. Nat. Immunol. *7*, 859–867.

Heyd, F., Carmo-Fonseca, M., and Möröy, T. (2008). Differential isoform expression and interaction with the P32 regulatory protein controls the subcellular localization of the splicing factor U2AF26. J. Biol. Chem. *283*, 19636–19645.

Holmes, C., and Stanford, W.L. (2007). Concise review: stem cell antigen-1: expression, function, and enigma. Stem Cells *25*, 1339–1347.

Van Den Hoogenhof, M.M.G., Pinto, Y.M., and Creemers, E.E. (2016). RNA Splicing regulation and dysregulation in the heart. Circ. Res. *118*, 454–468.

Hsu, T.Y.-T., Simon, L.M., Neill, N.J., Marcotte, R., Sayad, A., Bland, C.S., Echeverria, G. V, Sun, T., Kurley, S.J., Tyagi, S., et al. (2015). The spliceosome is a therapeutic vulnerability in MYC-driven cancer. Nature *525*, 384–388.

Hudson, B.P., Martinez-Yamout, M. a, Dyson, H.J., and Wright, P.E. (2004). Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. Nat. Struct. Mol. Biol. *11*, 257–264.

Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2014). U2AF1 mutations alter splice site recognition in hematological malignancies.

Genome Res. *124*, 14–26.

Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell *150*, 1107–1120.

Inoue, D., Bradley, R.K., and Abdel-wahab, O. (2016). Spliceosomal gene mutations in myelodysplasia : molecular links to clonal abnormalities of hematopoiesis. Genes Dev. *30*, 989–1001.

Jenkins, J.L., and Kielkopf, C.L. (2017). Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. Trends Genet. *33*, 336–348.

Jeremiah Brian Shepard (2004). Dissertation: Characterization of U2AF26, a paralog of the splicing factor U2AF35. Univ. Texas Southwest. Med. Cent. 124.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. Gene *514*, 1–30.

Kellenberger, E., Stier, G., and Sattler, M. (2002). Induced folding of the U2AF35 RRM upon binding to U2AF65. FEBS Lett. *528*, 171–176.

Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. Nat. Rev. Genet. *11*, 345–355.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A Novel Peptide Recognition Mode Revealed by the X-Ray Structure of a Core U2AF35 / U2AF65 Heterodimer. Cell *106*, 595–605.

Kielkopf, C.L., Lücke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. Genes Dev. *18*, 1513–1526.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C.W., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E., et al. (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell *27*, 617–630.

Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. Annu. Rev. Biochem. *79*, 213–231.

Koh, C.M., Bezzi, M., Low, D.H.P., Ang, W.X., Teo, S.X., Gay, F.P.H., Al-Haddawi, M., Tan, S.Y., Osato, M., Sabò, A., et al. (2015). MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. Nature *523*, 96–100.

Kon, A., Yamazaki, S., Nannya, Y., Kataoka, K., Ota, Y., Nakagawa, M.M., Yoshida, K., Shiozawa, Y., Morita, M., Yoshizato, T., et al. (2017). Physiological *Srsf2* P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. Blood.

Kornblihtt, A.R. (2015). Transcriptional control of alternative splicing along time : Ideas change , experiments remain. RNA *21*, 670–672.

Kralovicova, J., Knut, M., Cross, N.C.P., and Vorechovsky, I. (2015). Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3’ splice-site organization and

activity of U2AF-related proteins. Nucleic Acids Res. *43*, 3747–3763.

Krawczak, M., Reiss, J., and Cooper, D.N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. Hum. Genet. *90*, 41–54.

Kuznetsova, I.M., Turoverov, K.K., and Uversky, V.N. (2014). What macromolecular crowding can do to a protein. Int. J. Mol. Sci. *15*, 23090–23140.

Lee, S.C.W., Dvinge, H., Kim, E., Cho, H., Micol, J.B., Chung, Y.R., Durham, B.H., Yoshimi, A., Kim, Y.J., Thomas, M., et al. (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat. Med. *22*, 672–678.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L., and Lafyatis, R. (2002). Stability of a PKCI-1 -related mRNA is controlled by the splicing factor ASF / SF2 : a novel function for SR proteins. Genes Dev. *16*, 594–607.

Li, Y.I., Geijn, B. Van De, Raj, A., Knowles, D. a, Petti, A. a, Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. Science. *352*, 600–604.

Lin, X., Miller, J.W., Mankodi, A., Kanadia, R.N., Yuan, Y., Moxley, R.T., Swanson, M.S., and Thornton, C.A. (2006). Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. Hum. Mol. Genet. *15*, 2087–2097.

Madan, V., Kanojia, D., Li, J., Okamoto, R., Sato-Otsubo, A., Kohlmann, A., Sanada, M., Grossmann, V., Sundaresan, J., Shiraishi, Y., et al. (2015). Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nat. Commun. *6*, 6042.

Makishima, H., Visconte, V., Sakaguchi, H., Jankowska, A.M., Kar, S.A., Jerez, A., Przychodzen, B., Bupathi, M., Guinta, K., Afable, M.G., et al. (2012). Plenary paper Mutations in the spliceosome machinery , a novel and ubiquitous pathway in leukemogenesis. *119*, 3203–3210.

Martinez, N.M., and Lynch, K.W. (2013). Control of alternative splicing in immune responses : many regulators , many predictions , much still to learn. Immunol. Rev. *253*, 216–236.

Martinez, N.M., Pan, Q., Cole, B.S., Yarosh, C.A., Babcock, G.A., Heyd, F., Zhu, W., Ajith, S., Blencowe, B.J., and Lynch, K.W. (2012). Alternative splicing networks regulated by signaling in human T cells. RNA *18*, 1029–1040.

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

McManus, C.J., and Graveley, B.R. (2011). RNA structure and the mechanisms of alternative splicing. Curr. Opin. Genet. Dev. *21*, 373–379.

Merendino, L., Guth, S., Bilbao, D., Concepcion, M., and Juan, V. (1999). Inhibition of msl-2 splicing by Sex- lethal reveals interaction between U2AF35 and the 3`splice site AG. Nature *402*, 838–841.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science. 1593–1599.

Michel, S.L.J., Guerrerio, A.L., and Berg, J.M. (2003). Selective RNA binding by a single CCCH zinc-binding domain from Nup475 (tristetraprolin). Biochemistry *42*, 4626–4630.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. (2006). An interaction between U2AF 65 and CF Imlinks the splicing and 3′ end processing machineries. EMBO J. *25*, 4854–4864.

Mollet, I., Barbosa-Morais, N.L., Andrade, J., and Carmo-Fonseca, M. (2006). Diversity of human U2AF splicing factors. FEBS J. *273*, 4807–4816.

Mortensen, M., Soilleux, E.J., Djordjevic, G., Tripp, R., Lutteropp, M., Sadighi-Akha, E., Stranks, A.J., Glanville, J., Knight, S., W. Jacobsen, S.-E., et al. (2011). The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. J. Exp. Med. *208*, 455–467.

Mupo, A., Seiler, M., Sathiaseelan, V., Pance, A., Yang, Y., Agrawal, A.A., Iorio, F., Bautista, R., Pacharne, S., Tzelepis, K., et al. (2016). Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. Leukemia *31*, 720–727.

Muto, Y., and Yokoyama, S. (2012). Structural insight into RNA recognition motifs: Versatile molecular Lego building blocks for biological systems. Wiley Interdiscip. Rev. RNA *3*, 229–246.

Naftelberg, S., Schor, I.E., Ast, G., and Kornblihtt, A.R. (2015). Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. Annu. Rev. Biochem. *84*, 165–198.

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

Obeng, E.A., Chappell, R.J., Seiler, M., Chen, M.C., Campagna, D.R., Schmidt, P.J., Schneider, R.K., Lord, A.M., Wang, L., Gambe, R.G., et al. (2016). Physiologic Expression of Sf3b1K700ECauses Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. Cancer Cell *30*, 404–417.

Okeyo-Owuor, T., White, B.S., Chatrikhi, R., Mohan, D.R., Kim, S., Griffith, M., Ding, L., Ketkar-Kulkarni, S., Hundal, J., Laird, K.M., et al. (2015). U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. Leukemia *29*, 909–917.

Pacheco, T.R., Moita, L.F., Gomes, A.Q., Hacohen, N., and Carmo-Fonseca, M. (2006a). RNA Interference Knockdown of hU2AF35 Impairs Cell Cycle Progression and Modulates Alternative Splicing of Cdc25 Transcripts. Mol. Biol. Cell *17*, 4187–4199.

Pacheco, T.R., Coelho, M.B., Desterro, J.M.P., Mollet, I., and Carmo-Fonseca, M. (2006b). In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. Mol. Cell. Biol. *26*, 8183–8190.

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al. (2013). Clinical and biological implications of driver

mutations in myelodysplastic syndromes. Blood *122*, 3616–3627.

Park, S.M., Ou, J., Chamberlain, L., Simone, T.M., Yang, H., Virbasius, C.-M., Ali, A.M., Zhu, L.J., Mukherjee, S., Raza, A., et al. (2016). U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3′ End Formation. Mol. Cell *35*, 479–490.

Poulos, M.G., Batra, R., Li, M., Yuan, Y., Zhang, C., Darnell, R.B., and Swanson, M.S. (2013). Progressive impairment of muscle regeneration in muscleblind-like 3 isoform knockout mice. Hum. Mol. Genet. *22*, 3547–3558.

Preußner, M., Wilhelmi, I., Schultz, A.-S., Finkernagel, F., Michel, M., Möröy, T., and Heyd, F. (2014). Rhythmic U2af26 alternative splicing controls PERIOD1 stability and the circadian clock in mice. Mol. Cell *54*, 651–662.

Preußner, M., Goldammer, G., Neumann, A., Haltenhof, T., Rautenstrauch, P., Müller-McNicoll, M., and Heyd, F. (2017). Body Temperature Cycles Control Rhythmic Alternative Splicing in Mammals. Mol. Cell *67*, 1–14.

Przychodzen, B., Jerez, A., Guinta, K., Sekeres, M. a., Padgett, R., Maciejewski, J.P., and Makishima, H. (2013). Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. Blood *122*, 999–1006.

Qiu, J., Zhou, B., Thol, F., Zhou, Y., Chen, L., Shao, C., DeBoever, C., Hou, J., Li, H., Chaturvedi, A., et al. (2016). Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators. RNA *22*, 1535–1549.

Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron *87*, 14–27.

Rudner, D.Z., Kanaar, R., Breger, K.S., and Rio, D.C. (1996). Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects. PNAS *93*, 10333–10337.

Rudner, D.Z., Breger, K.S., Kanaar, R., Melissa, D., and Rio, D.C. (1998a). RNA Binding Activity of Heterodimeric Splicing Factor U2AF : at Least One RS Domain Is Required for High-Affinity Binding RNA Binding Activity of Heterodimeric Splicing Factor U2AF : at Least One RS Domain Is Required for High-Affinity Binding. Mol. Cell. Biol. *18*, 4004–4011.

Rudner, D.Z., Breger, K.S., and Rio, D.C. (1998b). Molecular genetic analysis of the heterodimeric splicing factor U2AF: The RS domain on either the large or small Drosophila subunit is dispensable in vivo. Genes Dev. *12*, 1010–1021.

Rudner, D.Z., Kanaar, R., Breger, K.S., and Donald, C. (1998c). Interaction between Subunits of Heterodimeric Splicing Factor U2AF Is Essential In Vivo Interaction between Subunits of Heterodimeric Splicing Factor U2AF Is Essential In Vivo. Mol. Cell. Biol. *18*, 1765–1773.

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

Sanford, J.R., Ellis, J.D., Cazalla, D., and Cáceres, J.F. (2005). Reversible phosphorylation

differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. PNAS *102*, 15042–15047.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003). Structural Basis for the Molecular Recognition between Human Splicing Factors U2AF 65 and SF1 / mBBP. Mol. Cell *11*, 965–976.

Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H., et al. (2014). Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat. Struct. Mol. Biol. *21*, 997–1005.

Sharma, A., and Lou, H. (2011). Depolarization-mediated regulation of alternative splicing. Front. Neurosci. *5*, 1–6.

Shen, H., Zheng, X., Luecke, S., and Green, M.R. (2010). The U2AF35-related protein Urp contacts the 3′ splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. Genes Dev. *24*, 2389–2394.

Shepard, J., Reick, M., Olson, S., and Graveley, B.R. (2002). Characterization of U2AF26, a splicing factor related to U2AF35. Mol. Cell. Biol. *22*, 221–230.

Shin, C., and Manley, J.L. (2004). Cell signalling and the control of pre-mRNA splicing. Nat. Rev. Mol. Cell Biol. *5*, 727–738.

Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E.J., Okeyo-Owuor, T., et al. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell *27*, 631–643.

Shirai, C.L., White, B.S., Tripathi, M., Tapia, R., Ley, J.N., Ndonwi, M., Kim, S., Shao, J., Carver, A., Saez, B., et al. (2017). Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. Nat. Commun. *8*, 14060.

Singh, R., Valcarcel, J., and Green, M. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science. *268*, 1173–1176.

Soares, L.M.M., Zanier, K., Mackereth, C., and Valcárcel, J. (2014). Intron Removal Requires Proofreading of U2AF/3' Splice Site Recognition by DEK. Science. *1961*, 1961–1965.

Soemedi, R., Cygan, K.J., Rhine, C.L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J., and Fairbrother, W.G. (2017). Pathogenic variants that alter protein code often disrupt splicing. Nat. Genet. *49*, 848–855.

Sperling, A.S., Gibson, C.J., and Ebert, B.L. (2017). The genetics of myelodysplastic syndrome: From clonal haematopoiesis to secondary leukaemia. Nat. Rev. Cancer *17*, 5–19.

Stanford, W.L., Haque, S., Alexander, R., Liu, X., Latour, a M., Snodgrass, H.R., Koller, B.H., and Flood, P.M. (1997). Altered proliferative response by T lymphocytes of Ly-6A (Sca-1) null mice. J. Exp. Med. *186*, 705–717.

Sterne-Weiler, T., Martinez-Nunez, R.T., Howard, J.M., Cvitovik, I., Katzman, S., Tariq, M.A., Pourmand, N., and Sanford, J.R. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. Genome Res. *23*, 1615–1623.

Taliaferro, J.M., Lambert, N.J., Sudmant, P.H., Dominguez, D., Merkin, J.J., Alexis, M.S., Bazile, C.A., and Burge, C.B. (2016). RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. Mol. Cell *64*, 294–306.

Terenzi, F., and Ladd, A.N. (2010). Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity. RNA Biol. *7*, 43–55.

Tronchère, H., Wang, J., and Fu, X.D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. Nature *388*, 397–400.

Turunen, J.J., Niemelä, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: Splicing by the minor spliceosome. Wiley Interdiscip. Rev. RNA *4*, 61–76.

Twyffels, L., Gueydan, C., and Kruys, V. (2011). Shuttling SR proteins: More than splicing factors. FEBS J. *278*, 3246–3255.

de Vries, H., Rüegsegger, U., Hübner, W., Friedlein, A., Langen, H., Keller, W., Amrani, N., Minet, M., Wyers, F., Dufour, M., et al. (2000). Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. *19*, 5895–5904.

Vuong, C.K., Black, D.L., and Zheng, S. (2016). The neurogenetics of alternative splicing. Nat. Rev. Neurosci. *17*, 265–281.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Wang, D., Guo, Y., Wu, C., Yang, G., Li, Y., and Zheng, C. (2008). Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. BMC Genomics *9*, 44.

Weatheritt, R.J., Sterne-Weiler, T., and Blencowe, B.J. (2016). The ribosome-engaged landscape of alternative splicing. Nat. Struct. Mol. Biol. *23*, 1117–1123.

Webb, C.J., and Wise, J.A. (2004). The Splicing Factor U2AF Small Subunit Is Functionally Conserved between Fission Yeast and Humans. Mol. Cell. Biol. *24*, 4229–4240.

Will, C.L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T., and Lührmann, R. (2004). The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. RNA *10*, 929–941.

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

Wu, T., and Fu, X.-D. (2015). Genomic functions of U2AF in constitutive and regulated splicing. RNA Biol. *12*, 479–485.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832–835.

Yan, C., Wan, R., Huang, G., and Shi, Y. (2016). Structure of a yeast activated spliceosome at 3.5 A resolution. Science. *353*, 904–911.

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., Sun, S., Yang, F., Shen, Y. a., Murray, R.R., et al. (2016). Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. Cell *164*, 805–817.

Yeo, G., and Burge, C.B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. J. Comput. Biol. *11*, 377–394.

Yoshida, K., and Ogawa, S. (2014). Splicing factor mutations and cancer. Wiley Interdiscip. Rev. RNA *5*, 445–459.

Yoshida, H., Park, S., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., Tsuda, K., Kuwasako, K., Unzai, S., Muto, Y., et al. (2015). A novel 3′ splice site recognition by the two zinc fingers in the U2AF small subunit. Genes Dev. *29*, 1649–1660.

Yoshida, K., Sanada, M., Shiraishi, Y., and Nowak, D. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69.

Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. PNAS *86*, 9243–9247.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609–614.

Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. PNAS *112*, E4726–E4734.

Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, a I., and Green, M.R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. PNAS *89*, 8769–8773.

Zhu, Y., Wang, X., Forouzmand, E., Jeong, J., Qiao, F., Sowd, G.A., Engelman, A.N., Xie, X., Hertel, K.J., and Shi, Y. (2018). Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. Mol. Cell *69*, 62–74.

Zolotukhin, A.S., Tan, W., Bear, J., Smulevitch, S., and Felber, B.K. (2002). U2AF participates in the binding of TAP (NXF1) to mRNA. J. Biol. Chem. *277*, 3935–3942.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. Nature *402*, 835–838.

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

ABBREVIATIONS

ss	splice site
A	adenine
AML	acute myeloid leukemia
Bp	base pairs
BPS	branch point sequence
C	cytosine
CHX	cycloheximide
CP	cleavage and polyadenylation
DNA	deoxyribonucleic acid
E	exon
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
G	guanine
hnRNP	heterogeneous nuclear ribonucleoprotein particle
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
kDa	kilo Dalton
Ly6a	lymphocyte antigen 6 a
MDS	myelodysplastic syndromes
mRNA	messenger ribonucleic acid
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
PCR	polymerase chain reaction
PMA	phorbol myristate acetate
RNA	ribonucleic acid
RNAPII	RNA polymerase II
RNP	ribonucleoprotein particle
RRM	RNA recognition motif
RS	serine-arginine rich
RT-PCR	Reverse transcriptase PCR
SD	standard deviation
SRSF2	serine/arginine-rich splicing factor 2
SF3B1	splicing factor 3b subunit 1
SELEX	systematic evolution of ligands by exponential enrichment
SF	splicing factor
siRNA	small interfering RNA
snRNP	small nuclear ribonucleoprotein particle

T	thymine
U2AF	U2 auxiliary factor
U	uracil
UHM	U2AF homology motif
UTR	untranslated region
Wt	wild type
ZnF	zinc finger
ZRSR	zinc finger RNA binding motif and serine/arginine rich
Δ	delta

LIST OF FIGURES

DANKSAGUNG

Zunächst bedanke ich mich herzlich bei meinem Betreuer Prof. Dr. Florian Heyd für die Möglichkeit meine Doktorarbeit in seiner Arbeitsgruppe anfertigen zu können. Du hast Dir stets viel Zeit genommen um über meine Ergebnisse zu diskutieren, bist immer optimistisch geblieben und konntest mich damit immer motivieren.

Außerdem bedanke ich mich bei Prof. Dr. Markus C. Wahl für die Übernahme des Zweitgutachtens.

Des Weiteren gilt mein Dank Dr. Bernd Timmermann für die Durchführung der RNA-Sequenzierung.

Ein ganz besonderer Dank geht an alle aktuellen und ehemaligen Kollegen der AG Heyd, Chakrabarti und Wahl für die hilfreichen Diskussionen und aufbauenden Gespräche, lustigen Mittagessen, Grillabende, Feierabendbiere, Geburtstage, SO36-Besuche usw. Ich habe die Zeit mit Euch sehr genossen und werde sie so schnell nicht vergessen!

Ich danke Alexander Neumann für seine bioinformatischen Analysen der Patienten und der RNA-Sequenzierung. Danke, dass Du immer viel Geduld mit mir hattest und immer ruhig geblieben bist, obwohl ich oft alles auf einmal haben wollte und Du vieles mehrmals erklären musstest. Auch danke ich Dir für das kritische Korrekturlesen meiner Arbeit.

Mein spezieller Dank gilt auch Dr. Marco Preußner für seine vielen hilfreichen Anregungen und Diskussionen, die Zusammenarbeit am U26Δ7-Projekt und das kritische Korrekturlesen dieser Arbeit.

Ich danke auch Francesca De Bortoli für die Beantwortung meiner Fragen zu Proteinstrukturen und für die Hilfe bei der Erstellung der Strukturvorhersagen. Außerdem danke ich Dir natürlich für die unzähligen kulinarischen Highlights.

Ein weiterer Dank gilt Antje Grünwald für das Präparieren unzähliger RNAs und Minis und das über den Jordan bringen vieler Mäuse. Auch danke ich Karin Hesse für ihre Hilfe bei allen administrativen Fragen.

Zu guter Letzt danke ich meiner Familie und meinen Freunden, die mir immer mit Rat und Tat zur Seite standen und mich in den letzten Jahren unterstützt haben.

CURRICULUM VITAE

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

PUBLICATION I

Herdt O*, Neumann A*, Timmermann B, Heyd F (2017). The cancer-associated U2AF35 470A>G (Q157R) mutation creates an in-frame alternative 5′ splice site that impacts splicing regulation in Q157R patients. RNA 23, 1796-1806.

* These authors contributed equally to this work.

https://doi.org/10.1261/rna.061432.117

PUBLICATION II

Herdt O, Preußner M, Heyd F (2018). The zinc finger domains in U2AF35 and U2AF656 control stability, localization, interaction with U2AF65 and play a role in translation. NAR, *in preparation for re-submission.*

The zinc finger domains in U2AF35 and U2AF656 control stability, localization, interaction with U2AF65 and play a role in translation

Olga Herdt[1], Marco Preußner[1] and Florian Heyd[1*]

[1]Freie Universität Berlin, Institute of Chemistry and Biochemistry, Laboratory of RNA Biochemistry, Takustrasse 6, 14195 Berlin, Germany.

*Corresponding Author: florian.heyd@fu-berlin.de

Phone: +49 30 83862938

FAX: +49 30 838-4-62938

ABSTRACT

Recent work has associated point mutations in both zinc fingers (ZnF) of the spliceosome component U2AF35 with malignant transformation and splicing defects caused by these mutations have been analyzed in numerous studies. However, apart from a structural approach in yeast, little is known about the functionality of the U2AF35 ZnF domains in general. Here we have analyzed key functionalities of mammalian U2AF35 and U2AF26 and show that ZnF2 is required for protein stability and stable interaction with U2AF65. Skipping of U2AF26 exon 7 is strongly induced upon activation of primary mouse T cells and leads to disruption of ZnF2. Consistently, U2AF26ΔE7 shows reduced stability and interaction with U2AF65 and, as shown previously, is localized in the cytoplasm. MS2 tethering assays as well as cell culture based analyses suggest that cytoplasmic U2AF26ΔE7 is involved in controlling translation when localized to the 5'UTR of target mRNAs. This regulation requires ZnF1 and parts of the RRM within U2AF26ΔE7 and provides a new connection between a core splicing factor and a cytoplasmic functionality. Altogether, our work reveals unexpected functions of U2AF35 and U2AF26 and its ZnF domains, thereby contributing to a better understanding of their role and regulation in mammalian cells.

INTRODUCTION

Splicing is an essential step during pre-mRNA processing that removes introns and joins exons to yield the mature mRNA with a continuous open reading frame. Splicing is catalyzed by the spliceosome that assembles de novo on every intron through binding to several sequence motives on the pre-mRNA (Wahl et al., 2009). For example, the 3' splices site (3'ss) is marked by a polypyrimidine tract followed by an AG-dinucleotide directly at the intron-exon boundary. The 3'ss is recognized by the U2 auxiliary factor (U2AF) heterodimer composed of the large subunit U2AF65 binding to the polypyrimidine tract and the small subunit U2AF35 binding to the AG-dinucleotide (Mackereth et al., 2011; Merendino et al., 1999; Wu et al., 1999; Zamore et al., 1992; Zorio and Blumenthal, 1999). Structural analysis of a minimal U2AF heterodimer identified a 28-residue segment in U2AF65, directly upstream of its RRMs, and two short 13- and 8-residue fragments in U2AF35 as interacting regions between both proteins (Kielkopf et al., 2001). The recent structure of full length yeast U2AF35 (U2AF23) with a short fragment of yeast

U2AF65 (U2AF59) revealed an additional interaction surface in the C-terminus of U2AF35 directly downstream of its second zinc finger (Yoshida et al., 2015). This additional U2AF heterodimer interaction surface was shown to play a role in the expression of U2AF35 splicing variants (Kralovicova and Vorechovsky, 2016).

U2AF35 consists of an U2AF homology motif (UHM) flanked by two zinc finger domains (ZnFs) and a C-terminal RS-domain important for subcellular localization (Gama-Carvalho et al., 2001; Kielkopf et al., 2004; Webb and Wise, 2004; Wu and Fu, 2015; Zhang et al., 1992). This architecture is highly conserved across species and several paralogs are found in mouse and human (Mollet et al., 2006). One example is the smaller paralog U2AF26 that lacks parts of the C-terminal RS-domain but shares 89% of the amino acids in its N-terminus with U2AF35 (Heyd et al., 2006, 2008; Shepard et al., 2002). U2AF26 is itself alternatively spliced and gives rise to at least two different isoforms in addition to the full length protein (Heyd et al., 2008). The U2AF26 splice variants lack either exon 7 or exon 6 and 7 and were shown to localize to the cytoplasm in contrast to the nuclear full length U2AF26 (Heyd et al., 2008; Preußner et al., 2014). Exclusion of exon 6 and 7 changes the reading frame of the mRNA and leads to a new prolonged C-terminus in U2AF26 that influences circadian rhythms in mice (Preußner et al., 2014). The function of cytoplasmic U2AF26 lacking exon 7 is so far unknown.

Although the role of U2AF35 in directly binding the AG at the 3' splice site has been discovered more than 15 years ago, the protein domain that makes direct contact to the RNA remained elusive. In early studies, it was reported that a minimal U2AF heterodimer composed out of the RRM of U2AF35 and a short N-terminal fragment of U2AF65 is sufficient to bind an RNA containing a 3' splice site consensus sequence (Kielkopf et al., 2001). A comprehensive mutational analysis provided first indications that, besides the RRM, both zinc fingers of yeast U2AF35 are important for viability and that all three domains contribute to RNA binding (Webb and Wise, 2004). Recently, the crystal structure of the full length yeast U2AF35 homolog U2AF23 was solved showing that RNA-binding of U2AF35 is mediated through its ZnF domains. The 'RRM' however serves as a scaffold for the ZnF domains and as a protein-protein interaction surface (Yoshida et al., 2015). As this work is based on *in vitro* assays and a structure of the yeast homolog U2AF23 with a conservation of 60% to mouse U2AF35 in the N-terminus, functional implications of this structure remain to be confirmed in the mammalian system. Recently, cancer-associated mutations in U2AF35 were identified exclusively occurring in either

ZnF1 (S34F/Y) or the ZnF2 (Q157R/P) of U2AF35 and not in the other domains (Graubert et al., 2012; Yoshida et al., 2011). Although several studies have shown that these mutations affect the binding specificity of U2AF35 to the 3'ss (Herdt et al., 2017; Ilagan et al., 2014; Okeyo-Owuor et al., 2015a; Przychodzen et al., 2013; Shirai et al., 2015), additional roles of the ZnF domains in controlling U2AF35 function remain largely unknown.

In the present work we have analyzed several key functions of U2AF35 using ZnF deletions or cancer-associated point mutations in mammalian cells. We show that the second zinc finger is required for protein stability and stable interaction with U2AF65, whereas the point mutations behave like the wt protein in these assays. We have furthermore quantified the effect of the ZnF deletion mutants on splicing of several U2AF35-dependent exons and show loss of function for ZnF-deleted U2AF35. We obtained similar results with the paralog U2AF26 and show that full length U2AF26 can fully restore splicing defects caused by U2AF35 knockdown. We further validate our observation with the alternative splice variant U2AF26ΔE7 that lacks the ZnF2, localizes to the cytoplasm and is upregulated upon activation of primary mouse T cells. We use an MS2-based reporter assay to show a potential role of U2AF26ΔE7 in translational regulation through ZnF1 and the RNP2 motif. Generation of a mouse T cell line stably expressing U2AF26ΔE7 then identified lymphocyte antigen 6a (Ly6a) as an endogenous target of U2AF26ΔE7. Our work provides evidence for a new function of U2AF26ΔE7 in the cytoplasm and demonstrates the importance of the ZnF domains for U2AF35 and U2AF26 function beyond RNA binding.

MATERIALS AND METHODS

Cell culture, transfection and treatments

HEK293T and HeLa cells were cultured in DMEM High glucose (Biowest) containing 10% FCS (Biochrom) and 1% Penicillin/Streptomycin (Biowest). EL4 cells were cultivated in RPMI medium (Biowest) with the same additives. Transfections of plasmid and siRNA were performed using RotiFect (Carl Roth) following manufacturer's instructions. For protein stability assays, $1.5x10^5$ HEK293T cells in a 12 well plate were transfected with 0.4 µg expression vectors for Flag-tagged U2AF35 and U2AF26 in the presence or absence of 0.4 µg HA-U2AF65. 24 hours after transfection, cells were treated with 30 ng/mL Cycloheximide (Sigma) for 4, 8 or 24 hours

and lysed as described below. For rescue experiments, knockdown of U2AF35 was performed by transfecting $0.5x10^5$ HEK293T cells in a 12 well plate with 20 pmol siRNAs (siU2AF35: GAAAGUGUUGUAGUUGAUUGA; siCTRL: UUCUCCGAACGUGUCACGU). After 24 hours, knockdown was rescued by transfection of 0.8 µg expression vectors for Flag-tagged U2AF35 and U2AF26 expression constructs for an additional 48 hours. For mRNA stability assays, cells were treated with 5 µg/mL Actinomycin D (Sigma).

Constructs

U2AF35-Flag, U2AF35-Q157R-Flag, U2AF35-Q157P-Flag, U2AF26-Flag, U2AF26ΔE7-Flag and U2AF65-HA protein expression constructs have been described previously (Herdt et al., 2017; Heyd et al., 2008); these plasmids served as template for insertions of deletions and point mutations using standard cloning procedures. For MS2 tethering assays U2AF26 deletions were cloned into pEF-MS2 plasmid, encoding an N-terminal MS2 tag. To introduce the MS2 binding loop, we designed oligonucleotides encoding the MS2 binding loop and overhangs for HindIII (5'UTR) or XbaI (3'UTR), annealed the oligonucleotides (5'UTR: fwd: agcttgcgatcaccatcagggatcgca, rev: agcttgcgatccctgatggtgatcgca; 3'UTR: fwd: ctagagcgatcaccatcagggatcgct, rev: ctagagcgatccctgatggtgatcgct) and cloned them into pGL3 luciferase (Promega) using either HindIII (5'UTR) or XbaI (3'UTR). We obtained several constructs with different amounts and orientations of the MS2 binding loop. For this study we used constructs containing either three loops in forward or in reverse orientation for the 5'UTR, and one loop in forward or reverse orientation for the 3'UTR. All constructs were confirmed by sequencing. Primer sequences are provided in the Supplementary table 1.

RNA, RT-PCR and RT-qPCR

RNA extraction, radioactive RT-PCR, Phosphor imager quantification and RT-qPCR was done as previously described (Wilhelmi et al., 2016). Results of endogenous alternative splicing represent the mean value of at least three independent experiments with the according standard deviation. Significance, if not differently labeled, was calculated by Student's unpaired t-Test: *$p<0.05$, **$p<0.01$, ***$p<0.001$. Primer sequences are provided in the Supplementary table 1.

Immunoblotting, immunoprecipitation (IP) and antibodies

Cells were lysed with a buffer containing 60 mM Tris pH 7.5, 30 mM NaCl, 1 mM EDTA, 1% TritionX-100 and protease inhibitors. SDS-page and immunoblotting were performed according to standard protocols. For isolation of nuclear and cytoplasmic proteins, cells were incubated in low salt buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl and protease inhibitors) for cytoplasmic proteins followed by incubation in high salt buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol and protease inhibitors) to isolate nuclear proteins. For IPs, Flag- and HA-tagged proteins were overexpressed in HEK293T cells for 48 hours. 100 µg lysate was pre-incubated with protein A/G-plus agarose (Santa Cruz) in 400 µL RIPA buffer containing 400 mM or 1200 mM NaCl, 2% BSA and protease inhibitors in the presence or absence of 40 µg RNase A (Roth) and 100 units RNase T1 (Thermo Scientific). Afterwards, IP was performed with the pre-cleaned lysates using anti-Flag® M2 affinity gel (Sigma). After protein binding, beads were washed four times with RIPA buffer (10 mM Tris HCl pH 8, 400 mM/1200 mM NaCl, 2 mM EDTA, 1% NP-40, 5 mg/mL Sodium deoxycholate) and then boiled in 2x SDS loading dye. Protein expression and IPs were analyzed by SDS-page, blotting to a PVDF membrane and detection with the following antibodies: α-Flag (Cell Signaling, 2368), α-HA (Santa Cruz, sc-7392), α-hnRNPL (Santa Cruz, sc-32317), α-GAPDH (Epitope Biotech Inc., L001), α-Firefly Luciferase HRP (GeneTex, GTX20635), α-Vinculin (Santa Cruz, sc-5573).

Secondary structure prediction

Secondary structure prediction was performed with the PSIPRED online tool (Buchan et al., 2013).

Luciferase assays

For luciferase assays $1.5x10^5$ HEK293T or $1.0x10^5$ HeLa cells were seeded into a 12-well plate using DMEM medium containing 10% FBS. Cells were transfected with Lipofectamin 2000 according to the manufacturer's instructions, using 0.6 µg of the different pEF-MS2 constructs, 0.2 µg of the different pGL3 luciferase reporter constructs and 0.01 µg of a Renilla control vector (Promega). As a control, the pEF-MS2 constructs were replaced by GFP or an C-terminal GFP tagged variant of U2AF26ΔE7 (Heyd et al., 2008). 48 hours after transfection cells were harvested using the passive lysis buffer from p.j.k. according to the manufacturer's instructions. Luciferase activity (beetle juice, p.j.k.) and renilla activity (renilla juice, p.j.k) were measured in

technical duplicates by an Orion L microplate luminometer (Berthold Detection Systems). Average luciferase expression was normalized to averaged renilla expression. Data were normalized to the pGL3 control luciferase and pEF-MS2 expressing cells. Depicted data represent the average and standard deviation of several independent transfections.

Primary T cell preparation

Spleens of C57Bl6 mice where squeezed through a 70 µmeter cell strainer into complete medium (RPMI, 10% FCS, 1% Penicillin/ Streptomycin, 1x NEAA, 0,1 mM ß-Mercaptoethanol, 1 mM Pyruvat, 2 mM Glutamine) to get a single cell suspension. Cells were washed with PBS++ (1x PBS, 2% FCS/ 2 mM EDTA) followed by incubation in 1 mL red blood cell lysis buffer (Sigma) for 5 min at room temperature. After washing with 5 mL 1x PBS/ 2% FCS/ 2 mM EDTA, cells were counted and $5x10^7$ cells were used to purify T cells with the Dynabeads™ Untouched™ Mouse T Cells Kit (Thermo Fischer) following the manufacturer's protocol. For T cell stimulation, 12 well plates were coated with 1 µg mouse α-CD3 antibody (BD Pharmingen) in 500 µL 1xPBS for 2 hours at 37 °C followed by washing with 1 mL 1xPBS. $1x10^6$ purified T cells were seeded in 1 mL complete medium in the coated 12 well plate and were additionally treated with either 20 ng/µL PMA or 1 µg mouse α-CD28 antibody (BD Pharmingen) or were treated with 1 µL DMSO as control. 48 hours after the stimulation total RNA was isolated.

Stable EL4 cells

For generation of stable EL4 cells 5 µg of linearized pCMV-U2AF26ΔE7-Flag plasmid was transfected into $2x10^6$ EL4 cells by electroporation with the Nucleofector II (Amaxa Biosytems) electroporation system following the manufacturer's protocol. 48 hours after transfection, clones were selected with 4 mg/ml G418 (Sigma) in RPMI and clones were expanded.

Cell surface staining and flow cytometry

For EL4 cell surface staining $1x10^5$ EL4 cells were stained in 100 µL 1x PBS with 2% FCS and 2 mM EDTA with 1.5 µg/mL antibody for 15 minutes at 4°C. After washing with 1 mL 1xPBS with 2% FCS and 2 mM EDTA cells were analyzed by flow cytometry with a guava easyCyte 8 cytometer (Millipore). The mean signal intensity was calculated. Antibodies used for flow cytometry: anti CD3e-FITC, CD8a-FITC, CD90.1-PerCP-Vio700, CD45-FITC, CD28-PE, CD199(CCR9)-PE and Ly6a-PE from MACS-Miltenyi Biotec.

RESULTS

ZnF2 stabilizes U2AF35 and U2AF26

Despite the identification of U2AF35 as the AG-binding protein more than 15 years ago and the recent discovery of several cancer-associated mutations within U2AF35, surprisingly little is known about the contribution of the different protein domains to U2AF35 function. The ZnF domains have been suggested to mediate RNA-protein contacts and they also represent hotspots for somatic mutations associated with malignant transformation. We therefore cloned U2AF35 constructs that lack either the first (ZnF1) or the second (ZnF2) domain or introduced the cancer-associated mutation S34F in ZnF1 and Q157R and Q157P in ZnF2 of U2AF35 (Figure 1A). We expressed these U2AF35 constructs as Flag-tagged proteins in HEK293T cells and addressed a role in regulating protein stability, interaction with U2AF65 and in controlling alternative splicing.

To analyze stability of the different U2AF35 mutants we treated transfected cells with the translation inhibitor Cycloheximide (CHX) for different time points and analyzed protein expression by western blot. The full length U2AF35 protein (fl) has a protein half-life above 24 hours that was not affected by co-expression of U2AF65 (Figure 1B and C). Removal of the first zinc finger led to a slightly reduced basal expression level and a reduction of the half-life to less than 8 hours. Deletion of the second zinc finger had a strong impact on the basal expression (Supplementary Figure S1A) and the stability was reduced to a half-life below 4 hours (Figure 1B and C). This is not due to reduced solubility of the ZnF2-deleted U2AF35 as we do not observe an enrichment of this variant in the pellets of cell debris after whole cell extract preparation compared to the other U2AF35 mutants (Supplementary Figure S1B). These results suggest the second zinc finger of U2AF35 to play a role in controlling protein abundance and stability. Interestingly, both, reduced basal expression and decreased stability, were rescued by co-expression of U2AF65, suggesting that the formation of a U2AF heterodimer stabilizes the structure. Similar results were observed for its shorter paralog U2AF26 that, apart from its RS-rich domain, has 89% sequence conservation to U2AF35 (Supplementary Figure S1C). This is consistent with *in vitro* data showing that the RRM of U2AF35 adapts an ordered structure only upon binding of U2AF65 (Kellenberger et al., 2002) and suggests that the second ZnF contributes to the formation of a stabilized protein fold in the absence of U2AF65 interaction. In the ΔZnF1 construct we consistently observed a faster migrating protein species that, like the ΔZnF2 mutant

showed decreased basal expression and stability that was partially rescued by co-expression of U2AF65. This protein likely is the result of the usage of an internal ATG; based on size, it could be the methionine at position 74. This finding supports the idea that several parts of the protein have to be present, i.e. the RRM and the second ZnF, to allow a stable protein fold in cells in the absence of U2AF65.

Given the prominent role of the ZnF domains in controlling expression and stability of U2AF35, we further analyzed cancer-associated mutations falling within these domains, namely S34F (c.101C>T) in ZnF1 and Q157R (c.470A>G) and Q157P (c.470A>C) in ZnF2. Cancer-associated U2AF35 mutations have been primarily analyzed with respect to splicing defects and RNA-binding specificity but additional features conferred by these mutations have not been characterized in detail. We therefore tested the stability of the S34F, Q157R and Q157P mutants in the presence or absence of U2AF65. In these assays we observed slightly reduced basal expression for all three point mutants similar to the complete ZnF1 deletion and a protein half-life above 24 hours (Supplementary Figure S1A and Figure 1C). Again, the reduced basal expression of the S34F, Q157R and Q157P mutations was completely rescued by co-expression of U2AF65 suggesting that the cancer-associated mutations interfere with proper expression of the U2AF35 monomer in conditions where U2AF65 is limiting for the formation of the heterodimeric U2AF complex.

ZnF2 is required for stable U2AF heterodimer formation

Although the co-expression experiments suggested that all U2AF35 mutants were able to interact with U2AF65, we analyzed this interaction in more detail, as it is crucial for U2AF function. To this end, we performed CoIP experiments with U2AF65 and the different U2AF35 mutations and ZnF deletions. All U2AF35 constructs showed a similar interaction with U2AF65 under standard conditions in the presence and absence of RNase (Figure 2). Interestingly, this interaction was almost completely lost in more stringent (high salt) conditions for the mutant lacking ZnF2 (Figure 2A, bottom). Again we observed the same pattern for the paralog U2AF26 (Supplementary Figure S2A). These data is in agreement with the U2AF35 crystal structure from yeast, which shows the presence of a U2AF35-U2AF65 helical interaction surface in the C-terminal region of U2AF35, which is independent of the interaction mediated by the RRMs

(Yoshida et al., 2015). Although the C-terminus of mouse and yeast U2AF35 is not well conserved, a helix in the C-terminus of mouse U2AF35 directly downstream of the second zinc finger is predicted that may function as additional U2AF65 interaction surface (Supplementary Figure S2B). Deletion of the second ZnF in U2AF35 likely alters the location, orientation or fold of this interaction surface, which reduces the interaction strength and results in loss of heterodimer formation under stringent conditions. The disease-associated mutations S34F in the first and Q157R/P in the second ZnF did not affect the interaction with U2AF65, suggesting that they do not interfere with the C-terminal interaction domain. Together these data show a crucial role of the second ZnF domain of U2AF35 and U2AF26 for protein stability and formation of the U2AF heterodimer.

Both ZnFs are indispensable for splicing regulation

The structure of yeast U2AF35 showed that both ZnF domains of U2AF35 are positioned on its RRM and cooperatively bind to the target RNA sequence (Yoshida et al., 2015). To directly test a role of the ZnF domains in splicing regulation we used an established knock-down complementation assay with quantitative read-out, in which missplicing of exons sensitive to loss of endogenous U2AF35 can be rescued by expression of U2AF35 variants (Herdt et al., 2017; Preußner et al., 2014). Endogenous U2AF35 mRNA was knocked down to around 10% and replaced by the respective ectopically expressed mutants (Figure 3A and B). For our analysis we have chosen eight exons that, based on previous studies, show a substantial change in exon inclusion upon U2AF35 knock down (Herdt et al., 2017; Shao et al., 2014). For seven out of eight exons we observed decreased exon inclusion upon knock-down of U2AF35 and one exon showed the opposite effect (Figure 3C and Supplementary Figure S3), which is in agreement with previously published work (Herdt et al., 2017; Shao et al., 2014). Overexpression of U2AF35fl reversed this effect in all cases and led to an almost complete rescue of the splicing defect in six out of eight cases (TPD52L2, EIF4A2, CHECK2, CEP164, RIPK2 and THYN1) thus validating our experimental system. In contrast, mutants with deletion of either the first or the second ZnF domain completely lost the ability to rescue splicing defects induced by knock-down of endogenous U2AF35 (Figure 3C). For the effect of cancer-associated point mutants on alternative splicing we refer to previously published work (Herdt et al., 2017; Ilagan et al., 2014; Okeyo-Owuor et al., 2015b; Przychodzen et al., 2013; Shirai et al., 2015). Our data, showing that ZnF

deletions completely lose the ability to rescue splicing defects, are in agreement with the recent yeast U2AF35 structure, showing that both ZnF domains are required to make contact with the RNA (Yoshida et al., 2015). Stabilization of unstable ZnF-deleted U2AF35 mutants by co-expressing U2AF65 had no influence on the rescue abilities of U2AF35 (Figure 3D) ruling out that reduced protein levels of the mutants are causing the inability to rescue. U2AF35 knock-down was also rescued through the full length paralog U2AF26 but not by U2AF26 lacking either the first or the second zinc finger (Figure 3C). This is in agreement with *in vitro* splicing assays for two model RNAs were U2AF26 was shown to fully substitute for U2AF35 function in splicing regulation (Shepard et al., 2002).

The naturally occurring U2AF26ΔE7 behaves similar as the ZnF2 deletion mutant

Previous studies have shown that U2AF26 itself is alternatively spliced (13,14). Interestingly, exclusion of exon 7 in U2AF26 leads to the removal of three of the four zinc complexing residues of ZnF2 and 20 additional amino acids downstream of the ZnF2. We asked whether this endogenously occurring U2AF26 variant lacking a large part of the second zinc finger possesses similar functional impairments as observed for the artificial deletion of the full ZnF2 (Figure 4A). Skipping of exon 7 led to a slightly reduced basal expression and reduced stability of the U2AF26ΔE7 protein, but the effects were milder than for the full ZnF2 deletion (Figure 4B and Supplementary Figure 1C). Both basal protein expression and stability are rescued by co-expression of U2AF65 similar as seen for the full ZnF2 deletion. Binding of U2AF26ΔE7 to U2AF65 is similar as for full length U2AF26 in low salt conditions, but abolished in the presence of high salt (Figure 4C), thus resembling behavior of the ZnF2 deletion. This is consistent with the potential interaction helix in the C-terminal part of U2AF26 being absent in the isoform lacking exon 7 (Supp. Fig. S2B). U2AF26ΔE7 is also not able to rescue miss-splicing upon U2AF35 knock-down, which is probably due to both the predominant expression of this variant in the cytoplasm (Heyd et al., 2008) and the destroyed second ZnF2 (Figure 4D). These results confirm the functional studies made by artificially deleting ZnF2 in U2AF35 and U2AF26 in an endogenously occurring U2AF26 variant.

U2AF26ΔE7 regulates translation of a reporter when tethered to the 5'UTR

Skipping of U2AF26 exon 7 has a severe impact on the known function of U2AF26 in splicing regulation and in addition, the U2AF26ΔE7 protein was shown to mainly localize to the cytoplasm (Heyd et al., 2008). Therefore we next addressed a potential role of U2AF26ΔE7 on cytoplasmic RNAs. We set up a luciferase reporter assay where we cloned three MS2-stem loops upstream of the firefly-luciferase gene and recruited MS2-tagged U2AF26ΔE7 to the reporter (Figure 5A). The MS2-protein alone and the MS2-tagged U2AF26ΔE7 are both localized in the cytoplasm (Supplementary Figure S4A). Binding of MS2-U2AF26ΔE7 to the forward orientated MS2-loops in the 5' UTR of the reporter leads to a substantial and significant increased luciferase expression in HEK293T (Figure 5A) and HeLa cells (Supplementary Figure S4B). A construct with the reverse complementary sequences in the MS2 loops which cannot be bound by the MS2-protein was unaffected by the expression of MS2-U2AF26ΔE7 (Figure 5A and Supplementary Figure S4B). Furthermore, reporter constructs with insertions of forward and reverse orientated MS2-loops in the 3' UTR were unaffected by the expression of MS2-U2AF26ΔE7 (Supplementary Figure S4C). The increase in luciferase expression was mediated by binding of U2AF26ΔE7 through the fused MS2-domain as we did not detect a change in luciferase expression when we overexpressed a GFP-tagged U2AF26ΔE7 construct (Supplementary Figure S4D). Furthermore, the increased luciferase expression was not due to increased RNA turnover or altered mRNA export, as we observed similar levels of total and cytoplasmic luciferase mRNA in the presence of MS2-U2AF26ΔE7 compared to MS2 alone (Figure 5B). As an additional read out we used western blot analysis that shows a strong increase in Luciferase protein expression for the forward orientated MS2-loops when MS2-U2AF26ΔE7 is present (Figure 5C). As mRNA abundance and localization are not affected by the expression of MS2-tagged U2AF26ΔE7, increased expression of the reporter is most likely mediated at the level of translation, thus linking U2AF26ΔE7 with translational control. To identify the domain within U2AF26ΔE7 that is involved in controlling translation, we cloned MS2-tagged deletions of U2AF26ΔE7 and analyzed their activity in the luciferase reporter assay (Figure 5D). Deletion of ZnF1 reduced the reporter activity and additional deletion of the RNP2 fully abolished the activation of the reporter gene (Figure 5E). Expression of the ZnF1 alone partially activated the reporter and the RNP2 alone had an even stronger activation potential. The N-terminus including both domains showed a similar activation as the RNP2 alone (Figure 5E). These results were confirmed by western blot

analysis where a strong increase in luciferase expression was only observed in constructs including the ZnF1 and RNP2 domains (Supplementary Figure S4E). For the reporter with the reverse MS2-loop sequence no such regulation was detected, confirming specificity of the assay (Figure 5E and Supplementary Figure S4E). These results indicate that binding of U2AF26ΔE7 to the 5' UTR of target mRNAs might impact on translation efficiency of the bound mRNA through the N-terminal ZnF1 and RNP2 domains of U2AF26ΔE7.

U2AF26ΔE7 increases Ly6a expression in T cells

In a previous study we have observed that U2AF26 regulates alternative splicing of the T cell surface marker CD45 in cooperation with the transcription factor Gfi1 and thereby impacts on T cell activation (Heyd et al., 2006). Interestingly, in this study we observed the expression of several protein species in western blot analysis stained with a specific U2AF26 antibody, which are regulated upon activation of primary mouse T cells. We asked whether these protein species correspond to U2AF26 splice variants and purified primary T cell from Bl6-mice. We stimulated the T cells with CD3 antibody alone or in combination with CD28 antibody or the phorbol ester PMA (phorbol 12-myristate 13-acetate) and analyzed alternative splicing of U2AF26 by splicing sensitive RT-PCR. For all three stimuli we observed a strong increase of the U2AF26 variant lacking exon 7 whereas the second known isoform, U2AF26ΔE67, was only mildly affected (Figure 6A and B). The total amount of full length U2AF26 mRNA was unaffected, although the ratio is shifted towards the ΔE7 variant in activated T cells, suggesting an increase in mRNA expression (Figure 6A and B). To analyze the function of U2AF26ΔE7 in a model that can separate the effects of T cell activation from the effect specifically mediated by U2AF26ΔE7, we generated mouse EL4 T cells stably expressing Flag-tagged U2AF26ΔE7. To this end we generated single clones that stably express Flag-tagged U2AF26ΔE7 or used clones that show no expression as control (Figure 6C). In total we have analyzed five clones with stable expression of Flag-tagged U2AF26Δ7 and two clones without detectable amounts of U2AF26ΔE7 (Supplementary Figure S5B). Our previous analysis in Figure 5 points to a function of U2AF26ΔE7 in translational regulation. We therefore asked whether the expression of seven known T cell surface markers, CD3, CD8, CD90.1, CD45, CD28, CD199 or Ly6a is affected in our EL4 cells with stable U2AF26ΔE7 expression. The first six proteins showed no change in surface expression in the five U2AF26ΔE7 positive clones compared to the control cells

(Supplementary Figure S5A, note that CD8 is not expressed). However, for Ly6a all five U2AF26ΔE7 positive clones showed an increase in surface expression when compared to control (Figure 6D and Supplementary Figure S6A). The two EL4 cell lines with highest U2AF26ΔE7 expression (#3 and #7) showed a strong and significant increase in Ly6a expression. Analysis of the total mRNA abundance revealed a slight increase in Ly6a expression (Figure 6E). Furthermore, we also addressed Ly6a mRNA stability by treating the stable clones with the transcription inhibitor Actinomycin D for four and eight hours. In this experiment a slightly increased stability of the Ly6a mRNA in the U2AF26ΔE7 positive clones 3 and 7 compared to the control clones was observed (Figure 6F). As the effect of U2AF26ΔE7 expression on Ly6a mRNA abundance and stability is not significant and small compared to the 3-4 fold increase in protein expression, we suggest that U2AF26ΔE7 controls increased Ly6a expression at least partially at the level of translation.

DISCUSSION

In this report we show the significance of both ZnF domains of U2AF35 for key functionalities in mammalian cells: preserving U2AF35 stability, supporting interaction with U2AF65 and being crucial for the regulation of alternative splicing. In addition, we suggest a role of a cytoplasmic U2AF26 isoform in controlling translation in activated T cells thus providing evidence for a new function of U2AF26.

Our data show that ZnF2 plays a crucial role in stabilizing U2AF35, probably by allowing formation of a more stable protein structure, which in the absence of the zinc finger requires interaction with U2AF65 to form. Furthermore, we suggest ZnF2 to properly position a helix that acts as an additional interaction surface with U2AF65, which was shown to be present in the C-terminus of yeast U2AF35 (Yoshida et al., 2015). Yoshida et al. also describe reduced expression of recombinant yeast U2AF35 lacking the very C-terminal helix directly downstream of the second ZnF, whereas deletion of the first ZnF had no effect on protein expression and stability. In the present work, we have observed similar results for mouse U2AF35 in cell culture. Yoshida et al. also show that binding of the AG dinucleotide at the 3`ss is accomplished through both ZnFs and not through the RRM. In line with this, ZnF-deleted U2AF35 were unable to rescue splicing

of U2AF35-dependent alternative exons in cells. With our study we have thus confirmed the structure and in vitro based conclusions from yeast U2AF35 in cell culture for mouse U2AF35.

In parallel we performed the same experiments with the U2AF35 paralog U2AF26 and obtained similar results indicating that U2AF26 can functionally substitute U2AF35 in cells. This confirms *in vitro* data showing that recombinant U2AF26 interacts with U2AF65 and that U2AF26 can substitute U2AF35 in *in vitro* splicing assays (Shepard et al., 2002). Based on SELEX experiments a previous study suggested that U2AF26 and U2AF35 possess different RNA-binding specificities that differ at the +1 position located directly downstream of the AG-dinucleotide (Jeremiah Brian Shepard, 2004). However, in our knockdown and complementation assay U2AF26 rescues splicing of all eight U2AF35-dependent exons independent of the base at the +1 position (Guanine for ASUN, THYN1, EIF4A2; Cytosine for CHEK2, TPD52L2; Adenine for CEP164, RNF10; Thymine for RIPK2). Our results thus indicate that both proteins bind different 3' splice sites with similar affinity, although rescue efficiencies were slightly different between U2AF35 and U2AF26, which could indicate subtle binding preferences (e.g. RIPK2 in Fig. 3C).

Interestingly, we could verify our results with an alternative splice variant of U2AF26 (U2AF26ΔE7) that lacks the largest part of the second zinc finger and is endogenously expressed in different mouse cell lines and tissues (Heyd et al., 2008; Preußner et al., 2014, 2017). In this U2AF26 variant, exon 7 is excluded from the U2AF26 mRNA leading to cytoplasmic localization of the protein through loss of a non-canonical nuclear localization signal (Heyd et al., 2008).

We further focused our study on the function of cytoplasmic U2AF26ΔE7. There are several examples of splicing factors that localize to the cytoplasm upon alternative splicing and have functions in cytoplasmic mRNA processing. The alternative splicing regulator muscleblind-like 1 (MBNL1) plays a crucial role in the development of skeletal and heart muscles and is itself alternatively spliced during development (Lin et al., 2006; Terenzi and Ladd, 2010). MBNL1 lacking exon 5 localizes to both the nucleus and the cytoplasm in contrast to nuclear MBNL1 full length protein (Lin et al., 2006; Terenzi and Ladd, 2010). So far the function of cytoplasmic MBNL1ΔE5 is unknown. However, its paralog MBNL3 that lacks the region encoded by MBNL1 exon 5 is also localized into the nucleus and the cytoplasm. Cytoplasmic MBNL3 was shown to be associated with polysomes and to bind 3' UTRs of genes involved in cell growth and

proliferation (Poulos et al., 2013). Another example is the RNA-binding protein Quaking (Qk) in which alternative splicing of the C-terminus localizes the Qk6 and Qk7 splice variants in both the cytoplasm and the nucleus whereas the Qk5 splice variant is localized into the nucleus (Fagg et al., 2017). The authors show that the cytoplasmic Qk6 promotes its own translation and represses Qk5 translation through 3'UTR binding, although mechanistic details are still missing. MS2 tethering assays revealed a regulatory role of U2AF26ΔE7 when bound to the 5' UTR of a model mRNA. As U2AF26ΔE7 lacks the largest part of the ZnF2, RNA binding might be accomplished by the ZnF1 alone. RNA binding of a single ZnF was shown for one ZnF domain of the tristetraprolin protein (Michel et al., 2003), which, interestingly, has two CCCH-type ZnFs similar as U2AF26. Binding U2AF26ΔE7 to target RNA might also be mediate indirectly through recruitment by other RNA-binding proteins.

The most well-known RBP family that fulfills nuclear as well as cytoplasmic functions is the SR-protein family. SR-proteins contain one or two RNA recognition motives and a C-terminal Arginine-Serine-rich domain which functions as a protein interaction platform and is important for shuttling between cytoplasm and nucleus. Direct functions of cytoplasmic SR-proteins where shown in the regulation of mRNA stability and translation (reviewed in (Twyffels et al., 2011)). For example, SRSF1 was shown to bind to the 3'UTR of protein-kinase-C-interacting protein mRNA and to induce its degradation (Lemaire et al., 2002). Furthermore, an association of SRSF1 and SRSF7 with monosomes or polysomes was shown that appeared to be regulated by phosphorylation (Sanford et al., 2005). The RRM2 of SRSF1 seems to play an important role in controlling translation, probably through protein-protein interactions mediated by this domain (Sanford et al., 2005). We obtained similar results for U2AF26ΔE7 showing that deletion of the RRM2 and the ZnF1 leads to the loss of translational activation (Figure 5E). The identification of RRM2/ZnF1 interacting proteins will be important to unravel the mechanism by which cytoplasmic U2AF26ΔE7 regulates translation when bound to the 5'UTR of target mRNAs. As U2AF26 contains several RS-repeats in its C-terminal domain that are not affected by the skipping of exon 7 it will also be very interesting to analyze whether these repeats are phosphorylated and impact on translational regulation by U2AF26ΔE7.

Interestingly, we identified an upregulation of U2AF26ΔE7 expression during activation of primary mouse T cells (Figure 6A). Activated T cells were shown to undergo many splicing changes which we just start to understand functionally (Martinez et al., 2012; Schultz et al., 2016;

Wilhelmi et al., 2016). Utilizing mouse T cells stably overexpressing U2AF26ΔE7 we identified Ly6a to be upregulated upon U2AF26ΔE7 expression (Figure 6D). Ly6a (lymphocyte antigen 6 a) also known as Sca-1 (stem cell antigen-1) is a cell surface marker expressed on hematopoietic stem cells and on primary CD4-positive T cells where its expression is strongly upregulated upon activation (Holmes and Stanford, 2007). It was suggested that Ly6a has both a positive and negative role in T cell response upon antigen stimulation probably mediated through TCR signaling (Flood et al., 1990; Hanson et al., 2003; Henderson et al., 2002; Stanford et al., 1997). However, mice lacking Ly6a expression seem to develop normal primary and memory T cells and are unaffected in the response to virus infection *in vivo* (Whitmire et al., 2009). U2AF26fl was shown to reduce T cell activation by promoting the formation of a less active CD45 protein isoform through alternative splicing regulation (Heyd et al., 2006). U2AF26ΔE7 might further affect the T cell response by increasing the translation of Ly6a and potentially other cell surface markers.

Taken together, our data show that the ZnF domains control key functionalities of U2AF35 and U2AF26 and connect this core splicing factor to the regulation of translation. Whether the link to translation is also relevant for cancer-associated U2AF35 mutations is an important question to be addressed in future studies.

FUNDING

This work was supported by Emmy-Noether-Fellowship by the Deutsche Forschungsgemeinschaft (DFG) [He5398/3].

ACKNOWLEGMENTS

We would like to thank members of the Heyd lab for discussion and comments on the manuscript.

AUTHOR CONTRIBUTIONS

O.H. and M.P. performed experiments. F.H., O.H. and M.P. designed the study, planned the experiments and analyzed the data. F.H. and O.H. wrote the manuscript. F.H. initiated and supervised the work.

CONFLICT OF INTEREST

The authors report no conflict of interest.

REFERENCES

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. *41*, 349–357.

Fagg, W.S., Liu, N., Fair, J.H., Shiue, L., Katzman, S., Donohue, J.P., and Ares Jr, M. (2017). Autogenous cross-regulation of Quaking mRNA processing and translation balances Quaking functions in splicing and translation. Genes Dev. *31*, 1–16.

Flood, P.M., Dougherty, J.P., and Ron, Y. (1990). Inhibition of Ly-6A antigen expression prevents T cell activation. J. Exp. Med. *172*, 115–120.

Gama-Carvalho, M., Carvalho, M.P., Kehlenbach, A., Valcarcel, J., and Carmo-Fonseca, M. (2001). Nucleocytoplasmic shuttling of heterodimeric splicing factor U2AF. J. Biol. Chem. *276*, 13104–13112.

Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2012). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. *44*, 53–57.

Hanson, P., Mathews, V., Marrus, S.H., and Graubert, T.A. (2003). Enhanced green fluorescent protein targeted to the Sca-1 (Ly-6A) locus in transgenic mice results in efficient marking of hematopoietic stem cells in vivo. Exp. Hematol. *31*, 159–167.

Henderson, S.C., Kamdar, M.M., and Bamezai, A. (2002). Ly-6A.2 expression regulates antigen-specific CD4+ T cell proliferation and cytokine production. J. Immunol. *168*, 118–126.

Herdt, O., Neumann, A., Timmermann, B., and Heyd, F. (2017). The cancer-associated U2AF35 470A>G (Q157R) mutation creates an in-frame alternative 5' splice site that impacts splicing regulation in Q157R patients. RNA J. *23*, 1796–1806.

Heyd, F., ten Dam, G., and Möröy, T. (2006). Auxiliary splice factor U2AF656 and transcription factor Gfi1 cooperate directly in regulating CD45 alternative splicing. Nat. Immunol. *7*, 859–867.

Heyd, F., Carmo-Fonseca, M., and Möröy, T. (2008). Differential isoform expression and interaction with the P32 regulatory protein controls the subcellular localization of the splicing factor U2AF656. J. Biol. Chem. *283*, 19636–19645.

Holmes, C., and Stanford, W.L. (2007). Concise review: stem cell antigen-1: expression, function, and enigma. Stem Cells *25*, 1339–1347.

Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2014). U2AF1 mutations alter splice site recognition in hematological malignancies.

Genome Res. *124*, 14–26.

Jeremiah Brian Shepard (2004). Dissertation: Characterization of U2AF656, a paralog of the splicing factor U2AF35. Univ. Texas Southwest. Med. Cent. 124.

Kellenberger, E., Stier, G., and Sattler, M. (2002). Induced folding of the U2AF35 RRM upon binding to U2AF65. FEBS Lett. *528*, 171–176.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A Novel Peptide Recognition Mode Revealed by the X-Ray Structure of a Core U2AF35 / U2AF65 Heterodimer. Cell *106*, 595–605.

Kielkopf, C.L., Lücke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. Genes Dev. *18*, 1513–1526.

Kralovicova, J., and Vorechovsky, I. (2016). Alternative splicing of *U2AF1* reveals a shared repression mechanism for duplicated exons. Nucleic Acids Res. *2*, gkw733.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L., and Lafyatis, R. (2002). Stability of a PKCI-1 -related mRNA is controlled by the splicing factor ASF / SF2 : a novel function for SR proteins. Genes Dev. *16*, 594–607.

Lin, X., Miller, J.W., Mankodi, A., Kanadia, R.N., Yuan, Y., Moxley, R.T., Swanson, M.S., and Thornton, C.A. (2006). Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. Hum. Mol. Genet. *15*, 2087–2097.

Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcárcel, J., and Sattler, M. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature *475*, 408–411.

Martinez, N.M., Pan, Q., Cole, B.S., Yarosh, C. a, Babcock, G. a, Heyd, F., Zhu, W., Ajith, S., Blencowe, B.J., and Lynch, K.W. (2012). Alternative splicing networks regulated by signaling in human T cells. RNA J. *18*, 1029–1040.

Merendino, L., Guth, S., Bilbao, D., Concepcion, M., and Juan, V. (1999). Inhibition of msl-2 splicing by Sex- lethal reveals interaction between U2AF35 and the 3`splice site AG. Nature *402*, 838–841.

Michel, S.L.J., Guerrerio, A.L., and Berg, J.M. (2003). Selective RNA binding by a single CCCH zinc-binding domain from Nup475 (tristetraprolin). Biochemistry *42*, 4626–4630.

Mollet, I., Barbosa-Morais, N.L., Andrade, J., and Carmo-Fonseca, M. (2006). Diversity of human U2AF splicing factors. FEBS J. *273*, 4807–4816.

Okeyo-Owuor, T., White, B.S., Chatrikhi, R., Mohan, D.R., Kim, S., Griffith, M., Ding, L., Ketkar-Kulkarni, S., Hundal, J., Laird, K.M., et al. (2015a). U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. Leukemia *29*, 909–917.

Okeyo-Owuor, T., White, B.S., Chatrikhi, R., Mohan, D.R., Kim, S., Griffith, M., Ding, L., Ketkar-Kulkarni, S., Hundal, J., Laird, K.M., et al. (2015b). U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. Leukemia *29*, 909–917.

Poulos, M.G., Batra, R., Li, M., Yuan, Y., Zhang, C., Darnell, R.B., and Swanson, M.S. (2013). Progressive impairment of muscle regeneration in muscleblind-like 3 isoform knockout mice. Hum. Mol. Genet. *22*, 3547–3558.

Preußner, M., Wilhelmi, I., Schultz, A.-S., Finkernagel, F., Michel, M., Möröy, T., and Heyd, F. (2014). Rhythmic U2AF656 alternative splicing controls PERIOD1 stability and the circadian clock in mice. Mol. Cell *54*, 651–662.

Preußner, M., Goldammer, G., Neumann, A., Haltenhof, T., Rautenstrauch, P., Müller-McNicoll, M., and Heyd, F. (2017). Body Temperature Cycles Control Rhythmic Alternative Splicing in Mammals. Mol. Cell *67*, 1–14.

Przychodzen, B., Jerez, A., Guinta, K., Sekeres, M. a., Padgett, R., Maciejewski, J.P., and Makishima, H. (2013). Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. Blood *122*, 999–1006.

Sanford, J.R., Ellis, J.D., Cazalla, D., and Cáceres, J.F. (2005). Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. PNAS *102*, 15042–15047.

Schultz, A.-S., Preußner, M., Bunse, M., Karni, R., and Heyd, F. (2016). Activation-dependent TRAF3 exon 8 alternative splicing is controlled by CELF2 and hnRNP C binding to an upstream intronic element. Mol. Cell. Biol. *37*, e00488–1.

Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H., et al. (2014). Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat. Struct. Mol. Biol. *21*, 997–1005.

Shepard, J., Reick, M., Olson, S., and Graveley, B.R. (2002). Characterization of U2AF656, a splicing factor related to U2AF35. Mol. Cell. Biol. *22*, 221–230.

Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E.J., Okeyo-Owuor, T., et al. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell *27*, 631–643.

Stanford, W.L., Haque, S., Alexander, R., Liu, X., Latour, a M., Snodgrass, H.R., Koller, B.H., and Flood, P.M. (1997). Altered proliferative response by T lymphocytes of Ly-6A (Sca-1) null mice. J. Exp. Med. *186*, 705–717.

Terenzi, F., and Ladd, A.N. (2010). Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity. RNA Biol. *7*, 43–55.

Twyffels, L., Gueydan, C., and Kruys, V. (2011). Shuttling SR proteins: More than splicing factors. FEBS J. *278*, 3246–3255.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Webb, C.J., and Wise, J.A. (2004). The Splicing Factor U2AF Small Subunit Is Functionally Conserved between Fission Yeast and Humans. Mol. Cell. Biol. *24*, 4229–4240.

Whitmire, J.K., Eam, B., and Whitton, J.L. (2009). Mice deficient in stem cell antigen-1 (Sca1, Ly-6A/E) develop normal primary and memory CD4+ and CD8+ T-cell responses to virus infection. Eur. J. Immunol. *39*, 1494–1504.

Wilhelmi, I., Kanski, R., Neumann, A., Herdt, O., Hoff, F., Jacob, R., Preußner, M., and Heyd, F. (2016). Sec16 alternative splicing dynamically controls COPII transport efficiency. Nat. Commun. *7*, 12347.

Wu, T., and Fu, X.-D. (2015). Genomic functions of U2AF in constitutive and regulated splicing. RNA Biol. *12*, 479–485.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832–835.

Yoshida, H., Park, S., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., Tsuda, K., Kuwasako, K., Unzai, S., Muto, Y., et al. (2015). A novel 3′ splice site recognition by the two zinc fingers in the U2AF small subunit. Genes Dev. *29*, 1–12.

Yoshida, K., Sanada, M., Shiraishi, Y., and Nowak, D. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609–614.

Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, a I., and Green, M.R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. PNAS *89*, 8769–8773.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. Nature *402*, 835–838.

FIGURES

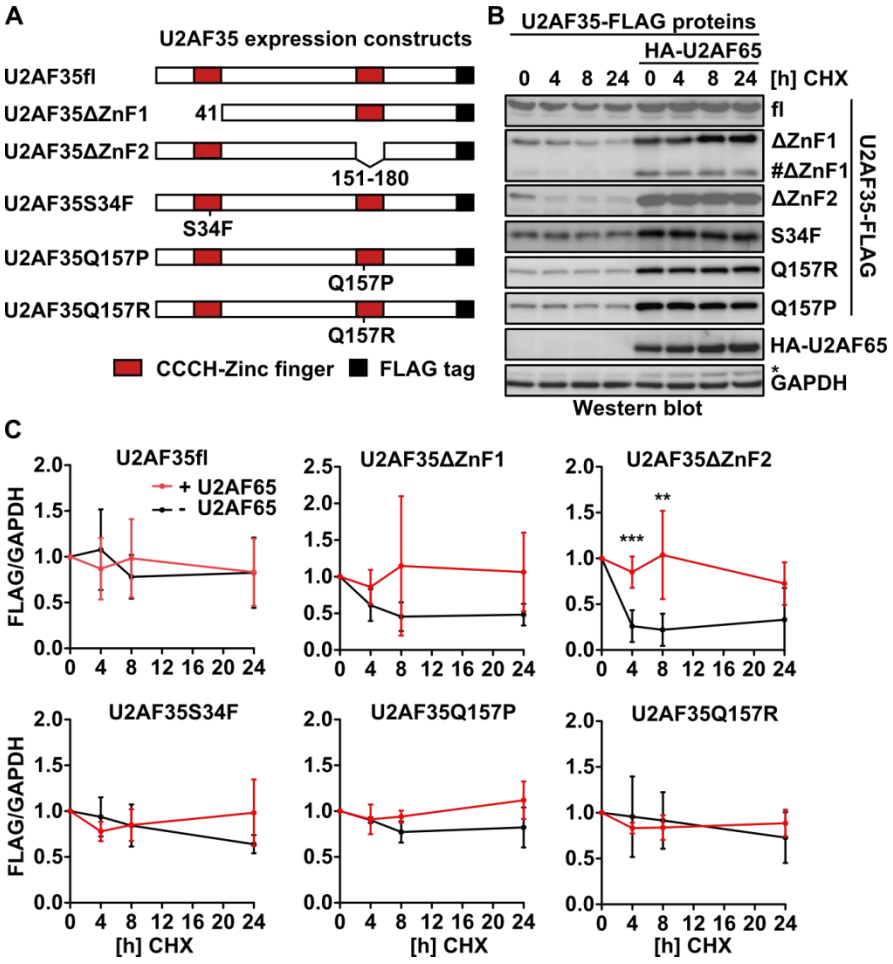


Figure 1: The second zinc finger of U2AF35 is required for protein stability.

(A) Scheme of Flag-tagged U2AF35 expression constructs with deletions or point mutations in the zinc finger domains. **(B)** Deletion of the second zinc finger destabilizes U2AF35. HEK293T cells were transfected with the constructs depicted in (A) in the presence or absence of HA-U2AF65 and incubated with Cycloheximide (CHX) for the indicated times. Protein expression was analyzed through western blotting with antibodies against Flag- and HA-tag. GAPDH was used as loading control. ΔZnF1: product of an alternative translational start site. **(C)** Quantifications of experiments shown in (B). Shown is the mean Flag signal normalized to GAPDH ± SD; n≥3; **p<0.01; ***p<0.001.

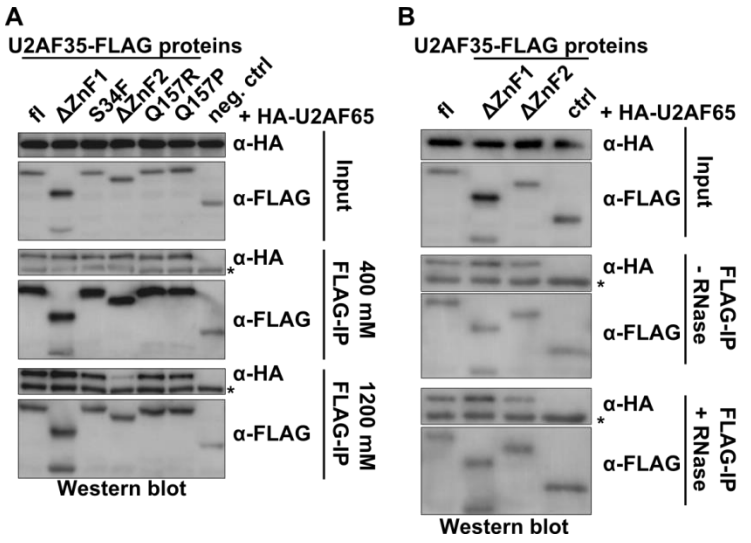


Figure 2:. The second zinc finger of U2AF35 is important for stable interaction with U2AF65.

(A) HEK293T cells were transfected with Flag-tagged U2AF35 expression constructs with deletions or mutations in the zinc fingers and with full length HA-tagged U2AF65. 48 hours after transfection α-Flag immunoprecipitation (IP) was performed in the presence of 400 or 1200 mM NaCl. Input and IP were analyzed with α-HA and α-Flag antibody. fl: full length; ctrl: Flag-tagged control protein. **(B)** Co-IP as in (A) with 400 mM NaCl in the presence or absence of RNase A and T1.

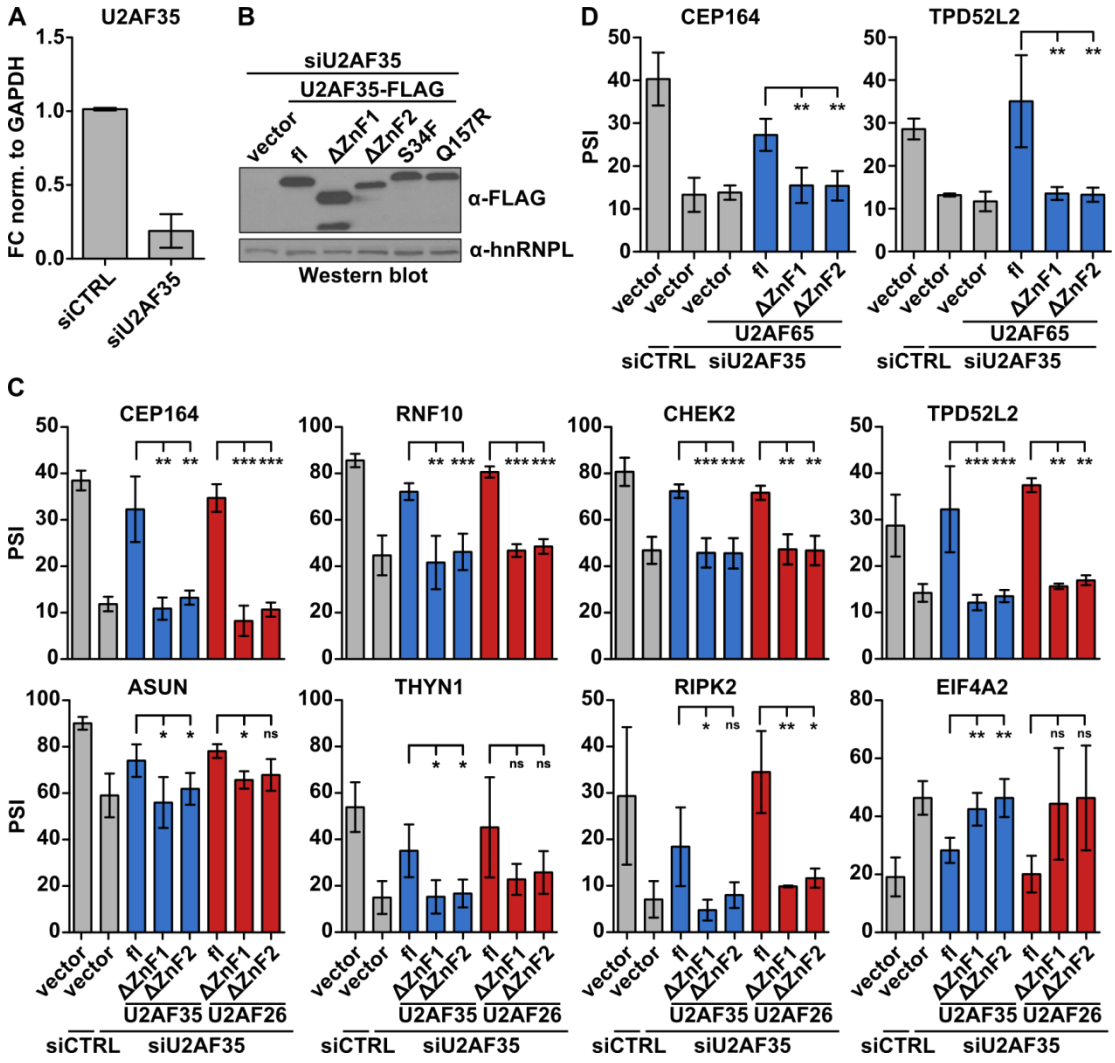


Figure 3: Both ZnFs are indispensable for U2AF35-dependant regulation of alternative splicing.

(A) Expression of Flag-tagged mouse U2AF35 mutants in HEK293T cells in the presence of siRNA against endogenous human U2AF35. **(B)** Quantitative RT-PCR showing the knock down efficiency of endogenous human U2AF35 in HEK293T cells 72 hours after transfection. **(C, D)** Quantifications of splicing sensitive radioactive RT-PCRs that are exemplarily shown in the Supplementary Figure S3. HEK293T cells were transfected with siCTRL or siU2AF35 for 24 hours and rescued with U2AF35 or U2AF26 mutants for an additional 48 hours prior to total RNA preparation. Alternative splicing of U2AF35-dependent cassette exons was analyzed in the absence **(C)** or in the presence **(D)** of HA-U2AF65. Shown is the mean percent spliced in (PSI) value ±SD. n=4; *p<0,05; **p<0,01; ***p<0,001.

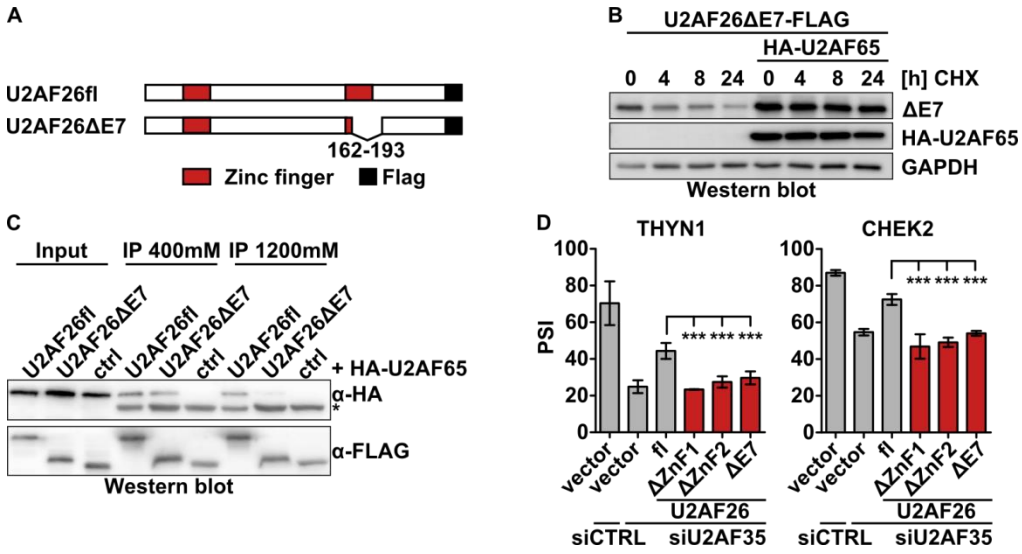


Figure 4: The naturally occurring U2AF26ΔE7 behaves similar as the ZnF2 deletion mutant.

(A) Scheme of Flag-tagged U2AF26fl and U2AF26ΔE7 expression constructs. **(B)** HEK293T cells were transfected with the constructs depicted (A) in the presence or absence of HA-U2AF65 and incubated with Cycloheximide (CHX) for the indicated times. Protein expression was analyzed through western blot with antibodies against Flag- and HA-tag. GAPDH was used as loading control. **(C)** Co-immunoprecipitation of Flag-tagged U2AF26fl, U2AF26ΔE7 and control protein with HA-tagged U2AF65 in the presence of 400 or 1200 mM NaCl. **(D)** Knockdown and complementation assay as described in Figure 3C (mean PSI ± SD). n=4 ; *p<0.05.

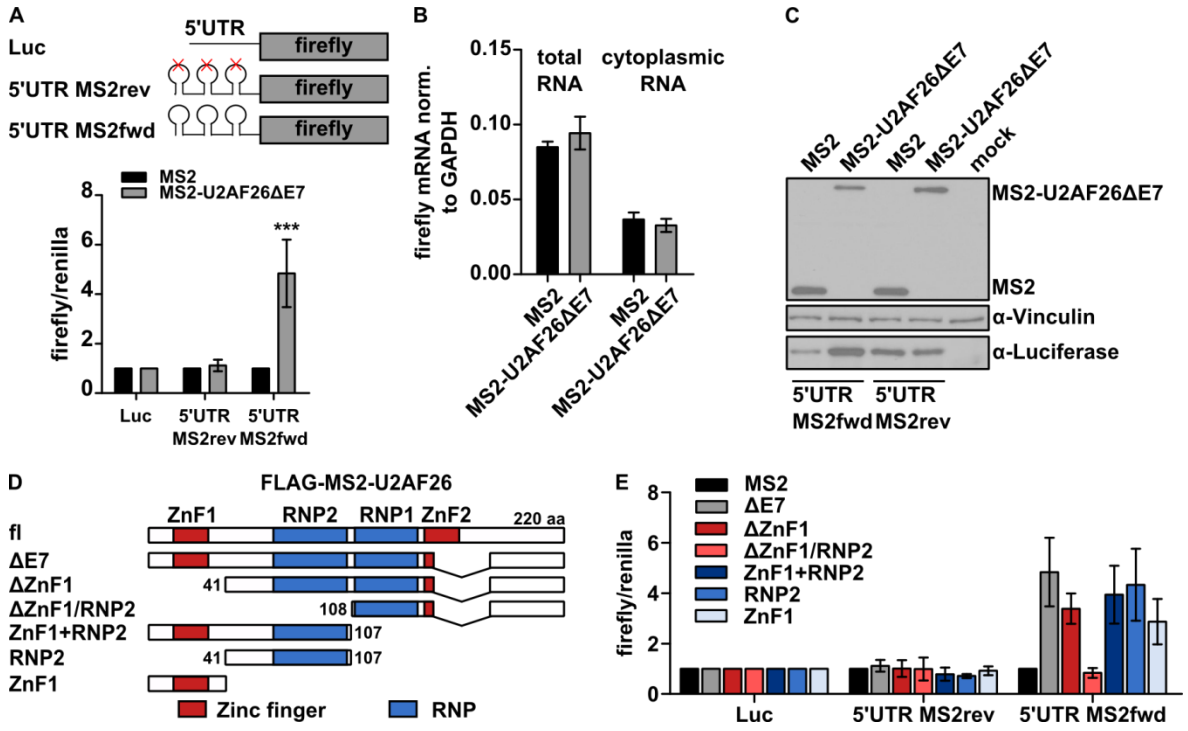


Figure 5: Cytoplasmic MS2-U2AF26ΔE7 regulates translation in a MS2 tethering assay.

(A) Luciferase assay with different firefly luciferase reporter constructs in presence of MS2 protein or MS2-tagged U2AF26ΔE7. The 5' untranslated region (UTR) of the firefly gene was either unmodified (Luc) or modified with MS2 binding loops in reverse (rev) or forward (fwd) orientation. Firefly luciferase signal was normalized to a cotransfected renilla control and MS2-alone signal (mean ± SD). n=9, one sample t-test, ***p<0.001. **(B)** mRNA expression of firefly luciferase with MS2fwd loops in the 5'UTR in total or cytoplasmic RNA in the presence of MS2- or MS2-U2AF26ΔE7. Expression was analyzed by RT-qPCR and normalized to GAPDH (mean ± SD). n=3. **(C)** Western blot analysis of HEK293T cells transfected with firefly luciferase constructs and MS2- or MS2-U2AF26ΔE7. MS2-proteins were detected via the Flag-tag and Vinculin was used as loading control. **(D)** Scheme of Flag-tagged MS2-U2AF26 expression constructs used for luciferase assays to identify translation regulatory domain. **(E)** Luciferase assay as in **(A)** with expression constructs depicted in **(D)**. Shown is the normalized luciferase mean signal ± SD. n≥4.

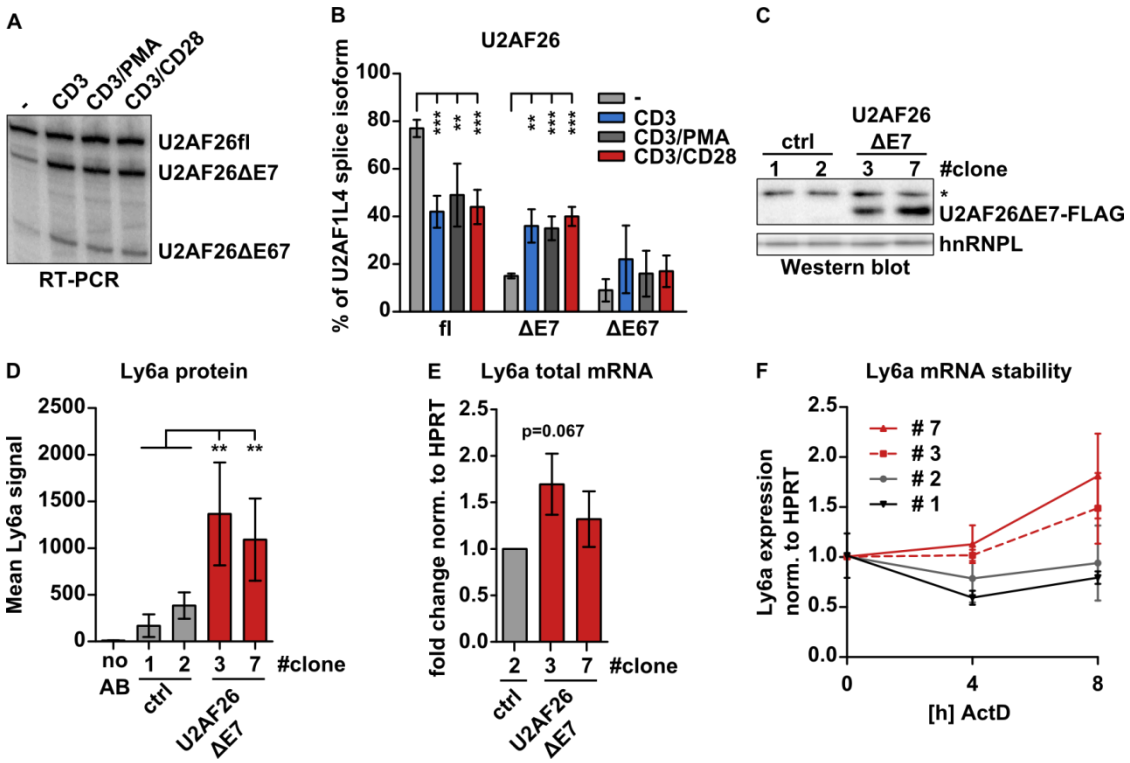


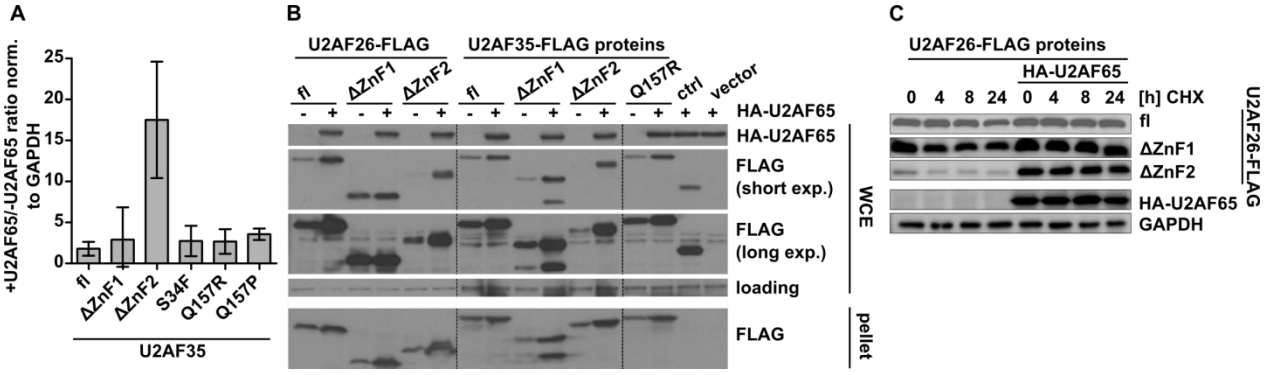
Figure 6: U2AF26ΔE7 increases Ly6a expression in T cells.

(A) Primary mouse T cells were activated with the depicted stimuli for 48 hours. Alternative splicing of U2AF26 was analyzed by radioactive RT-PCR with primers in exon 4 and exon 8. **(B)** Quantification of U2AF26 splice isoforms shown in **(A)**. n=4; **p<0.01; ***p<0.001. **(C)** α-Flag western blot analysis of EL4 single cell clones stably expressing Flag-tagged U2AF26ΔE7 and control clones without expression. hnRNPL was used as loading control. **(D)** Surface expression of Ly6a protein on stable EL4 cells analyzed by flow cytometry. Depicted is the mean Ly6a signal intensity ± SD. n≥3; **p<0.01. **(E)** Fold change of Ly6a mRNA expression in EL4 cell clones normalized to HPRT (mean ± SD). One sample T-test. n=3. **(F)** Ly6a mRNA stability in EL4 stable clones treated with Actinomycin D for 4 and 8 hours. Expression was normalized to HPRT (mean ±SD). #1: n=2; #2/3/7: n=4.

SUPPLEMENTARY TABLES AND FIGURES

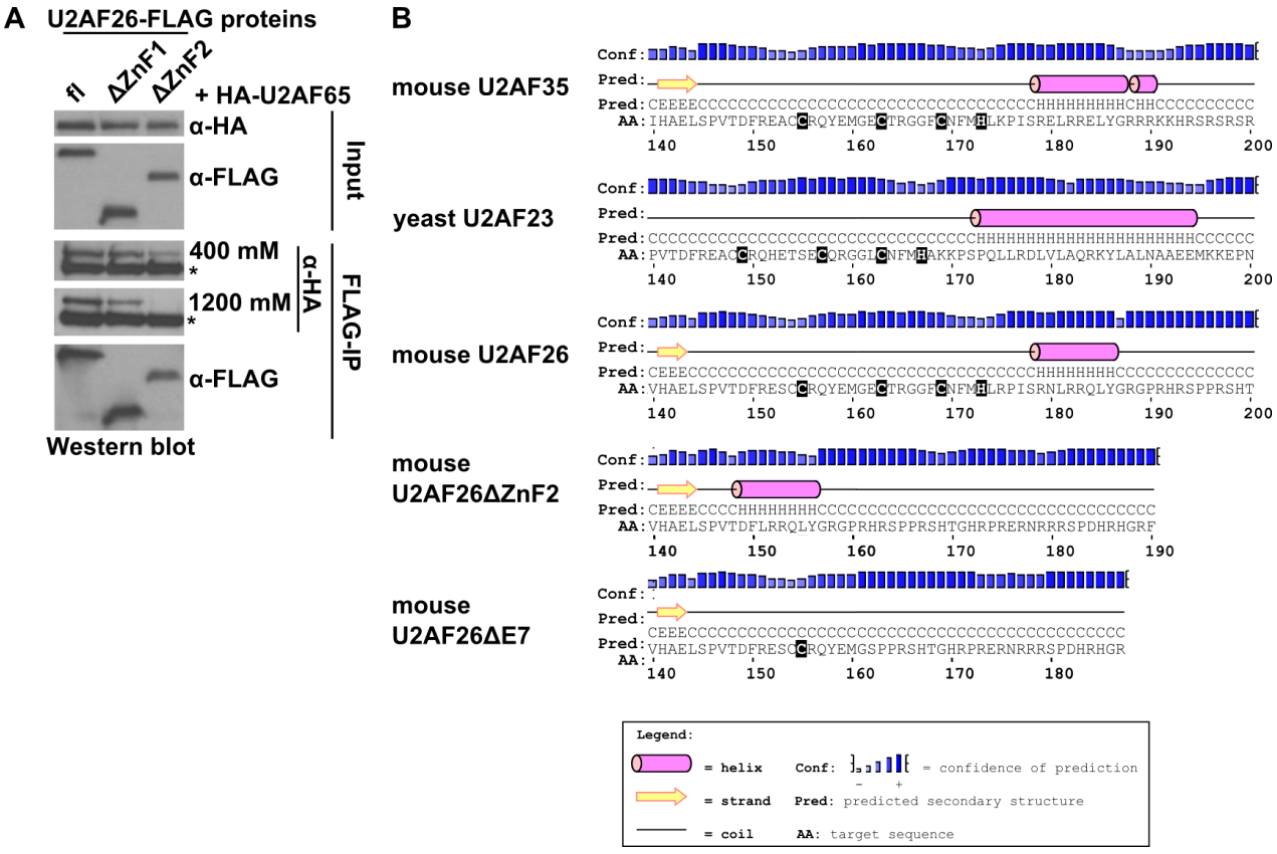
Cloning	forward	reverse
U2AF35fl	GCGGATCCatggcggaatacttggcctccatc	GCCTCGAGtcagaatcgtccagatctctcccg
U2AF35ΔZnF1	GCCTCGAGATGACCTTTAGCCAGACCATTGCCCTCT	GCCTCGAGtcagaatcgtccagatctctcccg
U2AF35ΔZnF2	ccccagtaactgacttcCTACGACGGGAGCTG	CAGCTCCCGTCGTAGgaagtcagttactgggg
U2AF35S34F	ggagacagatgtTTTcggttgcac	gtgcaaccgAAAacatctgtctcc
U2AF35Q157R	cctgctgccgcCGGtatgaaatgg	ccatttcataCCGgcggcagcagg
U2AF35Q157P	cctgctgccgcCCGtatgaaatgg	ccatttcataCGGgcggcagcagg
U2AF26fl	GCCTCGAGATGGCTGAATATTTAGCTTCGATATTCGGG	GCGGATCCGAAGCGACCATGCCGG
U2AF26ΔZnF1	GCAAGCTTACTTTCAGCCAGACCATAGTC	GCGGATCCGAAGCGACCATGCCGG
U2AF26ΔZnF2	ctcctgtcactgacttcTTGCGCCGGCAGCTC	GAGCTGCCGGCGCAAgaagtcagtgacaggag
RT-qPCR	forward	reverse
U2AF35	GACCACCTGGTGGGGAAC	GGATCGGCTGTCCATTAAAC
GAPDH	CTTCGCTCTCTGCTCCTCCTGTTCG	ACCAGGCGCCCAATACGACCAAAT
Radioactive RT-PCR	forward	reverse
THYN1	ACAAGAGGAAAGGGCCCTAA	CCTCCAGGCTCAAAACAAAA
CHEK2	CAGCTCTCAATGTTGAAACAGAA	CTGCACAGCCAAGAGCATC
ASUN	TTGCACGTCCTTAGCAGTTC	TTAGGGCCCTTTCCTCTTGT
TPD52L2	GGGAGCTGAAACAGAACCTG	TCTCCAAGCTTCCTGCTGAT
RIPK2	CATTCCCTACCACAAACTCG	AGGATGCGAAATCTCAATGG
CEP164	TCAGCCTCTTGGGTTTAGGA	GAGGGCTTGCAGGGAGTAG
EIF4A2	GGTATGATGTGATTGCTCAAGCTCAG	GTCTCCTTGAACTCAATCTCCAACTG
SETX	TCACACGAGCCAAGTACAGC	GCACAGGCTTGAGTTTCAGA
U2AF35fl	GCGGATCCatggcggaatacttggcctccatc	GCCTCGAGtcagaatcgtccagatctctcccg
U2AF35ΔZnF1	GCCTCGAGATGACCTTTAGCCAGACCATTGCCCTCT	GCCTCGAGtcagaatcgtccagatctctcccg
U2AF35ΔZnF2	ccccagtaactgacttcCTACGACGGGAGCTG	CAGCTCCCGTCGTAGgaagtcagttactgggg

Supplementary Table 1: Primers used in this study to perform splicing sensitive RT-PCR, quantitative RT-PCR and cloning of expression constructs.

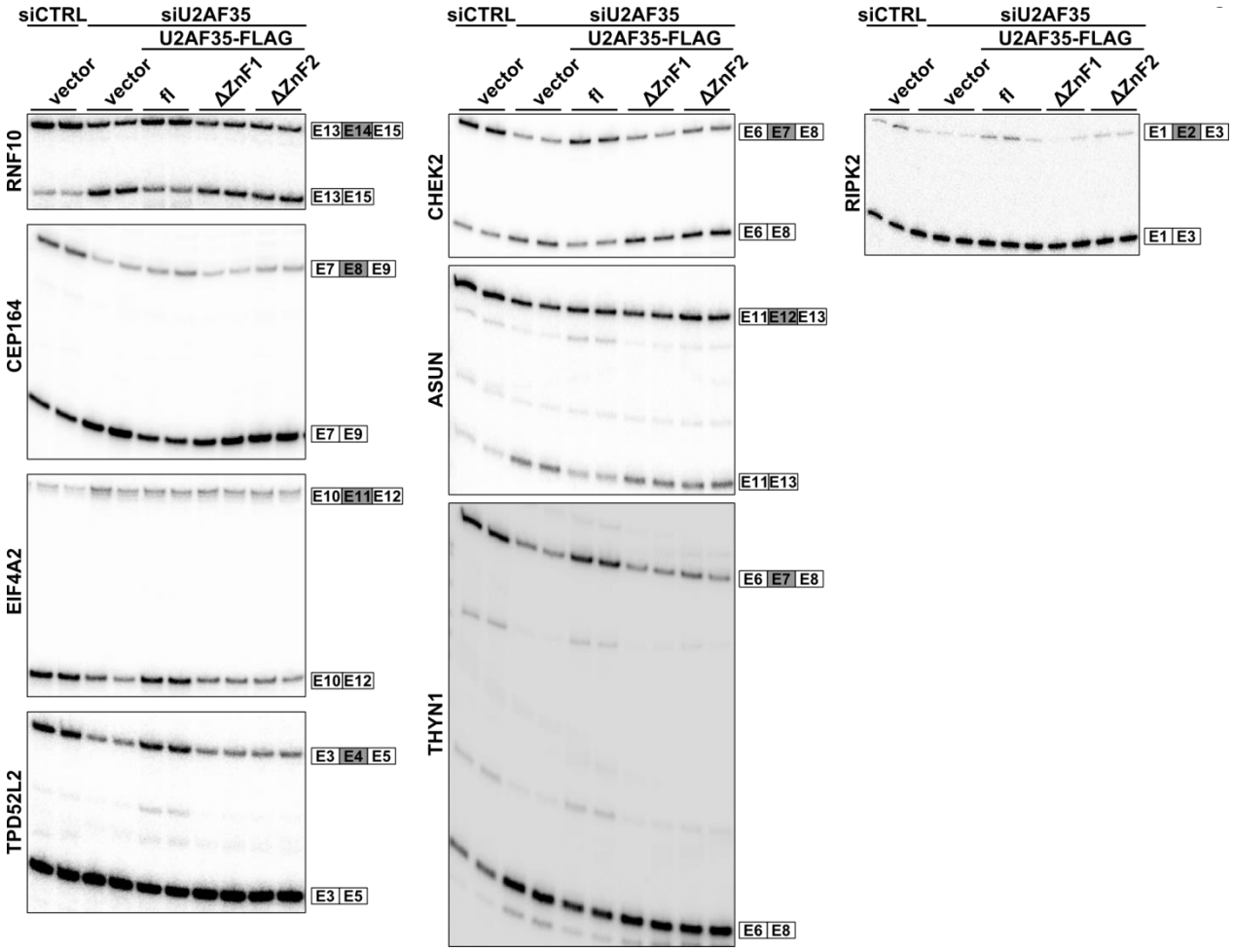


Supplementary Figure S1:. Solubility and stability of U2AF35 and U2AF26. **(A)** The basal level of U2AF35 mutants is increased in the presence of U2AF65. Protein expression was quantified for 0 h CHX, normalized to GAPDH and to the expression in the absence of U2AF65. **(B)** Presence of U2AF35 and U2AF26 deletion and point mutants in the supernatant of a whole cell extract (WCE) or in the cell pellet

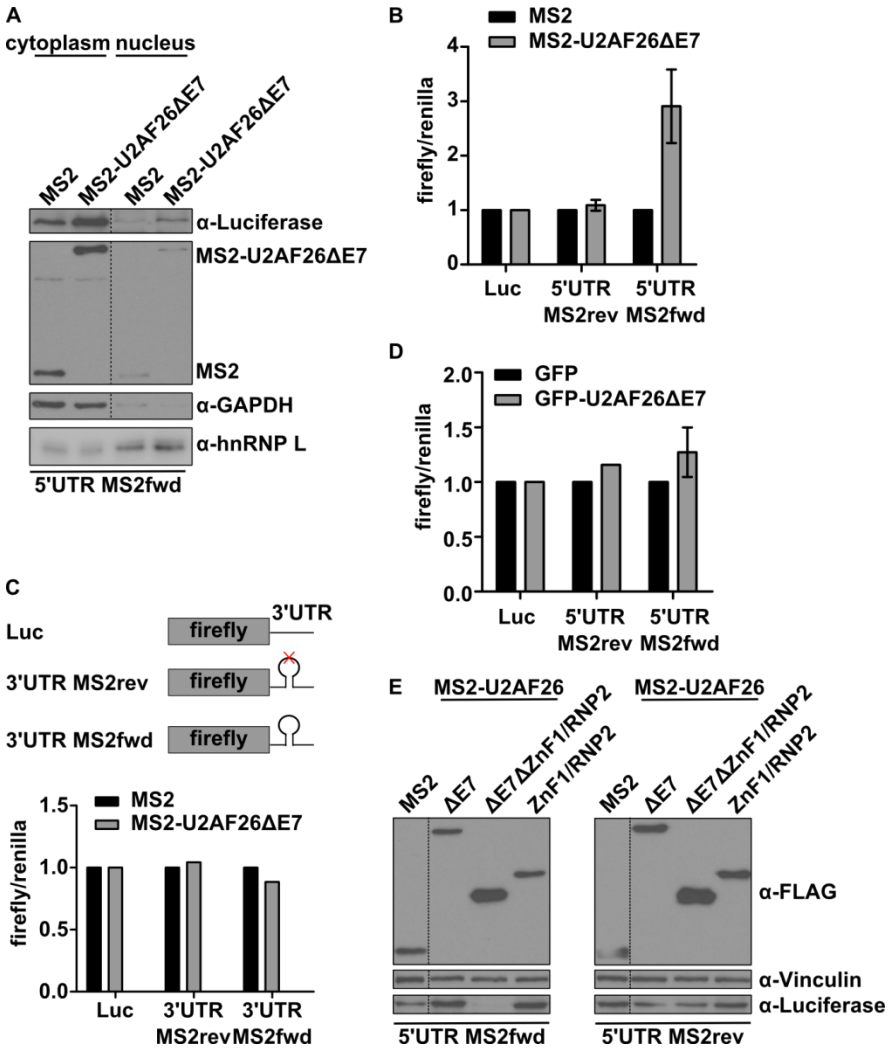
of cell debris post lysis. HEK293T cells were transfected, lysed and supernatant was analyzed as WCE via western blot. The debris pellets were resuspended and boiled in 2xSDS loading buffer and analyzed via western blot. **(C)** Protein stability experiment as in Figure 1B with Flag-tagged U2AF26fl and zinc finger deletion mutants.



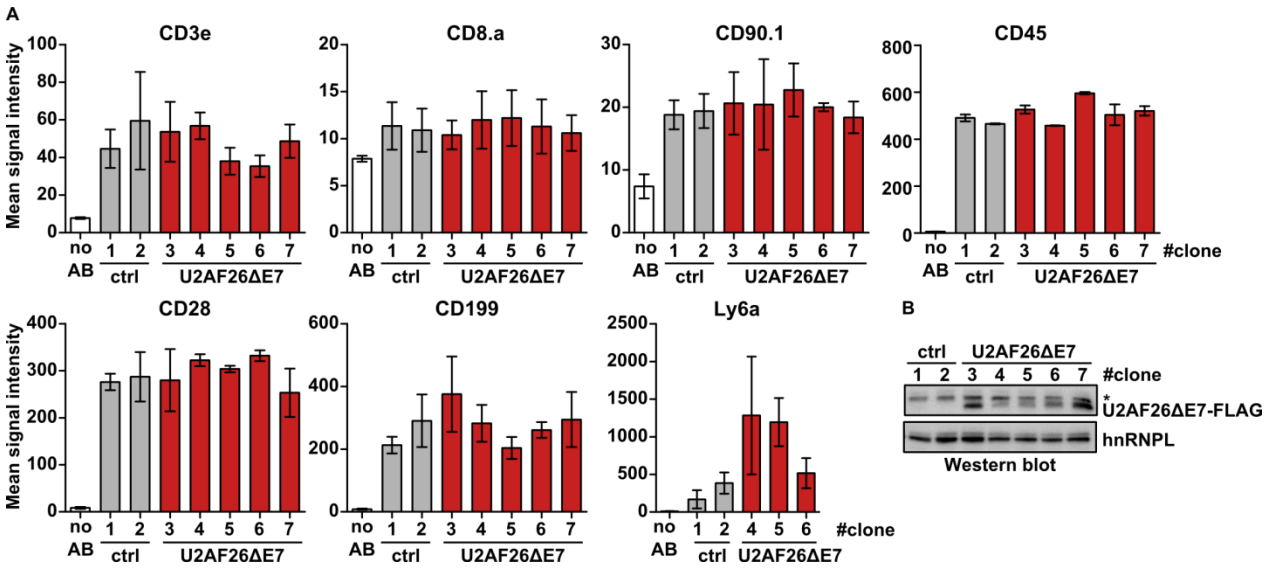
Supplementary Figure 2: Interaction of U2AF26 with U2AF65 and secondary structure prediction of U2AF35 and U2AF26 C-terminus. **(A)** HEK293T cells were transfected with Flag-tagged U2AF26 expression constructs with deletions of either one of the zinc fingers and with full length HA-tagged U2AF65. 48 hours after transfection cells were lysed and α-Flag immunoprecipitation (IP) was performed in the presence of 400 or 1200 mM NaCl. Input and IP were analyzed with α-HA and α-Flag antibody. **(B)** Secondary structure prediction of the second zinc finger and the downstream region of mouse full length U2AF35 and U2AF26, yeast U2AF23 and mouse U2AF26ΔZnF2 and U2AF26ΔE7. Residues in black boxes coordinate the zinc in the zinc finger domain.



Supplementary Figure S3: Both ZnFs are indispensable for U2AF35-dependant regulation of alternative splicing. Examples of splicing sensitive radioactive RT-PCRs for the eight analyzed cassette exons quantified in Figure 3C. Constitutive exons are depicted in white and alternative exons in grey boxes. Primers for splicing sensitive RT-PCRs where design to bind in the constitutive exons flanking the alternative exon.



Supplementary Figure S4: Cytoplasmic U2AF26ΔE7 regulates expression of luciferase when tethered to the 5'UTR. **(A)** MS2 and MS2-tagged U2AF26ΔE7 localize to the cytoplasm. Transiently transfected HEK293T cells were separated in cytoplasmic and nuclear fractions and analyzed via western blotting. GAPDH was used as cytoplasmic and hnRNP L as nuclear loading control. **(B)** Luciferase assay as in Figure 5A but in HeLa cells. n≥3. **(C)** Luciferase assay as in Figure 5A but with single MS2-loops in forward and reverse orientation in the 3'UTR. n=1 **(D)** Luciferase assay with MS2-loops in the 5'UTR and GFP alone or GFP-tagged U2AF26ΔE7. n=2. **(E)** MS2-tagged U2AF26 deletion mutants were transfected together with luciferase reporter with MS2-loops in the 5'UTR in HEK293T cells. Cell lysates were analyzed by western blotting using α-Flag and α-Luciferase antibody. α-Vinculin was used as loading control.



Supplementary Figure S5: Expression of surface proteins on EL4 cells with U2AF26ΔE7 expression. **(A)** Surface expression of CD3e, CD8.a, CD90.1, CD45, CD28, CD199 and Ly6a protein on stable EL4 cells analyzed by flow cytometry. The first bar shows unstained cells (no antibody). Depicted is the mean signal intensity ± SD. n≥3. **(B)** α-Flag western blot analysis of EL4 single cell clones stably expressing Flag-tagged U2AF26ΔE7 and control clones without expression. hnRNPL was used as loading control.