



Published in final edited form as:

Leukemia. 2015 April ; 29(4): 909–917. doi:10.1038/leu.2014.303.

U2AF1 Mutations Alter Sequence Specificity of pre-mRNA Binding and Splicing

Theresa Okeyo-Owuor¹, Brian S. White^{1,2}, Rakesh Chatrikhi³, Dipika R. Mohan¹, Sanghyun Kim¹, Malachi Griffith², Li Ding², Shamika Ketkar-Kulkarni¹, Jasreet Hundal², Kholiswa M. Laird³, Clara L. Kielkopf³, Timothy J. Ley^{1,2}, Matthew J. Walter¹, and Timothy A. Graubert^{1,4}

¹Department of Internal Medicine, Division of Oncology, Washington University, Saint Louis, MO, USA

²The Genome Institute, Washington University School of Medicine, Saint Louis, MO, USA

³Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, NY, USA

Abstract

We previously identified missense mutations in the *U2AF1* splicing factor affecting codons S34 (S34F and S34Y) or Q157 (Q157R and Q157P) in 11% of patients with *de novo* myelodysplastic syndromes (MDS). Although the role of U2AF1 as an accessory factor in the U2 snRNP is well established, it is not yet clear how mutations affect splicing or contribute to MDS pathophysiology. We analyzed splice junctions in RNA-seq data generated from transfected CD34+ hematopoietic cells and found significant differences in the abundance of known and novel junctions in samples expressing mutant U2AF1 (S34F). For selected transcripts, splicing alterations detected by RNA-seq were confirmed by analysis of primary *de novo* MDS patient samples. These effects were not due to impaired U2AF1 (S34F) localization as it co-localized normally with U2AF2 within nuclear speckles. We further found evidence in the RNA-seq data for decreased affinity of U2AF1 (S34F) for uridine (relative to cytidine) at the e-3 position immediately upstream of the splice acceptor site and corroborated this finding using affinity binding assays. These data suggest that the S34F mutation alters U2AF1 function in the context of

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding author: Timothy A. Graubert, MD Massachusetts General Hospital Cancer Center 10 North Grove Street, LWH 204 Boston, MA 02114 tgraubert@partners.org Phone: 617-643-0670.

⁴current address: Massachusetts General Hospital Cancer Center

AUTHOR CONTRIBUTIONS

The study was designed by: TOO, TAG, MJW

RNA affinity assays by: KML, RC, CLK

Sub-cellular localization and junction validation experiments performed by: TOO

Minigene assays performed by: DM, TOO, SK

Bioinformatics analysis was performed by: BSW, MG, LD, SKK, JH

The manuscript was written and edited by: TOO, BSW, MG, TJL, CLK, TAG, MJW

All co-authors reviewed and approved the submission.

CONFLICTS OF INTEREST

The authors declare no competing financial interests.

Supplementary information is available at *Leukemia's* website.

specific RNA sequences, leading to aberrant alternative splicing of target genes, some of which may be relevant for MDS pathogenesis.

INTRODUCTION

Recent studies have revealed that core spliceosome components are targets of recurrent mutation in a variety of hematopoietic malignancies. Splicing factor mutations, particularly in *SF3B1*, *U2AF1* and *SRSF2*, are present in approximately 50% of MDS cases, making them the most common class of mutations in this disease.¹⁻⁶ These mutations are also common in acute myeloid leukemia (AML), occurring with a frequency of ~14%,⁷ and *SF3B1* is the second most frequently mutated gene in chronic lymphocytic leukemia.⁸⁻¹⁰

U2AF1 encodes the 35 kDa auxiliary factor for the U2 pre-mRNA splicing complex and recognizes the 3' AG dinucleotide at the splice acceptor site in a pre-mRNA intron.^{11, 12} *U2AF1* has four domains: a *U2AF* homology motif (UHM), two zinc finger (ZnF) domains, and an arginine-serine (RS) domain.¹³ *U2AF1* heterodimerizes with *U2AF2* through its UHM domain,^{13,14} and *U2AF2* in turn binds the pre-mRNA as a complex with SF1.¹⁵ This *U2AF1* interaction leads to the recruitment and stabilization of *U2AF* binding to degenerate pre-mRNA polypyrimidine (Py) tracts.¹⁶ *U2AF1* also interacts directly with serine-arginine (SR) splice factors *SRSF1* and *SRSF2*,¹⁷ and interacts either directly or indirectly with other factors during spliceosome assembly.¹⁸

11 distinct mutations have been reported in *U2AF1*, including 9 missense mutations (resulting in A26V, S34F/Y, R35L, R156H/Q, Q157P/R, or G213A substitutions) and 2 frame-shift mutations (affecting codons Q157 or E159).^{1, 2, 6, 19-22} Most of these mutations occur within the two ZnF domains of *U2AF1*, with S34 and Q157 being the most commonly mutated residues. In our previous analysis in MDS, the S34F substitution was the most common (66.7%), followed by Q157P (16.7%), S34Y (11%) and Q157R (5.6%) out of 18 total *U2AF1* mutations.^{1, 22}

We previously reported that mutant *U2AF1* (S34F) causes increased exon skipping and cryptic/alternative splice site utilization in minigene assays.¹ In addition, other groups have observed differential splicing resulting from exon inclusion and skipping in AML patient samples with S34F (n=4) or S34Y (n=2) mutations.²⁰ Overexpression of *U2AF1* (S34F) suppresses growth and proliferation, and increases the rate of apoptosis in HeLa cells *in vitro*. Hematopoietic stem cells (CD34⁻KSL) expressing the S34F, Q157P or Q157R mutations have reduced capability to reconstitute recipient mouse bone marrow.² Collectively, these studies provide evidence that the S34F mutation affects not only *U2AF1* function but may also alter hematopoiesis.

While there have been attempts to understand the global impact of the S34F mutant on splicing and gene expression using transcriptome sequencing (i.e., RNA-seq), the genetic heterogeneity of primary MDS and AML poses challenges for discovery of target genes and no consistently dysregulated genes have been reported. It is also unclear what role the ZnF domains play and whether mutations in the S34 and Q157 residues possibly alter *U2AF1* interactions with RNA or result in mis-localization.

We performed transcriptome sequencing (RNA-seq) on CD34+ hematopoietic cells to assess the global impact of mutant U2AF1 on splicing and gene expression. We found that U2AF1 (S34F) affects pre-mRNA splicing of a large number of target genes, some of which are known oncogenic drivers, and preferentially skips splice acceptor sites immediately adjacent to uridine at the -3 position. We also determined the effect of the S34F substitution on sub-cellular localization and on the RNA binding specificity of U2AF1 for a representative affected splice site and U and C-variants at the e-3 position. Finally, we examined the effect of a spectrum of *U2AF1* mutations on splicing activity.

METHODS

RNA sequencing

Human hematopoietic mononuclear cells (MNCs) were separated from cord blood using density gradient centrifugation (Ficoll Paque, GE Healthcare). CD34+ cells were isolated from MNCs using the CD34 MicroBead kit (Miltenyi Biotec) on an autoMACs magnetic separator. These cells were cultured in SFEMII media (Stemcell Technologies) supplemented with IL-3, SCF, FLT-3 and TPO cytokines. WT and S34F *U2AF1* cDNAs were generated, as previously described,¹ and then cloned into pcDNA3.1-Ires-GFP (PIG) to create PIG-U2AF1 (WT or S34F). CD34+ cells then were transfected with PIG-U2AF1 (WT or S34F) using the Nucleofector Kit for Human CD34+ Cells (Lonza). GFP+CD34+ cells were sorted 24 hours later, followed by RNA extraction using the RNeasy kit (Qiagen). Ribosomal RNA was depleted (Ribozero, Epicenter), followed by cDNA preparation and Illumina library production. Sequencing was performed on the HiSeq2000 platform (Illumina). Bioinformatics analysis of RNA-seq data is described in the supplementary material.

RNA-seq validation

RT-PCR followed by gel electrophoresis was carried out using RNA isolated from independent CD34+ samples, transfected and purified as described above. RNA extraction and cDNA preparation from patient samples has been previously described.¹ Primers used for validation can be found in **Supplementary Table 5** and were designed to span the splice junction such that both the canonical and alternatively spliced isoforms are amplified. Quantitative RT-PCR (qRT-PCR) to quantify mRNA expression was performed using Taqman 2X Universal mix on the 7300 Real-Time PCR system (Applied Bioscience) and analyzed using the relative quantification of comparative C_T method.

RNA affinities of purified U2AF1 protein complexes

Fluorescence anisotropy changes were monitored during titration of fluorescein-labeled RNAs with purified protein complexes comprising U2AF2 (residues 85-471 at the C-terminus of NCBI RefSeq NP_001012496, isoform b), SF1 (residues 1-255 of NCBI RefSeq NP_004621) with either WT U2AF1 (residues 1-193 of NCBI RefSeq NP_006749) or the S34F mutant. Proteins were full-length with the exception of the nonspecific U2AF RS domains, and for SF1, a proline-rich C-terminal domain. The protein complex purification is explained in supplementary material. The 5'-labeled fluorescein RNAs (sequences DEK-“skipped”(UAG): 5'- UAAGAAAUACUAAAUAAUUUCUAG AAAAGAGUCUCA;

DEK-“skipped”-CAG: 5'- UAAGAAUACUAAAUAAUUUCCAG
AAAAGAGUCUCA; DEK-“spliced”(CAG) 5'-
AAUUGUGAUUUUUUUUUUCCCCAG GAAAGGGGCAGA; DEK-“spliced”-UAG:
5'- AAUUGUGAUUUUUUUUUUCCCCAG GAAAGGGGCAGA; the three nucleotides
preceding 3' splice site junction are underlined) were synthesized and purified
(ThermoScientific Dharmacon). Fluorescence anisotropy changes were measured at 520 nm
following excitation at 490 nm using a Fluoromax-3 (Horiba Ltd.) equipped with
microcuvette (Starna Cells Inc.). An RNA stock (0.75 mM) was diluted to 25 nM and the
protein complex stocks (20 μ M) were diluted to the final concentrations shown in
Supplementary Figure 5B. The protein and RNA buffer composition for the binding
experiments was 25 mM HEPES pH 6.8, 150 mM NaCl, 25 μ M ZnCl₂ and 1 mM TCEP.
The apparent equilibrium dissociation constants were fit, as previously described.²³

Minigene constructs and transfection

To create the MIG (MND Ires GFP)-U2AF1-Flag plasmid, U2AF1-Flag cDNA was amplified from the p3X-Flag-U2AF1 plasmid (obtained from the Kinji Ohno laboratory, Nagoya, Japan) and cloned into the MND-Ires-GFP (MIG) vector.²⁴ The S34F, S34Y, S34F/Q157R, Q157R and Q157P mutations were generated by site-directed mutagenesis (Life Technologies) using the WT MIG-U2AF1-Flag construct as a template. 293T cells were co-transfected with each MIG-U2AF1-Flag expression plasmid described above and either the *GHI* or *FMRI* minigene reporter constructs described previously.^{1, 25} GFP⁺ cells were sorted 48 hours later, followed by RNA extraction and RT-PCR as previously described.¹ Amplicons were visualized by polyacrylamide gel electrophoresis and quantified by densitometry (ImageJ). Three independent experiments were performed for each assay and analyzed using the Student's *t*-test. Lysates were made from transfected 293T cells and immunoblotting was performed using rabbit polyclonal anti-U2AF1 antibody (Abcam) to confirm over-expression.

Sub-cellular localization

Constructs containing the S34F mutant allele were generated by site-directed mutagenesis of p3X-Flag-U2AF1. 293T cells were transfected with either p3X-Flag-U2AF1 (WT) or p3X-Flag-U2AF1 (S34F). 24 hours later, the cells were fixed in 2-4% formaldehyde-PBS for 20 min at room temperature, and then washed with PBS. Cells were permeabilized with 0.5% (wt/vol) Triton X-100/PBS for 10 minutes and blocked with 1% goat serum/0.3% TritonX-100/PBS for 1 hour. The following primary antibodies were used for fluorescence microscopy: mouse monoclonal antibodies against Smith Antigen snRNP family (Y12; Abcam) or U2AF2 (Sigma), and Alexa Fluor 555-conjugated anti-mouse (Sigma) as a secondary antibody. Monoclonal anti-Flag M2-FITC (Sigma) used to detect Flag-tagged U2AF1 and TOPRO (Life Technologies) used as a nuclear counterstain. Images were acquired using a Zeiss LSM510 Meta laser scanning confocal microscope (Carl Zeiss, Thornwood, NY) equipped with a 63X, 1.4 numerical aperture, Zeiss Plan Apochromat oil objective at 2.5 zoom and captured using Zeiss LSM510 software.

RESULTS

The S34F mutation affects pre-mRNA splicing

We transfected primary human CD34+ hematopoietic cells with *U2AF1* (either WT or the S34F mutant) and performed RNA-seq to comprehensively determine the effects of the S34F mutant on pre-mRNA splicing and gene expression. We isolated cells 24 hours after transfection in order to identify immediate splicing changes induced by mutant U2AF1, while minimizing alterations that may occur with prolonged time in culture or as a consequence of secondary adaptations to altered splicing. Total reads of the raw sequence output ranged from 300 to 500 million per sample (**Supplementary Figure 1A**). Unique reads mappable to the human transcriptome were similar across all samples except for one outlier (S34F sample in replicate R3), whose replicate was excluded from further analysis (**Supplementary Figure 1B**). The distribution of mapped bases (coding, untranslated (UTR), intergenic, intronic and ribosomal bases) was similar for all 8 samples, with coding and UTR comprising 60-80% of the bases (**Supplementary Figure 1C**). As expected, reads mapped to *U2AF1* exon 2 demonstrated a G>A substitution only in cells transfected with the mutant cDNA (**Supplementary Figure 1D**). In these samples, mutant U2AF1 represented 85- 97% of total *U2AF1* expression, after normalizing for total mapped reads in each sample. Subsequent analyses capitalize on the paired experimental design (i.e., the same pool of CD34+ cells transfected with either WT or mutant U2AF1, repeated in 3 biological replicates). The ratio of total *U2AF1* expression (measured by FPKM) between mutant and WT samples were consistent across pairs (5.78, 2.49, 3.66 for R1, R2, and R4, respectively).

Though unsupervised clustering using adjusted expression levels (see Supplementary Methods) of 17,390 genes segregated samples based on genotype (mutant vs WT samples; **Supplementary Figure 2A**), lengths of dendrogram branches connecting samples within a genotype are similar to lengths of branches connecting samples of different genotypes. This indicates that the samples do not strongly cluster by genotype and that there is no strong, acute, global, gene-level effect induced by U2AF1 (S34F). To focus on those relatively few genes that were affected by U2AF1 (S34F), we applied edgeR, a statistical approach based on total counts of reads mapped to a gene that incorporates the experiment's paired design for improved power. This revealed that 1,296 genes were differentially expressed between paired WT U2AF1 and U2AF1 (S34F) samples (FDR < 5%). Hierarchical clustering of these differentially expressed genes (supervised at the 5% FDR cutoff) segregated mutant and WT samples, as expected (**Supplementary Figure 2B**).

We next analyzed junction-level expression (using edgeR and total reads mapped to a junction) to assess the effect of U2AF1 (S34F) on global splicing activity. We discovered that mutant U2AF1 alters pre-mRNA splicing of expressed junctions in 6% (959/15,687, FDR<5%) of genes (Figure 1, Supplementary Figure 3A-B and Supplementary Table 2). Of the 959 genes alternatively spliced junctions, 241 (FDR<5%) were also differentially expressed at the gene level in S34F samples compared to WT (**Supplementary Figure 3B** and **Supplementary Table 2**). These expression differences involved known splice junctions (present in Ensembl, release 67), and novel junctions resulting from: 1) known splice acceptor and donor sites (novel junction involving a donor and acceptor that are

individually present in Ensembl transcripts, but which have not been observed in combination), 2) known splice acceptor and novel (not present in Ensembl) splice donor sites, 3) novel splice acceptor and known splice donor sites, or 4) novel splice acceptor and donor sites (**Supplementary Figure 3A**). Both the known and novel junctions reflected an increase or decrease in mRNA isoforms due to alternative splicing, including exon skipping, alternative 3' or 5' splice site usage, and events that involved usage of both the alternative 5' and 3' splice sites (**Supplementary Figure 3C; Supplementary Table 2**). Next, we identified altered junctions differing by at least 4-fold in mutant vs. WT expressing samples. We discovered significant differences in the abundance of 544 out of 202,401 total junctions ($FDR < 5\%$, $|\log_2(\text{fold change})| > 2$) in cells expressing U2AF1 (S34F), compared with WT [known splice junctions (280/170,321; 0.16%); novel splice junctions (264/31,180; 0.85%)] (**Figure 1A-D; Supplementary Table 2**). Reads from novel junctions resulting from both novel acceptor and donor sites were not analyzed further.

We selected 24 junctions with higher expression ($FDR < 5\%$; $|\log_2(\text{fold change})| > 2$) in U2AF1 (S34F) samples for validation in independently prepared CD34+ cells. Of these, we validated 10/20 known junctions (**Figure 2A-B** and **Supplementary Figure 4A-C**, left panels) and 4/4 novel junctions (**Supplementary Figure 4D**). 10 of the validated changes were further tested in primary clinical samples [$n=6$ MDS bone marrow samples with U2AF1 (S34F) vs. $n=6$ MDS controls with no U2AF1 splicing mutations] and 5 were confirmed, including junctions resulting from alternative splicing due to exon skipping in *DEK* and *SMN1* and 3' alternative site usage in *SERPIN8B*, *KIAA1033*, and *IFI44* (**Figure 2A-B** and **Supplementary Figure 4A-B**, right panels). As further independent validation of the 544 junctions, we assessed their differential expression between 6 S34(F/Y) U2AF1 mutant AML patient samples and 108 control AML samples in publically-available TCGA RNA-seq data.²⁶ Other than the S34 variants in the 6 mutant samples, none of the samples had nucleotide mutations or copy number alternations in U2AF1 or any of 272 other splicing factors. 10 junctions had statistically significant ($FDR < 5\%$) dysregulation with concordant fold change, i.e., in the same direction as detected in CD34+ cells ($p\text{-value} < 10^{-3}$; **Supplementary Table 3**).

It has been shown previously that U2AF1 is not required for splicing of introns containing strong Py tracts,²⁷ but it is important for splicing of introns that contain short or weak Py tracts.^{11, 28, 29} To determine whether there were sequence differences in the 3' splice sites of junctions affected by U2AF1 (S34F), we analyzed sequences flanking (from -20 to +3 bps) the 3' AG dinucleotide (the binding site for U2AF1). We focused on exon skipping and alternative 3' site usage events, as these represent the majority of dysregulated events (**Supplementary Figure 3C**). An exon-skipping event was defined as skipping of a single cassette exon (**Figure 3A**). An alternative 3' usage event resulted in partial loss of an exon due to skipping of the canonical junction and usage of a single alternative 3' splice site down-stream of the canonical 3' splice site (**Figure 3B**). When we analyzed the 3' splice sites of skipped exons, we found that there was a high frequency of uridine (U) at position e-3 in exons that were skipped more by U2AF1 (S34F) compared to those skipped more in WT [S34F>WT: 81/111 (73%); WT>S34F: 10/68 (15%); $p=8.5 \times 10^{-15}$] (**Figure 3A**), consistent with previously reported findings.^{20, 30} The 3' canonical junctions skipped more

by U2AF1 (S34F) also had a higher frequency of U at position e-3 [S34F>WT: 64/101 (63%); WT>S34F: 22/101 (22%); $p=3.7 \times 10^{-9}$] (**Figure 3B**, left panel), and resulted in increased 3' alternative cryptic splice site usage, consistent with a recent report.³⁰ We then analyzed the 3' alternative cryptic splice site to determine whether there was a sequence preference for the S34F mutant. As seen with skipped exons, there was a higher frequency of uridine (U) at the e-3 position of the skipped junctions expressed more in U2AF1 (S34F) compared to WT [71/101 (70%); WT>S34F: 38/101 (38%); $p=5.2 \times 10^{-6}$] (**Figure 3B**, right panel). No other apparent differences in base preferences were noted in the Py tract preceding the 3' AG dinucleotide of the skipped exons or junctions. As a control, we analyzed junctions that showed no evidence for differential expression by U2AF1 (S34F) [$|\log_2(\text{fold change})| < 0.001$] (**Figure 3C**). We did not observe any differences in sequences at the 5' splice site of skipped exons, or at 5' sites of exons with alternative 5' or 3' splice sites (**Supplementary Figure 3D** and data not shown), consistent with the known activity of U2AF1 which is restricted to 3' splice sites. Collectively, these data suggest that the 3' splice sites that are more frequently skipped by U2AF1 (S34F) are enriched for U at e-3, while alternative sites utilized more frequently by the mutant are enriched for C at e-3. We examined junctions that were validated in patient samples and found that all validated junctions (5/5) had U at position e-3 at skipped junctions, suggesting that this consensus sequence enriches for true positive junctions that are preferentially skipped by U2AF1 (S34F) (**Supplementary Table 4**).

The S34F mutation alters the affinity of pre-mRNA to U2AF1

The enhanced exon skipping and alternative 3' splice site usage seen with the U2AF1 (S34F) mutant could be due to altered binding of U2AF1 to canonical 3' AG (acceptor site). To explore this further, we compared the RNA affinities of WT and S34F-mutant U2AF1 for splice sites from a representative affected transcript, the *DEK* oncogene, where a U at position e-3 (i.e., UAG) was skipped in favor of splicing into a CAG splice site. The *DEK*-“skipped” RNA oligo comprised the intron/exon region of the splice acceptor site of exon 3 (skipped by S34F mutant U2AF1) whereas the *DEK*-“spliced” RNA oligo comprised the downstream intron/exon 4 sequence that is preferentially spliced into by the mutant (**Figure 4A**). To better emulate the context of the assembling spliceosome, we used ternary complexes of either wild-type U2AF1 (residues 1-193) or the S34F mutant with accessory protein subunits for the early stage of 3' splice site recognition: U2AF2 (residues 85-471 at the C-terminus) and SF1 (residues 1-255) (**Figure 4B**, **Supplementary Figure 5A**). Both of the U2AF proteins were full length with the exception of the nonspecific RS domains. For SF1, a C-terminal proline-rich domain thought to interact with 5' splice site subunits was truncated. The RNA oligos were fluorescein-labeled and the RNA affinities were determined from anisotropy changes during titration with the purified protein complexes (**Supplementary Figure 5B**). Consistent with S34F-dependent skipping of the corresponding *DEK* splice site, the *DEK*-“skipped” RNA oligo bound less avidly to mutant U2AF1 (S34F) compared with WT (**Figure 4C**). The nearly identical binding of mutant or WT U2AF1 to the *DEK*-“spliced” RNA oligo also was consistent with the similar splicing of the downstream splice site in normal and mutant U2AF1 samples.

Based on the preferential, S34F-dependent skipping of junctions with U at position e-3 and splicing of junctions with C at position e-3 in the presence of mutant U2AF1 (S34F), we next tested the RNA sequence discrimination of the WT and S34F mutant protein complexes for the C or U e-3 variants of the *DEK* splice sites (**Figure 4A**). Alteration of the UAG in the skipped *DEK* sequence to CAG (*DEK*-“skipped”-CAG, **Figure 4A**) increased the affinity for both the U2AF1 protein complexes and to a significantly greater extent for the S34F-mutant (2-fold and 9-fold higher affinity for the WT and S34F proteins, respectively) (**Figure 4C**). Conversely, alteration of the CAG in the downstream, spliced *DEK* sequence to UAG (*DEK*-“spliced”-UAG, **Figure 4A**) decreased the affinity for both the U2AF1 protein complexes with a significantly greater penalty for the S34F mutant (factors of 2 and 6 affinity decrease for the WT and S34F proteins, respectively) (**Figure 4C**). These data demonstrate that the S34F mutation alters the sequence specificity of the ternary U2AF1 complex in favor of binding splice sites comprising C at e-3 and discriminates against splice sites comprising U at the e-3 position.

The S34F mutation does not affect U2AF1 localization within nuclear speckles

U2AF1 is diffusely distributed in the nucleoplasm and localizes within the nuclear speckles (sites of spliceosome assembly).^{31, 32} Using fluorescence immunocytochemistry on U2AF1-Flag (WT or S34F) transfected 293T cells, we found that U2AF1 (S34F) localized normally (**Supplementary Figure 6A**). Furthermore, U2AF1 (S34F) co-localized with U2AF2 and Smith Antigen family of snRNP proteins (Y12 antibody) in a similar pattern as WT (**Supplementary Figure 6B**), suggesting that the altered splicing activity of U2AF1 (S34F) was not due to abnormal trafficking of the mutant protein.

Specific effects of U2AF1 mutations in alternative splicing

Apart from the S34F substitution, there are 10 other reported somatic mutations in *U2AF1* that may affect its function. Since mutations at codons S34 and Q157 are the most common, we utilized *GHI* and *FMRI* minigenes^{1, 25} to determine the effect of substitutions at these positions on splicing activity. The *GHI* or *FMRI* minigene was transiently co-transfected with either MND-IRES-GFP (MIG; empty vector) or MIG-U2AF1-Flag (WT, S34F, S34Y, S34F/Q157R, Q157R or Q157P alleles) into 293T cells. Over-expression was confirmed by qRT-PCR (**Figure 5A**; **Supplementary Figure 7A**) and Western blot analysis (**Figure 5B**; **Supplementary Figure 7B**). Isoform a represents the canonical isoform, and isoforms b and c (if present) are the alternatively spliced isoforms. In *GHI*, isoform b results from exon skipping and in *FMRI* isoforms b and c result from alternative 3' splice site usage. U2AF1 (S34F) yielded the most significant increase in alternative splicing activity for both *GHI* and *FMRI* (**Figure 5C** and **D**, respectively), consistent with previously published results.¹ U2AF1 (S34Y) modestly enhanced relative expression of the alternative isoforms for both minigenes. Conversely, there was a reduction in relative expression of the alternative isoforms for both minigenes in cells expressing the Q157R (**Figure 5C** and **D**) and Q157P mutations (**Supplementary Figure 7C** and **D**). In cells expressing the S34F/Q157R double mutant (in which both the mutant S34F and Q157R occur on one allele, discovered in one patient with MDS¹), *GHI* splicing was indistinguishable from WT (**Figure 5C**) and there was a modest reduction in the relative expression of the alternative isoform of *FMRI*

(**Figure 5D**). Next, we examined the effects of different U2AF1 mutations on the splicing of endogenous *DEK*. Exon skipping in endogenous *DEK* was increased in 293T samples expressing S34F/Y, S34F/Q157R and Q157R mutants relative to WT (**Figure 5E**). As observed with the *GHI* and *FMRI* minigenes, the S34F mutant caused the most robust increase in alternative isoform b expression. However, unlike the *GHI* and *FMRI* minigenes, expression of the Q157R mutant resulted in increased alternative isoform splicing compared to WT expressing cells for *DEK*. Interestingly, analysis of TCGA AML RNA-seq data demonstrated that junctions differentially expressed in a sample with Q157P mutant U2AF1 vs. WT U2AF1 ($|\log_2(\text{fold change})| > 1$) do not share the signature at flanking sequences (C>U at e-3) that we detected in junctions differentially spliced by S34F U2AF1 (FDR <5%; **Supplementary Figure 8**). These results provide evidence that splicing is altered in a mutation-specific manner, but variable between specific junctions tested, with the most potent effects caused by U2AF1 (S34F) mutant.

DISCUSSION

Recent discoveries of recurrent mutations in core components of the pre-mRNA splicing complex highlight the potential importance of these genes in myelodysplastic syndromes. In this study, we used RNA-seq to detect global changes in splicing activity and isoform expression and discovered differences in expression of known and novel splice junctions. Our study is the first to analyze the impact of a mutant splicing factor on the transcriptome in a controlled genetic background in primary hematopoietic cells. So a consistent signature has not been defined for genes whose splicing is clearly affected by U2AF1 (S34F). Published results are conflicting, with one study reporting ~130 unspliced exons due to intron retention in mutant U2AF1-transfected HeLa cells² while another (utilizing patient samples with a spectrum of splicing mutations) found no evidence for genome-wide intron retention, although these mutations did perturb *TET2* splicing.⁶ Analysis of TCGA AML RNA-seq data revealed a predominance of exon skipping in 35 exons dysregulated by U2AF1 (S34F).²⁰ The affected genes were involved in mitosis and RNA processing. Our results include a substantially larger number of alternative splicing events that were impacted by U2AF1 (S34F). Of the 35 genes that had significant exon skipping or retention in the data set from Prychodzen, *et al.*,²⁰ only one gene (*STRAP*) overlaps with our results. When we validated the 544 junctions that were differential expressed in our RNA-seq data with the TCGA RNA-seq data, we found 10 genes that overlapped with our results, including *STRAP* and *CCT6A*. Although the *DEK* junction was not among these, the same *DEK* junction did have a *p*-value of 0.06 and concordant fold change in the TCGA data. Several characteristics of our experimental design may explain these largely non-overlapping results, including: (1) the use of CD34+ cells which have less genetic heterogeneity, compared to primary patient samples, (2) use of paired samples (S34F vs. WT), minimizing variability within biological replicates, and (3) differences in analytical approaches.

We found a higher frequency of uridine at the e-3 nucleotide in 3' splice junctions that were skipped more often in U2AF1 (S34F) samples, including both exon skipping events and alternative 3' splice site usage. These findings confirm previous observations related to exon skipping and alternative splice site usage.^{20, 30} This implies that there is selective or altered

U2AF1 (S34F) binding to pre-mRNA at these junctions, possibly resulting from an increased affinity of U2AF1 (S34F) for C over U, relative to WT U2AF1, which we corroborated experimentally by determining the affinities of purified U2AF2/U2AF1 (WT or S34F) heterodimers for the affected splice sites of a representative example, *DEK*. Consistent with the S34F-affected splice site junctions, the S34F mutation selectively inhibited binding of the mutant U2AF1 complexes to the *DEK* splice site variants comprising U at the e-3 position. Conversely, substitution with a C at this position significantly increased U2AF1 (S34F) binding.

While it is not known which U2AF1 domain or domains directly bind RNA, the altered RNA binding affinity of the S34F mutant U2AF1 suggests that the ZnF domains are critical. The S34F mutation is in the first of these domains. In *S. pombe*, substitution of highly conserved cysteine residues (C18 and C149) to alanine of the ZnF domains results in lethality, and an RNA three-hybrid screen showed that these two ZnF domains are independently required for RNA binding of the U2AF heterodimer in yeast.³³ Furthermore, the U2AF1 ZnF domains are predicted to be structurally similar to the murine and human ZFP36 family (both CX8CX5CX3H zinc fingers), which are known to bind RNA,^{34, 35} and our data suggest that the ZnF1 domain of U2AF1 may have this function. Rather than simply ablating function, the capacity of the S34F mutation to specifically discriminate between U and C at the e-3 position of the splice site junction implicates this residue in hydrogen bonds with the N3 or O4 positions of the pyrimidine base.

While mutant U2AF1 alters pre-mRNA splicing, the overall change in the total number of known and novel junctions is minimal, affecting (at an FDR <5%) fewer than 1% (1,594/201,501) of expressed junctions in 6% of analyzed genes. In addition, the proportion of isoforms induced by U2AF1 (S34F) expression is a small fraction of the total transcript for those genes. Collectively, these findings raise the possibility that disease pathogenesis induced by mutant U2AF1 may also be mediated through non-pre-mRNA splicing mechanisms that should be explored in future studies.

Finally, we discovered that the S34 and Q157 *U2AF1* mutations have different effects on pre-mRNA splicing using *GHI* and *FMRI* minigenes. We detected an increase in relative expression of the alternative isoforms for both minigenes with S34 mutants, and a decrease in relative expression of these isoforms with Q157 mutants relative to WT, indicating that these mutations affect U2AF1-dependent splicing function differently depending on their location and the reporter junctions being queried (S34; ZnF1 and Q157; ZnF2). In fact, when splicing of endogenous *DEK* was analyzed following expression of U2AF1 (Q157R), a similar but less robust increase in alternative splicing was observed as seen with expression of U2AF1 (S34F). Junction analysis of one Q157P sample in the TCGA AML RNA-seq data adds credence to these findings. The differentially expressed junctions in this Q157P sample did not share the same signature at flanking sequences that were detected in junctions differentially spliced by S34F U2AF1 (C>U at e-3). These results suggest that there are differences in alternative splicing induced by S34F vs. Q157R/P mutations, but highlight the need for an unbiased transcriptome analysis (RNA-seq) comparing expression of U2AF1 S34F vs. Q157R/P in order to elucidate the full extent of these similarities and differences. The splicing alterations detected with specific U2AF1 mutants may be shared

by other splicing factors with recurrent mutations in myeloid neoplasms such as *SF3B1*, *ZRSR2*, *SRSF2* and *SF1* and common changes could serve as novel therapeutic targets across these diseases. Identifying important potential target genes that are affected by U2AF1 mutants is necessary for further understanding of their role in MDS and other cancers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Support was provided by Barnes-Jewish Hospital Foundation (7603-55 to BSW), an NIH/NCI SPORE in Leukemia (P50CA171063 to TAG, MJW, and BSW), an NIH/NCI grant (1K12CA167540 to BSW), a Clinical and Translational Award from the NIH National Center for Advancing Translational Sciences (UL1 TR000448 to BSW), a Howard Hughes Medical Institute Physician-Scientist Early Career Award and Leukemia and Lymphoma Society Scholar Award (MJW) and NIH R01 GM070503 (CLK). Technical assistance was provided by the Alvin J. Siteman Cancer Center High Speed Cell Sorting core (supported in part by P30CA91842), Molecular Microbiology Imaging Facility (MMIF), the Genome Technology Access Center (GTAC) and the Genome Institute. Additional technical support provided by Masayo Izumi. The authors are grateful to Dr. Christopher R. Cabanski for assistance with edgeR and Dr. Cara Lunn Shirai for helpful scientific discussions.

REFERENCES

1. Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nature Genetics*. Jan; 2012 44(1):53–57. [PubMed: 22158538]
2. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Sato Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011; 478(7367)
3. Papaemmanuil E, Cazzola M, Boultonwood J, Malcovati L, Vyas P, Bowen D, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *The New England Journal of Medicine*. Oct 13; 2011 365(15):1384–1395. [PubMed: 21995386]
4. Mian SA, Smith AE, Kulasekararaj AG, Kizilers A, Mohamedali AM, Lea NC, et al. Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto oncogenes mutated in myelodysplastic syndrome. *Haematologica*. Jul; 2013 98(7):1058–1066. [PubMed: 23300180]
5. Visconte V, Makishima H, Jankowska A, Szpurka H, Traina F, Jerez A, et al. SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia*. Mar; 2012 26(3): 542–545. [PubMed: 21886174]
6. Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Abu Kar S, Jerez A, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood*. Apr 5; 2012 119(14):3203–3210. [PubMed: 22323480]
7. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England Journal of Medicine*. May 30; 2013 368(22):2059–2074. [PubMed: 23634996]
8. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics*. Jan; 2012 44(1):47–52. [PubMed: 22158541]
9. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *The New England Journal of Medicine*. Dec 29; 2011 365(26):2497–2506. [PubMed: 22150006]
10. Rossi D, Brusca A, Spina V, Rasi S, Khiabani H, Messina M, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*. Dec 22; 2011 118(26):6904–6908. [PubMed: 22039264]

11. Wu S, Romfo CM, Nilsen TW, Green MR. Functional recognition of the 3[prime] splice site AG by the splicing factor U2AF35. *Nature*. 1999; 402(6763):832–835. 12/16/print. [PubMed: 10617206]
12. Ruskin B, Zamore PD, Green MR. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell*. 1988; 52(2):207–219. 1/29/. [PubMed: 2963698]
13. Zhang M, Zamore PD, Carmo-Fonseca M, Lamond AI, Green MR. Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proceedings of the National Academy of Sciences*. Sep 15; 1992 89(18):8769–8773. 1992.
14. Zamore PD, Green MR. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proceedings of the National Academy of Sciences*. Dec 1; 1989 86(23):9243–9247. 1989.
15. Berglund JA, Abovich N, Rosbash M. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev*. Mar 15; 1998 12(6):858–867. [PubMed: 9512519]
16. Rudner DZ, Breger KS, Kanaar R, Adams MD, Rio DC. RNA Binding Activity of Heterodimeric Splicing Factor U2AF: at Least One RS Domain Is Required for High-Affinity Binding. *Molecular and Cellular Biology*. Jul 1; 1998 18(7):4004–4011. 1998. [PubMed: 9632785]
17. Wu JY, Maniatis T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*. 1993; 75(6)
18. Prigge JR, Iverson SV, Siders AM, Schmidt EE. Interactome for auxiliary splicing factor U2AF65 suggests diverse roles. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. 2009; 1789(6–8):487–492. 6//; [PubMed: 19540372]
19. Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. April 5. 2012; 119(14):3211–3218. 2012. [PubMed: 22343920]
20. Przychodzen B, Jerez A, Guinta K, Sekeres MA, Padgett R, Maciejewski JP, et al. Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood*. Aug 8; 2013 122(6):999–1006. 2013. [PubMed: 23775717]
21. Thol F, Kade S, Schlarman C, Löffeld P, Morgan M, Krauter J, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. Apr 12; 2012 119(15):3578–3584. 2012. [PubMed: 22389253]
22. Walter MJ, Shen D, Shao J, Ding L, White BS, Kandoth C, et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia*. 2013; 27(6):1275–1282. [PubMed: 23443460]
23. Jenkins JL, Shen H, Green MR, Kielkopf CL. Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF65. *The Journal of Biological Chemistry*. Nov 28; 2008 283(48):33641–33649. [PubMed: 18842594]
24. Kobayashi H, Carbonaro D, Pepper K, Petersen D, Ge S, Jackson H, et al. Neonatal gene therapy of MPS I mice by intravenous injection of a lentiviral vector. *Molecular therapy : the Journal of the American Society of Gene Therapy*. May; 2005 11(5):776–789. [PubMed: 15851016]
25. Fu Y, Masuda A, Ito M, Shinmi J, Ohno K. AG-dependent 3'-splice sites are predisposed to aberrant splicing due to a mutation at the first nucleotide of an exon. *Nucleic Acids Research*. May 1; 2011 39(10):4396–4404. 2011. [PubMed: 21288883]
26. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England Journal of Medicine*. May 30; 2013 368(22):2059–2074. [PubMed: 23634996]
27. Reed R. The organization of 3' splice-site sequences in mammalian introns. *Genes & Development*. Dec 1; 1989 3(12b):2113–2123. 1989. [PubMed: 2628164]
28. Pacheco TR, Coelho MB, Desterro JMP, Mollet I, Carmo-Fonseca M. In Vivo Requirement of the Small Subunit of U2AF for Recognition of a Weak 3' Splice Site. *Molecular and Cellular Biology*. Nov 1; 2006 26(21):8183–8190. 2006. [PubMed: 16940179]

29. Guth S, Martínez C, Gaur RK, Valcárcel J. Evidence for Substrate-Specific Requirement of the Splicing Factor U2AF35 and for Its Function after Polypyrimidine Tract Recognition by U2AF65. *Molecular and Cellular Biology*. Dec 1; 1999 19(12):8263–8271. 1999. [PubMed: 10567551]
30. Brooks AN, Choi PS, de Waal L, Sharifnia T, Imielinski M, Saksena G, et al. A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. *PLoS ONE*. 2014; 9(1):e87361. [PubMed: 24498085]
31. Chusainow J, Ajuh PM, Trinkle-Mulcahy L, Sleeman JE, Ellenberg J, Lamond AI. FRET analyses of the U2AF complex localize the U2AF35/U2AF65 interaction in vivo and reveal a novel self-interaction of U2AF35. *RNA*. Aug 1; 2005 11(8):1201–1214. 2005. [PubMed: 16043505]
32. Gama-Carvalho M, Krauss RD, Chiang L, Valcárcel J, Green MR, Carmo-Fonseca M. Targeting of U2AF65 to Sites of Active Splicing in the Nucleus. *The Journal of Cell Biology*. Jun 2; 1997 137(5):975–987. 1997. [PubMed: 9166400]
33. Webb CJ, Wise JA. The splicing factor U2AF small subunit is functionally conserved between fission yeast and humans. *Mol Cell Biol*. May; 2004 24(10):4229–4240. [PubMed: 15121844]
34. Lai WS, Kennington EA, Blackshear PJ. Interactions of CCCH Zinc Finger Proteins with mRNA: NON-BINDING TRISTETRAPROLIN MUTANTS EXERT AN INHIBITORY EFFECT ON DEGRADATION OF AU-RICH ELEMENT-CONTAINING mRNAs. *Journal of Biological Chemistry*. Mar 15; 2002 277(11):9606–9613. 2002. [PubMed: 11782475]
35. Liang J, Song W, Tromp G, Kolattukudy P, Fu M. Genome-Wide Survey and Expression Profiling of CCCH-Zinc Finger Family Reveals a Functional Module in Macrophage Activation. *PLoS ONE*. 2008; 3(8):1371.

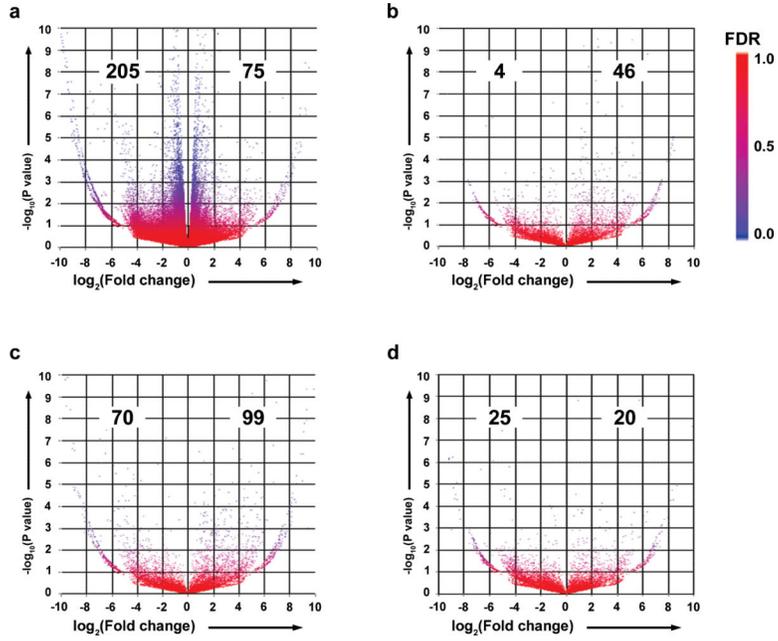


Figure 1. Overexpression of U2AF1 (S34F) alters the distribution of known and novel splice junctions

Expression of splice junctions from RNA-seq of primary human CD34+ cells transfected with either WT or S34F mutant U2AF1. **(a)** Known splice junctions. **(b, c and d)** Novel splice junctions: **(b)** Known splice acceptor/donor in novel combination; **(c)** Novel splice acceptor/known splice donor sites; **(d)** Known splice acceptor/novel splice donor sites. Numbers in each panel indicate significantly altered (FDR<5%) splice junctions with a $|\log_2(\text{fold change})| > 2$.

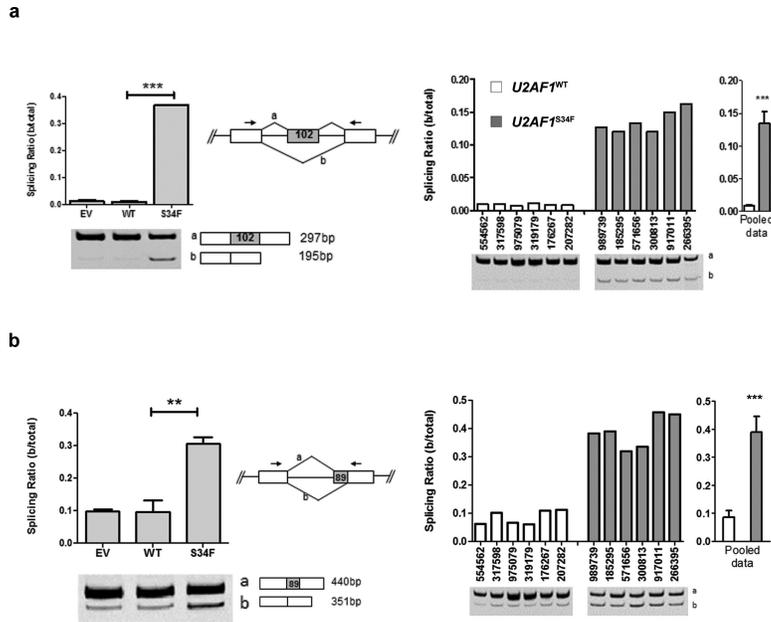


Figure 2. Exon skipping and alternative splice site usage increased in U2AF1 (S34F) mutant samples

U2AF1 (S34F) caused increased exon skipping in (a) *DEK* and increased utilization of a 3' alternative splice site in (b) *SERPIN8B* in transfected CD34+ cells (left panels) and primary MDS samples (right panels). Arrows indicate primers used for RT-PCR. Data in the left panels represent mean +/- SD of 3 technical replicates, repeated in 3 biological replicates with similar results. **P<0.01, ***P<0.001.

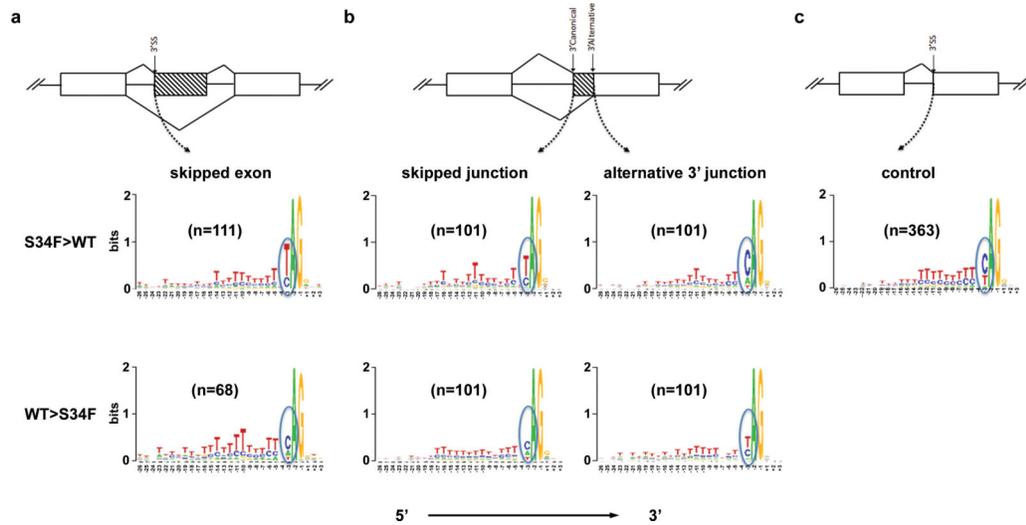


Figure 3. The U2AF1 (S34F) mutant recognizes the uridine at position e-3 of the 3' splice site less frequently

(a) Schematic representation of an exon skipping event (top) and Logos nucleotide sequence analysis at the 3' splice sites of skipped exons are shown. Exons skipped more often with U2AF1 (S34F) have an increased frequency of U at the e-3 position (middle), compared with exons skipped more often with WT U2AF1 (lower). (b) Schematic representation of alternative cryptic site usage showing the 3' canonical and alternative sites (top). Part of a cassette exon is skipped due to alternative site usage. Logos plot of nucleotide sequences at the 3' splice sites of either the skipped (left panels) or alternative (right panels) junction. The U at e-3 leads to less recognition of the 3' splice site of the canonical junction by U2AF1 (S34F), resulting in an increase in alternative site usage. Consistently, a U at e-3 of the alternative 3' junction is recognized less often by U2AF1 (S34F) than by U2AF1 (WT). (c) 3' splice sites of junctions not altered by U2AF1 [S34F; i.e., $|\log_2(\text{fold change})| < 0.001$] are shown as control. n is the number of splice junction events used in the Logos analysis and includes both known and novel junctions. The frequency of each nucleotide at a position is represented by the height of each letter. 'Skipped Junction' refers to canonical splice acceptor sites that are skipped in lieu of downstream cryptic splice acceptor sites.

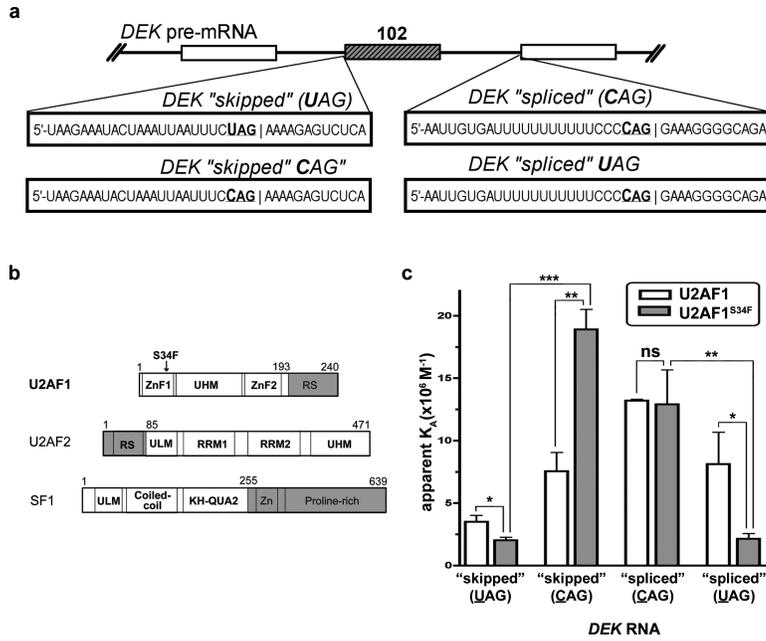


Figure 4. Comparison of the WT and S34F U2AF1 affinities for splice sites from a representative affected transcript, the DEK oncogene

(a) DEK exon (exon 3) and RNA sequences used in the experiment. (b) Boundaries of recombinant protein constructs used in this study (white). The U2AF2 and SF1 subunits are included. Both U2AF proteins are full-length with the exception of the nonspecific RS domain (gray). KH-QUA2, K-Homology Quaking motif; UHM, U2AF homology motif; ULM, U2AF ligand motif; ZnF, zinc finger. (c) Bar graph of DEK-“skipped”(UAG), DEK-“skipped”-CAG, DEK-“spliced”(CAG), or DEK-“spliced”-UAG RNA affinities of either wild-type U2AF1 (white) or the S34F mutant (gray) in purified U2AF2/SF1- complexes. Average values and standard deviations of three experiments are given. ns, not significant ($P>0.3$); * $P<0.03$, ** $P<0.003$, *** $P<0.0003$.

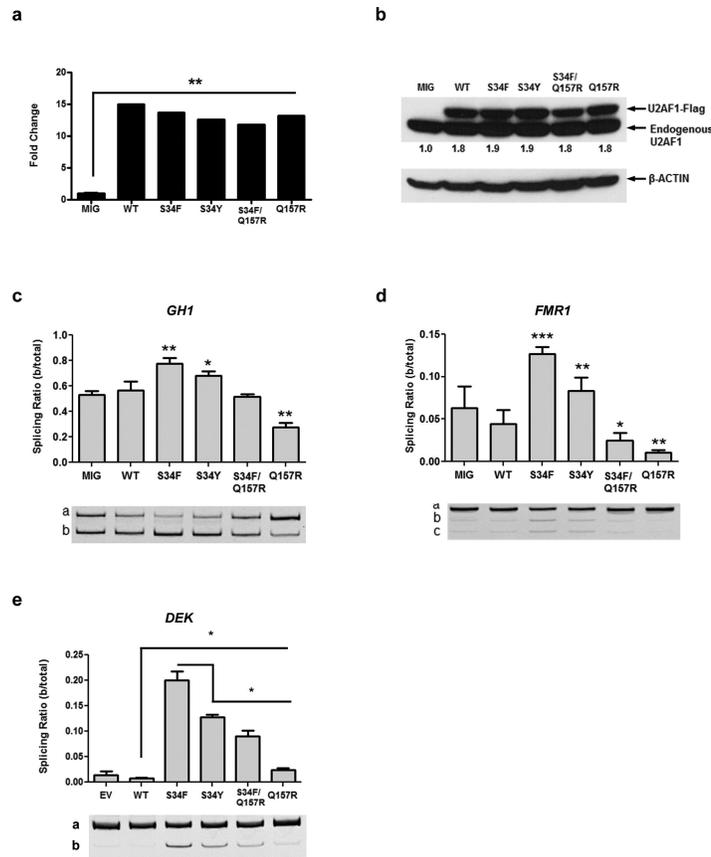


Figure 5. Recurrent missense mutations in U2AF1 have specific effects on alternative splicing (a) *U2AF1* mRNA expression was increased >10-fold in cells transfected with WT or mutant alleles, compared to MIG control. Error bars (too small to visualize) represent SD of 3 technical replicates; repeated in 3 biological replicates for each construct with similar results. (b) Exogenous Flag-tagged U2AF1 protein abundance of the different mutants in 293T cells compared with MIG empty vector control. Numbers beneath the blot are the sum of U2AF1-Flag and endogenous U2AF1 values in each case, normalized to MIG, obtained by densitometry analysis. (c) The U2AF1 S34 mutants (S34F>S34Y) increased exon skipping in the *GHI* minigene. Exon skipping was decreased by the Q157R mutant, and unaffected with an allele containing both mutations (S34F/Q157R), compared to WT U2AF1. (d) Similar mutation-specific effects were seen with cryptic splicing of the *FMR1* minigene. (e) Exon skipping of endogenous *DEK* was increased by the U2AF1 mutants relative to WT in 293T cells but decreased by the S34Y, S34F/Q157R and Q157R mutants, compared to the S34F mutant. The splicing ratio (b/total) = expression of isoform b/(the sum of the expression of isoforms a and b). The mean \pm SD of 3 replicates are shown; repeated in 3 separate experiments with similar results. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.