**Wild-type Splicing Factor U2AF1 inhibits splicing associated with a recurrent U2AF1 mutant in human lung cancers and is required for cell survival**

Dennis Liang Fei[1,2,*], Hayley Motowski[1], Rakesh Chatrikhi[3], Sameer Prasad[1], Jovian Yu[1], Shaojian Gao[4], Clara L. Kielkopf[3,*], Robert K. Bradley[5,6,*], Harold Varmus[1,2*]

1. Cancer Biology Section, Cancer Genetics Branch, National Human Genome Research Institute, Bethesda, MD 20892, United States
2. Department of Medicine, Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, United States
3. Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, NY 14642, United States
4. Genetics Branch, National Cancer Institute, Bethesda, Maryland, 20892, United States
5. Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, United States
6. Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States

* Correspondence: D.L.F. dlf2002@med.cornell.edu; C.L.K. clara_kielkopf@urmc.rochester.edu; R.K.B. rbradley@fredhutch.org; H.V. varmus@med.cornell.edu

**Abstract**

The splicing factor gene, *U2AF1*, is recurrently mutated in a variety of human cancers, including lung adenocarcinomas. The most frequent U2AF1 mutant, U2AF1 p.Ser34Phe (S34F), induces specific changes in splicing that we collectively refer to as "S34F-associated splicing", but it is unclear how these splicing changes are regulated. Moreover, while a wild-type *U2AF1* allele is retained in all cancers expressing a *U2AF1* mutation, the functional significance of the retained wild-type allele is unknown.

Our analysis of published data on human lung adenocarcinomas indicates that lung adenocarcinomas carrying a *U2AF1 S34F* allele exhibit a wide range of mutant to wild-type *U2AF1* (S34F:WT) mRNA ratios, which can be partially attributed to copy number variation at the *U2AF1* locus. S34F:WT mRNA ratios, rather than absolute levels of *U2AF1 S34F* or total *U2AF1* mRNA, correlate positively with the magnitude of S34F-associated splicing in lung adenocarcinoma transcriptomes. To examine the effect of S34F:WT ratios on S34F-associated splicing directly, we modified a human bronchial epithelial cell line so that *U2AF1 S34F* is expressed at one of the two endogenous *U2AF1* loci and the S34F:WT mRNA ratio approximates one. By altering the levels of mutant or wild-type U2AF1 in this engineered cell line, we show that the degree of S34F-associated splicing is proportional to the ratio of S34F:WT gene products and not to absolute levels of either the mutant or wild-type factor. Further, we show that in nearly all cases, S34F-associated splicing alterations are largely explained by the different RNA binding affinities of recombinant protein complexes containing wild-type or mutant U2AF1. Together, these observations suggest that wild-type U2AF1 is a negative regulator of S34F-associated splicing, at least in part through differential

binding to 3´ splice sites. Finally, we show that the *U2AF1 S34F* allele does not behave like some oncogenes: the mutated gene does not induce cell transformation, and lung adenocarcinoma cell lines are not dependent on it for growth *in vitro* or *in vivo*. Wild-type *U2AF1*, however, is absolutely required for cell survival, regardless of whether the cells carry the *U2AF1 S34F* allele.

We conclude that wild-type *U2AF1* has two important functions in *U2AF1 S34F*-expressing lung cancers: it controls the magnitude of S34F-associated splicing, and it is essential for cell survival.

## Introduction

Somatic mutations in genes encoding four splicing factors (*U2AF1*, *SF3B1*, *SRSF2* and *ZRSR2*) have recently been reported in many human cancers. They are found in up to 50% of myelodysplastic syndromes (MDS) and related neoplasms, and, albeit at a lower frequency, in a variety of solid tumors, including lung adenocarcinomas (LUADs) [1-9]. These splicing factor mutations are considered likely driver events of carcinogenesis because they occur at higher-than-expected frequencies, affect specific residues of the encoded proteins (with the exception of ZRSR2) even in different tumors, and are often found in so-called “founder” tumorigenic clones [10-12]. The frequent occurrence of these mutations in human cancers suggests that a property conferred by the mutant alleles confers a selective advantage on cancer cells. However, that property and the mechanism by which mutant splicing factors play an oncogenic role remain poorly understood [1,13,14].

Among the four commonly mutated splicing factor genes in myeloid neoplasms, only *U2AF1* is recurrently mutated in LUADs [3,9]. The only highly recurrent missense mutation of *U2AF1* in LUAD affects codon 34 and always changes the conserved serine in a zinc knuckle motif to phenylalanine (p.Ser34Phe, or S34F). This striking mutational consistency suggests a critical, yet unknown, role for *U2AF1 S34F* during lung carcinogenesis. In addition, the wild-type (WT) *U2AF1* allele is always retained in cancers with common *U2AF1* mutations, including *U2AF1 S34F* [2]. However, the functional significance of WT *U2AF1* in cells with mutant *U2AF1* is not known.

U2AF1 is a component of the U2 small nuclear ribonucleoprotein auxiliary factor complex (U2AF) [15,16]. During early spliceosome assembly, U2AF recognizes

sequences at the 3´ end of introns to facilitate the recruitment of U2 small nuclear ribonucleoprotein to the 3´ splice junction; the recruitment occurs in conjunction with recognition of the intronic branch point by splicing factor 1 (SF1) [17,18]. Unlike U2AF2, which is an essential subunit of U2AF, U2AF1 is thought to be required for splicing of only a subset of introns [19-21]. *In vitro* crosslinking assays show that U2AF1 contacts the AG dinucleotide at the intron-exon boundary and other sequences flanking the 3´ splice junction [19,22,23].

Consistent with the critical role U2AF1 plays in RNA splicing, *U2AF1* mutations are known to be associated with specific alterations in RNA splicing [24-27]. For example, the most frequent *U2AF1* mutation, *U2AF1 S34F*, affects many types of alternative splicing, most prominently the inclusion of cassette exons in mRNA. (Here we collectively refer to the changes found in cancers bearing the S34F mutation as "S34F-associated splicing.") Although the identities of cassette exons affected by *U2AF1 S34F* vary among different reports, the affected exons generally share one prominent feature: distinct nucleotides are enriched at the 3´ splice sites immediately preceding cassette exons that are alternatively used in cells with the *U2AF1 S34F* allele [24-27].

Despite the specific features of splicing associated with mutant *U2AF1*, little is known about the molecular basis underlying these changes and how they are regulated. A computational model of the structure of the U2AF1:RNA complex suggested that Ser34 is a critical residue that contacts RNA [24]. Consistent with this model, U2AF1 S34F exhibited altered affinity relative to the wild-type protein for an RNA oligonucleotide derived from an S34F-associated cassette exon [25]. However, it is not

known whether such differences in RNA-binding affinity account for most S34F-associated splicing alterations.

Here, we combine evidence from several different experimental approaches to show that WT U2AF1 is an inhibitor of S34F-associated splicing in lung epithelial and cancer cells. Analyses of the transcriptomes of primary LUAD samples as well as isogenic lung cells in culture indicate that the ratio of mutant to WT U2AF1 gene products is a critical determinant of the magnitude of S34F-associated splicing. S34F-associated splicing alterations can be explained largely, but not entirely, by differences in the relative affinities of mutant versus wild-type U2AF1 for the relevant 3´ splice sites. Moreover, we show that proliferation of cancer cells with *U2AF1 S34F* is critically dependent on WT *U2AF1*. Overall, we have revealed two pivotal roles of WT *U2AF1* in cells with *U2AF1 S34F*: it inhibits S34F-associated splicing and supports cell growth.

## Results

### S34F-associated splicing correlates with S34F:WT mRNA ratios in LUAD

Before undertaking experiments to study the effects of mutant U2AF1 in cultured cells, we first examined 512 transcriptomes from primary human LUADs published by The Cancer Genome Atlas (TCGA; [9]). Thirteen of these tumors harbor the most common *U2AF1* mutation, S34F. (Two others carry rare mutations of unknown significance, S65I and G216R, and were not considered further.) Previous reports have indicated that the *U2AF1 S34F* mutation affects the likelihood of inclusion of so-called "cassette exons" into mature mRNAs [24-27]. Therefore, we identified all the cassette exons whose inclusion was increased or decreased by ten percent or more in each tumor with the *U2AF1 S34F* mutation, relative to the average inclusion of each cassette exon across all of tumors without a *U2AF1* mutation. We then sought to identify enrichment of particular nucleotide sequences at the 3' splice site immediately upstream of the cassette exons that were promoted or repressed in association with *U2AF1 S34F*, represented by "sequence logos" as shown in Figures 1A and S1. To confirm that any identified 3' splice site alterations were specific to tumors with a *U2AF1* mutation, we also performed the same analysis on 19 tumors lacking a *U2AF1* mutation as controls. In addition, we used DNA and mRNA copy number data from the affected tumors to identify potential correlations between S34F-associated splicing and the ratio of mutant to WT U2AF1.

As illustrated by data from patient 7903 in Figure 1A, over two hundred cassette exons were included more frequently and a similar number were included less frequently in this *U2AF1*-mutant tumor. Notably, as illustrated by the sequence logos,

the nucleotide distribution at the -3 position (boxed) of promoted and repressed exons was different from that observed upstream of the much larger number of unaffected exons: A replaced T as the second most common nucleotide preceding the promoted exons; T was more common than C in the sequence preceding the repressed exons. This pattern was observed in nine of the thirteen tumors with the *U2AF1 S34F* allele (Supplemental Figures S1A). This pattern has been observed previously in comparisons between transcriptomes carrying the *U2AF1 S34F* allele with WT transcriptomes [24-27]. Therefore, we refer to this pattern as typical S34F-associated splicing. In the other four tumors with the *U2AF1 S34F* mutant, this typical S34F-associated pattern of the nucleotide distributions at the -3 position were partially or completely lost (Supplemental Figure S1B), resulting in sequence logos similar to those from tumors lacking a *U2AF1* mutation, where variations in inclusion are presumably stochastic (Supplemental Figure S1C). Thus, this type of sequence logo, as shown for tumor from patient 7727 in Figure 1A, was called “quasi-WT”.

To explain why transcriptomes of some tumors showed a typical S34F-associated pattern and others exhibited a quasi-WT pattern, we estimated the levels of mutant and total *U2AF1* mRNA based on available data from the tumors to determine the S34F:WT mRNA ratio. Strikingly, tumors with the quasi-WT pattern had low S34F:WT mRNA ratios (ranging from 0.27 to 0.31), whereas all but one tumor with the S34F-associated pattern had a higher ratio (0.43 or more) (Figure 1B). Absolute *U2AF1 S34F* or total *U2AF1* mRNA levels were not different between these two groups of tumors (Supplemental Figures S2A and S2B). The wide range of S34F:WT ratios (ranging from 0.26 to 0.82) could not be explained by contamination of tumor cells with

non-tumor cells, since the proportion of tumor nuclei reported for these samples did not correlate with S34F:WT mRNA ratios (Figure 1C) or with the S34F-associated splicing pattern (Supplemental Figures S2C). On the other hand, variations in *U2AF1* DNA copy number (CNV) correlated with the numbers of total *U2AF1* transcripts (Supplemental Figure S2D). Seven of the 13 *U2AF1 S34F* mutant samples showed either copy number loss or gain at the *U2AF1* locus, suggesting that CNV might account, at least in part, for the varying ratios of mutant and WT *U2AF1* mRNA in LUAD samples. Unbalanced allelic expression or proportions of tumor subclones might also contribute to the variable S34F:WT mRNA ratios, although these possibilities could not be easily determined using the available LUAD datasets.

We also examined the effects of different *U2AF1* S34F:WT mRNA ratios in these LUAD tumors by assessing their relationship to the inclusion of a few S34F-associated cassette exons that were reported previously [24,26,27]. This included two mRNAs with less (*ASUN* and *STRAP*) and one with more (*ATR*) inclusion of a cassette exon in the presence of *U2AF1 S34F*. Tumors with the highest ratios of mutant to WT U2AF1 mRNA showed the lowest inclusion levels of the cassette exon in *STRAP* mRNA, whereas tumors without a U2AF1 mutation had the highest level of inclusion (Figure 1D). Similar correlations were observed between inclusion levels of the *ASUN* and *ATR* cassette exons and the S34F:WT mRNA ratios (Supplemental Figures S3E and S3I). We also tested for a correlation between the inclusion of these cassette exons and levels of *U2AF1 S34F* mRNA, total *U2AF1* mRNA, or percent tumor nuclei. None of these analyses showed a relationship as strong as that observed with the S34F:WT mRNA ratio (Supplemental Figure S3B-D, F-H and J-L). These results further support

the notion that the S34F:WT mRNA ratio is the best predictor of the magnitude of S34F-associated splicing in human LUAD.

**Creation of isogenic lung cell lines that recapitulate features of S34F-associated splicing in LUAD.**

The results presented in the preceding section, based on analysis of LUAD tumors with the *U2AF1 S34F* mutation, suggest that the extent of S34F-associated splicing is a function of the S34F:WT mRNA ratio. To directly test this hypothesis, we developed a cell line that allows manipulation of WT and mutant U2AF1 gene product levels and measurement of the corresponding effects on RNA splicing.

The human bronchial epithelial cell line (HBEC3kt) was derived from normal human bronchial tissue and immortalized by introduction of expression vectors encoding human telomerase reverse transcriptase (*hTERT*) and cyclin-dependent kinase-4 (*CDK4*) [28]. To create a *U2AF1 S34F* allele at an endogenous locus in HBEC3kt cells, we adopted a published genomic DNA editing approach [29], using the *PiggyBac* transposon that leaves no traces of exogenous DNA at the locus (Figure 2A, see Materials and Methods for details). We identified three cell clones at intermediate stages of gene editing after screening more than 50 primary transfectants (Supplemental Figures S4A and S4B). Sanger sequencing of these intermediate alleles revealed that one of the three clones carried the desired S34F missense sequence, while two clones were WT (Supplemental Figure S4C). WT intermediates were expected because a homologous sequence between the S34F point mutation and the drug cassette in the vector can serve as the 5´ homology arm for recombination

(designated as 5´ HA#2 in Figure 2A). From the final clones derived from these intermediates (after transposition to remove the *PiggyBac* element), we chose one subclone from each of the two WT intermediate clones (referred to as WT1 and WT2 cells) and two subclones from the sole mutant intermediate clone (referred to as MUT1a and MUT1b cells) for all subsequent experiments with isogenic HBEC3kt cells (Supplemental Figure S4D). The MUT and WT cells all expressed similar levels of *U2AF1* mRNA and protein (Supplemental Figure S4E). Using high-throughput mRNA sequencing (RNA-seq) and allele-specific RT-qPCR, we observed similar levels of WT and S34F mutant *U2AF1* mRNAs in the mutant cells (Figures 2B and S5), consistent with heterozygosity at the *U2AF1* locus.

To examine how the engineered *U2AF1 S34F* allele affects mRNA splicing, we first assayed the inclusion levels of 20 cassette exons that were previously reported to be associated with mutant *U2AF1* in both LUAD and AML (acute myeloid leukemia) [26]. We confirmed that all 20 of these cassette exons were indeed S34F-dependent, using RT-qPCR with isoform-specific primers (Figures 2C and S6). For example, the cassette exons of *ASUN* and *STRAP* mRNAs were more frequently skipped in *U2AF1* mutant cancer transcriptomes, as shown above. Similarly, these mRNAs showed an increased ratio of “short” (exon-skipped) to “long” (exon-included) isoforms in MUT cells as compared to WT cells. We further evaluated the global pattern of use of cassette exons in MUT and WT cells using RNA-seq (Supplemental Table S1). Unsupervised cluster analysis grouped MUT and WT cells separately (Figure 2D). When the 3´ splice site sequences preceding the affected cassette exons were used to assemble sequence logos, typical S34F-associated pattern at the -3 position was observed with data from

MUT cells (Figure 2E). Overall, these results indicate that we successfully created clonal HBEC3kt cells isogenic for *U2AF1 S34F* and that the MUT cells carry out a typical S34F-associated splicing program.

**The ratio of S34F:WT gene products controls S34F-associated splicing in isogenic lung cells.**

We next used our isogenic cell model with the *U2AF1* mutation to test the hypothesis that the S34F:WT mRNA ratio, rather than absolute levels of the mutant or wild-type mRNA, controls S34F-associated splicing. We examined two specific predictions. First, we reasoned that changing the levels of U2AF1 S34F, while keeping the S34F:WT ratio constant, should not affect S34F-associated splicing. Second, we predicted that changing the level of WT U2AF1 while keeping the level of U2AF1 S34F constant (e.g., allowing the S34F:WT ratio to change) should alter the inclusion of S34F-dependent cassette exons. We tested these predictions in the isogenic HBEC3kt cells by manipulating levels of WT or mutant *U2AF1* gene products and measuring the subsequent changes in S34F-associated splicing.

We first reduced the amounts of both mutant and WT *U2AF1* mRNA concordantly in MUT1a cells, keeping the S34F:WT mRNA ratio constant. This was achieved by transducing MUT1a cells with short hairpin RNAs (shRNAs) that target regions of the *U2AF1* transcripts distant from the S34F missense mutation. The same shRNAs were also introduced in WT1 cells as a control. Allele-sensitive RT-qPCR confirmed that the mRNA ratio of the mutant and WT *U2AF1* remained constant (Supplemental Figure S7A), while the overall U2AF1 mRNA and protein levels were

reduced by more than 90% (Figures 3A, bottom panel, and S7B). Knockdown of total U2AF1 in both MUT1a and WT1 cells did not cause significant changes in recognition of the *ASUN* or *STRAP* cassette exons, two splicing events that are strongly associated with *U2AF1 S34F*, in either cell line (Figure 3A, upper panels). Similar results were obtained for two additional S34F-associated cassette exons in *USP25* and *AXL* that exhibit increased inclusion in cells expressing *U2AF1 S34F* (Supplemental Figures S7C and S7D).

We next confirmed that our *U2AF1* knockdown was sufficient to alter known U2AF1-dependent splicing events. We studied a competing 3' splice site event in *SLC35C2* mRNA, in which the recognition of an intron-proximal over an intron-distal 3´ splice site depends on the level of U2AF1 independent of a *U2AF1* mutation [30]. Knockdown of total *U2AF1* mRNA in either WT1 or MUT1a cells lines reduced the use of this U2AF1-dependent 3´ splice site in both WT1 and MUT1a cells (Supplemental Figures S7E and S7F). Thus, reduction of *U2AF1* to levels that affected U2AF1-dependent alternative splicing, while preserving the S34F:WT ratio, did not affect S34F-associated splicing in MUT1a cells.

We next altered the S34F:WT ratio by overexpressing mutant or WT *U2AF1* in WT1 and MUT1a cells and examining the subsequent changes in the recognition of *ASUN* and *STRAP* cassette exons. These cassette exons are preferentially excluded in cells expressing *U2AF1 S34F*. Increasing the amount of U2AF1 S34F protein, and hence increasing the S34F:WT ratio in either cell type, further enhanced skipping of these cassette exons (Figure 3B). Conversely, decreasing the S34F:WT ratio in MUT1a

cells by increasing the production of WT U2AF1 protein reduced the extent of exon skipping to near the levels seen in WT1 cells (Figure 3B).

We also altered the production of WT U2AF1 mRNA and protein in MUT1a cells by disrupting the endogenous WT *U2AF1* locus with the CRISPR-Cas9 system. Single-guide RNAs (sgRNAs) designed to match either the WT or mutant *U2AF1* sequences were shown to selectively disrupt either reading frame, generating indels (insertions and deletions) at the *U2AF1* locus and thereby changing the S34F:WT ratios (Supplemental Figures S8 and S9). Since WT *U2AF1* is required for the growth of these cells (as shown below; Figure 6), we extracted RNA and protein from cells six days after transduction with sgRNA-WT/Cas9, when depletion of WT U2AF1 was incomplete (Figure 3C). Selective disruption of WT *U2AF1* increased the S34F:WT mRNA ratio in MUT1a cells; as predicted, the extent of exon skipping was further increased in *ASUN* and *STRAP* mRNAs (Figure 3C). Notably, the degree of exon skipping induced by mutant cDNA was similar to that caused by disrupting the WT *U2AF1* allele (compare Figures 3C with 3B), even though the absolute protein levels of U2AF1 S34F were different in the two experiments. These results suggest that WT U2AF1 antagonizes the activity of U2AF1 S34F by a competitive mechanism. In other words, the S34F-associated pattern of splicing can be enhanced in a manner independent of the concentration of either form of the protein, as long as the S34F:WT ratio is increased.

**Disruption of WT U2AF1 globally enhances S34F-associated splicing.**

Results in the preceding sections are based on studies of a few well-documented S34F-sensitive cassette exons that likely serve as surrogates for the widespread effects

of U2AF1 S34F on the splicing profile. To determine whether these results reflect general rules governing S34F-associated splicing, we used RNA-seq to measure the effects of U2AF1 S34F on the global recognition of cassette exons and on other types of alternative splicing that are S34F-associated. When WT *U2AF1* was diminished by CRISPR/Cas9-mediated disruption in MUT1a cells (see Figure 3C), the S34F-associated changes in inclusion (Figure 4A) or skipping (Figure 4B) of cassette exons were enhanced. These global results are consistent with our measurements of individual cassette exons by RT-qPCR (Figure 3C). We also observed a slight enhancement of the pattern of S34F-associated nucleotides at the -3 position in sequence logos when the WT *U2AF1* allele was disrupted (Figure 4C), consistent with the observed association between the S34F:WT mRNA ratio and typical S34F-associated sequence logos computed from LUAD tumor data (Figure 1).

We observed similar results when we extended our analyses to other types of alternative splicing associated with *U2AF1 S34F* (Supplemental Figure S10). We used RT-qPCR to validate five splicing alterations that are sensitive to ablation of WT *U2AF1* in the presence of *U2AF1 S34F* (Supplemental Figure S11). Two of the five events involve incorporation of cassette exons (in *ATR* and *MED15*), two involve competition for use of alternative 3´ splice sites (in *CASP8* and *SRP19*), and one involves a choice between two mutually exclusive exons (in *H2AFY*). We conclude that the apparent inhibition of the activity of U2AF1 S34F by WT U2AF1 on S34F-associated splicing is not restricted to cassette exons.

So far, our analyses have been restricted to alterations in splicing. We further asked whether S34F-associated changes in gene expression were also influenced by

changes in the S34F:WT mRNA and protein ratios. We identified genes that showed at least a ten percent increase or decrease in expression in MUT1a cells compared to the expression in WT1 cells. Increases in the S34F:WT mRNA and protein ratios by disrupting the WT *U2AF1* locus, further increased or decreased expression of similar numbers of these genes, suggesting that the changes in gene expression were not determined by the ratio of S34F:WT gene products (Supplemental Figure S12). This is in contrast to our findings in similar experiments that measured alternative splicing (see Figures 4 and S10 for comparison).

**Changes in RNA splicing correlate with relative binding affinities of mutant and WT U2AF1 complexes.**

Results presented above are consistent with a model in which mutant and WT U2AF1 compete to determine the outcome of a subset of splicing events; yet the nature of this competition has not been defined. One possible explanation depends on the relative RNA affinities of protein complexes containing mutant or WT U2AF1 for binding to the alternative 3´ splice sites. This hypothesis is consistent with the role of U2AF1 in facilitating exon recognition, and with a recent study showing that a reduced affinity of mutant U2AF1 complexes for a proximal 3´ splice site was associated with skipping of this exon in cells containing the mutant splicing factor [25].

To test this hypothesis, we determined the RNA binding affinities of purified proteins by fluorescence anisotropy assays, in which fluorescein-labeled RNA oligonucleotides are titrated with recombinant U2AF1-containing complexes. The resultant fluorescence anisotropy changes were fit to obtain the apparent equilibrium

binding affinities (Supplemental Figure S13; see Materials and Methods for details). The protein complexes comprised either WT or S34F mutant U2AF1 (residues 1-193), with two additional proteins, U2AF2 (residues 85-471 of isoform b) and SF1 (residues 1-255), that together are required for 3´ splice site recognition (Figure 5A). RNA oligonucleotides (33 - 35 nucleotides in length) were derived from the proximal or distal 3´ splice site junctions of cassette exons or of competing 3´ splice sites that generate new 5´ ends of exons (Figure 5B).

The binding affinities of mutant and WT U2AF1 protein complexes for the oligonucleotides at six pairs of alternative 3´ splice sites were tested; each of these sites has been shown to display increased or decreased use in our isogenic cell lines (Supplemental Table S1), and in LUAD transcriptomes when the thirteen *U2AF1* mutant samples were compared as a group with all WT samples (Supplemental Table S2). Results of the RNA binding assay and the observed splicing alterations by U2AF1 S34F are summarized in Supplemental Table S3. Overall, altered binding affinities of U2AF1-containing complexes to the proximal 3´ splice site could account for four of the six S34F-associated alternative splicing events (*CEP164*, *FMR1*, *ZFAND1*, *FXR1*) (Figures 5C - 5F). For example, mutant U2AF1 complex had a 4-fold lower affinity for the proximal 3´ splice site sequence upstream of the *CEP164* cassette exon (Figure 5C), consistent with preferential skipping of the cassette exon in the presence of *U2AF1 S34F*. Similarly, the mutant U2AF1 complex showed nearly 5-fold greater affinity than the WT complex for the proximal 3´ splice site in *ZFAND1* pre-mRNA at the upstream border of a cassette exon that is preferentially incorporated into *ZFAND1* mRNA in the presence of *U2AF1 S34F* (Figure 5E).

The two remaining tested events (*ATR* and *MED15*) show increased inclusion of a cassette exon in the presence of *U2AF1 S34F*, but the affinities for binding of corresponding proximal 3´ splice sites were not consistent with the observed changes in splicing. That is, the affinities of WT and S34F complexes were not significantly different with the *ATR* oligonucleotide and were lower, not higher, for mutant U2AF1 complexes than that for WT U2AF1 complexes with the *MED15* oligonucleotide (Figures 5G and 5H, first and second columns). However, for the *ATR* pre-mRNA, we found that the relative affinities of mutant and WT U2AF1 complexes for oligonucleotides representing the distal 3´ splice sites (located downstream of the cassette exon) could explain the observed increased inclusion of the cassette exon. Given co-transcriptional splicing in the 5´-to-3´ direction, downstream (as opposed to upstream) 3´ splice sites could compete as splicing acceptors for a given 5´ donor splice site when transcription is relatively rapid as compared to splicing. Since the *U2AF1* mutation reduced binding to the distal more than to the proximal 3´ splice site (Figure 5G, third and fourth columns), a “net gain” in recognition of the proximal 3´ splice site could explain the observed S34F-associated exon inclusion in *ATR* mRNAs. For the *MED15* pre-mRNA, the S34F-induced splicing changes were inconsistent with a simple RNA affinity model, suggesting that additional mechanisms control recognition of the *MED15* 3´ splice sites.

**HBEC3kt and LUAD cells are not dependent on U2AF1 S34F for growth, but WT U2AF1 is absolutely required.**

Other than its effect on RNA splicing, the consequences of the *U2AF1 S34F* mutation on cell behavior are largely unknown. Recurrent mutations, such as *U2AF1*

*S34F*, are considered likely to confer a selective advantage to cells in which they occur when expressed at physiologically relevant levels. However, mutant HBEC3kts cells (MUT1a and MUT1b) do not exhibit obvious phenotypic properties of neoplastic transformation --- such as a growth advantage over control cells (Supplemental Figure S14) or an ability to grow in an anchorage-independent manner --- traditionally observed in cultured cells expressing well-known oncogenes, like mutant *RAS* genes.

Another attribute of some well-known oncogenes, such as *BCR-ABL* fusion in chronic myeloid leukemia or mutant *EGFR* or *KRAS* in LUAD, is the dependence on sustained expression of those oncogenes for the maintenance of cell growth or viability. To determine whether LUAD cells harboring a pre-existing *U2AF1 S34F* mutation are dependent upon (or "addicted to") the mutant allele, we searched the COSMIC database for LUAD cell lines with the *U2AF1 S34F* mutation [31]. We found two such lines (H441 and HCC78) and confirmed the retention of the *U2AF1 S34F* allele by Sanger sequencing and allele-specific RT-qPCR (Supplemental Figure S15A). We used the CRISPR-Cas9 system to selectively disrupt the WT or mutant *U2AF1* sequences. We then assessed the impact of inactivating the *U2AF1* alleles on the clonogenic growth of the two LUAD lines with the *U2AF1* mutation. In addition, we performed similar experiments with the LUAD cell line A549 (WT for U2AF1) and the MUT1a cell line.

In all instances, loss of the mutant allele did not impair cell growth. Only one line (H441) exhibited altered growth, in the form of a two-fold increase in clonogenicity (Figure 6A). Successful disruption of the *U2AF1 S3F* allele was confirmed by restoration of a normal RNA splicing profile in sub-clonal cells derived from the clonogenic assay

colonies (see Supplemental Figures S15 – S18, Tables S4 and S5, and text below). In contrast, loss of the WT allele completely inhibited clonogenic growth in all tested cell lines, regardless of whether the line carried the *U2AF1 S34F* allele or not. A rescue experiment confirmed that loss of cell growth was due to loss of WT *U2AF1* expression. The loss of clonogenic capacity after disrupting endogenous *U2AF1* in A549 cells was prevented by first transducing them with a form of *U2AF1* cDNA (Figure 6B), which is not predicted to be the target for sgRNA-WT/Cas9 (Supplemental Figure S8). Overall, these findings indicate that WT *U2AF1* is required for clonogenic growth of cells, including lung cancer cell lines, that the *U2AF1 S34F* mutant is unable to compensate for elimination of the *WT* allele, and that LUAD cells with the *U2AF1 S34F* mutation are not dependent on the mutant allele for growth.

To examine the effect of *U2AF1 S34F* on tumor growth *in vivo*, we derived H441 and HCC78 cells transduced with sgRNA-S34F/Cas9 or sgRNA-GFP/Cas9 as polyclonal pools (Supplemental Figure S15) or as clones (Supplemental Figure S16). The cell lines were verified to either carry or not carry the *U2AF1 S34F* allele, and we confirmed that the sgRNA-S34F/Cas9-transduced cells lost the S34F-associated splicing program (Supplemental Figure S17 and S18).

We inoculated these subclonal cell lines subcutaneously in nude mice and monitored xenograft tumor growth. The H441-derived cell lines, in which the *U2AF1 S34F* allele was disrupted, were able to establish tumors *in vivo*, at rates similar to that observed with tumor cells carrying the mutant allele (Supplemental Table S6). The HCC78-derived cell lines failed to grow palpable tumors after xenografting within the observation period, so the requirement for *U2AF1 S34F in vivo* could not be tested in

that line. These experiments show that *U2AF1 S34F* is dispensable for growth of these LUAD cell lines *in vivo*, a result consistent with the clonogenicity assays shown in Figure 6. We conclude that *U2AF1 S34F* appears to be neither sufficient nor necessary for lung cell transformation in these assays. In contrast, the WT *U2AF1* allele is required for cell viability, a conclusion consistent with the retention of a WT allele in human cancers carrying common *U2AF1* mutations.

## Discussion

### Mechanisms underlying S34F-associated splicing

In this report, we have examined the consequences of a *U2AF1* splicing factor mutation, S34F, that is highly recurrent in a variety of human cancers. Although we remain unable to account for the apparent selective advantage conferred on cells that harbor this mutation, we have observed several important effects on the RNA splicing apparatus, shown that the magnitude of the effects reflects the relative abundance of mutant and WT U2AF1, and demonstrated that cell viability is dependent on retention of a WT allele even in the presence of mutant *U2AF1*.

Earlier work revealed that *U2AF1 S34F* induces alterations in the use of cassette exons; collectively, the affected genes display altered sequences at 3' splice acceptor sites and we use the altered sequences as the signature of "S34F-associated splicing" [14,24-27]. In our analysis of human LUAD transcriptomes, we found that most tumors with the *U2AF1 S34F* allele showed the anticipated pattern of S34F-associated splicing. However, four such transcriptomes did not, and all four of these transcriptomes had low S34F:WT mRNA ratios. Further analysis confirmed that the S34F:WT mRNA ratio correlated with the extent of S34F-associated splicing in LUAD transcriptomes (Figure 1).

We tested the hypothesis that the relative amount of mutant to WT U2AF1, rather than the absolute amount of U2AF1 S34F gene products, controls S34F-associated splicing. For this purpose, we made an isogenic lung epithelial cell line in which *U2AF1 S34F* is engineered to be expressed from one of the two endogenous U2AF1 loci. By changing the levels of WT and mutant U2AF1 individually or simultaneously, we found

that the extent of S34F-associated splicing was altered only when the ratio of S34F:WT gene products was disturbed. We further performed biochemical studies *in vitro*, showing that pre-spliceosome complexes containing mutant or WT U2AF1 have different affinities for functionally relevant splice sites; these differences could explain most, but not all, observed S34F-associated changes in RNA splicing. We conclude from these results that WT and mutant U2AF1 act competitively to control splicing, possibly through differential binding to the affected splice sites.

Our findings about the effects of U2AF1 S34F on splicing in lung cells, combined with previous findings, reveal some general features of S34F-associated splicing:

1. Molecular consequences of U2AF1 S34F on pre-mRNA splicing in different cell types are very similar. Many S34F-associated splicing events, as well as the overall consensus splice site sequence alterations, are nearly identical in lung epithelial and LUAD cells and in *U2AF1*-mutant AML or MDS cells [24-27]. This is a remarkable result, considering that lung and myeloid neoplasms originate from different cell lineages; that the cells harbor largely different sets of “driver” oncogenes [9,32]; and that alternative splicing is usually considered to be tissue-specific [33]. Because of these similarities, findings from our study are likely to be applicable to the mechanisms of action of mutant and WT U2AF1 in other cell types, including hematopoietic cells.

2. U2AF1 S34F acts primarily by competing with WT U2AF1. U2AF1 is thought to be dispensable for splicing of a subset of pre-mRNA substrates, based on biochemical assays *in vitro* [19-21]. We and others have shown that U2AF1 S34F affects splicing of only a small set of pre-mRNAs [24-27]. Therefore, at least two scenarios could account for the altered activity of the mutant protein. First, U2AF1 S34F could affect splicing in a

manner independent of WT U2AF1 (for example, it could mediate splicing at a U2AF1-independent site); the amount of U2AF1 S34F should then determine the extent of altered splicing. Second, U2AF1 S34F might compete with WT U2AF1---for instance, for binding to certain sites in pre-mRNA---to alter splicing; in this case, alterations in splicing would depend on the ratio of mutant to WT U2AF1. Our results with isogenic lung cell lines, analysis of LUAD transcriptomes, and *in vitro* RNA binding assays clearly favor the second scenario.

3. The main functional difference between purified mutant and WT U2AF1 proteins appears to reside in the binding affinities for some 3´ splice sites. This conclusion is consistent with the location of the missense mutation in codon 34, which normally encodes a conserved amino acid in a zinc knuckle motif that is most likely in direct contact with RNA [24]. We found that recombinant mutant and WT U2AF1 form complexes with U2AF2 and SF1 at similar efficiencies. Therefore, mutant U2AF1 appears equally capable of recruiting these co-factors for splicing. These results are consistent with a previous report that nuclear extracts of cells overexpressing mutant *U2AF1* can catalyze specific *in vitro* splicing reactions as or more efficiently than nuclear extracts derived from cells overexpressing WT *U2AF1* in a sequence-dependent manner [24]. Mutant U2AF1 could also compensate for the loss of WT U2AF1 for the inclusion of some U2AF1-dependent cassette exons [34]. Together, these results suggest that U2AF1 S34F is functionally equivalent to WT U2AF1 once it binds to a 3' splice site, and support our hypothesis that the primary functional difference between WT and mutant U2AF1 lies in their respective RNA-binding affinities.

4. Both proximal and distal 3´ splice sites are likely relevant to the consequences of *U2AF1 S34F* expression for cassette exon recognition. Differential binding of mutant versus WT U2AF1 complexes to the proximal 3´ splice site predicts differential splicing in the presence of the *U2AF1 S34F* mutation for four alterative splicing events that we measured (Figures 5C – 5F). This finding is consistent with a previous report [25]. However, in the example of the *ATR* cassette exon, which is more often included in the presence of *U2AF1 S34F*, mutant and WT U2AF1 complexes exhibit similar affinities for the 3´ splice site immediately preceding the cassette exon, but the relative affinity of the mutant complex for the distal, downstream 3´ splice site is reduced (Figure 5G). We interpret this to mean that the relative affinity of U2AF1 for the upstream versus downstream 3' splice site contributes to preferential recognition of each splice site in mutant versus WT cells. This potential mechanism requires mutant U2AF1 to open a window of opportunity for splicing to the downstream 3´ splice site, which could occur either through alteration of splicing or transcription rates. Although recent results indicate that splicing and transcription are closely coupled in time and space [35], a wealth of data supports the sensitivity of splice site choice to kinetic competition [36,37]. In support of this notion, a recent report showed that U2AF1 S34F inhibited co-transcriptional splicing of a model pre-mRNA, deferring the splicing reaction to a post-transcriptional phase [36]. If U2AF1 S34F delays splicing, competition between proximal and distal 3´ splice sites would be facilitated. Further studies will be needed to fully resolve the contribution of kinetic competition to the effects of common splicing factor mutations on alternative splicing.

5. The -3 nucleotide does not alone determine the altered binding affinity of

mutant U2AF1. This is best illustrated by the binding assay results for RNA oligonucleotides with an A in the -3 position. Since the consensus motif for U2AF1 recognition of 3' splice sites includes a pyrimidine at the -3 position (yAG|r) [19], it was striking to find an enrichment of A at this position preceding exons that show increased inclusion in cells with U2AF1 S34F. We determined binding affinities for the 3´ splice sites of three such cassette exons (*ZFAND1*, *ATR* and *MED15*). The affinities of complexes with mutant U2AF1 for these oligonucleotides were increased, decreased, or did not change relative to WT U2AF1. Therefore, sequences flanking the -3 nucleotide must also influence the altered binding of mutant U2AF1 complex.

6. Features that cannot be described in binding assays with RNA oligonucleotides contribute to mechanisms for S34F-associated splicing. For example, the RNA binding affinity results in Figure 5H are inconsistent with the observed increased inclusion of *MED15* cassette exons in the presence of U2AF1 S34F. This suggests that additional mechanisms are required to fully explain the splicing alterations affected by U2AF1 S34F. Most obviously, the binding reaction *in vivo* might be affected by more than the oligonucleotides used in our *in vitro* assays, including secondary or tertiary RNA structure. Alternatively, recent findings suggest that U2AF1 might suppress, rather than facilitate, splice site recognition in some pre-mRNAs [38]. This non-canonical activity of U2AF1 could account for the observed, S34F-associated splicing pattern for the *MED15* cassette exon. As a third option, mRNAs encoding other splicing factors undergo alternative splicing in the presence of U2AF1 S34F [14,24], and the resulting isoforms might affect the inclusion of the *MED15* cassette exon indirectly in the presence of U2AF1 S34F. Future studies are needed to test these possibilities.

**How are cells with the U2AF1 S34F mutation selected during oncogenesis?**

The *U2AF1 S34F* mutation is recurrently found in LUAD, other solid tumors, and myeloid disorders, suggesting that the mutant allele confers a physiological property that provides a selective advantage during neoplasia. However, *U2AF1 S34F* does not transform HBEC3kt cells in culture or affect cell proliferation. Moreover, targeted inactivation of *U2AF1 S34F* in LUAD cell lines does not diminish, and in one case even increases, clonogenic growth in culture (Figure 6). These results are consistent with conclusions from a previous report in which cell proliferation was not affected in an erythroleukemic cell line, K562, when *U2AF1 S34F* was expressed at moderate level as a transgene [24].

The lack of a testable cellular phenotype is a major hindrance to understanding the functional significance of mutant *U2AF1*. Cell proliferation is only one of the many hallmarks of cancer, so careful examination of other cell properties may be needed to establish the presumptive role of *U2AF1 S34F* in carcinogenesis. The isogenic cell lines we have generated from LUAD cells or from the HBEC3kt line may prove to be useful tools to test these properties.

It is possible that mutant splicing factors will influence cell behavior only in certain cell types, since alterations in the splicing pattern that generate pro-oncogenic mRNAs might occur only in the context of a particular transcriptional program that is highly lineage-restricted. This possibility might be best pursued in an animal model. It was recently reported that induction of a *U2af1 S34F* transgene in a genetically engineered mouse caused increased proliferation of hematologic progenitor cells and apoptosis in

monocytes [14]. Since the S34F:WT ratio is the critical determinant of the effects of U2AF1 S34F on splicing, we have recently engineered and begun to characterize a mouse strain that expresses the mutant U2af1 from the endogenous locus in an Cre-inducible fashion.

**Can cancer cells carrying the U2AF1 S34F mutation be targeted therapeutically?**

The *U2AF1 S34F* mutation is an intriguing if enigmatic target for therapeutics against multiple cancer types. The work reported here provides some guidance for thinking about those prospects.

*Targeting WT U2AF1.* We have shown that the WT *U2AF1* allele is absolutely required for the growth of lung epithelial and LUAD cells that carry a mutant *U2AF1* allele (Figure 6). This result indicates that mutant U2AF1 cannot complement a deficiency of U2AF1 and explains why tumors are never homozygous for the *U2AF1 S34F* mutation. Furthermore, the frequent occurrence of a low ratio of mutant:WT *U2AF1* mRNA, accompanied by increased copies of *U2AF1* alleles, suggests that there may be selection for a favorable ratio of mutant to WT U2AF1 in addition to the likely selection, perhaps at an earlier stage of tumorigenesis, for the U2AF1 S34F mutant. It has recently been shown that cancer cells harboring recurrent *SF3B1* mutations also depend on WT SF3B1 for growth [39]. Together, these results provide a rationale for targeting WT splicing factors (or the splicing machinery more generally) in cancers harboring splicing factor mutations, assuming that differences between normal cells and cancer cells with mutations in splicing factors provide a therapeutic window to allow selective killing of cancer cells.

*Strategies that depend on the properties of mutant U2AF1*. Although we do not understand how mutant splicing factors drive carcinogenesis, there are several hypothetical ways in which the presence of a mutant factor might be exploited therapeutically. As mentioned above, mutations like *U2AF1 S34* might make cancer cells vulnerable to inhibitors of the normal splicing machinery. Alternatively, it is possible that one or a few of the changes in the transcriptome can change the nature or abundance of proteins to promote neoplasia; if so, proteins identified as contributory to the cancer phenotype might be targeted. Finally, the altered transcriptome in cancer cells with mutant splicing factors might create a context in which secondary changes in function of some unrelated gene(s) might have detrimental effects on cell viability; such changes, resulting from enhanced or diminished function of normal genes, can now be sought with high throughput screens for synthetic lethality in appropriate isogenic cell lines of the sort described here.

## Conclusion

Our study is the first to use isogenic lung cell lines to investigate the molecular and phenotypic properties of a common splicing factor mutation, *U2AF1 S34F*. We show that the relative amounts of mutant and WT U2AF1 determine the extent of S34F-associated splicing. Combining these results with those from an *in vitro* binding assay, we propose that U2AF1 S34F affects splicing principally by competing with WT U2AF1 for binding at the affected 3´ splice sites. We also use genetically engineered and edited cells to show that the WT *U2AF1* allele, but not the mutant allele, is required for cell growth. These results highlight the essential role that WT U2AF1 plays in cancer cells

with a *U2AF1* mutation, and the cell lines should be helpful in future attempts to identify the functional contributions of mutant *U2AF1* to neoplasia.

## Materials and Methods

### *Cell culture, reagents, immunoblots and lentiviruses*

The HBEC3kt cell line was a gift from Dr. John Minna (University of Texas Southwestern Medical Center). This cell line and all its derivatives were grown in Keratinocyte serum-free media supplemented with epidermal growth factor and bovine pituitary extract (Invitrogen) as previously described [28]. The H441 and A549 LUAD cell line was purchased from ATCC (American Type Culture Collection). The HCC78 LUAD cell line was purchased from DMSZ (German Collection of Microorganisms and Cell Cultures GmbH). H441, HCC78, A549 and their derivatives were grown in RPMI1640 (Lonza) supplemented with 10% fetal bovine serum (HyClone) and antibiotics (Penicillin-Streptomycin, Gibco). All cell cultures were incubated in a 37 degree humidified incubator with 5% $CO_2$. Immunoblots were performed as previously described [40]. The primary antibodies used in the study are: rabbit anti U2AF1 (1:5000, # NBP1-19121, Novus), rabbit anti GFP (1:5000, #A-11122, Invitrogen), mouse anti ACTB (1:5000, Clone 8H10D10, Cell Signaling). Lentiviruses were produced and titered in HEK293T cells as previously described [41]. An MOI (multiplicity of infection) of 1 – 5 were used for all assays.

### *Plasmid DNA*

*Plasmids for engineering the U2AF1 S34F mutation*. The plasmids used for creating the endogenous S34F mutation in HBEC3kt cells were custom designed and synthesized by Transposagen. They include the donor vector (pU2AF1-S34F_Target),

the TALEN (Transcription Activator-like Effector Nucleases)-expressing vectors (pU2AF1-FXTN, pU2AF1-RXTN) and the *PiggyBac*-expressing vector (pCMV-Pbo).

*shRNA constructs*. The shRNA constructs for *U2AF1* knockdown were obtained from OpenBiosystems (shU2AF1#1, TRCN0000001155; shU2AF1#4, TRCN0000001158). The control shRNAs (shScbr and shGFP) were described previously [41].

*Lentiviral constructs for overexpression of transgenes*. The lentiviral constructs for overexpressing mutant or WT U2AF1 were constructed in multiple steps. Firstly, the open reading frame of mouse *U2af1* (GenBank BC115479) was PCR amplified from a cDNA clone and cloned into the pENTR/D-TOPO vector (Invitrogen) following vendor's instruction. Because mouse U2af1 (RefSeq NP_077149) is identical to human U2AF1 (RefSeq NP_006749) except lacking a Gly in the Gly rich domain (amino acid residues 212-222), a codon for Gly (GGA) was inserted into the corresponding *U2af1* DNA sequence via site-directed mutagenesis (Agilent). The S34F (TCT to TTT) missense mutation was also created by site-directed mutagenesis. The resulting WT and S34F mutant *U2af1* constructs were then cloned into a lentiviral destination vector, pLenti-CMV-Blast-DEST (Addgene # 17451, [42]), via LR Gateway reaction (Invitrogen). Entry vectors expressing enhanced green fluorescence protein (pENTR1A-GFP-N2, Addgene #19364, [42]) or DsRed-Express 2 [43] was used to clone the GFP or DsRed-Express 2 sequence into the same destination vector.

*pU2AF1-WT-S34F-e2*. This plasmid was used to serve as a reference to quantify the S34F:WT *U2AF1* ratio via quantitative PCR (qPCR). It was constructed by inserting a WT *U2AF1* exon 2 DNA fragment (GRCh37/hg19, Chr21: 44524425-

44524512) or the same sequence with the Ser34 codon mutated to Phe34 (TCT to TTT) into the miniTK-Luc plasmid [40]. One copy of the WT U2AF1 sequence was inserted 5´ of the *Luciferase* gene sequence via the Kpn I site. One copy of the mutant U2AF1 sequence was inserted 3´ of the *Luciferase* gene sequence via the BamH I site.

*CRISPR/Cas9 constructs*. The plasmids for expressing sgRNA-WT/Cas9 and sgRNA-S34F/Cas9 were constructed by inserting the relevant sgRNA sequences (WT, GTCATGGAGACAGGTGCTCT; S34F, GTCATGGAGACAGGTGCTTT) into lentiCRISPR (Addgene #49535, [44]) via a method described by the depositing investigator. The same lentiviral CRISPR/Cas9 vector expressing an sgRNA against GFP was obtained from Addgene (#51760, [44]).

All plasmids mentioned above were sequence verified. They will be deposited to Addgene upon manuscript acceptance.

***Genome editing approaches***

The creation of endogenous U2AF1 S34F missense mutation was conducted as described in Figure 2A following an established protocol [29]. The donor vector (pU2AF1-S34F_Target) contains the S34F missense mutation (TCT to TTT) flanked by homology arm sequences (HA) of roughly 1000 bp on each side of the point mutation. A PGK-Hygro∆TK drug cassette, flanked by *PiggyBac* recognition sequences (ITR, inverted terminal repeat), was located 3´ of the S34F mutation at an endogenous AATT site that is essential for *PiggyBac* recognition. The homologous recombination was facilitated by cutting the *U2AF1* genomic DNA by a pair of TALENs, which bind to the *U2AF1* genomic sequences located in intron 2 (underlined):

TAAGTGTGTTCTTTTATTTAAATAATAGTGAGGCAGGTGATCGACTTCCA

The donor vector and the TALENs were co-transfected into HBEC3kt cells in 6-well dishes by Lipofectamine 2000 (Invitrogen). The cells were transferred to 150 mm dishes and cultured in growth media containing Hygromycin B (50 µg/ml, Invitrogen). Cell clones (≥ 50 cells) were harvested using cloning rings (Thermo Fisher Scientific) and expanded. Genomic DNA from these clones were harvested and screened for mutant and WT intermediate clones by genomic PCR and Southern blot as described in Supplemental Figure S4 (P1-F, TGTCCTAATTCATCAGAGATCG; P1-R, TAAACCTCGATATACAGACCGATA). The correct intermediate cell clones were co-transfected with plasmids expressing an excision-only *PiggyBac* transposase (pCMV-Pbo), to remove the PGK-HygroΔTK drug cassette, and eGFP (pEGFP-C1). GFP positive cells were sorted on a BD FACSAria and cultured in the presence of Ganciclovir (10 µg/ml, Sigma). Cell clones were isolated by cloning rings and expanded. The DNA and RNA from these cells were used to confirm the removal of the drug cassette (by genomic PCR using the P1-F and P1-R primers as described in Supplemental Figure S4), as well as expression of *U2AF1 S34F* (by allele-specific qPCR and RNA-seq).

Disruption of the WT or mutant *U2AF1* alleles from H441, HCC78 and MUT1a cells were conducted by transducing these cells with lentiviruses expressing sgRNA-WT/Cas9 or sgRNA-S34F/Cas9. Virus-infected cells were selected by puromycin (1 µg/ml) except in HBEC3kt-derived cells since HBEC3kt cells were intrinsically resistant to puromycin [28]. Polyclonal and clonal cells were derived from these infected cells. CRISPR-mediated induction of InDels in the *U2AF1* locus was confirmed by

fluorescence PCR (data not shown) [45]. The disruption of the WT or mutant U2AF1 allele was further confirmed by changes in the ratios of S34F:WT gene products and subsequent changes in S34F-associated splicing.

***RNA extraction, reverse transcription, and quantitative PCR (RT-qPCR)***

RNA was extracted using Trizol (Molecular Research Center or Invitrogen) followed by a column cleanup step using an RNeasy kit (Qiagen). cDNA was synthesized from 1-2 µg total RNA using a High-Capacity cDNA Reverse Transcription (RT) kit (Applied Biosystems). Quantitative PCR (qPCR) was done in a 7900HT real-time PCR system (Applied Biosystems) (other than those described in Supplemental Figure S11). The mRNA level of total *U2AF1* and *GAPDH* was measured by inventoried Taqman assays (Hs01597469_m1, Hs99999905_m1, Applied Biosystems) in a standard Taqman qPCR master mix (Applied Biosystems). The allele-sensitive S34F/WT SNP Taqman assay were custom synthesized and was characterized in Supplemental Figure S5.

Splicing alterations were measured by isoform-specific primers (listed in Supplemental Table S7) in a SYBR Green qPCR master mix (Affymetrix). These primers were designed following a previously described method using a web tool: http://designs.lgfus.ca/ [46]. PCR efficiency of each primer set was estimated by performing the qPCR assay using at least four 10-fold serial dilutions of the cDNA template. The slope of the standard curve was then translated into a PCR efficiency value by the formula: $(10^{(-1/slope)} – 1) \times 100\%$. The PCR end products were separated on a 2% agarose gel to check specificity. The specificity of a subset of the primers was

further confirmed by Sanger sequencing using the PCR end products. All the primers used in the study have a PCR efficiency of 85% or more and are specific for the target they measure.

RT-qPCR results were quantified using a relative quantification method ($\Delta\Delta$Ct) in the RQ Manager 1.2 software (Applied Biosystems). In most cases, the reference sample was cells that underwent control treatment. For allele-sensitive S34F/WT SNP Taqman assay, the reference sample was the plasmid DNA pU2AF1-WT-S34F-e2, which contains one copy of each of the *U2AF1* WT and S34F mutant sequences.

### *mRNA sequencing*

High throughput mRNA sequencing (RNA-seq) was conducted in the Sequencing Facility of the National Cancer Institute. RNA quality was assessed by 2100 Bioanalyzer (Agilent). Total RNA with good integrity values (RIN > 9.0) was used for poly A selection and library preparation using the Illumina TruSeq RNA library prep kit. Two or four samples were pooled per lane and ran on the HiSeq2000 or HiSeq2500 instrument using TruSeq V3 chemistry. All samples were sequenced to the depth of at least 100 million pair-end 101 bp reads per sample.

### *Bioinformatics*

Analysis of RNA-seq data from the TCGA LUAD cohort as well as engineered HBEC3kt, H441, and HCC78 cell lines was performed as previously described [24]. A brief description follows below.

MISO v2.0 annotations were used for all splicing analyses [47], as was a set of annotated constitutive junctions defined as junctions with no evidence of alternative splicing in the UCSC knownGene database [48]. A genome annotation for read mapping was created by merging transcript annotations from MISO v2.0 annotations [47], the UCSC knownGene database [48], and the Ensembl 71 database [49]. Reads were first mapped to these transcripts using RSEM [50]. Remaining unaligned reads were then mapped to a database of possible junctions between all 5´ and 3´ splice sites of those transcripts, and then subsequently to the GRCh37/hg19 human genome assembly, using TopHat [51].

Differentially spliced events were identified using MISO to quantify reads supporting distinct isoforms and/or junction-spanning reads. These read counts were then subjected to Wagenmakers's Bayesian alternative to the binomial proportion test [52]. Differentially spliced events were defined as those that exhibited a difference in absolute isoform ratio of at least 10% with an associated Bayes factor of at least 5. All analyses were restricted to events with at least 20 relevant reads in the samples being compared.

Sequences logos were created using the seqLogo package from Bioconductor [53]. Cluster analysis was performed using Ward's method. The information of percent tumor nuclei for each LUAD sample that carries a *U2AF1 S34F* mutation was downloaded from TCGA data portal (http://cancergenome.nih.gov/). For samples with more than one percent tumor nuclei value, an average value was calculated and used for comparisons.

Quantification of the number of RNA-seq reads specific to S34F mutant or WT *U2AF1* (Figure 2B) was conducted by first extracting sequence reads mapped to the U2AF1 locus from the alignment bam files. A 21-mer oligonucleotide matching the S34F mutant cDNA, in which the mutant nucleotide was centered and flanked by 10 bp sequences on each side (GACAGGTGCTTTCGGTTGCAC), together with the corresponding 21-mer WT sequence (GACAGGTGCTCTCGGTTGCAC), was used to BLAST against the extracted sequence reads. Reads with exact match to the mutant 21-mer sequence were considered to be *U2AF1 S34F*-specific, whereas those matched to the WT 21-mer sequence were considered as WT *U2AF1*-specific.

The reported RNA-seq data from lung cell lines will be deposited into the NCBI GEO database upon manuscript acceptance.

**RNA affinities of purified U2AF1 complexes**

Sequences of synthetic 5′-labeled fluorescein RNAs (GE Healthcare Dharmacon) are given in the Supplementary Figure S13. Both the protein complex (see Supplemental Materials and Methods for purification details) and the RNA stocks were prepared and diluted in the same buffer used for size-exclusion chromatography (25 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 6.8, 150 mM NaCl, 3% glycerol, 20 μM $ZnCl_2$, 3 mM βME (β-mercaptoethanol) and 0.5 mM TCEP (tris(2-carboxyethyl)phosphine)). Fluorescence anisotropy changes were monitored over a concentration series of purified U2AF1 protein complexes (final concentrations shown in Supplementary Figure S13) mixed with fluorescein-labeled RNAs (20 nM final concentration) in 384-well flat bottom assay plates (Corning). Fluorescence polarization

changes were measured at 520 nm following excitation at 490 nm using Envision high-throughput plate reader (PerkinElmer). The data were fit by non-linear regression to obtain the apparent equilibrium dissociation constant ($K_D$) using the following equation, where $X$ is the total protein concentration, [RNA] is the total RNA concentration, $r$ is the observed anisotropy at the $i^{th}$ titration, $r_B$ is the anisotropy at zero protein concentration, and $r_F$ is the anisotropy at saturating protein concentration:

$$r = r_F + \frac{r_B - r_F}{2[RNA]}\,(K_D + X + [RNA]) - \sqrt{(K_D + X + [RNA])^2 - (4[RNA]X)}$$

The reciprocal of the $K_D$ value provided the apparent equilibrium affinity constant ($K_A$).

***Clonogenic assay***

Cells were seeded into six-well dishes at $1x10^5$ cells per dish and allowed to grow overnight before lentiviral infection. Two days after virus infection, 1000 live cells were seeded in a 100 mm dish in growth media supplemented with puromycin (1 µg/ml) for selecting infected cells and growing colonies. Growth media were changed once a week for three week and the colonies were stained with a staining solution (0.03% methylene blue, 20% methanol in water) for 5 min. Clonogenicity was defined as colony numbers formed as a percent of those in control cells.

***Statistics***

All experiments were independently performed at least three times unless otherwise stated. Statistical significance was determined by two-tailed Student's *t* test or otherwise stated. In all analyses, $p$ values ≤ 0.05 are considered statistically significant.

**Acknowledgements**

We thank Ms. Jackie Idol (NHGRI) and Ursula Harper (NHGRI) for technical assistance, Ms. Danielle Miller-O'Mard (NHGRI) and the NHGRI transgenic mouse core for mouse husbandry, Dr. Heidi Dvinge (FHCRC) for help with the LUAD data analysis, Dr. Matthew J. Walter (Wash.U.) for sharing the *CEP164* and *FMR1* splice site sequences prior to publication. We thank members of the Varmus lab, Drs. Janine Ilagan (FHCRC) and Paul Liu (NHGRI) for helpful discussions during the course of the study. HV was supported by the Intramural Program at the National Institutes of Health and is now supported by the Meyer Cancer Center at Weill Cornell Medicine. RKB is supported by the Edward P. Evans Foundation, Ellison Medical Foundation (AG-NS-1030-13), NIH/NHLBI (R01 HL128239), and NIH/NIDDK (R01 DK103854). CLK is supported by the Edward P. Evans Foundation and NIH/NIGMS (R01 GM070503). The results published here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

**References**

1. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478: 64–69. doi:10.1038/nature10496

2. 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. 2012;44: 53–57. doi:10.1038/ng.1031

3. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012;150: 1107–1120. doi:10.1016/j.cell.2012.08.029

4. Furney SJ, Pedersen M, Gentien D, Dumont AG, Rapinat A, Desjardins L, et al. SF3B1 Mutations Are Associated with Alternative Splicing in Uveal Melanoma. Cancer Discovery. 2013;3: 1122–1129. doi:10.1158/2159-8290.CD-13-0330

5. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. N Engl J Med. 2011;365: 2497–2506. doi:10.1056/NEJMoa1109016

6. Waterfall JJ, Arons E, Walker RL, Pineda M, Roth L, Killian JK, et al. High prevalence of MAP2K1 mutations in variant and IGHV4-34-expressing hairy-cell leukemias. Nature Genetics. 2014;46: 8–10. doi:10.1038/ng.2828

7. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490: 61–70. doi:10.1038/nature11412

8. Biankin AV, Waddell N, Kassahn KS, Gingras M-C, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 2012;491: 399–405. doi:10.1038/nature11547

9. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511: 543–550. doi:10.1038/nature13385

10. Yoshida K, Ogawa S. Splicing factor mutations and cancer. Wiley Interdiscip Rev RNA. 2014;5: 445–459. doi:10.1002/wrna.1222

11. 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: 1275–1282. doi:10.1038/leu.2013.58

12. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122: 3616–27– quiz 3699. doi:10.1182/blood-2013-08-518886

13. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SCW, Ramakrishnan A, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell. Elsevier Inc; 2015;27: 617–630. doi:10.1016/j.ccell.2015.04.006

14. Shirai CL, Ley JN, White BS, Kim S, Tibbitts J, Shao J, et al. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell. Elsevier Inc; 2015;27: 631–643. doi:10.1016/j.ccell.2015.04.008

15. Ruskin B, Zamore PD, Green MR. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell. 1988;52: 207–219.

16. Zamore PD, Green MR. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci U S A. 1989;86: 9243–9247.

17. Krämer A, Utans U. Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. The EMBO Journal. 1991;10:

1503–1509.

18. Berglund JA, Abovich N, Rosbash M. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes & Development. 1998;12: 858–867.

19. Wu S, Romfo CM, Nilsen TW, Green MR. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature. 1999;402: 832–835. doi:10.1038/45590

20. Guth S, Tange TØ, Kellenberger E, Valcárcel J. Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. Molecular and Cellular Biology. 2001;21: 7673–7681. doi:10.1128/MCB.21.22.7673-7681.2001

21. Guth S, Martínez C, Gaur RK, Valcárcel J. Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). Molecular and Cellular Biology. 1999;19: 8263–8271.

22. Merendino L, Guth S, Bilbao D, Martínez C, Valcárcel J. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature. 1999;402: 838–841. doi:10.1038/45602

23. Zorio DA, Blumenthal T. Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. Nature. 1999;402: 835–838. doi:10.1038/45597

24. Ilagan JO, Ramakrishnan A, Hayes B, Murphy ME, Zebari AS, Bradley P, et al. U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res. 2015;25: 14–26. doi:10.1101/gr.181016.114

25. Okeyo-Owuor T, White BS, Chatrikhi R, Mohan DR, Kim S, Griffith M, et al. U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. Leukemia. 2015;29: 909–917. doi:10.1038/leu.2014.303

26. 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. Ast G, editor. PLoS ONE. 2014;9: e87361. doi:10.1371/journal.pone.0087361.s008

27. 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. 2013;122: 999–1006. doi:10.1182/blood-2013-01-480970

28. Ramirez RD, Sheridan S, Girard L, Sato M, Kim Y, Pollack J, et al. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. Cancer Res. 2004;64: 9027–9034. doi:10.1158/0008-5472.CAN-04-3703

29. Yusa K. Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyBac transposon. Nat Protoc. 2013;8: 2061–2078. doi:10.1038/nprot.2013.126

30. Kralovicova J, Vorechovsky I. Allele-specific recognition of the 3' splice site of INS intron 1. Hum Genet. 2010;128: 383–400. doi:10.1007/s00439-010-0860-1

31. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43: D805–11. doi:10.1093/nar/gku1075

32. The Cancer Genome Atlas Research Network. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N Engl J Med. 2013. doi:10.1056/NEJMoa1301689

33. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. Genome Biol. 2004;5: R74. doi:10.1186/gb-2004-5-10-r74

34. Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, et al. Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat Struct Mol Biol. 2014;21: 997–1005. doi:10.1038/nsmb.2906

35. Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell. 2016;165: 372–381. doi:10.1016/j.cell.2016.02.045

36. Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR. Kinetic competition during the transcription cycle results in stochastic RNA processing. Elife. 2014;3. doi:10.7554/eLife.03939

37. Shcherbakova I, Hoskins AA, Friedman LJ, Serebrov V, Corrêa IR, Xu M-Q, et al. Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. Cell Rep. 2013;5: 151–165. doi:10.1016/j.celrep.2013.08.026

38. Kralovicova J, Knut M, Cross NCP, Vorechovsky I. 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. 2015;43: 3747–3763. doi:10.1093/nar/gkv194

39. Zhou Q, Derti A, Ruddy D, Rakiec D, Kao I, Lira M, et al. A Chemical Genetics Approach for the Functional Assessment of Novel Cancer Genes. Cancer Res. 2015;75: 1949–1958. doi:10.1158/0008-5472.CAN-14-2930

40. Fei DL, Li H, Kozul CD, Black KE, Singh S, Gosse JA, et al. Activation of Hedgehog signaling by the environmental toxicant arsenic may contribute to the etiology of arsenic-induced tumors. Cancer Res. 2010;70: 1981–1988. doi:10.1158/0008-5472.CAN-09-2898

41. Fei DL, Sanchez-Mejias A, Wang Z, Flaveny C, Long J, Singh S, et al. Hedgehog Signaling Regulates Bladder Cancer Growth and Tumorigenicity. Cancer Res. 2012;72: 4449–4458. doi:10.1158/0008-5472.CAN-11-4123

42. Campeau E, Ruhl VE, Rodier F, Smith CL, Rahmberg BL, Fuss JO, et al. A versatile viral system for expression and depletion of proteins in mammalian cells. PLoS ONE. 2009;4: e6529. doi:10.1371/journal.pone.0006529

43. Strack RL, Strongin DE, Bhattacharyya D, Tao W, Berman A, Broxmeyer HE, et al. A noncytotoxic DsRed variant for whole-cell labeling. Nat Methods. 2008;5: 955–957. doi:10.1038/nmeth.1264

44. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014;343: 84–87. doi:10.1126/science.1247005

45. Carrington B, Varshney GK, Burgess SM, Sood R. CRISPR-STAT: an easy and reliable PCR-based method to evaluate target-specific sgRNA activity. Nucleic Acids Res. 2015;43: e157. doi:10.1093/nar/gkv802

46. Brosseau JP, Lucier JF, Lapointe E, Durand M, Gendron D, Gervais-Bird J, et al. High-throughput quantification of splicing isoforms. RNA. 2010;16: 442–449. doi:10.1261/rna.1877010

47. Katz Y, Wang ET, (null), Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7: 1009–1015. doi:10.1038/nmeth.1528

48. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, et al. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. 2013;41: D64–9. doi:10.1093/nar/gks1048

49. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, et al. Ensembl 2013. Nucleic Acids Res. 2013;41: D48–55. doi:10.1093/nar/gks1236

50. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12: 323. doi:10.1186/1471-2105-12-323

51. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25: 1105–1111. doi:10.1093/bioinformatics/btp120

52. Wagenmakers E-J, Lodewyckx T, Kuriyal H, Grasman R. Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. Cogn Psychol. 2010;60: 158–189. doi:10.1016/j.cogpsych.2009.12.001

53. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al.

Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5: R80. doi:10.1186/gb-2004-5-10-r80

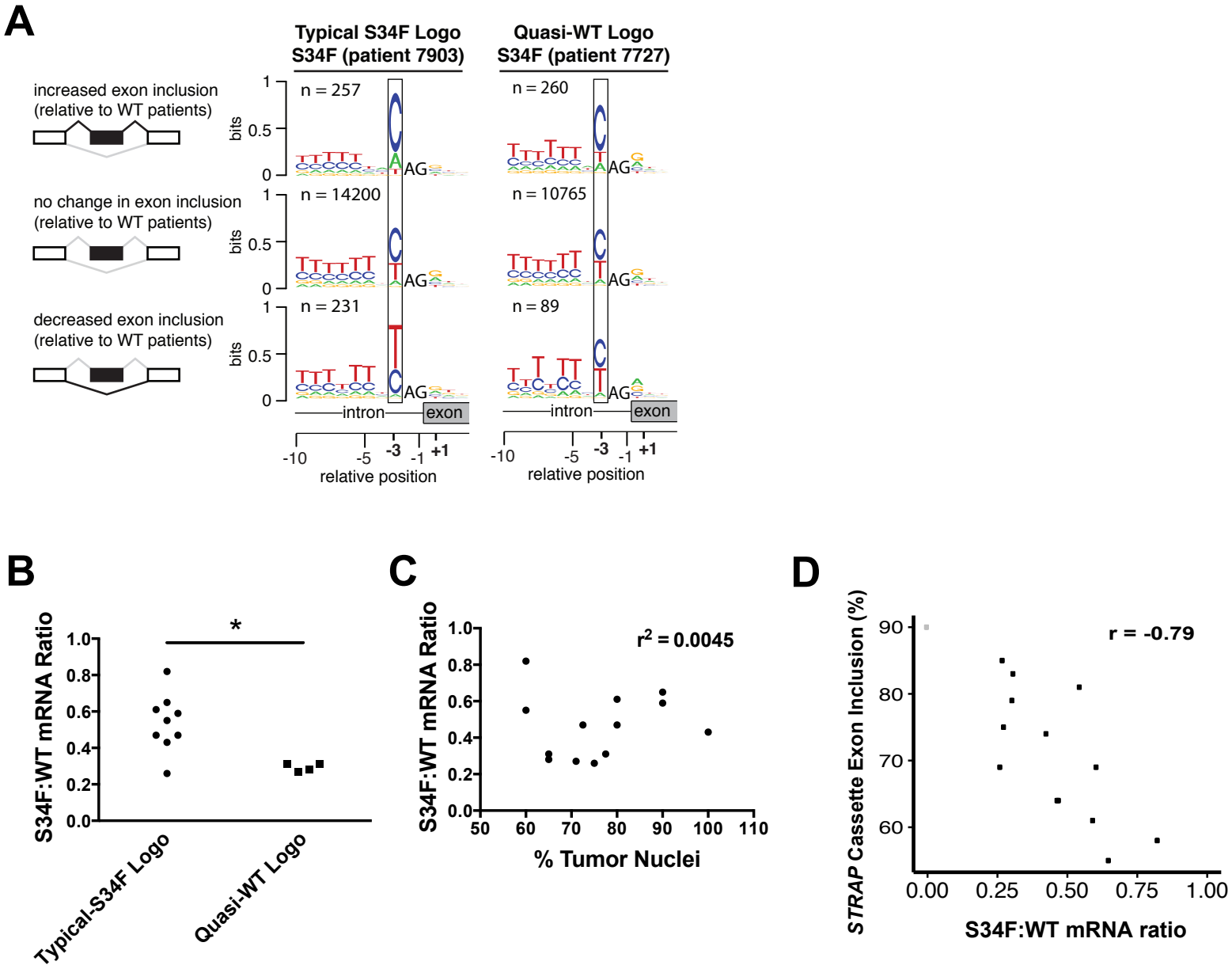
A
Typical S34F Logo
S34F (patient 7903)
Quasi-WT Logo
S34F (patient 7727)
increased exon inclusion
(relative to WT patients)
no change in exon inclusion
(relative to WT patients)
decreased exon inclusion
(relative to WT patients)
n = 257
n = 260
n = 14200
n = 10765
n = 231
n = 89
bits
intron
exon
relative position
B
S34F:WT mRNA Ratio
Typical-S34F Logo
Quasi-WT Logo
C
r² = 0.0045
% Tumor Nuclei
D
r = -0.79
STRAP Cassette Exon Inclusion (%)
S34F:WT mRNA ratio

Figure 1

**Figure 1. S34F-associated splicing correlates with S34F:WT mRNA ratios in LUAD.**
**(A)**. Representative sequence logos depicting consensus 3´ splice site sequences preceding cassette exons for two representative LUADs with the *U2AF1 S34F* mutation computed from published TCGA transcriptomes. In both cases, changes in the use of cassette exons were determined by comparisons with an average transcriptome from LUADs without *U2AF1* mutations. Boxes highlight nucleotides found preferentially at the -3 position. The nucleotide frequencies preceding exons that are more often included or more often skipped in tumors with mutant U2AF1 differ from the genomic consensus in the tumor from patient #7903, with a high S34F:WT mRNA ratio, but not in the tumor from patient #7727, with a low ratio. (Sequence logos from transcriptomes of every S34F-mutant LUAD are presented in Supplemental Figure S1.) The scheme of alternative splicing with a cassette exon (black box) is shown on the left side of the sequence logo. Black lines represent S34F-promoted splices. The analysis was restricted to introns with canonical GT-AG U2-type splice sites. The vertical axis represents the information content in bits. n is the number of cassette exon sequences used to construct the logo. **(B)**. U2AF1 mutant LUAD transcriptomes harboring a S34F stereotypical sequence logo have relatively high S34F:WT mRNA ratios. U2AF1 mutant LUAD samples were sub-grouped based on the type of sequence logos. The asterisk represents a statistically significant change by student's t test. **(C)**. S34F:WT mRNA ratios do not correlate with tumor purity in LUAD tumors with the S34F mutation. Tumor purity is represented by the percent of tumor nuclei in each LUAD sample (derived from TCGA clinical data) and plotted against the S34F:WT mRNA ratio. The $r^2$ value of linear regression is shown. **(D)**. Inclusion of the *STRAP* cassette exon correlates with the S34F:WT mRNA ratio. The twelve mutant LUAD transcriptomes in the TCGA database with a sufficient number of informative reads to calculate the inclusion frequency for the *STRAP* cassette exon are represented as black dots. The median inclusion level of the same cassette exon for all transcriptomes from tumors without the mutation is shown as a grey dot near the ordinate axis. r = Pearson's association coefficient.

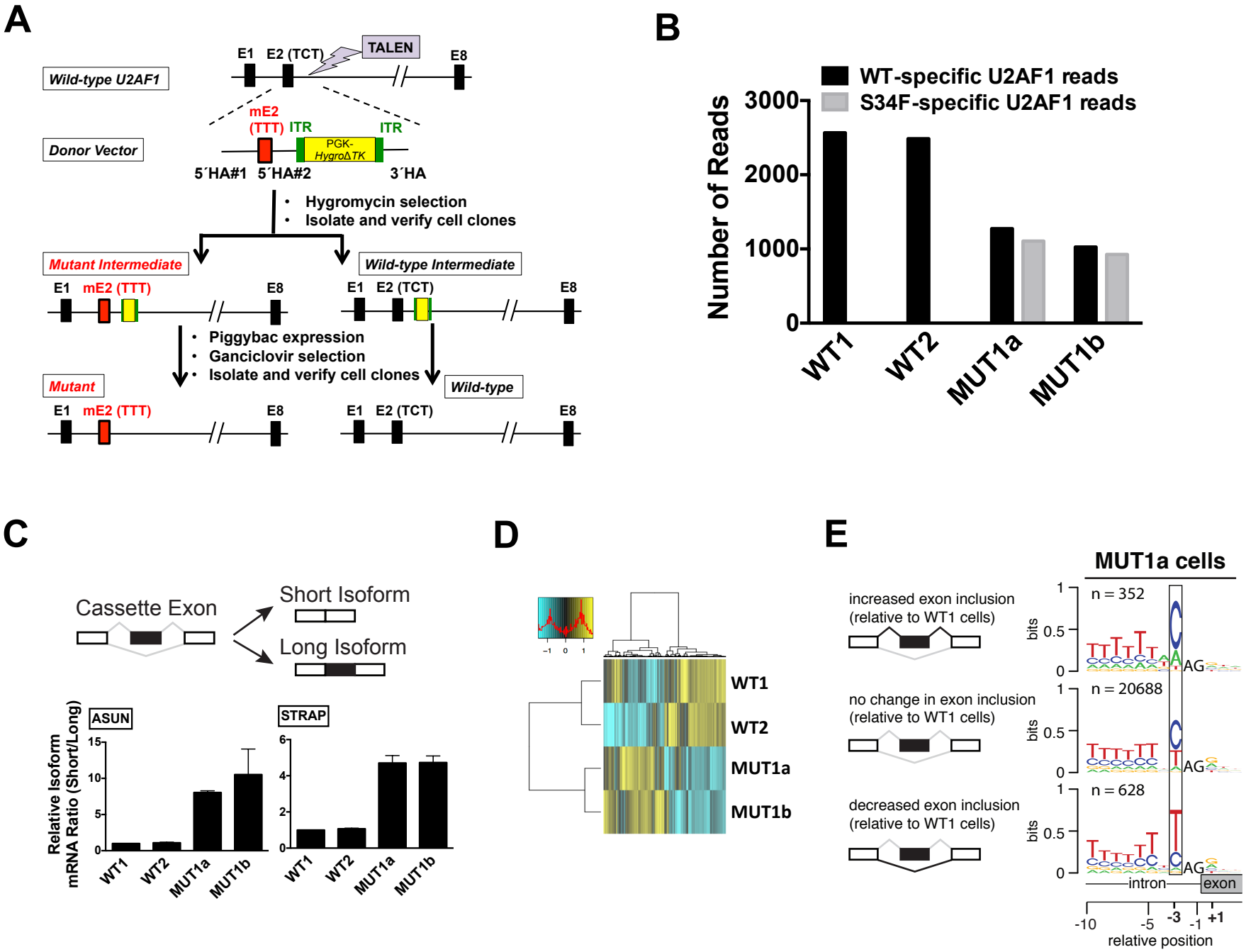
A
Wild-type U2AF1
E1
E2 (TCT)
TALEN
E8
Donor Vector
mE2 (TTT)
ITR
ITR
PGK-HygroΔTK
5´HA#1
5´HA#2
3´HA
Hygromycin selection
Isolate and verify cell clones
Mutant Intermediate
Wild-type Intermediate
Piggybac expression
Ganciclovir selection
Isolate and verify cell clones
Mutant
Wild-type
B
WT-specific U2AF1 reads
S34F-specific U2AF1 reads
Number of Reads
3000
2000
1000
0
WT1
WT2
MUT1a
MUT1b
C
Cassette Exon
Short Isoform
Long Isoform
ASUN
STRAP
Relative Isoform mRNA Ratio (Short/Long)
D
WT1
WT2
MUT1a
MUT1b
E
MUT1a cells
increased exon inclusion (relative to WT1 cells)
no change in exon inclusion (relative to WT1 cells)
decreased exon inclusion (relative to WT1 cells)
n = 352
n = 20688
n = 628
bits
AG
intron
exon
-10
-5
-3
-1
+1
relative position

Figure 2

**Figure 2. Creation of isogenic lung cell lines that recapitulate features of S34F-associated splicing.**
**(A)**. Strategy to create a TCT to TTT point mutation (S34F) at the endogenous *U2AF1* locus in HBEC3kt cells. TALEN, transcription activator-like effector nuclease; E, exon; mE, mutant (S34F) exon; ITR, inverted terminal repeat; HA, homology arm. See Results and Materials and Methods for details. **(B)**. MUT1a and MUT1b cells contain similar levels of mutant and WT *U2AF1* mRNA. The number of reads specific for mutant and WT *U2AF1* was obtained from RNA-seq, using poly A-selected RNA from the four cell lines. **(C).** The S34F-associated cassette exons in *ASUN* and *STRAP* mRNAs showed increased skipping in MUT cell lines. (Top) Scheme of alternative splicing with a cassette exon (black box) to generate short and long isoforms in which the cassette exon is skipped or included. (Bottom) Alternative splicing of cassette exons in *ASUN* and *STRAP* mRNAs, measured by RT-qPCR using isoform-specific primers. The short/long isoform ratio in WT1 cells was arbitrarily set to 1 for comparison. **(D)**. Heat map depicting the inclusion levels of all cassette exons. Dendrograms were constructed from an unsupervised cluster analysis based on all cassette exons that showed at least a 10% change in use between WT and MUT1 lines. **(E).** Sequence logos from 3' splice sites preceding cassette exons with altered use in MUT1a cells display typical S34F-associated features. Logos were constructed as in Figure 1A based on the transcriptome of MUT1a cells in comparison with that of WT1 cells. Other comparisons of transcriptomes from a MUT cell line and a WT cell line yielded similar sequence logos (data not shown).

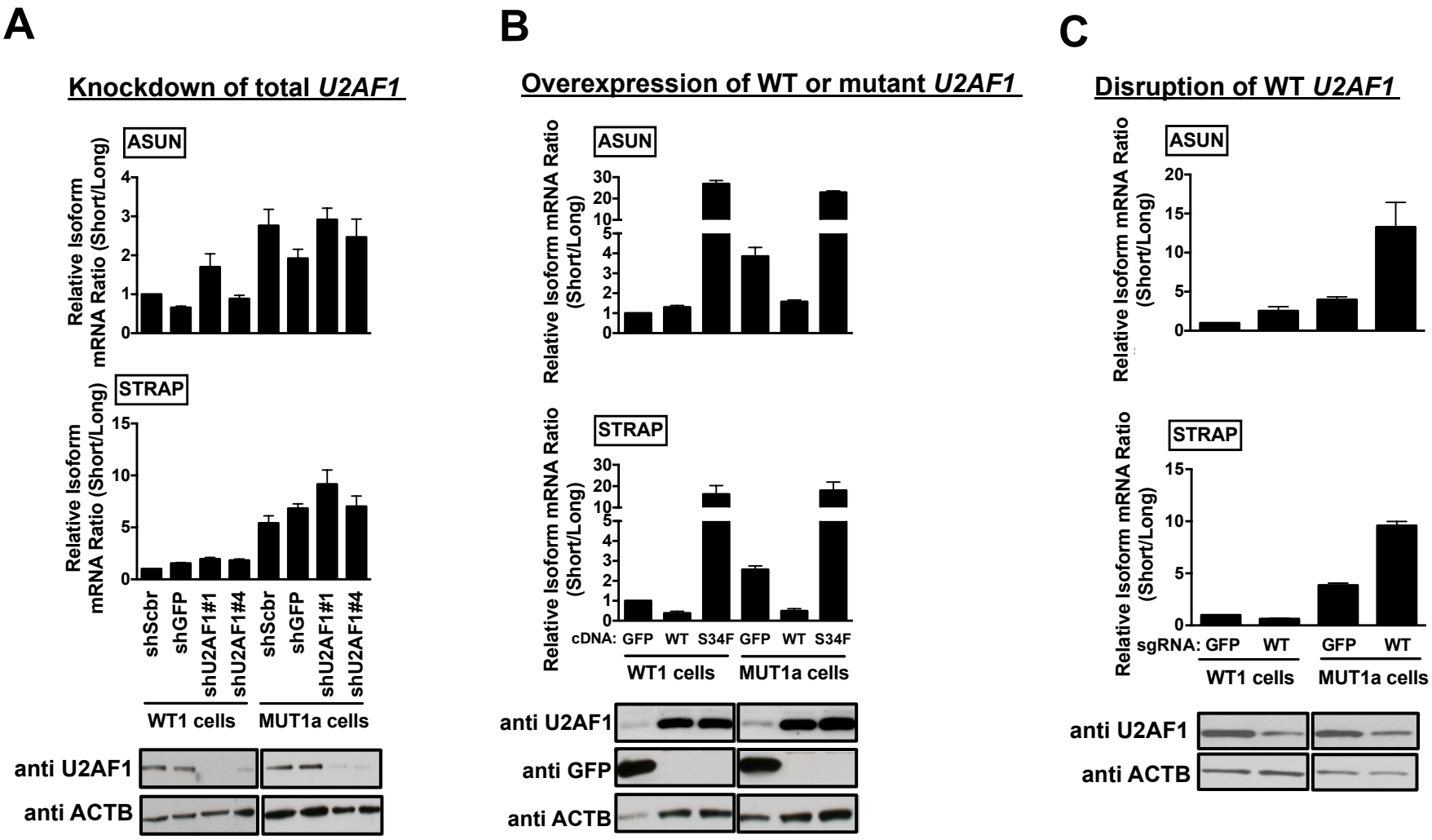
A
Knockdown of total U2AF1
ASUN
STRAP
Relative Isoform mRNA Ratio (Short/Long)
shScbr
shGFP
shU2AF1#1
shU2AF1#4
WT1 cells
MUT1a cells
anti U2AF1
anti ACTB
B
Overexpression of WT or mutant U2AF1
Relative Isoform mRNA Ratio (Short/Long)
cDNA: GFP WT S34F GFP WT S34F
anti GFP
C
Disruption of WT U2AF1
sgRNA: GFP WT GFP WT

Figure 3

**Figure 3. The ratio of S34F:WT U2AF1 gene products controls S34F-associated splicing in isogenic HBEC3kt cell lines.** **(A)**. Reduction of both mutant and WT U2AF1 RNA and protein does not affect S34F-associated splicing. WT1 and MUT1a cells were transduced with shRNAs against *U2AF1* (shU2AF1#1 and #4) or two control shRNAs. Total RNA and protein were harvested 4 days later. The frequencies of incorporation of cassette exon sequences in *ASUN* and *STRAP* mRNAs (top and middle panels) were determined by the relative short/long isoform ratios by RT-qPCR, as represented in Figure 2C. Immunoblots for U2AF1 and ACTB in total cell lysates are shown in the bottom panel. shScbr, scrambled shRNA; shGFP, shRNA against GFP. **(B)**. Overexpression of WT or mutant U2AF1 to change S34F:WT ratios alters S34F-sensitive splicing. WT1 and MUT1a cells were transduced with expression vectors encoding *GFP*, WT or mutant (S34F) *U2AF1* for 3 days before harvesting cells to quantify the level of splicing changes and proteins as in panel A. **(C)**. Disruption of WT *U2AF1* by gene editing to increase S34F:WT ratios enhances S34F-sensitive splicing. WT1 and MUT1a cells were transduced with either sgRNA-GFP/Cas9 or sgRNA-WT/Cas9. Total RNA and protein were harvested 6 days later for assays as in panel A.

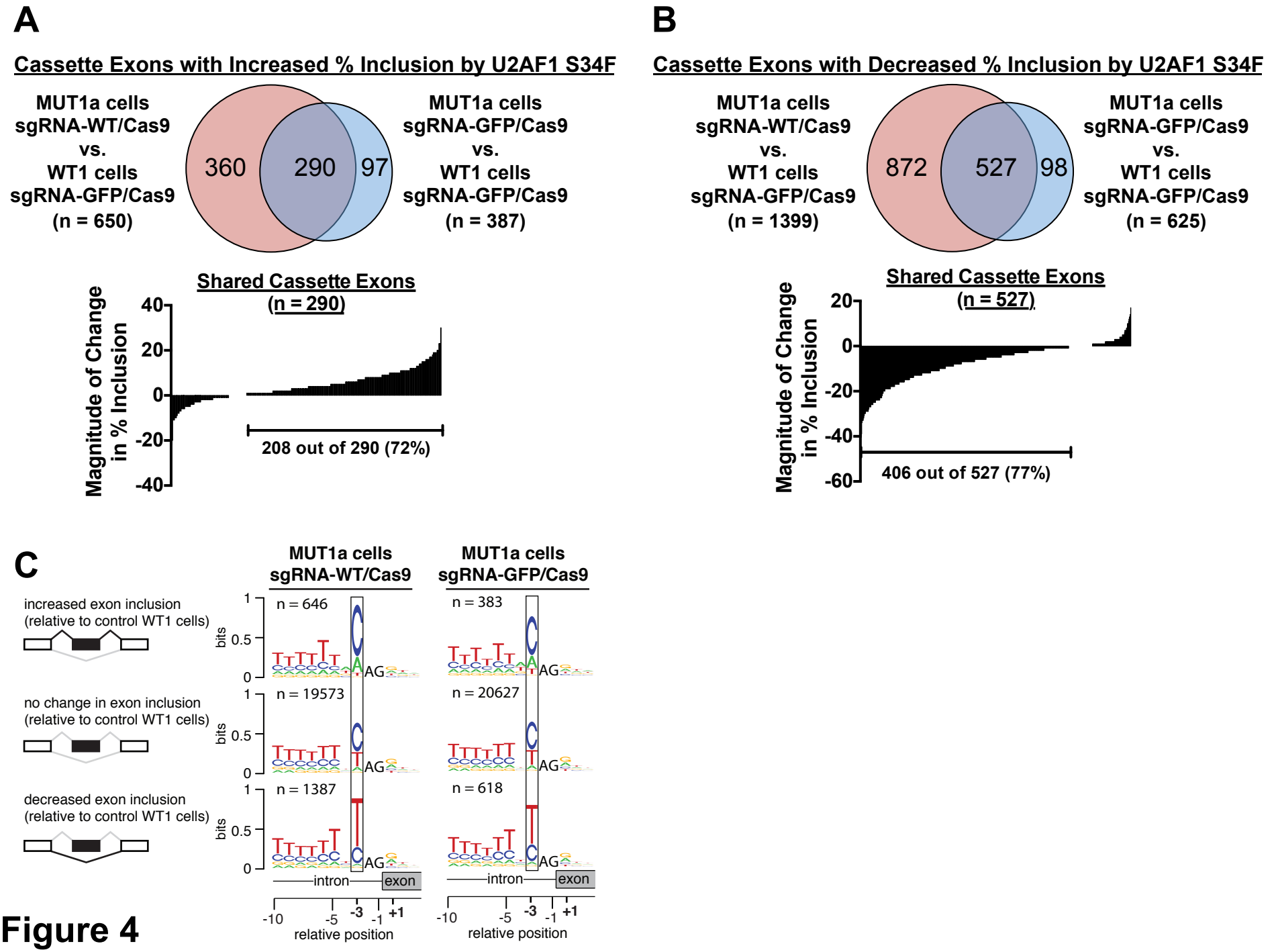
A
Cassette Exons with Increased % Inclusion by U2AF1 S34F
MUT1a cells
sgRNA-WT/Cas9
vs.
WT1 cells
sgRNA-GFP/Cas9
(n = 650)
360
290
97
MUT1a cells
sgRNA-GFP/Cas9
vs.
WT1 cells
sgRNA-GFP/Cas9
(n = 387)
Shared Cassette Exons
(n = 290)
Magnitude of Change
in % Inclusion
40
20
0
-20
-40
208 out of 290 (72%)
B
Cassette Exons with Decreased % Inclusion by U2AF1 S34F
MUT1a cells
sgRNA-WT/Cas9
vs.
WT1 cells
sgRNA-GFP/Cas9
(n = 1399)
872
527
98
MUT1a cells
sgRNA-GFP/Cas9
vs.
WT1 cells
sgRNA-GFP/Cas9
(n = 625)
Shared Cassette Exons
(n = 527)
Magnitude of Change
in % Inclusion
20
0
-20
-40
-60
406 out of 527 (77%)
C
increased exon inclusion
(relative to control WT1 cells)
no change in exon inclusion
(relative to control WT1 cells)
decreased exon inclusion
(relative to control WT1 cells)
MUT1a cells
sgRNA-WT/Cas9
MUT1a cells
sgRNA-GFP/Cas9
n = 646
n = 383
n = 19573
n = 20627
n = 1387
n = 618
bits
1
0.5
0
AG
intron
exon
-10
-5
-3
-1
+1
relative position

Figure 4

**Figure 4. Increasing the ratio of S34F:WT gene products by disrupting the WT *U2AF1* locus enhances S34F-associated splicing of cassette exons.**

**(A)**. Upper panel: The Venn diagrams show overlap of 290 cassette exons that display at least a ten percent increase in inclusion levels in MUT1a cells relative to levels in WT1 cells, with or without CRISPR/Cas9-mediated disruption of WT *U2AF1*. Lower panel: Increasing the S34F:WT ratio by disrupting WT *U2AF1* enhances the magnitude of S34F-associated splicing. The "waterfall" plot depicts changes in percent inclusion levels for all shared cassette exons identified from the Venn diagram when the WT *U2AF1* locus was disrupted. Each vertical bar represents one shared cassette exon. **(B)**. The analysis shown in Panel A was repeated for cassette exons showing ten percent or more decreased inclusion in MUT1a cells. **(C)**. Enhanced features of S34F-associated logos at 3' splice acceptor sites after disruption of WT *U2AF1*. Sequence logos were constructed as in Figure 1A, based on the indicated comparisons.

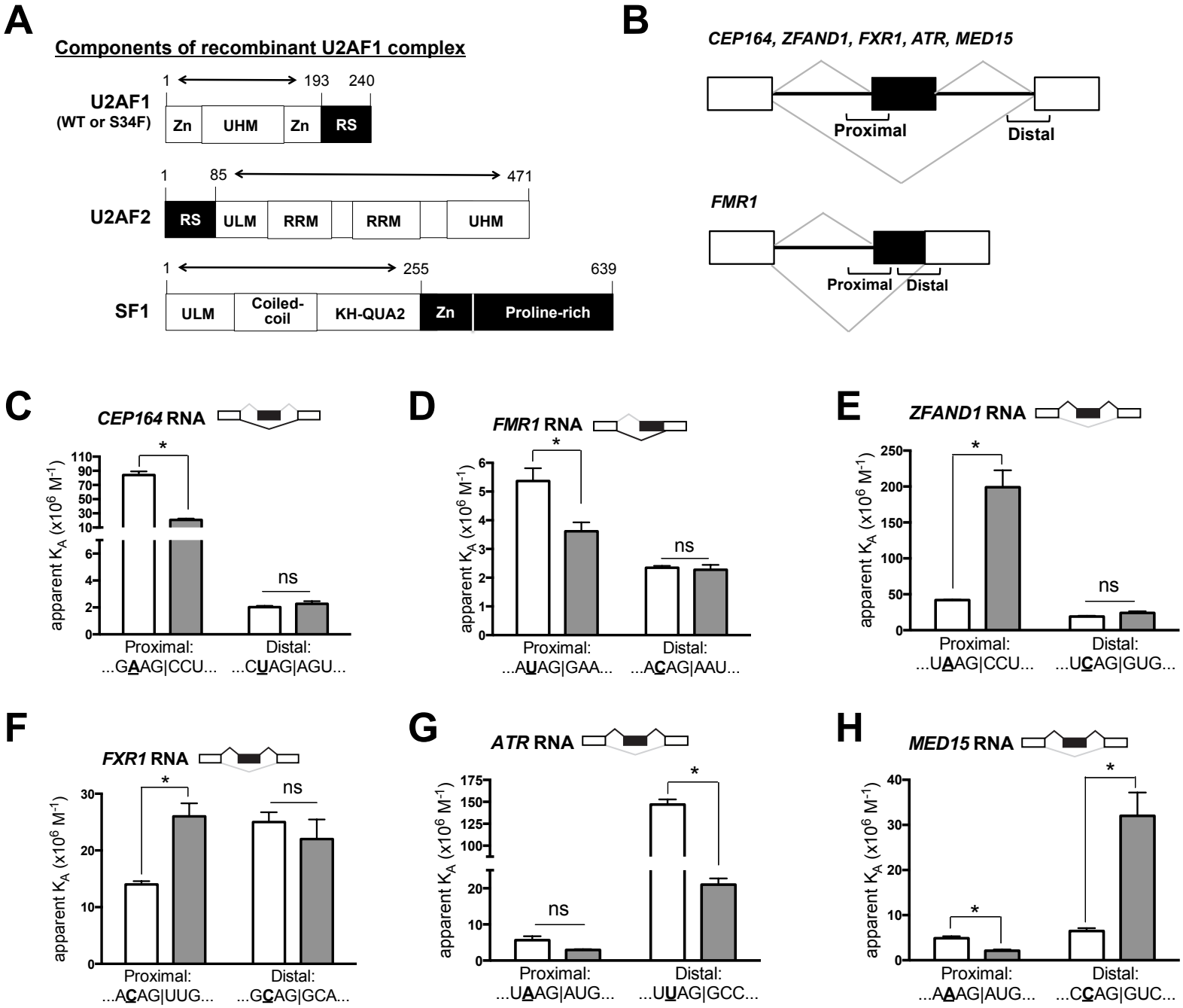
A
Components of recombinant U2AF1 complex
U2AF1
(WT or S34F)
1
193
240
Zn
UHM
Zn
RS
U2AF2
1
85
471
RS
ULM
RRM
RRM
UHM
SF1
1
255
639
ULM
Coiled-coil
KH-QUA2
Zn
Proline-rich
B
CEP164, ZFAND1, FXR1, ATR, MED15
Proximal
Distal
FMR1
Proximal
Distal
C
CEP164 RNA
apparent $K_A$ ($x10^6$ $M^{-1}$)
Proximal:
...GAAG|CCU...
Distal:
...CUAG|AGU...
ns
D
FMR1 RNA
apparent $K_A$ ($x10^6$ $M^{-1}$)
Proximal:
...AUAG|GAA...
Distal:
...ACAG|AAU...
ns
E
ZFAND1 RNA
apparent $K_A$ ($x10^6$ $M^{-1}$)
Proximal:
...UAAG|CCU...
Distal:
...UCAG|GUG...
ns
F
FXR1 RNA
apparent $K_A$ ($x10^6$ $M^{-1}$)
Proximal:
...ACAG|UUG...
Distal:
...GCAG|GCA...
ns
G
ATR RNA
apparent $K_A$ ($x10^6$ $M^{-1}$)
Proximal:
...UAAG|AUG...
Distal:
...UUAG|GCC...
ns
H
MED15 RNA
apparent $K_A$ ($x10^6$ $M^{-1}$)
Proximal:
...AAAG|AUG...
Distal:
...CCAG|GUC...

Figure 5

**Figure 5. Differential binding of mutant and WT U2AF1 complexes to RNA oligonucleotides explain most S34F-associated alterations in RNA splicing.**

**(A)**. Cartoons illustrate components of recombinant U2AF1 complexes used in the binding assay. Full-length proteins are shown but only partial sequences (denoted by bi-directional arrows) were used to make recombinant protein complexes. KH-QUA2, K-Homology Quaking motif; RRM, RNA recognition motif domain; RS, arginine/serine-rich domain; UHM, U2AF homology motif; ULM, U2AF ligand motif; ZnF, zinc finger domain. **(B)**. Scheme of alternative splicing patterns for cassette exons (top diagram) and 5´ extended exons from competing 3´ splice site selection (bottom). The brackets indicate the positions of RNA oligonucleotides used for the binding assays. Exons are shown as boxes: white boxes indicate invariant exonic sequences and black boxes denote sequences that are incorporated into mRNA (exonic) only when the proximal 3' splice sites are used. Introns are shown as solid lines. The grey lines represent possible splices. The names of characterized genes that conform to the patterns shown in the upper and lower cartoons are indicated. **(C - H)**. U2AF1 S34F and WT U2AF1 have different affinities ($K_A$'s), for relevant 3' splice site oligonucleotides. To accomplish the binding assays, the WT or mutant U2AF1 protein complexes were titrated into 5´ fluorescein-tagged RNA oligonucleotides over a range of concentrations as described in the Materials and Methods. RNA sequences from -4 to +3 relative to the 3´ splice sites (vertical lines) in proximal and distal positions are shown. The nucleotide at the -3 position is bolded and underlined. Empty bars, $K_A$ for WT U2AF1 complex; grey bar, $K_A$ for mutant U2AF1 complex. For ease of comparison between the affinity binding results with S34F-associated splicing, S34F-promoted splices, as determined from RNA-seq data, are shown on top of each bar graph in black lines. The fitted binding curves, full oligonucleotide sequences, and apparent equilibrium dissociation constants are shown in Supplemental Figure S13. The relative changes in affinity and use of proximal *versus* distal splice sites are summarized in Supplemental Table S3.

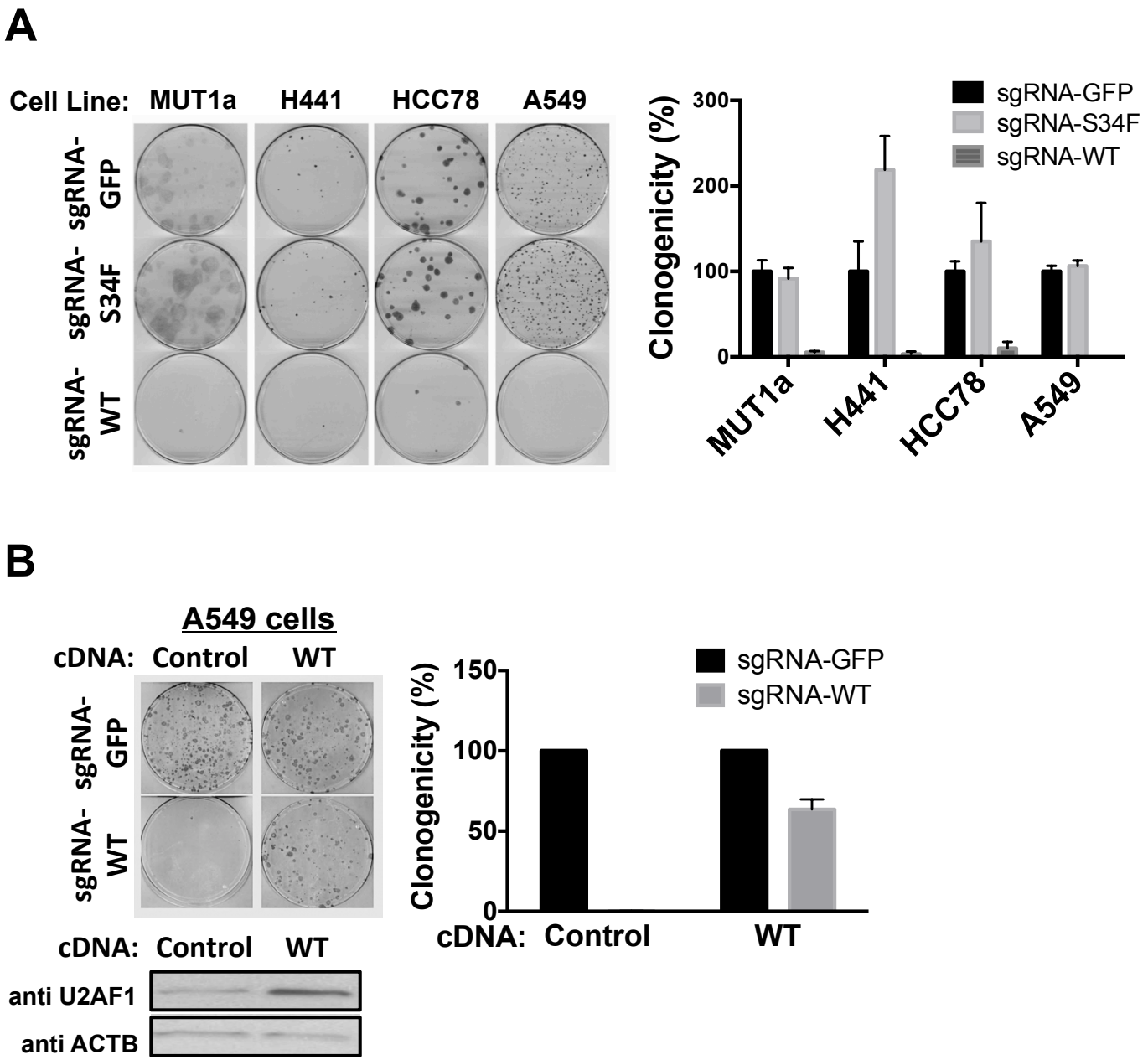
A
Cell Line: MUT1a H441 HCC78 A549
sgRNA-GFP
sgRNA-S34F
sgRNA-WT
Clonogenicity (%)
300
200
100
0
sgRNA-GFP
sgRNA-S34F
sgRNA-WT
MUT1a
H441
HCC78
A549
B
A549 cells
cDNA: Control WT
sgRNA-GFP
sgRNA-WT
cDNA: Control WT
anti U2AF1
anti ACTB
Clonogenicity (%)
150
100
50
0
sgRNA-GFP
sgRNA-WT
cDNA: Control WT

Figure 6

**Figure 6. WT but not mutant *U2AF1* is required for the clonogenic growth of the isogenic HBEC3kt cells and LUAD cell lines.**

**(A)**. Clonogenic growth assays after selective disruption of the WT or mutant and *U2AF1* allele. Left panel: The indicated cell lines were transduced with lentiviruses expressing sgRNA-GFP/Cas9, sgRNA-S34F/Cas9 or sgRNA-WT/Cas9, followed by plating 1000 cells in 100 mm tissue culture dishes. Cell colonies were stained with methylene blue and counted three weeks later. Right panel: Quantification of the clonogenic assay. The results are shown as percent clonogenicity by setting the number of control cell colonies (cells treated with Cas9/sgRNA-GFP) as 100% (n = 3). **(B)**. Rescue of growth inhibition by sgRNA-WT/Cas9 by overexpressing a form of WT *U2AF1* cDNA that is not predicted to be the target for sgRNA-WT/Cas9 (See Supplemental Figure S8 and Materials and Methods). Left panel: A549 cells were transduced with a control cDNA (DsRed-Express 2) or the *U2AF1* cDNA to establish polyclonal cell culture. The protein levels of U2AF1 and ACTB in these cells were examined by immunoblot (left bottom panel). These cells were subsequently transduced with either sgRNA-GFP/ Cas9 or sgRNA-WT/Cas9 followed by clonogenic assays as in panel A (left upper panel). Right panel: Quantification as in Panel **A** (n = 3).