U2AF1 mutations rescue deleterious exon skipping induced by KRAS mutations

David M. Walter[1,2,3], Katherine Cho[1,2], Smruthy Sivakumar[4], Iris T.-H. Lee[1,2], Anders B. Dohlman[1,2,3], Ethan Shurberg[1,2], Kevin X. Jiang[1,2], Akansha A. Gupta[1,2], Garrett M. Frampton[4], Matthew Meyerson[1,2,3]

1. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA
2. Cancer Program, Broad Institute of Harvard and MIT, Cambridge, MA
3. Department of Genetics, Harvard Medical School, Boston, MA
4. Foundation Medicine Inc., Boston, MA

Abstract

The mechanisms by which somatic mutations of splicing factors, such as U2AF1^{S34F} in lung adenocarcinoma, contribute to cancer pathogenesis are not well understood. Here, we used prime editing to modify the endogenous *U2AF1* gene in lung adenocarcinoma cells and assessed the resulting impact on alternative splicing. These analyses identified *KRAS* as a key target modulated by U2AF1^{S34F}. One specific *KRAS* mutation, G12S, generates a cryptic U2AF1 binding site that leads to skipping of *KRAS* exon 2 and generation of a non-functional *KRAS* transcript. Expression of the U2AF1^{S34F} mutant reverts this exon skipping and restores KRAS function. Analysis of cancer genomes reveals that U2AF1^{S34F} mutations are enriched in KRAS^{G12S}-mutant lung adenocarcinomas. A comprehensive analysis of splicing factor/oncogene mutation co-occurrence in cancer genomes also revealed significant co-enrichment of KRAS^{Q61R} and U2AF1^{I24T} mutations. Experimentally, KRAS^{Q61R} mutation leads to *KRAS* exon 3 skipping, which in turn can be rescued by the expression of U2AF1^{I24T}. Analysis of genomic and clinical patient data suggests that both *U2AF1* mutations occur secondary to *KRAS* mutation and are associated with decreased overall patient survival. Our findings provide evidence that splicing factor mutations can rescue splicing defects caused by oncogenic mutations. More broadly, they demonstrate a dynamic process of cascading selection where mutational events are positively selected in cancer genomes as a consequence of earlier mutations.

Main

Analysis of recurrent somatic mutations in cancers through next-generation sequencing has identified novel classes of cancer mutated genes, including genes encoding components of the splicing machinery[1]. Mutations in *U2AF1*, encoding the U2AF1 splicing factor that binds to an AG dinucleotide at the 3' splice site of RNA and recruits the U2 snRNP to these sites[2-4], are found in acute myeloid leukaemia (AML) and myelodysplastic syndrome (MDS)[1], lung adenocarcinoma (LUAD)[5] and pancreatic ductal adenocarcinoma (PDAC)[6]. Mutations in U2AF1 alter its affinity for AG dinucleotides depending on their nucleotide context, thereby disrupting normal splicing[7-10].

U2AF1^{S34F} is the most common amino acid substitution mutation in LUAD after those in *KRAS* and *EGFR* (Extended Data Fig. 1a), suggesting a powerful selective force for this substitution[11]. The frequency of U2AF1^{S34F} mutations does not correlate with tobacco smoking frequency across cancer types (Extended Data Fig. 1b), and does not appear to correlate with APOBEC activity or with mutation rate based on sequence context[11-14].

The mutation spectrum of *U2AF1* is distinct between AML/MDS, LUAD and PDAC. In AML/MDS, *U2AF1* mutations occur at one of two major hotspots, S34 or Q157 (Extended Data

Fig. 1c)[11]. In contrast, there is a single dominant hotspot of *U2AF1* mutations in LUAD at S34, with 99% of these mutations being S34F (Extended Data Fig. 1c)[11,15]. Furthermore, in PDAC, while S34 is the primary hotspot, there is a secondary mutational hotspot at I24[11]. These mutational patterns suggest that there are distinct positive selective pressures for U2AF1 mutations across different cancer types.

Despite the frequency of *U2AF1* mutations in LUAD and PDAC, the functional impact of these mutations on cancer cell behavior is poorly understood. The majority of work on *U2AF1* mutant function has been performed in the context of AML and MDS, hematological cancers with very different mutational and transcriptomic contexts[7,11,16-18]. Investigations on the functions of U2AF1 mutations in carcinomas have led to multiple proposed mechanisms, including the regulation of epithelial-mesenchymal transition (EMT), mRNA translation, MAPK signaling and stress granule formation[7,15,18,19]. However, the specific selective advantage conferred by U2AF1^{S34F} remains poorly understood, while the impact of U2AF1^{I24T} mutations is completely unknown.

To better delineate the function of *U2AF1* mutants, we combined prime editing of the endogenous locus with genomic analysis of cancer-derived mutations to uncover the selective pressure behind these mutations[20].

Generation of *U2AF1*-mutant lines by prime editing leads to differential alternative splicing

To identify the unique characteristics of U2AF1^{S34F} mutations that result in their selection in lung cancer, we applied twin prime editing (TwinPE)[20,21]. This allowed us to model a range of mutations at the S34 codon of endogenous *U2AF1* in LUAD cell lines and compare their impacts on alternative splicing to identify the unique characteristics of U2AF1^{S34F}. Using TwinPE we generated 6 U2AF1 mutations in A549 cells: a synonymous S34S mutation, S34C, S34A, S34Y, S34F and S34F(TTC) using an alternative codon for phenylalanine (Extended Data Table 1).

We performed RNA-sequencing on parental A549 cells as well as those modified at S34 (Extended Data Table 2). We then analysed the alternative splicing events associated with each amino acid substitution, and found that these were highest in cells harbouring U2AF1^{S34F} mutations (Extended Data Fig. 2a). The expression of the prime edited alleles for each engineered *U2AF1* variant was similar except for a slight reduction in expression of U2AF1$^{S34F(TTC)}$ (Extended Data Fig. 2b), associated with a slightly decreased splicing impact compared to U2AF1^{S34F} (Extended Data Fig. 2a). Alternative splicing events in U2AF1^{S34F}-mutant A549 cells showed significant overlap with those from U2AF1^{S34F}-mutant human LUAD samples (Extended Data Fig. 2c). In addition, the number of splicing events associated with each mutation positively correlated with the frequency of that mutation across human cancers (Extended Data Fig. 2d)[11,22,23]. U2AF1^{S34F} and U2AF1$^{S34F(TTC)}$ mutations resulted primarily in skipped exons and alternative 3' splice site usage (Extended Data Fig. 2e), and led to usage of CAG as opposed to TAG trinucleotides at the 3' splice site (Extended Data Fig. 3), as reported previously for this mutation[8,18].

KRAS^{G12S}-mutant cancers undergo *KRAS* exon 2 skipping which can be reversed by U2AF1^{S34F} mutations

To identify splicing events responsible for the positive selection of U2AF1^{S34F} mutations in lung cancer, we examined U2AF1^{S34F} and U2AF1$^{S34F(TTC)}$-specific events that occurred in known oncogenes and tumour suppressors[24,25]. We identified splicing alterations in multiple genes of interest including *AXL, KRAS* and *NCOR2* (Fig. 1a). Combining mutation data from

GENIE[11,26] with alternative splicing analysis, *KRAS* was notable due to its high mutation frequency in U2AF1^{S34F}-mutant LUAD (Extended Data Fig. 4a).

KRAS is known to undergo alternative splicing to produce two dominant protein isoforms: KRAS4A and KRAS4B. The predicted protein isoforms differ in C-terminal regions that control membrane localization, with KRAS4A containing a palmitoylation domain and KRAS4B functioning via a poly-lysine sequence[27,28], though both isoforms are capable of transforming cells[29]. In A549 cells with varying U2AF1 mutations, we found that *KRAS4A* was present in ~5-12% of transcripts, in line with previous studies, but did not differ according to U2AF1 mutation status (Extended Data Fig. 4b).

Instead, we observed that exon 2 of *KRAS* was skipped in a median of 18% of reads (range 12-32%) in parental or U2AF1^{S34S}-mutant A549 cells, while skipping was reduced to a median of 3% of reads (range 0-10%) in U2AF1^{S34F} and U2AF1$^{S34F(TTC)}$-mutant cells (Fig. 1b,c, Extended Data Fig. 4c,d). Interestingly, exon 2 of *KRAS* harbours the protein's translation start site as exon 1 is untranslated (Extended Data Fig. 4e). Exon 2 skipping is predicted to result in usage of a downstream alternative start codon in exon 3, omitting the first 66 amino acids of KRAS which include the P-loop, Switch-I and part of Switch-II (Extended Data Fig. 4f). While the effects of *KRAS* exon 2 skipping have not been studied, a similar event has been identified in *HRAS* whereby exon 2 skipping is thought to lead to decreased signalling in the context of Costello syndrome[30].

To further probe the relationship between *KRAS* exon 2 skipping and U2AF1^{S34F}, we "back-edited" the S34F mutant allele to wildtype *U2AF1* and found that the frequency of the skipping event was restored to baseline levels (~18% of reads) (Fig. 1d). To confirm this splicing event in another cell line, we edited the *U2AF1* locus in NCI-H441 cells. However, in contrast to A549 cells, NCI-H441 cells showed no exon 2 skipping regardless of *U2AF1* mutational status (Fig. 1e). As distinct *HRAS* mutations lead to *HRAS* exon 2 skipping or inclusion[30], we examined the *KRAS* locus of the two cell lines, and noticed that while A549 cells harbour a KRAS^{G12S} mutation, NCI-H441 cells contain a KRAS^{G12V} mutation.

To better understand the link between *KRAS* mutation status and alternative splicing, we examined alternative splicing data from 993 cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE)[31]. We found that only KRAS^{G12S}-mutant cell lines such as A549 cells had appreciable levels of *KRAS* exon 2 skipping (Fig. 1f, Extended Data Fig. 4g). The same held true in 6008 cancers with reported *KRAS* splicing data from The Cancer Genome Atlas (TCGA)[32], where *KRAS* exon 2 skipping was minimal except in cancers with KRAS^{G12S} mutations (Fig. 1g, Extended Data Fig. 4h). Furthermore, we found a negative correlation between variant allele frequency (VAF) of KRAS^{G12S} and exon 2 inclusion, suggesting that tumours with a higher fraction of KRAS^{G12S}-mutant cells have a higher degree of exon 2 skipping (Extended Data Fig. 4i). The TCGA dataset also contained a KRAS^{G12S}-mutant case which had 100% *KRAS* exon 2 inclusion (Fig. 1g, Extended Data Fig. 4i, red dot). Remarkably, this case contained both KRAS^{G12S} and U2AF1^{S34F} mutations, consistent with our findings that U2AF1^{S34F} restores exon 2 inclusion.

In addition to *KRAS* exon 2 skipping, we examined the relative usage of *KRAS4A* and *KRAS4B* across KRAS-mutant cancer cells and patient samples. In CCLE data, we found that most *KRAS* mutations trended towards increased *KRAS4A* usage (Extended Data Fig. 4j). We also investigated *KRAS4A* usage in 7518 cancers with *KRAS4A/KRAS4B* splicing data from TCGA and found that each assessed *KRAS* mutation was associated with significantly higher usage of *KRAS4A* than samples with wildtype KRAS (Extended Data Fig. 4k). These data support previous findings that *KRAS* mutations in LUAD are associated with increased levels of *KRAS4A* usage[33].

KRAS^{G12S}-mutant RNA contains a cryptic U2AF1 binding site which leads to exon 2 skipping

We observed that the RNA sequence of the KRAS^{G12S} mutation resembles a U2AF1 binding site, creating a UAG trinucleotide predicted to be bound by wildtype U2AF1 protein but not U2AF1^{S34F} (Fig. 2a)[8]. Previous work has reported that expression of KRAS^{G12V} or KRAS^{Q61H} leads to decreased phosphorylation of SR splicing factors and global changes in alternative splicing[34]. Therefore, we wondered if the *KRAS* exon 2 skipping we observed was due to the KRAS^{G12S} amino acid substitution itself, or due to the creation of a cryptic U2AF1 binding site in the RNA. To test this, we performed prime editing on A549 U2AF1^{S34S} cells to mutate the homozygous KRAS^{G12S} sequence to another codon for serine (TCT) that does not resemble a U2AF1 binding site (Fig. 2a). RT-PCR and RNA-sequencing analysis of exon 2 skipping found that exon skipping was completely abolished in KRAS^{G12S}-mutant cells with a TCT codon (Fig. 2b,c). This demonstrates that the creation of a novel cryptic U2AF1 binding site is required for *KRAS* exon 2 skipping, as opposed to the G12S amino acid substitution itself.

As it seemed surprising that a cryptic U2AF1 binding site would cause exon skipping rather than alternative 3' splice site usage, we performed a thorough analysis of the RNA-Seq reads using Integrative Genomics Viewer (IGV 2.16.0)[35]. Alternative 3' splice site usage would correlate with a stepwise increase in read coverage near the G12 RNA sequence, similar to the 3'UTR of *CTNNB1*, where U2AF1^{S34F} leads to decreased usage of the distal 3' splice site (Extended Data Fig. 5). However, we observed relatively constant read coverage across *KRAS* exon 2 regardless of U2AF1 mutational status, confirming that KRAS^{G12S} leads to exon 2 skipping as opposed to alternative 3' splice site usage (Extended Data Fig. 5).

As the 3' splice site upstream of *KRAS* exon 2 is only 45 bp from the cryptic binding site at G12S, we hypothesized that exon 2 skipping might arise from steric hindrance caused by wildtype U2AF1 binding and recruiting the spliceosome complex to this alternative site. To test whether such a process was possible, we targeted a catalytically inactive version of the RNA binding CasRX protein (dCasRx)[36] to various locations spanning *KRAS* exon 2 in NCI-H2023 cells with wildtype *KRAS* (Extended Data Table 3). Indeed, binding of dCasRx to the canonical 3' splice site leads to *KRAS* exon 2 skipping, likely due to interference with normal U2AF1 binding (Fig. 2d,e). Remarkably, directing dCasRx to the G12 sequence also resulted in significant levels of exon 2 skipping, while binding ~60 bases downstream lessened the impact on exon 2 skipping, likely due to decreased steric hindrance further from the upstream 3' splice site (Fig. 2d,e). This suggests that aberrant U2AF1 binding to the G12S codon could lead to *KRAS* exon 2 skipping due to steric hindrance with the proximal 3' splice site.

KRAS exon 2 skipping leads to reduced MAPK signalling and cell growth

We hypothesized that *KRAS* exon 2 skipping might lead to decreased MAPK signalling, while full activation of KRAS signalling could be enhanced by exon 2 inclusion. To test the impact of *KRAS* exon 2 inclusion on cell growth and signalling, we employed splice switching oligonucleotides (SSOs) targeting *KRAS* exon 2 (Extended Data Table 4)[30,37]. Treatment of NCI-H23 cells (KRAS^{G12C}) with increasing concentrations of exon-skipping SSOs led to increased levels of exon 2 skipping (Fig. 3a,b). In turn, higher levels of *KRAS* exon 2 skipping led to decreased MAPK signalling as indicated by reduced levels of p-Erk, p-Akt, and full length KRAS protein (Fig. 3c). We also treated cells with increasing levels of exon-skipping SSOs and grew them on ultra-low attachment plates, where we observed even small increases in *KRAS* exon 2 skipping led to decreased cell growth (Fig. 3d). These findings indicate that *KRAS* exon 2 skipping limits KRAS signalling and is detrimental to cell proliferation.

There is positive selection for U2AF1^{S34F} mutations in KRAS^{G12S}-mutant cancers that is associated with decreased overall survival

Our data show that $KRAS^{G12S}$ mutations result in exon 2 skipping which can be reversed by $U2AF1^{S34F}$ mutations, and that exon 2 skipping leads to reduced MAPK signalling and cell growth. Therefore, we hypothesized that $U2AF1^{S34F}$ mutations would be favoured in $KRAS^{G12S}$-mutant cancers as a means of driving KRAS signalling and tumour growth. To determine whether this was the case, we examined a set of 76,917 LUAD cases including 62,009 from Foundation Medicine and 14,908 from AACR Project GENIE and calculated the odds ratio of co-occurrence between $U2AF1^{S34F}$ and varying *KRAS* mutations[11,23]. Remarkably, we found that $KRAS^{G12S}$ mutation had an odds ratio of 7.4 for $U2AF1^{S34F}$ mutations, far higher than any other *KRAS* mutation, with a p value of 4.8×10^{-73} (Fig. 4a, Extended Data Fig. 6a). In addition, we found significant co-occurrence of $U2AF1^{S34F}$ mutations and $KRAS^{G12S}$ mutations in pancreatic adenocarcinoma cases (Fig. 4b).

$U2AF1^{S34F}$ has been described as a truncal mutation in both haematological malignancies and LUAD[15,38,39]. However, as $U2AF1^{S34F}$ rescues cryptic exon 2 skipping in $KRAS^{G12S}$-mutant cancers, we could imagine that $U2AF1^{S34F}$ mutations might occur as secondary events after the acquisition of a $KRAS^{G12S}$ mutation. To test this possibility, we analysed the variant allele frequencies (VAF) of $KRAS^{G12S}$ and $U2AF1^{S34F}$ in 43 LUAD cases from the Foundation Medicine Inc. dataset which harboured both mutations and lacked copy number variations (CNVs) at either locus. We found a strong positive correlation between the two mutations, however the VAF for $KRAS^{G12S}$ was significantly higher than the VAF for $U2AF1^{S34F}$ across the tumours, suggesting that $U2AF1^{S34F}$ occurs as a secondary mutation in $KRAS^{G12S}$-mutant LUADs (Fig. 4c, Extended Data Fig. 6b).

Finally, we examined survival data of 5,777 patients from the MSK-CHORD dataset of lung, breast, colorectal, prostate or pancreatic cancer patients[40]. We stratified patients based on their $KRAS^{G12}$ and $U2AF1^{S34F}$ mutation status, and found that while patients with $KRAS^{G12S}$ mutations fared better than patients with other $KRAS^{G12}$ mutations, acquisition of secondary $U2AF1^{S34F}$ mutations was associated with a significant reduction in overall survival (Fig. 4d). The allele fraction and survival results, taken together, suggest that $U2AF1^{S34F}$ mutations can occur as secondary mutations in LUAD and other cancers as a means of promoting $KRAS^{G12S}$ signalling, leading to poor survival outcomes.

$KRAS^{Q61R/Q61L}$ mutations are associated with *KRAS* exon 3 skipping and are both associated with and rescued by $U2AF1^{I24T}$ mutation

We searched for additional splicing factor/oncogene relationships by examining mutations which caused exon skipping events within the mutated gene and which had a significant co-occurrence with mutations in the splicing factor genes *U2AF1*, *SF3B1* or *SRSF2* (Fig. 5a). Interestingly, the strongest relationship was between another pair of *KRAS* and *U2AF1* mutations: $KRAS^{Q61R}$ and $U2AF1^{I24T}$ (Fig. 5a).

Our computational analysis suggested that $KRAS^{Q61R}$ mutations lead to significant levels of *KRAS* exon 3 skipping. Like *KRAS* exon 2, exon 3 skipping is expected to result in non-functional KRAS due to a frameshift after the first 37 amino acids of the protein, resulting in loss of Switch-II and all subsequent domains (Extended Data Fig. 7a). To understand the impact of $KRAS^{Q61R}$ mutations on *KRAS* splicing more clearly, we examined exon 3 skipping in 6,532 TCGA samples with *KRAS* exon 3 skipping data[32]. Unlike KRAS mutations at G12, where only G12S resulted in significant exon skipping, multiple mutations at Q61 led to *KRAS* exon 3 skipping, including Q61R, Q61L and Q61H (Fig. 5b). Interestingly, work from Kobayashi *et al*. found that $KRAS^{Q61K}$ mutations lead to exon 3 skipping, however this is rescued by the acquisition of adjacent silent mutations[41], explaining the absence of exon 3 skipping in Q61K-mutant samples.

The I24T mutation is located within the first RNA-binding zinc finger domain of U2AF1, but there is no detectable record of published work on $U2AF1^{I24T}$ mutations. Based on this

information and our findings above, we hypothesized that U2AF1^{I24T} mutations may be able to alter splicing to restore *KRAS* exon 3 inclusion in cells with KRAS^{Q61} mutations. To test this concept, we attempted to mutate the endogenous *U2AF1* locus using prime editing, however this was not successful. Therefore, we cloned wildtype U2AF1, U2AF1^{S34F} and U2AF1^{I24T} into the pLX301 lentiviral backbone[42], and transduced the resulting vectors into KRAS^{Q61R}-mutant Panc 02.13 cancer cells. Remarkably, we found that expressing U2AF1^{I24T}, but not U2AF1^{S34F}, rescued *KRAS* exon 3 inclusion from ~66% of transcripts to ~83% of transcripts (Fig. 5c,d). Interestingly, we also observed that over-expression of wildtype U2AF1 resulted in a small but statistically significant increase in *KRAS* exon 3 inclusion (Fig. 5c,d).

We examined patient data from AACR Project GENIE and found that both U2AF1^{I24T} and KRAS^{Q61} mutations are over-represented in pancreatic cancer compared to LUAD[11]. Therefore, we interrogated genomic data of 31,530 pancreatic cancer cases from Foundation Medicine Inc. and 8,304 cases from AACR Project GENIE (v16.1) to determine the odds ratio of co-occurrence for U2AF1^{I24T} and various mutations at Q61. Remarkably, U2AF1^{I24T} and KRAS^{Q61R} had an odds ratio for co-occurrence of 46.6, while an odds ratio of 25 was calculated for the co-occurrence of U2AF1^{I24T} and KRAS^{Q61L} (Fig. 5e, Extended Data Fig. 7b). These values were dramatically higher than for the co-occurrence of any other Q61 mutation with either U2AF1^{I24T} or U2AF1^{S34F}. Additionally, there was a strong correlation between the magnitude of exon 3 skipping caused by each Q61 mutation and the likelihood of co-occurrence with U2AF1^{I24T}, suggesting that U2AF1^{I24T} mutations are selected for as a means of rescuing normal *KRAS* splicing (Extended Data Fig. 7c).

To further determine whether U2AF1^{I24T} mutations occur secondary to KRAS^{Q61} mutations, we analysed the VAF of KRAS$^{Q61R/L}$ and U2AF1^{I24T} mutations in 20 pancreatic cancer samples from Foundation Medicine Inc. and AACR Project GENIE which harboured both mutations and had no copy number alterations at the *KRAS* or *U2AF1* loci. Similar to KRAS^{G12S} and U2AF1^{S34F}, there was a positive correlation between the two mutations, and the VAF for U2AF1^{I24T} was significantly higher than the VAF for KRAS$^{Q61R/L}$, suggesting that U2AF1^{I24T} occurs as a secondary mutation in KRAS^{Q61}-mutant pancreatic cancer (Extended Data Fig. 7d,e).

KRAS exon 3 skipping is detrimental to cell growth, and the acquisition of secondary U2AF1^{I24T} mutations is associated with poor overall survival in KRAS$^{Q61R/L}$-mutant pancreatic cancer patients

We previously observed that *KRAS* exon 2 skipping restricts cell growth and hypothesized that the same may be true for exon 3. To test whether *KRAS* exon 3 skipping hindered cell growth as expected, we employed SSOs targeting exon 3. As before, delivery of increasing amounts of exon 3 SSO led to increased exon skipping as quantified by RT-PCR (Fig. 6a,b). Furthermore, we found that KRAS^{G12C}-mutant NCI-H23 cells were sensitive to *KRAS* exon 3 skipping, with relatively small increases in skipping leading to significant reductions in cell growth (Fig. 6c).

Finally, we again investigated patient survival data from the MSK-CHORD dataset[40], this time focusing on 192 patients with pancreatic cancers with KRAS^{Q61} and/or U2AF1^{I24T} mutations. We stratified patients based on tumour KRAS^{Q61} and U2AF1^{I24T} mutation status, and found that patients with KRAS$^{Q61R/L}$-mutant cancers had similar survival rates as patients whose cancers harboured other KRAS^{Q61} mutations (Fig. 6d). However, patients whose cancers had KRAS$^{Q61R/L}$ mutations and acquired secondary U2AF1^{I24T} mutations had a significantly lower overall survival (Fig. 6d). In total, these results provide a second example of splicing factor mutations that occur as a means of promoting KRAS signalling.

DISCUSSION

The work presented here describes a mechanism for splicing factor mutation in cancer: certain oncogenic mutations lead to splicing defects that limit their impact on pathogenesis, leading to cascading selection whereby secondary mutations in splicing factors that restore normal splicing are positively selected.

Although U2AF1 mutations have been implicated in several biological pathways that are important in cancer[7,15,18,19], it has proved difficult to discern the specific selective advantage that these mutations confer in solid tumours. By using a combination of cell modelling and computational investigations of somatic mutations that frequently occurred with mutations in *U2AF1*, we discovered that mutant U2AF1 can regulate the alternative splicing of the *KRAS* oncogene. Our findings suggest that as cancer cells acquire oncogenic KRAS^{G12S} or KRAS$^{Q61R/L}$ mutations, inadvertent exon skipping in *KRAS* leads to reduced KRAS signalling and cell growth. However, the acquisition of U2AF1^{S34F} or U2AF1^{I24T} mutations respectively, reduces exon skipping, driving increased expression of the full-length *KRAS* transcript and improved cellular fitness. As a result, *U2AF1* mutations are much more likely to occur in KRAS^{G12S} or KRAS$^{Q61R/L}$-mutant lung or pancreatic adenocarcinomas compared to cancers with other *KRAS* mutations, and are associated with significantly reduced overall patient survival.

Our findings describe a novel mechanism by which mutant U2AF1 may act in cancer: by correcting inappropriate transcript splicing that is an unintended and counter-selected consequence of oncogenic mutations. These are the first examples of splicing factor mutations acting to correct splicing errors due to cancer-promoting genomic changes. However, we speculate that this phenomenon may be more widespread. Uncovering additional splicing factor mutations which are selected for as a means of fixing oncogene mis-splicing, if such mutations occur, will require concerted computational and cell modelling efforts.

Despite the clear enrichment of U2AF1^{S34F} and U2AF1^{I24T} mutations in KRAS^{G12S} and KRAS$^{Q61R/L}$-mutant cancers respectively, the majority of *KRAS*-mutant cancers retain wildtype *U2AF1*, and vice versa. This indicates that additional unknown processes drive the positive selection of *U2AF1* mutations in cancer, separate from the ability of mutant U2AF1 to rescue splicing defects in *KRAS*. We and others have shown that U2AF1^{S34F} mutations result in additional alternative splicing and gene expression changes that may affect the growth of cancer cells. Whether these changes act individually or in concert through coordinated gene expression programs remains unknown. Therefore, further analysis of RNA-sequencing datasets, including the one produced here, to uncover unique splicing and gene expression changes induced by specific *U2AF1* variants will be valuable.

Recent work from Kobayashi *et al.* found that KRAS^{Q61K} mutations lead to *KRAS* exon 3 skipping and early protein termination, while 100% of KRAS^{Q61K}-mutant cancers harbour silent mutations in codon A59 or G60 which restore *KRAS* exon 3 inclusion[41]. We found limited *KRAS* exon 3 skipping in KRAS^{Q61K}-mutant TCGA samples, likely due to their acquisition of secondary silent mutations[41]. However, KRAS^{Q61R}, KRAS^{Q61L} or KRAS^{Q61H} mutations led to varying degrees of *KRAS* exon 3 skipping. Our findings suggest that *KRAS* exon 3 skipping can be tolerated to some degree, but rescue of exon 3 inclusion by means of secondary mutations in *KRAS* or *U2AF1* are strongly selected for depending on the magnitude of exon 3 skipping. As such, KRAS^{Q61K} leads to high degrees of exon 3 skipping and always leads to acquisition of a secondary silent mutation[41], while other KRAS^{Q61} mutations have less exon 3 skipping and instead acquire secondary U2AF1^{I24T} mutations.

Our understanding of why certain amino acid substitutions in KRAS are more favoured than others is quite limited. Recent work by Huynh and colleagues found that Q61H mutations were over-represented in PDAC, despite Q61R, Q61L and Q61H-mutant cell lines exhibiting similar levels of anchorage-independent growth[43]. We believe that our results partly explain the

overabundance of $KRAS^{Q61H}$ mutations, as $KRAS^{Q61H}$ leads to the least amount of exon 3 skipping among the 3 major Q61 mutations. This concept that mutations that result in deleterious exon skipping are less favourable in the absence of secondary mutations may also explain the relative infrequency of $KRAS^{G12S}$ mutations compared to other KRAS mutations at G12.

Our study also provides evidence for another level of regulation for KRAS signalling. Previous studies have reported that KRAS is a relatively weak and poorly optimized oncogene, with a high percentage of rare codons, and weak 3' splice sites, particularly for exon 2 [30,37,44]. Our research supports this narrative as exon 2 is readily skipped in the presence of G12S mutations, while exon 3 is also frequently skipped due to multiple mutations at Q61. One possibility for this poor optimization may be that this provides an additional mechanism of regulating KRAS in order to limit oncogene-induced senescence, keeping signalling in a perfect range to induce oncogenesis[44].

Finally, our work provides evidence for a dynamic cancer genome, and offers a new view of evolutionary dependencies where oncogenic mutations are co-selected to drive tumorigenesis[45]. Previous research on *U2AF1* mutations has suggested these occur early in tumour evolution[15,38,39]. However, the variant allele frequencies of $KRAS^{G12S}$ or $KRAS^{Q61R/L}$ are significantly higher than $U2AF1^{S34F}$ or $U2AF1^{I24T}$ in tumours harbouring both mutations. This is consistent with a model in which a first round of selection events occurs during tumour initiation ($KRAS^{G12S}$ or $KRAS^{Q61R/L}$), that in turn leads to the acquisition of secondary mutational events ($U2AF1^{S34F}$ or $U2AF1^{I24T}$) to compensate for deleterious splicing effects of cancer-causing mutations. Most broadly, our results suggest that first-order cancer-causing mutations may often lead to a cascade of subsequent events whose selection is a response to the first mutation, a process that we term "cascading selection".

Methods

Cell Lines

All cell lines including A549, NCI-H441, NCI-H23, NCI-H2023, Panc 02.13 and 293T cells were obtained from American Type Culture Collection (ATCC). A549, NCI-H441, NCI-H23, and NCI-H2023 cells were cultured in RPMI 1640 medium (Gibco, 11875093) supplemented with 10% foetal bovine serum (Sigma Aldrich, F2442) and 50 μg/ml gentamicin (Gibco, 15750060). Panc 02.13 were cultured in RPMI 1640 medium (Gibco, 11875093) supplemented with 15% FBS, 0.3 mg/ml human recombinant insulin (Sigma Aldrich, 91077C) and 50 μg/ml Gentamicin. 293T cells were cultured in DMEM (Gibco #11965118) supplemented with 10% FBS and 50 μg/ml Gentamicin. Cells were cultured at 37 °C with 5% carbon dioxide. Cells were passaged by washing with phosphate buffered saline (Gibco, 10010023) and trypsinization with 0.25% Trypsin-EDTA (Gibco, 25200114). Cell lines were verified by STR profiling (ATCC, 135-XV) and tested negative for mycoplasma by PCR (ABM, G238).

Prime Editing

The *U2AF1* and/or *KRAS* alleles of A549 and NCI-H441 cells were modified by twin prime editing using plasmids assembled following the cloning protocol described by Anzalone and colleagues in Supplementary Note 3 [20]. pU6-tevopreq1-GG-acceptor backbone was a gift from David Liu (Addgene plasmid #174038). Oligonucleotides encoding the spacer sequence, extension template and SpCas9 sgRNA scaffold were ordered from Integrated DNA Technologies (IDT) (Extended Data Table 1) and resuspended in QIAGEN elution buffer at 100 μM concentration. Spacer sequence and extension template oligonucleotides were designed using pegFinder or PRIDICT software[46,47]. Each oligonucleotide pair was annealed by combining 1 μl of top (forward) oligonucleotide, 1 μl of bottom (reverse) oligonucleotide, and 23 μl of annealing buffer (water supplemented with 10 mM Tris-HCl pH 8.5 and 50 mM NaCl), heating at 95 °C for 3 minutes then cooling to 22 °C at 0.1 °C per second. Spacer, extension template and sgRNA scaffold annealed oligonucleotide pairs were then cloned into the GG acceptor backbone via golden gate assembly as described by Anzalone and colleagues in Supplementary Note 3 [20].

Cloned plasmids were propagated in NEB Stable Competent *E. coli* (New England BioLabs, C3040) and grown on 50 μg/ml carbenicillin plates (Corning, MT46100RG). Plasmids were extracted using the QIAprep Spin Miniprep Kit (QIAGEN, 27106) and plasmid concentrations were quantified using a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). Plasmids were then submitted to Quintara Biosciences for Sanger sequencing, or Plasmidsaurus Inc. for whole plasmid long-read sequencing.

Plasmids:
pU6-tevopreq1-U2AF1-S34S-Forward
pU6-tevopreq1-U2AF1-S34S-Reverse
pU6-tevopreq1-U2AF1-S34F-Forward
pU6-tevopreq1-U2AF1-S34F-Reverse
pU6-tevopreq1-U2AF1-S34F(TTC)-Forward
pU6-tevopreq1-U2AF1-S34F(TTC)-Reverse
pU6-tevopreq1-U2AF1-S34Y-Forward
pU6-tevopreq1-U2AF1-S34Y-Reverse
pU6-tevopreq1-U2AF1-S34C-Forward

pU6-tevopreq1-U2AF1-S34C-Reverse
pU6-tevopreq1-U2AF1-S34A-Forward
pU6-tevopreq1-U2AF1-S34A-Reverse

pU6-tevopreq1-U2AF1-S34F-to-WT-Forward
pU6-tevopreq1-U2AF1-S34F-to-WT-Reverse
pU6-tevopreq1-U2AF1-S34F-to-S-Forward
pU6-tevopreq1-U2AF1-S34F-to-S-Reverse
pU6-tevopreq1-U2AF1-S34F-to-F(TTC)-Forward
pU6-tevopreq1-U2AF1-S34F-to-F(TTC)-Reverse
pU6-tevopreq1-U2AF1-S34F-to-Y-Forward
pU6-tevopreq1-U2AF1-S34F-to-Y-Reverse
pU6-tevopreq1-U2AF1-S34F-to-C-Forward
pU6-tevopreq1-U2AF1-S34F-to-C-Reverse
pU6-tevopreq1-U2AF1-S34F-to-A-Forward
pU6-tevopreq1-U2AF1-S34F-to-A-Reverse

pU6-tevopreq1-KRAS-G12S(TCT)-Forward
pU6-tevopreq1-KRAS-G12S(TCT)-Reverse

Twin prime editing (TwinPE) was carried out by transfecting human cell lines with pCMV-PEMax-P2A-BSD (Addgene plasmid #174821) alongside 2 complementary epegRNAs cloned into the pU6-tevopreq1-GG-acceptor backbone (Addgene plasmid #174038)[21]. Cells were plated at 2 × 10^5 cells per well in a 6-well plate, and 24 hours after plating transfections were conducted using the Mirus Bio TransIT-X2 Dynamic Delivery System (Mirus Bio LLC, MIR6004) according to the manufacturer's recommended protocol for 6-well plates. A549 cells were treated with 5 μl of TransIT-X2 reagent in 6 well plates, while NCI-H441 cells were treated with 7.5 μl of TransIT-X2 reagent. 500 ng each of the forward and reverse epegRNA vectors were used, along with 1.5 μg of pCMV-PEMax-P2A-BSD. Cells were then treated with 10 mg/ml Blasticidin S HCl (Gibco, A1113903) 24-hours post-transfection and changed to normal media 72 hours post-transfection.

Single Cell Cloning

Cells were counted 1 week after prime editing, and 10,000 cells were plated in well A1 of a 96-well plate before being subjected to array dilution[48]. After 2-4 weeks, wells with single colonies were trypsinized and half of the cell volume was passaged in 24-well plates. 1 mL PBS was added to the remaining half of cells before being centrifuged for 5 minutes at 340 × *g*. The PBS was aspirated, and the cells were resuspended in 10 μl Elution Buffer (QIAGEN) and boiled at 100°C for 15 minutes. Samples were then placed on ice for 3 minutes before being pelleted by centrifugation at 20,000 × *g*. for 1 minute. PCR amplification of the desired locus was performed with 12.5 μl Q5 High-Fidelity 2X Master Mix (New England BioLabs, M0494S) along with 7.5 μl of the isolated supernatant as template, and 2.5 μl each of 5 μM forward and 5 μM reverse PCR primers. Samples were then submitted to Quintara Biosciences for Sanger sequencing.

RNA-Sequencing

RNA for each cell clone was extracted using the RNeasy Plus Mini Kit (QIAGEN, 74134) as per the manufacturer's instructions. RNA concentration was quantified using the Qubit RNA Broad

Range Quantification Assay Kit (Invitrogen, Q10210) according to the manufacturer's protocol. RNA was then submitted to Novogene for mRNA-Sequencing including the following steps described in this paragraph. Sample quality was confirmed to have a RIN score >9 by BioAnalyzer. Messenger RNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, the first strand cDNA was synthesized using random hexamer primers, followed by the second strand cDNA synthesis using dTTP. The samples were then processed by end repair, A-tailing, adapter ligation, size selection, amplification, and purification. The library quality was checked with Qubit and real-time PCR for quantification and BioAnalyzer for size distribution detection. Paired-end clean reads were aligned to the reference genome GRCh38 using Hisat2 (v2.0.5)[49]. FeatureCounts (v1.5.0-p3) was used to count the reads numbers mapped to each gene[50].

Alternative splicing events were identified using rMATS (4.1.0)[51] and defined as significant according to an FDR < 0.05. Sashimi plots were viewed and generated using Integrative Genomics Viewer (v2.16.10)[52]. Filtering for known oncogenes and tumour suppressors was performed by examining genes that were defined as cancer genes by at least 4 sources in OncoKB[24,25].

KRAS exon 2 read coverage and fraction of mutant reads for *U2AF1* were analysed using Integrative Genomics Viewer (v2.16.10)[52]. The overlap of significant alternative splicing events between U2AF1^{S34F}-mutant A549 cells and human lung adenocarcinomas was quantified by performing a Fisher's exact text on alternative splicing events ($p<0.05$ by rank-sum test) present or absence in each dataset as determined by rMATS (A549 cells) or SplAdder (TCGA)[32].

To identify mutant U2AF1 3' splice site sequence preferences, significant exon splicing events were first determined using rMATS. The regions of interest were set as 10 bp preceding and following the first base of each skipped exon, as well as 10 bp preceding and following the first base of each included exon. The coordinates of these genomic regions were recorded in BED file format. Sequences corresponding to these regions were extracted from the Hg38 human reference genome fasta file using the getfasta function from bedtools (v2.26.0)[53], and graphics were generated using Weblogo (v.3.7.11)[54].

RT-PCR quantification of *KRAS* exon 2 or exon 3 skipping

RT-PCR primers to detect *KRAS* exon 2 or exon 3 skipping were designed using Primer-BLAST[55], with the forward and reverse primers being present on exons flanking the exon of interest. Cells were trypsinized and centrifuged at 340 × *g* for 5 min, followed by RNA extraction using RNeasy Plus Mini Kit (QIAGEN, 74134) as per manufacturer's protocol. RNA concentrations were quantified with the Qubit RNA BR Assay Kit (Invitrogen, Q10210), and 100 ng of RNA was used to make cDNA using the SuperScript IV VILO Master Mix (Invitrogen, 11756050). The samples were PCR amplified using Phusion Plus PCR Master Mix (Thermo Scientific, F631S), using a 60°C annealing temperature per the manufacturer's 3-step protocol. The PCR products were run on a 2% agarose gel and imaged with a ChemiDoc MP (Bio-Rad Laboratories, 12003154). Band intensities were quantified using the Analyze Gels feature in Fiji (ImageJ) and used to calculate the percent spliced in for *KRAS* exon 2 or exon 3.

Modeling steric hindrance and *KRAS* exon 2 skipping using dCasRx

pXR002: EF1a-dCasRx-2A-EGFP and pXR003: CasRx gRNA cloning backbone were gifts from Patrick Hsu (Addgene plasmid # 109050 and 109053)[36]. dCasRx sgRNAs were designed using the cas13design tool (https://cas13design.nygenome.org)[56,57], ordered from IDT and resuspended in QIAGEN elution buffer (Extended Data Table 3). Each oligonucleotide pair was annealed by combining 1 μl of top (forward) oligonucleotide, 1 μl of bottom (reverse) oligonucleotide, and 23 μl of annealing buffer (water supplemented with 10 mM Tris-HCl pH 8.5 and 50 mM NaCl), heating at 95 °C for 3 minutes then cooling to 22 °C at 0.1 °C per second. Annealed oligonucleotide pairs were then cloned into the CasRX gRNA cloning backbone via golden gate assembly, using BbsI-HF (NEB R3539S)[20]. Cloned plasmids were propagated in NEB Stable Competent *E. coli* (New England BioLabs, C3040) and grown on 50 μg/ml carbenicillin plates (Corning, MT46100RG). Plasmids were extracted using the QIAprep Spin Miniprep Kit (QIAGEN, 27106) and plasmid concentrations were quantified using a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). Plasmids were then submitted to Quintara Biosciences for Sanger sequencing, or Plasmidsaurus Inc. for whole plasmid long-read sequencing.

Plasmids:
pXR003-KRAS-Exon2-3'SS-1
pXR003-KRAS-Exon2-3'SS-2
pXR003-KRAS-G12-1
pXR003-KRAS-G12-2
pXR003-KRAS-Downstream-4
pXR003-KRAS- Downstream-5

To test the impact of dCasRx binding to different locations on the *KRAS* exon 2 RNA, NCI-H2023 with wildtype *KRAS* were plated at 200,000 cells per well on 6-well plates. Cells were transfected 24 hours later with 5 μg EF1a-dCasRx-2A-EGFP and 5 μg dCasRx sgRNA in 500 ml Opti-MEM with 20 μl TransIT-X2. 48 hours post-transfection, cells were trypsinized and centrifuged at 340 × *g* for 5 min, followed by RNA extraction using RNeasy Plus Mini Kit (QIAGEN, 74134). 100 ng of RNA was used to make cDNA using the SuperScript IV VILO Master Mix (Invitrogen, 11756050). Exon 2 skipping was then quantified by RT-PCR and the PCRs were run on a 2% agarose gel.

Splice Switching Oligonucleotide Treatment

For *KRAS* exon 2 skipping, SSOs designed and validated by the laboratory of Brage Andresen were used[37]. For *KRAS* exon 3 skipping, SSOs were designed using the eSkip-Finder application[58], using 2'OMe chemistry and a length of 25 bp. SSO sequences are available in Extended Data Table 4. NCI-H23 cells were plated at 2×10^6 cells per 10 cm plate. 24 hours after plating, cells were transfected using Transit-X2 transfection reagent according to the manufacturer's recommended protocol for 10 cm plates, with 45 μl of TransIT-X2 reagent. Each well was treated with 120 nM total SSO, comprising of a range of concentrations of either scramble SSO control, or *KRAS* exon 2 or exon 3 skipping SSO (120 nM scramble SSO alone, 6 nM *KRAS* SSO with 114 nM scramble SSO, 12 nM *KRAS* SSO with 108 nM scramble SSO, 30 nM *KRAS* SSO with 90 nM scramble SSO, 60 nM *KRAS* SSO with 60 nM scramble SSO, or 120 nM *KRAS* SSO alone) to keep the total amount of SSO delivery the same between conditions[30,37]. Cells were plated for cell growth assays 24 hours after transfection, while cells were collected for RNA or protein 72 hours following transfection.

Immunoblot Analysis

Cells were lysed with RIPA Buffer (Millipore Sigma, R0278-50ML) containing Halt Protease and Phosphatase Inhibitor Cocktail (100X) (Thermo Scientific, 78440) and quantified with the Pierce BCA Protein Assay Kit (Thermo Scientific, 23225). Following addition of NuPAGE 4X LDS Sample Buffer (Invitrogen, NP0008) and NuPAGE 10X Sample Reducing Agent (Invitrogen, NP0009), the samples were run on a NuPAGE 4-12% Bis-Tris, 1.5 MM, Mini Protein Gel (Invitrogen, NP0335BOX) in NuPAGE 20X MOPS SDS Running Buffer (Invitrogen, NP0001) at 125V for 105 min on ice. The gel was transferred on a nitrocellulose membrane in NuPAGE 20X Transfer Buffer (Invitrogen, NP00061) at 70V for 2 hrs on ice. Blocking and Immunostaining for HSP90 (BD Biosciences, 610418, 1:10,000), p-Erk (Cell Signaling, 4370S, 1:2000), Total-Erk (Cell Signaling, 9102S, 1:2000), p-Akt (Cell Signaling, 4060S, 1:1000), Total-Akt (Cell Signaling, 58295S, 1:500) and KRAS (Invitrogen, 11H35L14, 1:2500) were done in Intercept (TBS) Blocking Buffer (LI-COR Biosciences, 927-60001), with 0.1% Tween-20 for primary and secondary antibody blockings. LI-COR Odyssey Classic Imaging System (LI-COR Biosciences, ODY-9120) was used to image the blots.

Cell Proliferation Assays

Cells were plated on ultra-low attachment plates (Corning, 3474) at 1,000 cells per well in 100 μl of media for each desired time point to measure growth. At each time point, cells were moved to white-walled 96-well plates (Corning, 3903). 100 μl of CellTiter-Glo Luminescent Cell Viability Assay (Promega, G7572) diluted 1:1 in PBS was then added to each well. After a 10-minute incubation at room temperature while rocking, the plates were analysed using a SpectraMax M5 (Molecular Devices) plate reader using the included CellTiter-Glo protocol with a luminescence integration time of 500 ms. Luminescence values were normalized to data points from the earliest time point.

Cloning of lentiviral vectors

U2AF1^{S34F} cDNA was synthesized by IDT, and subsequently cloned into the pLX301 backbone (Addgene #25895) via Gibson assembly. In brief, the U2AF1^{S34F} cDNA was amplified using 2X Phusion Plus PCR Master Mix (Thermo Fisher, F631S) and primers with overlapping sequences to the pLX301 backbone following the CMV promoter. Gibson assembly primers were designed using the NEBuilder Assembly Tool (NEB) and contained a FLAG tag sequence added to the C-terminal end of U2AF1. The PCR product was run on a 1% agarose gel and gel purified using the QIAquick PCR & Gel Cleanup Kit (QIAGEN, 28506). The pLX301 backbone was linearized via restriction digested with BsrGI-HF (NEB, R3575S) for 1 hour at 37°C, and then run on a 1% agarose gel and gel purified. Gibson assembly was then carried out using linearized pLX301, U2AF1 PCR product, and NEBuilder Hifi DNA Assembly master mix (NEB, E2621S) using the manufacturer's recommended protocol.

The vector was then mutated to express wildtype U2AF1 and U2AF1^{I24T} via sequential site-directed mutagenesis. First, pLX301-U2AF1^{S34F}-FLAG was mutated to pLX301-U2AF1-FLAG, before being mutated again to form pLX301-U2AF1^{I24T}-FLAG. Site directed mutagenesis primers were designed using PrimerX (https://www.bioinformatics.org/primerx/) and site-directed mutagenesis was performed by PCR amplification with Platinum SuperFi II Master Mix (Thermo

Fisher, 12368010) followed by restriction digestion with DpnI (NEB, R0176S) for 1 hour at 37°C before bacterial transformation.

Plasmids:
pLX301-U2AF1-WT-FLAG
pLX301-U2AF1-S34F-FLAG
pLX301-U2AF1-I24T-FLAG

Lentivirus Production and Transduction

Lentivirus was produced as described previously[59]. Briefly, 8 × 10^6 293T cells were plated on 15 cm plates coated with 0.1% gelatin. 24 hours after plating, cells were transfected with 10 μg pLX301 lentiviral vector, 7.5 μg delta 8.9 plasmid, and 2.5 μg pCMV-VSV-G along with 80 μl PEIMax (Polysciences, 24765) in 1 ml of Opti-MEM I (Gibco, 31985062). 24 hours after transfection, the media was replaced with fresh DMEM supplemented with 25 mM HEPES (Gibco, 15630-080) and 3 mM caffeine (Sigma, C0750). Lentivirus-containing supernatant was collected from the cells at 48 and 72 hours following transfection and filtered through 0.45-μm filters (Thermo Scientific 723-2545).

Lentivirus was then concentrated by centrifugation through a sucrose cushion[60]. A 10% sucrose solution (50 mM Tris-HCl,100 mM NaCl, 0.5 mM EDTA) was carefully pipetted underneath the lentivirus-containing media at a 4:1 ratio of lentivirus-containing media to sucrose solution. The lentivirus was then centrifuged at 10,000 × g for 4 hours at 4°C. Media was aspirated and the pellet was resuspended in 100 μl of PBS.

For Panc 02.13 cell transduction, 35,000 cells were plated in 300 μl insulin-containing RPMI in 48-well plates. 24 hours after plating, cells were transduced via spinfection. Briefly, fresh RPMI containing 8 μg/ml polybrene (Santa Cruz Biotechnology, sc-134220) was added to each well. The titer of each lentivirus was estimated using Lenti-X GoStix Plus (Takara, 631280), and equal amounts of lentivirus was delivered to each well. The 48-well plate was then centrifuged at 1,000 × g for 2 hours at 30°C. 24 hours after spinfection, fresh RPMI containing 2 μg/ml puromycin (Gibco, A1113803) was added to each well. 48 hours after puromycin treatment, cells were trypsinized and collected for RNA isolation and RT-PCR.

Genomic Data

All data from the AACR Project GENIE Consortium[11] were obtained through the dedicated cBioPortal website[26]. Additional lung adenocarcinoma and pancreatic cancer sequencing data were provided by Foundation Medicine Inc., comprising 62,009 lung cancer cases and 31,530 pancreatic cancer cases with tissue biopsy-based comprehensive genomic profiling using FoundationOne®/ FoundationOne®CDx during routine clinical care[23]. In these cohorts, co-occurrence of *U2AF1* and *KRAS* mutations, as well as a comparison of variant allele frequencies for mutant *U2AF1* and *KRAS* in patients with both alterations were assessed. Alternative splicing data from CCLE was obtained from Ghandi *et al.*[31], and *KRAS* and *U2AF1* mutational data was obtained via cBioPortal[26]. Alternative splicing data from TCGA was obtained from Kahles *et al.*[32], and *KRAS* and *U2AF1* mutation data was obtained via cBioPortal[26]. Patient survival data and accompanying mutation data was obtained from the MSK-CHORD study[40], via the cBioPortal website[26].

Statistical Analysis

All statistical analyses were performed in GraphPad Prism 10 except where noted. Alternative splicing events were analysed, and FDR q values were calculated using rMATS[51]. A simple linear regression was performed for the correlation between the percentage of patients with U2AF1^{S34F} mutations and the odds ratio of a smoker developing the disease, the correlation between the number alternatively spliced genes and the relative frequency of each U2AF1 mutation across all cancers, the correlation between the fraction of exon 2 inclusion and the variant allele frequency for KRAS^{G12S} in TCGA, the correlation between the variant allele frequency of KRAS^{G12S} and U2AF1^{S34F} in the Foundation Medicine Inc. dataset, the correlation between the variant allele frequency of KRAS$^{Q61R/L}$ and U2AF1^{I24T} in the Foundation Medicine Inc. and AACR Project GENIE datasets, and the correlation between the odds ratio of co-occurrence with U2AF1^{I24T} for each KRAS mutation at Q61 and the fraction of *KRAS* exon 3 inclusion for each mutation.

To quantify the fraction of mutant reads for each *U2AF1* variant, the fraction of reads with each individual splicing event across cells with different *U2AF1* variants, the fraction of reads with *KRAS* exon skipping and *KRAS4A* isoform usage in CCLE and TCGA data, and the fraction of *KRAS* exon 2 inclusion in CCLE and TCGA samples with and without U2AF1^{S34F} and KRAS^{G12S} mutations, a one-way ANOVA with multiple comparisons using Dunnett correction was performed. The Student's t-test was used to analyse the fraction of RNA-sequencing reads showing *KRAS* exon 2 skipping in backedited A549 cells and engineered NCI-H441 cells, and the fraction of *KRAS* exon 2 skipping in U2AF1^{S34F}-mutant A549 cells by RT-PCR. For the comparison of the variant allele frequency of KRAS^{G12S} and U2AF1^{S34F}, and KRAS$^{Q61R/L}$ and U2AF1^{I24T}, a paired t-test was used.

A two-sided Fisher Exact test was used to compare the overlap in splicing events between engineered U2AF1^{S34F}-mutant A549 cells and U2AF1^{S34F}-mutant lung adenocarcinoma samples, for analysis of the frequency of mutations in U2AF1^{S34F}-mutant lung adenocarcinoma compared to all other lung adenocarcinoma cases, for the calculation of co-occurrence between U2AF1 and KRAS mutations from Foundation Medicine Inc. and AACR Project GENIE data, and for the identification of additional somatic mutations which co-occur with splicing factor mutations. Patient survival analyses were compared using the Gehan-Breslow-Wilcoxon test.

Data Availability

RNA-sequencing data will be available in the NCBI Gene Expression Omnibus (GEO). Unique plasmids will be available in Addgene. All other data generated and analysed during this study are included in the manuscript. Requests for further information should be directed to the lead contact.

Code Availability

This paper does not report original code.

Acknowledgements

We wish to thank Mitchell Leibowitz and all the members of the Meyerson laboratory for their advice and technical assistance; Mitchell in particular suggested the steric hindrance

experiment. Thank you to Leslie Gaffney for assistance with figures. Thank you to Peter Chen and David Liu for their guidance with prime editing approaches and for sharing detailed protocols with us. Thank you to Andrew Aguirre, Ben Lampson and Colleen Harrington for helpful discussions. This research was supported by Damon Runyon Cancer Research Foundation awards (D.M.W. DRG-2449-21 and A.B.D. DRG-2504-23.), an NIH R35 CA197568 grant (M.M.) and the American Cancer Society Research Professorship (M.M.).

Contributions

D.M.W. and M.M. conceived of the study. D.M.W. and K.C. designed and performed most of the experiments and analysed most of the data. S.S. and G.F. performed and analysed the genomics data from Foundation Medicine Inc. I.T.-H.L., E.S., A.B.D., K.X.J. and A.A.G. performed additional genomic analyses. D.M.W., K.C. and M.M. wrote the manuscript. All authors read and edited the manuscript.

Ethics Declarations

S.S. and G.F. are employees of Foundation Medicine Inc., a subsidiary of Roche, and have stock ownership in Roche. M.M. consults for and holds equity in Bayer and Delve Bio; is an inventor on patents licensed to Bayer and Labcorp; and receives research support from Bayer and Janssen, all outside the scope of the current work. M.M. was also a founder of Foundation Medicine with shares sold to Roche but has no continued financial relationship with the company at the time of manuscript submission.

References

1. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64-69. 10.1038/nature10496.
2. Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835. 10.1038/45590.
3. Warnasooriya, C., Feeney, C.F., Laird, K.M., Ermolenko, D.N., and Kielkopf, C.L. (2020). A splice site-sensing conformational switch in U2AF2 is modulated by U2AF1 and its recurrent myelodysplasia-associated mutation. Nucleic Acids Res *48*, 5695-5709. 10.1093/nar/gkaa293.
4. Gehring, N.H., and Roignant, J.Y. (2021). Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. Trends Genet *37*, 355-372. 10.1016/j.tig.2020.10.008.
5. Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell *150*, 1107-1120. 10.1016/j.cell.2012.08.029.

6. Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.-M., Gingras, M.-C., Miller, D.K., Christ, A.N., Bruxner, T.J.C., Quinn, M.C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47-52. 10.1038/nature16965.
7. Biancon, G., Joshi, P., Zimmer, J.T., Hunck, T., Gao, Y., Lessard, M.D., Courchaine, E., Barentine, A.E.S., Machyna, M., Botti, V., et al. (2022). Precision analysis of mutant U2AF1 activity reveals deployment of stress granules in myeloid malignancies. Mol Cell *82*, 1107-1122.e1107. 10.1016/j.molcel.2022.02.025.
8. Brooks, A.N., Choi, P.S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., Pedamallu, C.S., Sivachenko, A., Rosenberg, M., Chmielecki, J., et al. (2014). A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. PLoS One *9*, e87361. 10.1371/journal.pone.0087361.
9. Yoshida, H., Park, S.-Y., Sakashita, G., Nariai, Y., Kuwasako, K., Muto, Y., Urano, T., and Obayashi, E. (2020). Elucidation of the aberrant 3′ splice site selection by cancer-associated mutations on the U2AF1. Nature Communications *11*, 4744. 10.1038/s41467-020-18559-6.
10. Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E.J., Okeyo-Owuor, T., et al. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell *27*, 631-643. 10.1016/j.ccell.2015.04.008.
11. The AACR Project GENIE Consortium. (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov *7*, 818-831. 10.1158/2159-8290.Cd-17-0151.
12. Buisson, R., Langenbucher, A., Bowen, D., Kwan, E.E., Benes, C.H., Zou, L., and Lawrence, M.S. (2019). Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. Science *364*. 10.1126/science.aaw2872.
13. Zhou, Z., Zou, Y., Liu, G., Zhou, J., Wu, J., Zhao, S., Su, Z., and Gu, X. (2017). Mutation-profile-based methods for understanding selection forces in cancer somatic mutations: a comparative analysis. Oncotarget *8*, 58835-58846. 10.18632/oncotarget.19371.
14. Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., Shibata, T., et al. (2016). Mutational signatures associated with tobacco smoking in human cancer. Science *354*, 618-622. 10.1126/science.aag0299.
15. Esfahani, M.S., Lee, L.J., Jeon, Y.J., Flynn, R.A., Stehr, H., Hui, A.B., Ishisoko, N., Kildebeck, E., Newman, A.M., Bratman, S.V., et al. (2019). Functional significance of U2AF1 S34F mutations in lung adenocarcinomas. Nat Commun *10*, 5712. 10.1038/s41467-019-13392-y.
16. Zhao, Y., Cai, W., Hua, Y., Yang, X., and Zhou, J. (2022). The Biological and Clinical Consequences of RNA Splicing Factor U2AF1 Mutation in Myeloid Malignancies. Cancers (Basel) *14*. 10.3390/cancers14184406.
17. Fei, D.L., Zhen, T., Durham, B., Ferrarone, J., Zhang, T., Garrett, L., Yoshimi, A., Abdel-Wahab, O., Bradley, R.K., Liu, P., and Varmus, H. (2018). Impaired hematopoiesis and leukemia development in mice with a conditional knock-in allele of a mutant splicing factor gene U2af1. Proc Natl Acad Sci U S A *115*, E10437-e10446. 10.1073/pnas.1812669115.

18. Wheeler, E.C., Vora, S., Mayer, D., Kotini, A.G., Olszewska, M., Park, S.S., Guccione, E., Teruya-Feldstein, J., Silverman, L., Sunahara, R.K., et al. (2022). Integrative RNA-omics Discovers GNAS Alternative Splicing as a Phenotypic Driver of Splicing Factor–Mutant Neoplasms. Cancer Discovery *12*, 836-855. 10.1158/2159-8290.Cd-21-0508.
19. Palangat, M., Anastasakis, D.G., Fei, D.L., Lindblad, K.E., Bradley, R., Hourigan, C.S., Hafner, M., and Larson, D.R. (2019). The splicing factor U2AF1 contributes to cancer progression through a noncanonical role in translation regulation. Genes Dev *33*, 482-497. 10.1101/gad.319590.118.
20. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149-157. 10.1038/s41586-019-1711-4.
21. Anzalone, A.V., Gao, X.D., Podracky, C.J., Nelson, A.T., Koblan, L.W., Raguram, A., Levy, J.M., Mercer, J.A.M., and Liu, D.R. (2022). Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nature Biotechnology *40*, 731-740. 10.1038/s41587-021-01133-w.
22. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291-304.e296. 10.1016/j.cell.2018.03.022.
23. Milbury, C.A., Creeden, J., Yip, W.-K., Smith, D.L., Pattani, V., Maxwell, K., Sawchyn, B., Gjoerup, O., Meng, W., Skoletsky, J., et al. (2022). Clinical and analytical validation of FoundationOne®CDx, a comprehensive genomic profiling assay for solid tumors. PLoS One *17*, e0264138. 10.1371/journal.pone.0264138.
24. Chakravarty, D., Gao, J., Phillips, S., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precision Oncology, 1-16. 10.1200/PO.17.00011.
25. Suehnholz, S.P., Nissan, M.H., Zhang, H., Kundra, R., Nandakumar, S., Lu, C., Carrero, S., Dhaneshwar, A., Fernandez, N., Xu, B.W., et al. (2024). Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. Cancer Discovery *14*, 49-65. 10.1158/2159-8290.Cd-23-0467.
26. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov *2*, 401-404. 10.1158/2159-8290.CD-12-0095.
27. Nuevo-Tapioles, C., and Philips, M.R. (2022). The role of KRAS splice variants in cancer biology. Front Cell Dev Biol *10*, 1033348. 10.3389/fcell.2022.1033348.
28. Whitley, M.J., Tran, T.H., Rigby, M., Yi, M., Dharmaiah, S., Waybright, T.J., Ramakrishnan, N., Perkins, S., Taylor, T., Messing, S., et al. Comparative analysis of KRAS4a and KRAS4b splice variants reveals distinctive structural and functional properties. Science Advances *10*, eadj4137. 10.1126/sciadv.adj4137.
29. Voice, J.K., Klemke, R.L., Le, A., and Jackson, J.H. (1999). Four human ras homologs differ in their abilities to activate Raf-1, induce transformation, and stimulate cell motility. J Biol Chem *274*, 17164-17170. 10.1074/jbc.274.24.17164.

30. Hartung, A.-M., Swensen, J., Uriz, I.E., Lapin, M., Kristjansdottir, K., Petersen, U.S.S., Bang, J.M.V., Guerra, B., Andersen, H.S., Dobrowolski, S.F., et al. (2016). The Splicing Efficiency of Activating HRAS Mutations Can Determine Costello Syndrome Phenotype and Frequency in Cancer. PLOS Genetics *12*, e1006039. 10.1371/journal.pgen.1006039.
31. Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature *569*, 503-508. 10.1038/s41586-019-1186-3.
32. Kahles, A., Lehmann, K.V., Toussaint, N.C., Hüser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., and Rätsch, G. (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell *34*, 211-224.e216. 10.1016/j.ccell.2018.07.001.
33. Yang, I.S., and Kim, S. (2018). Isoform specific gene expression analysis of KRAS in the prognosis of lung adenocarcinoma patients. BMC Bioinformatics *19*, 40. 10.1186/s12859-018-2011-y.
34. Lo, A., McSharry, M., and Berger, A.H. (2022). Oncogenic KRAS alters splicing factor phosphorylation and alternative splicing in lung cancer. BMC Cancer *22*, 1315. 10.1186/s12885-022-10311-1.
35. Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192. 10.1093/bib/bbs017.
36. Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N., and Hsu, P.D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell *173*, 665-676.e614. 10.1016/j.cell.2018.02.033.
37. Andresen, B. (2019). RAS exon 2 skipping for cancer treatment. Google Patents.
38. Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2011). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet *44*, 53-57. 10.1038/ng.1031.
39. Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616-3627; quiz 3699. 10.1182/blood-2013-08-518886.
40. Jee, J., Fong, C., Pichotta, K., Tran, T.N., Luthra, A., Waters, M., Fu, C., Altoe, M., Liu, S.-Y., Maron, S.B., et al. (2024). Automated real-world data integration improves cancer outcome prediction. Nature. 10.1038/s41586-024-08167-5.
41. Kobayashi, Y., Chhoeu, C., Li, J., Price, K.S., Kiedrowski, L.A., Hutchins, J.L., Hardin, A.I., Wei, Z., Hong, F., Bahcall, M., et al. (2022). Silent mutations reveal therapeutic vulnerability in RAS Q61 cancers. Nature *603*, 335-342. 10.1038/s41586-022-04451-4.
42. Yang, X., Boehm, J.S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S.R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression library of human ORFs. Nat Methods *8*, 659-661. 10.1038/nmeth.1638.
43. Huynh, M.V., Hobbs, G.A., Schaefer, A., Pierobon, M., Carey, L.M., Diehl, J.N., DeLiberty, J.M., Thurman, R.D., Cooke, A.R., Goodwin, C.M., et al. (2022). Functional and biological

heterogeneity of KRAS(Q61) mutations. Sci Signal *15*, eabn2694. 10.1126/scisignal.abn2694.

44. Lampson, B.L., Pershing, N.L., Prinz, J.A., Lacsina, J.R., Marzluff, W.F., Nicchitta, C.V., MacAlpine, D.M., and Counter, C.M. (2013). Rare codons regulate KRas oncogenesis. Curr Biol *23*, 70-75. 10.1016/j.cub.2012.11.031.
45. Mina, M., Iyer, A., Tavernari, D., Raynaud, F., and Ciriello, G. (2020). Discovering functional evolutionary dependencies in human cancers. Nature Genetics *52*, 1198-1207. 10.1038/s41588-020-0703-5.
46. Chow, R.D., Chen, J.S., Shen, J., and Chen, S. (2021). A web tool for the design of prime-editing guide RNAs. Nature Biomedical Engineering *5*, 190-194. 10.1038/s41551-020-00622-8.
47. Mathis, N., Allam, A., Kissling, L., Marquart, K.F., Schmidheini, L., Solari, C., Balázs, Z., Krauthammer, M., and Schwank, G. (2023). Predicting prime editing efficiency and product purity by deep learning. Nature Biotechnology *41*, 1151-1159. 10.1038/s41587-022-01613-7.
48. Ryan, J.A. (2008). Cell cloning by serial dilution in 96 well plates. Corning.
49. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology *37*, 907-915. 10.1038/s41587-019-0201-4.
50. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930. 10.1093/bioinformatics/btt656.
51. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A *111*, E5593-5601. 10.1073/pnas.1419161111.
52. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology *29*, 24-26. 10.1038/nbt.1754.
53. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.
54. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190. 10.1101/gr.849004.
55. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134. 10.1186/1471-2105-13-134.
56. Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z., and Sanjana, N.E. (2020). Massively parallel Cas13 screens reveal principles for guide RNA design. Nature Biotechnology *38*, 722-727. 10.1038/s41587-020-0456-9.
57. Guo, X., Rahman, J.A., Wessels, H.-H., Méndez-Mancilla, A., Haro, D., Chen, X., and Sanjana, N.E. (2021). Transcriptome-wide Cas13 guide RNA design for model organisms and viral RNA pathogens. Cell Genomics *1*, 100001. https://doi.org/10.1016/j.xgen.2021.100001.
58. Chiba, S., Lim, Kenji Rowel Q., Sheri, N., Anwar, S., Erkut, E., Shah, Md Nur A., Aslesh, T., Woo, S., Sheikh, O., Maruyama, R., et al. (2021). eSkip-Finder: a machine learning-based

web application and database to identify the optimal sequences of antisense oligonucleotides for exon skipping. Nucleic Acids Research *49*, W193-W198. 10.1093/nar/gkab442.

59. Walter, D.M., Venancio, O.S., Buza, E.L., Tobias, J.W., Deshpande, C., Gudiel, A.A., Kim-Kiselak, C., Cicchini, M., Yates, T.J., and Feldser, D.M. (2017). Systematic In Vivo Inactivation of Chromatin-Regulating Enzymes Identifies Setd2 as a Potent Tumor Suppressor in Lung Adenocarcinoma. Cancer Res *77*, 1719-1729. 10.1158/0008-5472.CAN-16-2159.

60. Jiang, W., Hua, R., Wei, M., Li, C., Qiu, Z., Yang, X., and Zhang, C. (2015). An optimized method for high-titer lentivirus preparations without ultracentrifugation. Sci Rep *5*, 13875. 10.1038/srep13875.

61. Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., Airoldi, E.M., and Burge, C.B. (2015). Quantitative visualization of alternative exon expression from RNA-seq data. Bioinformatics *31*, 2400-2402. 10.1093/bioinformatics/btv034.

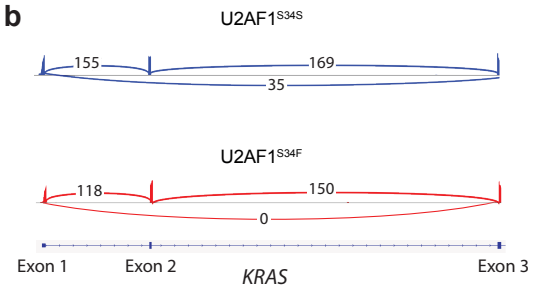
62. Rásó, E. (2020). Splice variants of RAS—translational significance. Cancer and Metastasis Reviews *39*, 1039-1049. 10.1007/s10555-020-09920-8.

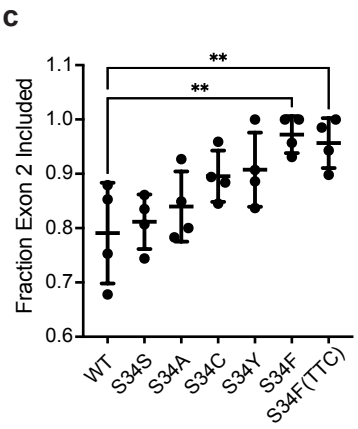
63. Kwan, A.K., Piazza, G.A., Keeton, A.B., and Leite, C.A. (2022). The path to the clinic: a comprehensive review on direct KRASG12C inhibitors. Journal of Experimental & Clinical Cancer Research *41*, 27. 10.1186/s13046-021-02225-w.

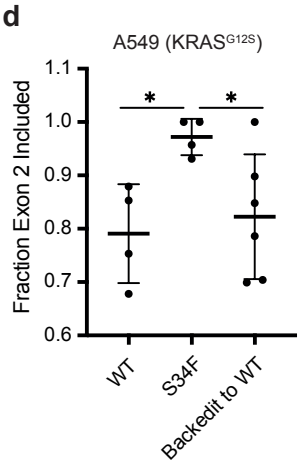
Figure 1

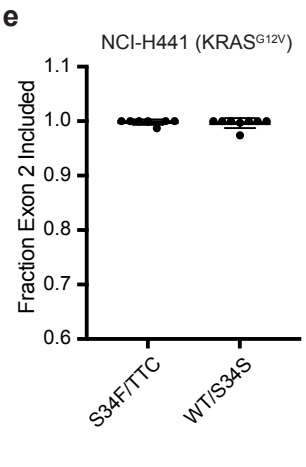
a

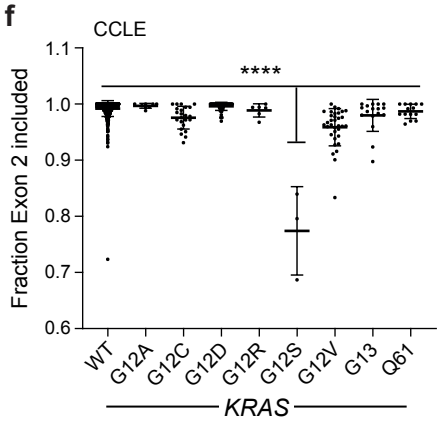
Gene	AS event	FDR	Delta mean PSI
AXL	Skipped Exon	7.38×10^{-8}	0.244
KRAS	Skipped Exon	5.85×10^{-6}	0.181
NCOR2	Skipped Exon	1.57×10^{-3}	0.146
KDM5A	Skipped Exon	2.86×10^{-3}	-0.143
DIS3	Skipped Exon	7.47×10^{-3}	0.182











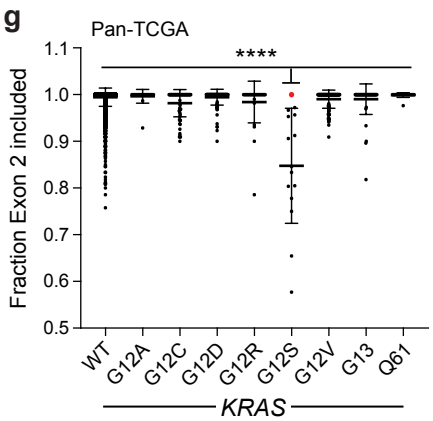


Fig. 1: Cancers with $KRAS^{G12S}$ mutations undergo *KRAS* exon 2 skipping which can be reversed by acquisition of $U2AF1^{S34F}$ mutations. a) Table of the top alternative splicing events unique to $U2AF1^{S34F}$ and $U2AF1^{S34F(TTC)}$ -mutant cells in known oncogenes and tumour suppressors as defined by their presence in OncoKB[24,25]. Table shows the gene, type of alternative splicing event, false discovery rate-corrected p value, and difference in percent spliced in (PSI) between $U2AF1^{S34F}$-mutant and parental A549 cells. b) Representative sashimi plots[61] for the alternative splicing of *KRAS* exon 2 in A549 cells with $U2AF1^{S34S}$ or $U2AF1^{S34F}$ mutations. Sashimi plot was generated using Integrative Genomics Viewer[52]. c) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion across parental A549 cells or those harbouring S34S, S34C, S34A, S34Y, S34F or S34F(TTC) mutations (n=4 clones for each). d) Quantification of fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion for parental A549 cells (n= 4 clones), $U2AF1^{S34F}$-mutant cells (n= 4 clones), or cells whose $U2AF1^{S34F}$ mutations have been reverted to wildtype by prime editing (n= 6 clones). e) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion for NCI-H441 cells with $U2AF1^{S34F}$ or $U2AF1^{S34F(TTC)}$ mutations (n= 7 clones) or wildtype U2AF1 or $U2AF1^{S34S}$ mutations (n=8 clones). f) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion for cell lines from the Cancer Cell Line Encyclopedia (CCLE)[31] with wildtype KRAS (n=835), $KRAS^{G12A}$ (n=8), $KRAS^{G12C}$ (n=23), $KRAS^{G12D}$ (n=53), $KRAS^{G12R}$ (n=6), $KRAS^{G12S}$ (n=3), $KRAS^{G12V}$ (n=34), $KRAS^{G13}$ (n=17), or $KRAS^{Q61}$ (n=14) mutations. g) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion for pan-cancer patient samples from The Cancer Genome Atlas (TCGA)[32] with wildtype KRAS (n=5570), $KRAS^{G12A}$ (n=24), $KRAS^{G12C}$ (n=66), $KRAS^{G12D}$ (n=120), $KRAS^{G12R}$ (n=30), $KRAS^{G12S}$ (n=15), $KRAS^{G12V}$ (n=109), $KRAS^{G13}$ (n=52), or $KRAS^{Q61}$ (n=22) mutations. Patient sample containing both a $U2AF1^{S34F}$ mutation and concurrent $KRAS^{G12S}$ mutation is marked in red.

Figure 2

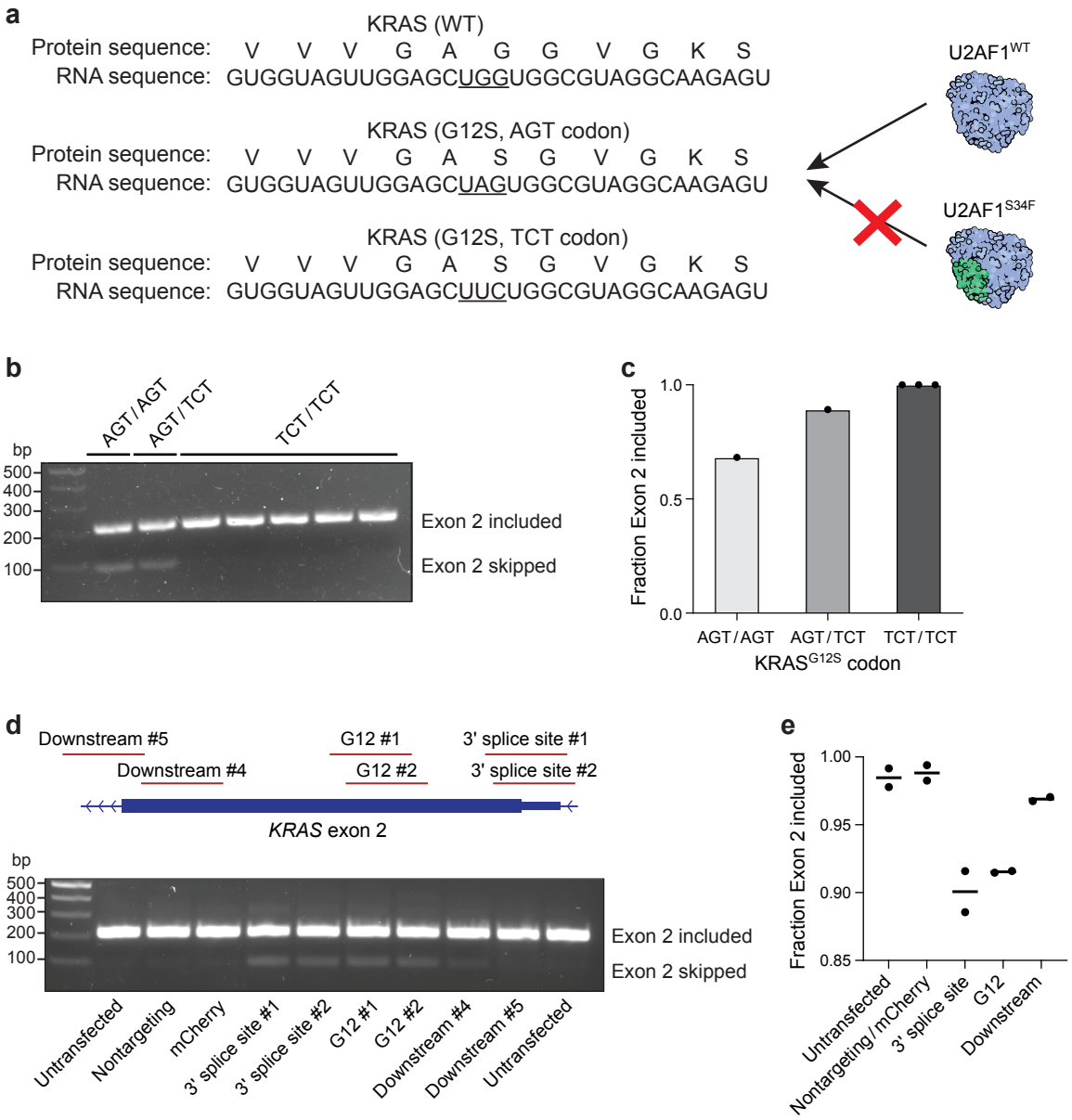


Fig. 2: The RNA sequence encoding KRAS^{G12S} creates a cryptic U2AF1 binding site which disrupts normal exon 2 inclusion. a) RNA and protein sequence for wildtype KRAS and KRAS^{G12S} indicating a cryptic U2AF1 binding site which is predicted to be bound by wildtype U2AF1 but not U2AF1^{S34F}. KRAS^{G12S} using an alternative codon for serine (TCT) does not resemble a U2AF1 binding site. b) RT-PCR detection of *KRAS* exon 2 skipping for A549 cells with KRAS^{G12S} mutations using the codons AGT/AGT, AGT/TCT or TCT/TCT (n= 5 clones). c) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion in A549 cells with AGT/AGT, AGT/TCT or TCT/TCT (n = 3 clones) codons. d) Schematic of dCasRx sgRNA locations and agarose gel for RT-PCR detection of *KRAS* exon 2 skipping in NCI-H2023 cells. Cells were transfected with dCasRx in addition to nontargeting sgRNA or sgRNAs targeting mCherry RNA, the 3' splice site of *KRAS* RNA (n=2), the G12 sequence of *KRAS* RNA (n=2), or downstream regions of *KRAS* RNA (n=2). e) Quantification of the fraction of *KRAS* exon 2 inclusion in untransfected NCI-H2023 cells, or following treatment with dCasRx alongside nontargeting or mCherry targeting sgRNA (n=2), 3' splice site sgRNA (n=2), G12 sgRNA (n=2) or downstream sgRNA (n=2).

Figure 3

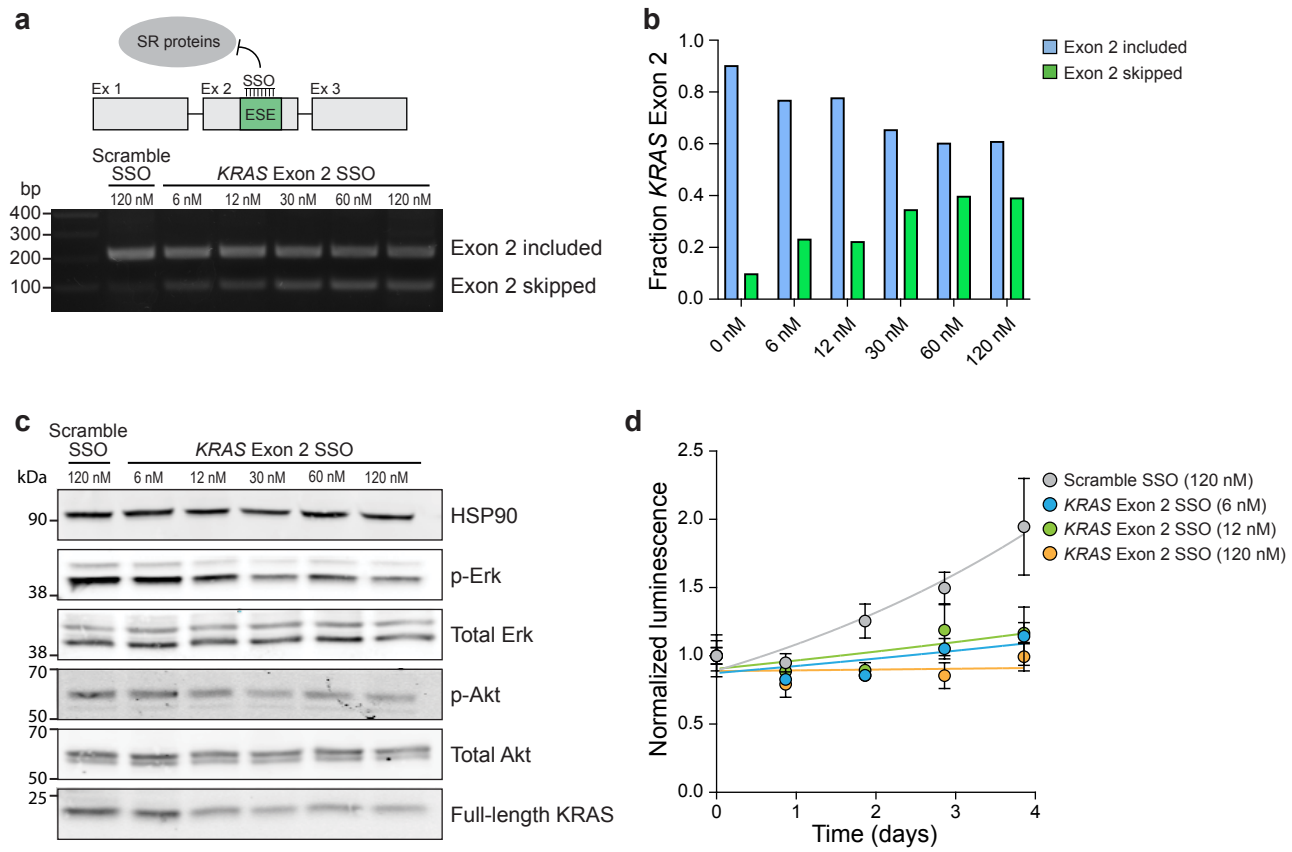


Fig. 3: *KRAS* exon 2 skipping limits MAPK signalling and cell growth. a) Schematic indicates the binding of splice switching oligonucleotides (SSOs) to exonic splicing enhancers (ESEs) in *KRAS* exon 2, blocking SR protein binding and inhibiting exon 2 inclusion. RT-PCR gel demonstrates the detection of *KRAS* exon 2 skipping for NCI-H23 cells treated with scramble SSO or increasing concentrations of *KRAS* exon 2 SSO. b) Quantification of the fraction of transcripts with *KRAS* exon 2 skipping or inclusion from a). c) Immunoblot analysis of HSP90, p-Erk, total Erk, p-Akt, total Akt, and full length KRAS protein in NCI-H23 cells treated with increasing concentrations of *KRAS* exon 2 skipping SSOs 3 days after treatment. d) Cell viability of NCI-H23 cells grown on ultra-low attachment plates and treated with increasing concentrations of *KRAS* exon 2 skipping SSOs over a period of 5 days as measured by CellTiter-Glo (n= 4 technical replicates at each time point).

Figure 4

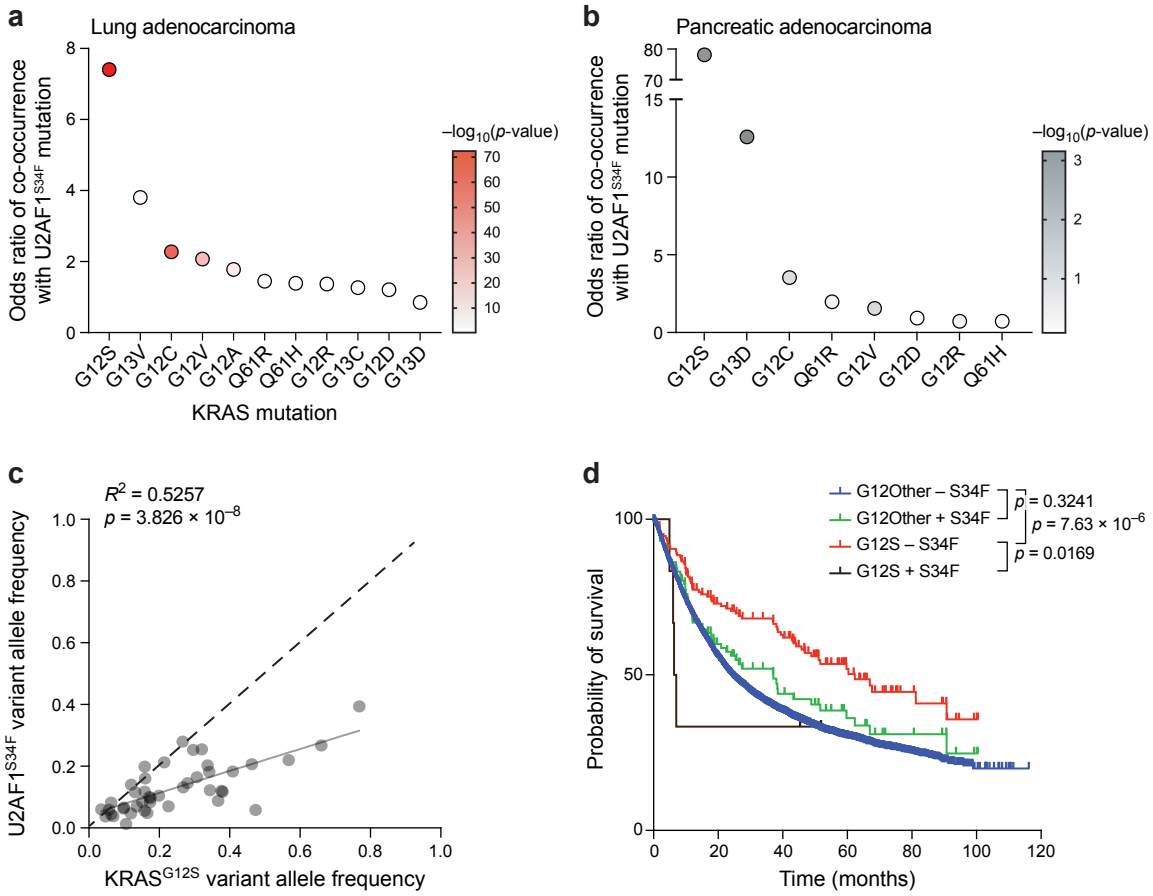
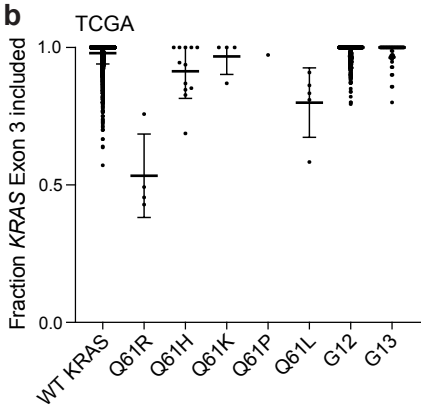


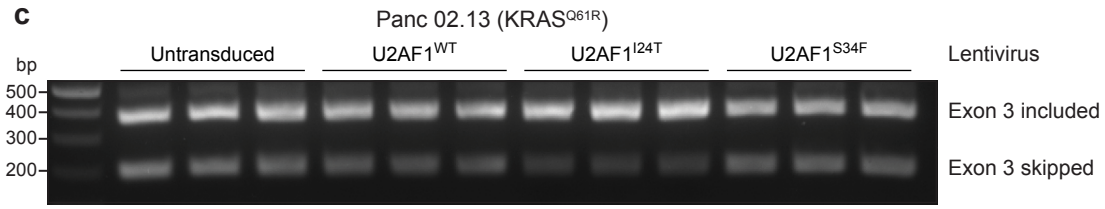
Fig. 4: U2AF1^{S34F} mutations co-occur secondary to KRAS^{G12S} and are associated with poor patient outcomes. a) Odds ratio of co-occurrence for U2AF1^{S34F} mutations and indicated KRAS mutations in lung adenocarcinoma patients using data from Foundation Medicine Inc. (n=62009 patients) and AACR Project GENIE (v14.0, n=14908 patients)[11]. b) Odds ratio of co-occurrence for U2AF1^{S34F} mutations and indicated KRAS mutations in pancreatic adenocarcinoma patients using data from AACR Project GENIE (v15.0, n=6528 patients)[11]. c) Correlation between the variant allele frequencies for KRAS^{G12S} and U2AF1^{S34F} in 43 patient samples from the Foundation Medicine Inc. dataset containing both mutations and lacking CNVs at either locus. Dashed line indicates a 1:1 ratio of KRAS^{G12S} and U2AF1^{S34F} allele frequencies. d) Overall patient survival from the MSK-CHORD study[40] comprising 5,777 patients with solid tumours with other KRAS^{G12} mutations in the absence (n=5,521) or presence (n=102) of U2AF1^{S34F} mutations, and patients with solid tumours with KRAS^{G12S} mutations in the absence (n=148) or presence (n=6) of U2AF1^{S34F} mutations.

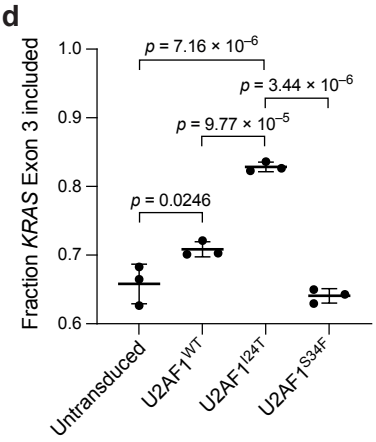
Figure 5

a

Mutation in TCGA	Co-occurring spling factor mutation	Delta mean PSI	Fisher exact test (FDR)
$KRAS^{Q61R}$	$U2AF1^{I24T}$	–0.446	1.45×10^{-12}
$KRAS^{G12S}$	$U2AF1^{S34F}$	–0.158	8.57×10^{-6}
$AXIN2^{G665Afs*24}$	$U2AF1^{G222\_G223del}$	–0.126	7.20×10^{-5}
$KRAS^{Q61L}$	$U2AF1^{I24T}$	–0.204	0.00132
$TCF7L2^{L200Sfs*25}$	$U2AF1^{G222\_G223del}$	–0.215	0.00923







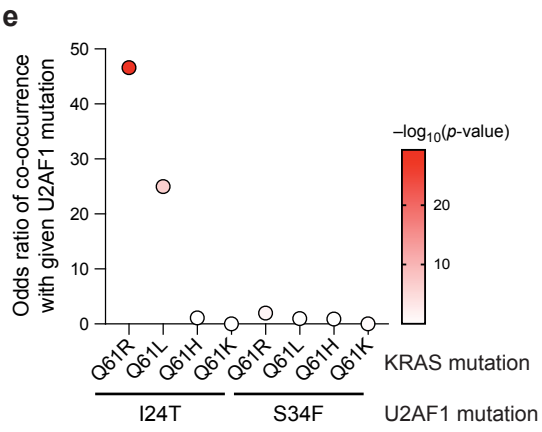


Fig. 5: $KRAS^{Q61R/L}$-mutant pancreatic cancers acquire secondary $U2AF1^{I24T}$ mutations that rescue inadvertent skipping of *KRAS* exon 3. a) Table of mutations associated with exon skipping events in TCGA data[32], and which co-occur with splicing factor mutations (data from AACR Project GENIE[11]). Threshold for exon skipping: delta mean PSI <-0.1. Threshold for co-occurrence: p value < 0.00001 as determined by Fisher's exact test. b) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 3 inclusion for pan-cancer patient samples from TCGA[32] with wildtype KRAS (n=6065), $KRAS^{Q61R}$ (n=4), $KRAS^{Q61H}$ (n=12), $KRAS^{Q61K}$ (n=4), $KRAS^{Q61P}$ (n=1), $KRAS^{Q61L}$ (n=5), $KRAS^{G12}$ (n=388), or $KRAS^{G13}$ (n=53) mutations. c) RT-PCR detection of *KRAS* exon 3 skipping for untransduced $KRAS^{Q61R}$-mutant Panc 02.13 cells (n= 3) or after delivery of lentivirus expressing wildtype U2AF1 (n=3), $U2AF1^{I24T}$ (n=3) or $U2AF1^{S34F}$ (n=3). d) Quantification of the fraction of *KRAS* exon 3 skipping in c). e) Odds ratio of co-occurrence for $U2AF1^{I24T}$ or $U2AF1^{S34F}$ mutations with $KRAS^{Q61R}$, $KRAS^{Q61L}$, $KRAS^{Q61H}$ or $KRAS^{Q61K}$ mutations in pancreatic cancers using data from Foundation Medicine Inc. (n=31,530 patients) and AACR Project GENIE (v16.1, n=8,304 patients)[11].

Figure 6

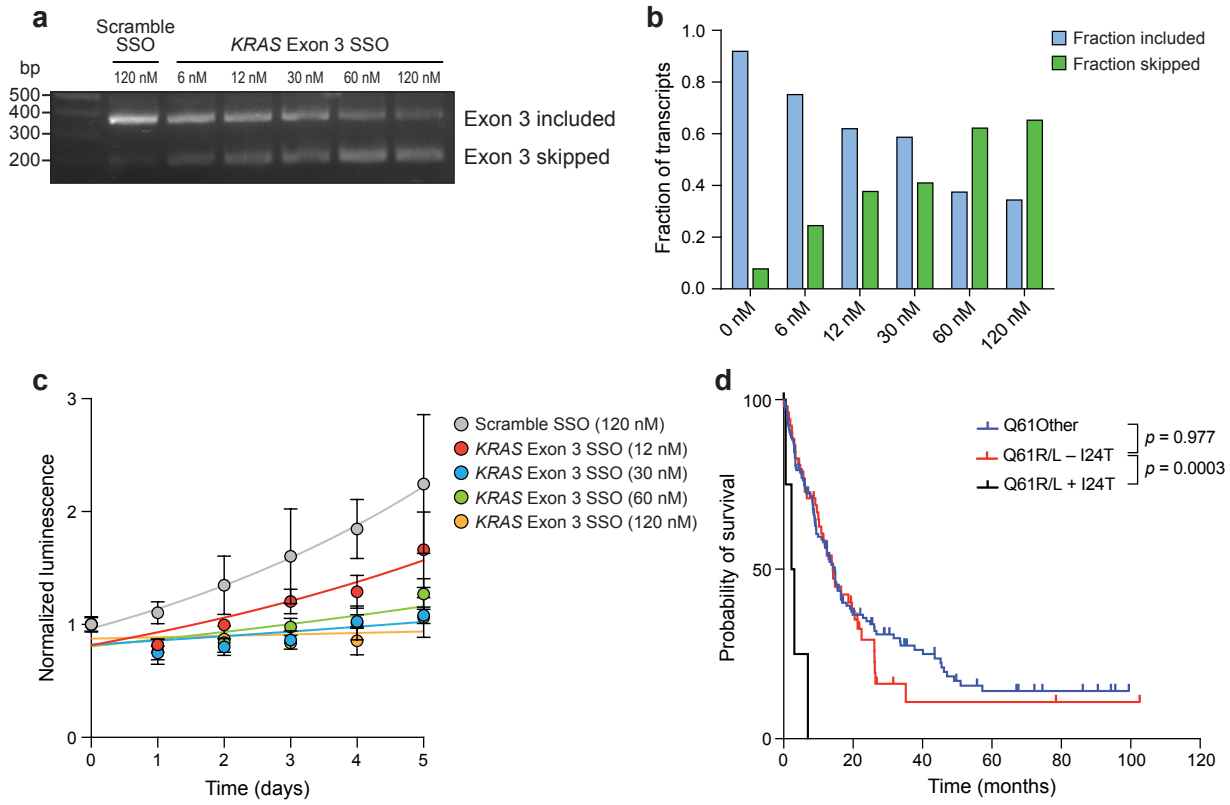
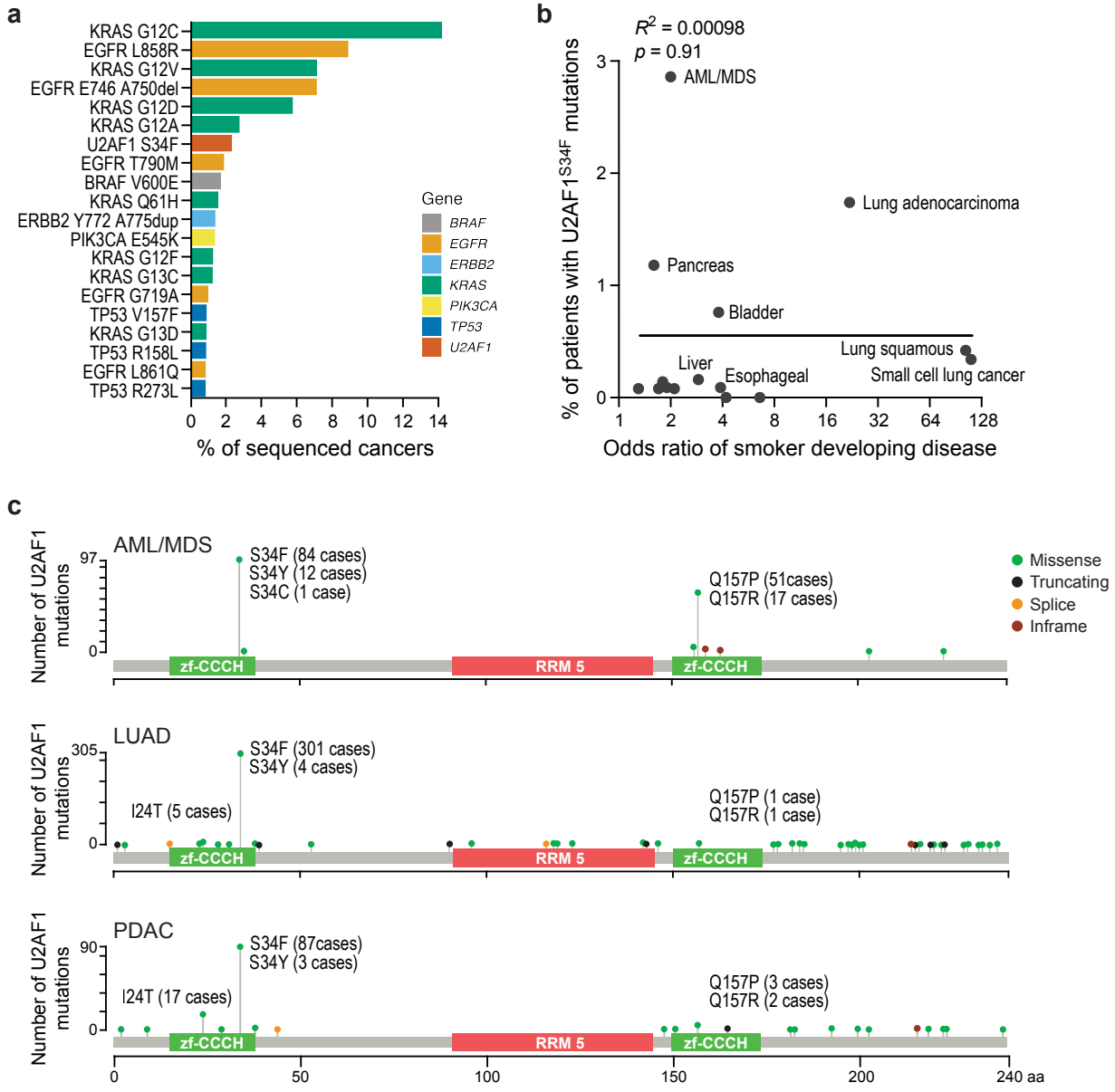


Fig. 6: *KRAS* exon 3 skipping leads to reduced cell growth, while acquisition of secondary U2AF1^{I24T} mutations in KRAS$^{Q61R/L}$-mutant pancreatic cancers is associated with reduced overall survival. a) RT-PCR detection of *KRAS* exon 3 skipping for NCI-H23 cells treated with scramble SSO or increasing concentrations of *KRAS* exon 3 SSO. b) Quantification of the fraction of transcripts with *KRAS* exon 3 skipping or inclusion from a). c) Cell viability of NCI-H23 cells grown on ultra-low attachment plates and treated with increasing concentrations of *KRAS* exon 3 skipping SSOs over a period of 5 days as measured by CellTiter-Glo (n= 4 technical replicates at each time point). d) Overall patient survival from the MSK-CHORD study comprising 192 pancreatic cancer patients with other KRAS^{Q61} mutations (n=135), or KRAS$^{Q61R/L}$ mutations in the absence (n=52) or presence (n=5) of U2AF1^{I24T} mutations.

Extended Data Fig. 1

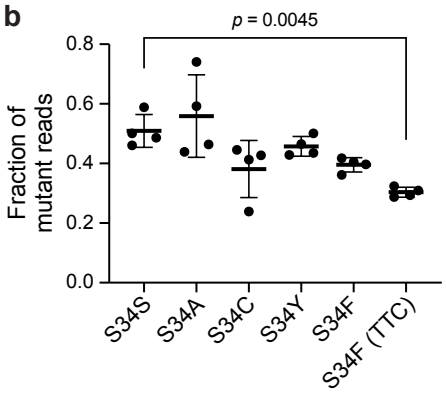


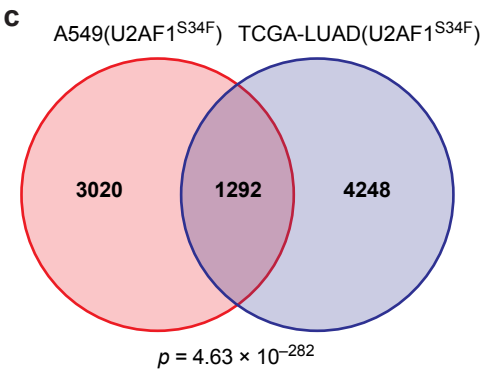
Extended Data Fig. 1: U2AF1^{S34F} mutations are over-represented in human lung adenocarcinomas. a) Top 20 most frequent hotspot mutations in human lung adenocarcinomas based on an analysis of data from AACR Project GENIE (v15.0)[11]. b) Comparison of the percentage of patients with U2AF1^{S34F} mutations across cancer types (AACR Project GENIE v15.0)[11] compared to the odds ratio of a smoker developing that disease[14]. c) Lollipop plot of mutations identified in the *U2AF1* gene in acute myeloid leukaemia/myelodysplastic syndrome (n=181 patients), pancreatic ductal adenocarcinoma (n= 131 patients) and lung adenocarcinoma (n=380 patients) based on data from AACR Project GENIE (v16.0)[11,26].

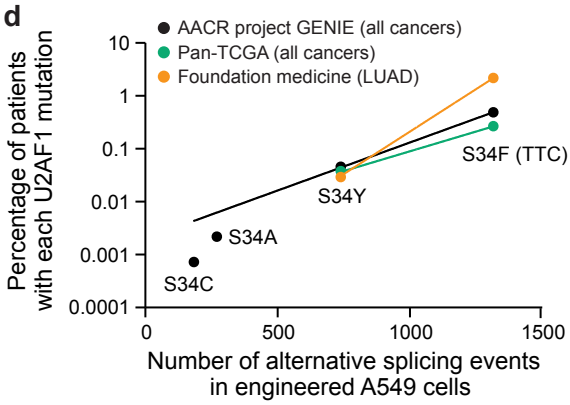
Extended Data Fig. 2

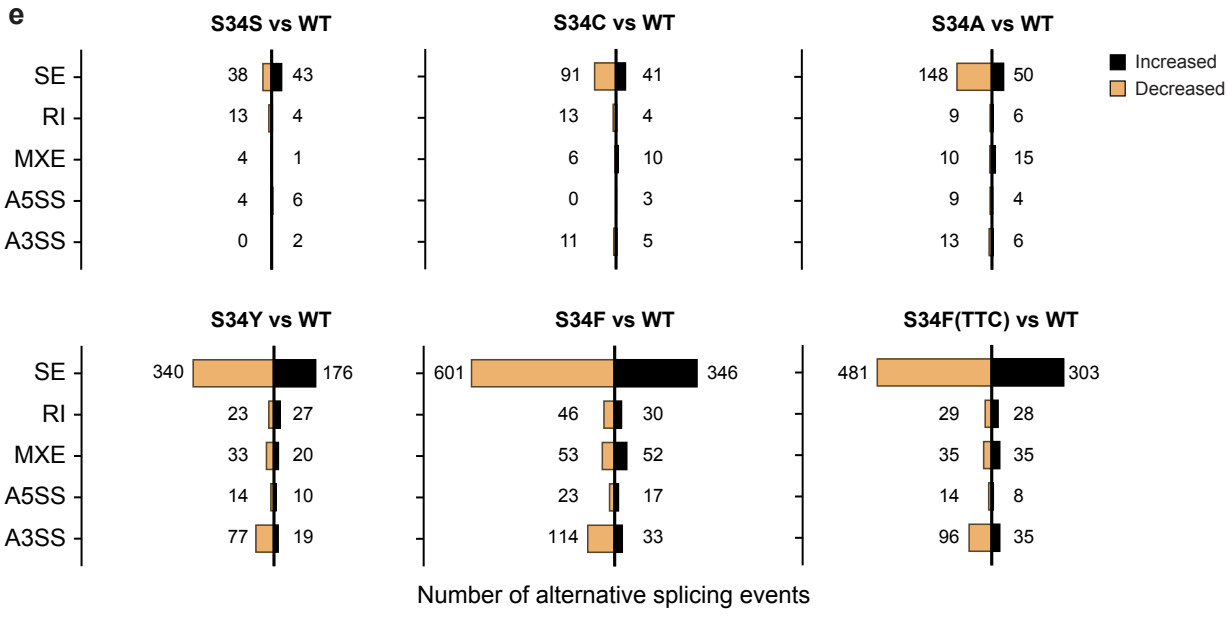
a

Engineered mutation in A549 cells	Number of AS events
S34S	115
S34C	184
S34A	270
S34Y	739
S34F (TTC)	1064
S34F	1315



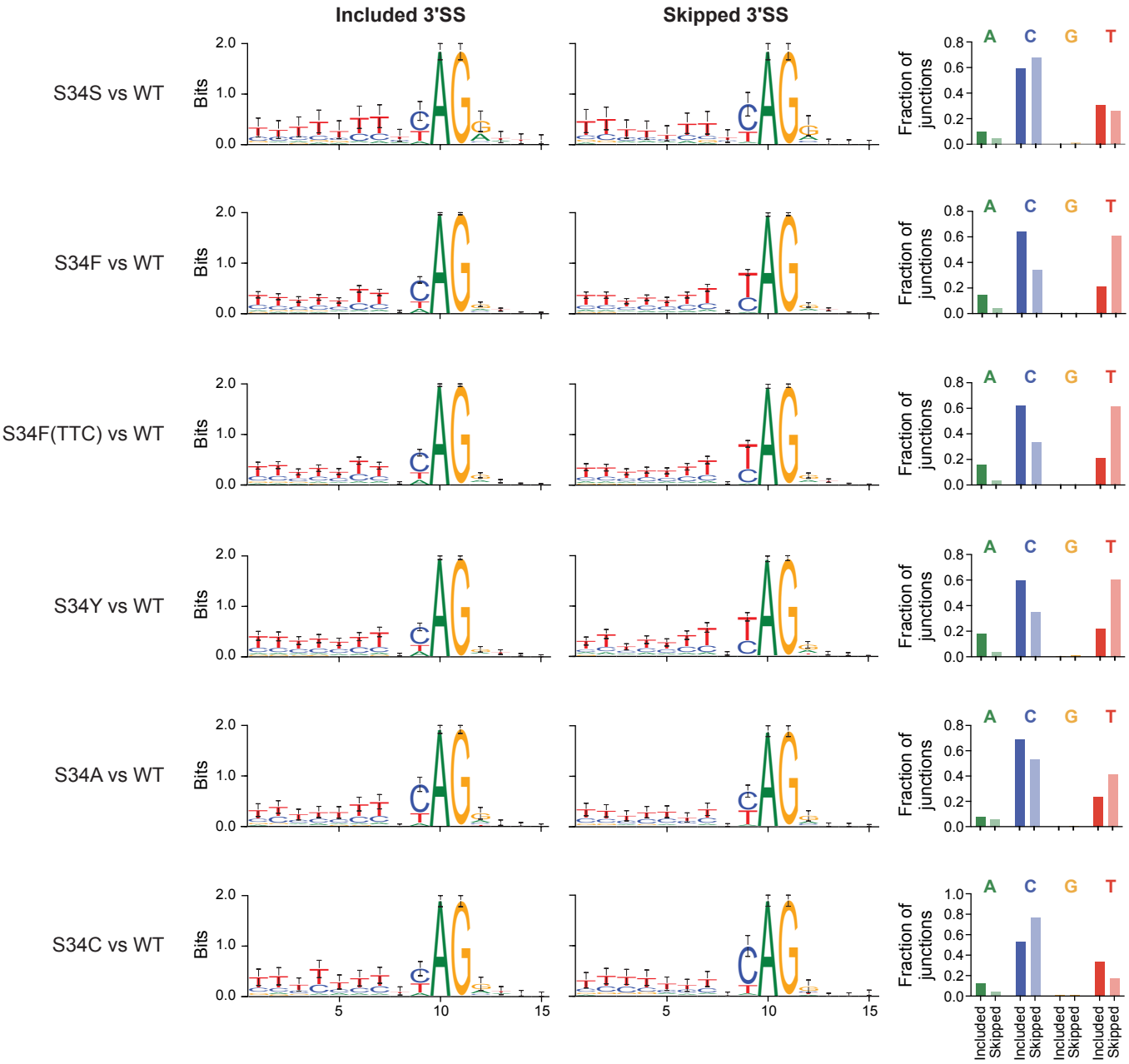






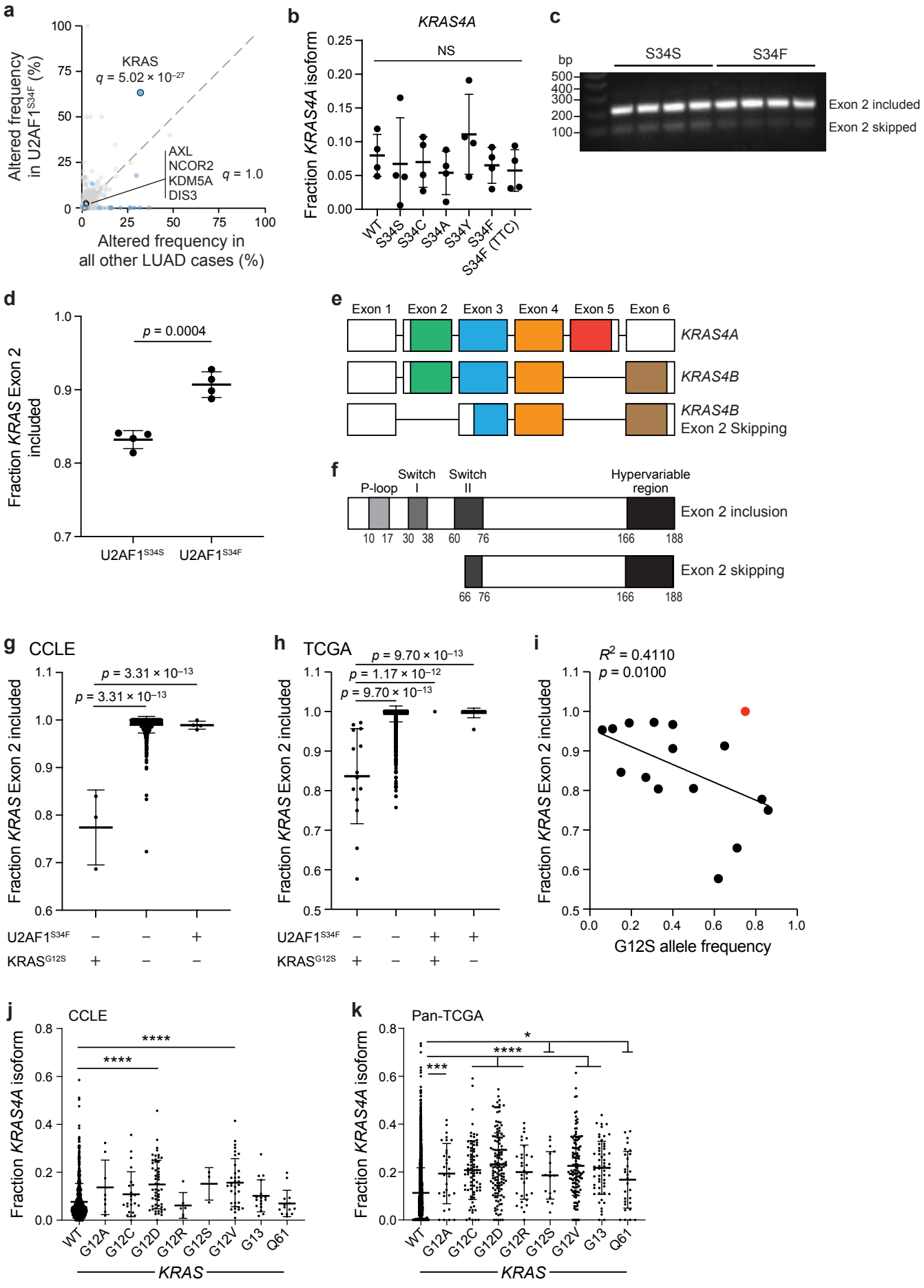
Extended Data Fig. 2: Alternative splicing analysis of A549 cells with varying U2AF1 amino acid substitutions. a) Number of significant alternative splicing events identified in the context of each engineered *U2AF1* mutation in A549 cells. Significant alternative splicing events were identified by comparing each variant to transcripts from parental A549 cells using rMATS (v4.1.0), and defined by an FDR q value < 0.05. b) Fraction of RNA-sequencing reads for each engineered U2AF1 variant in A549 cells (n=4 clones each). c) Overlap of significant alternative splicing events between mRNA from A549 cells with an engineered $U2AF1^{S34F}$ mutation and mRNA from $U2AF1^{S34F}$-mutant lung adenocarcinoma patient samples from TCGA[32]. d) Correlation between the number of alternative splicing events that occur in the presence of each amino acid substitution in A549 cells and the frequency of each substitution across all human cancers as determined by the AACR Project GENIE dataset (v15.1), TCGA Pan-Cancer Atlas, or Foundation Medicine Inc. LUAD data[11,22,23]. e) Number of significant alternative splicing events for A549 cells with S34S, S34C, S34A, S34Y, S34F or S34F(TTC) amino acid substitutions in U2AF1, compared to parental A549 cells as determined by rMATS (v4.1.0)[51]. Detected splicing events consist of 5 categories: skipped exons (SE), retained introns (RI), mutually exclusive exons (MXE), alternative 5' splice site usage (A5SS) and alternative 3' splice site usage (A3SS).

Extended Data Fig. 3



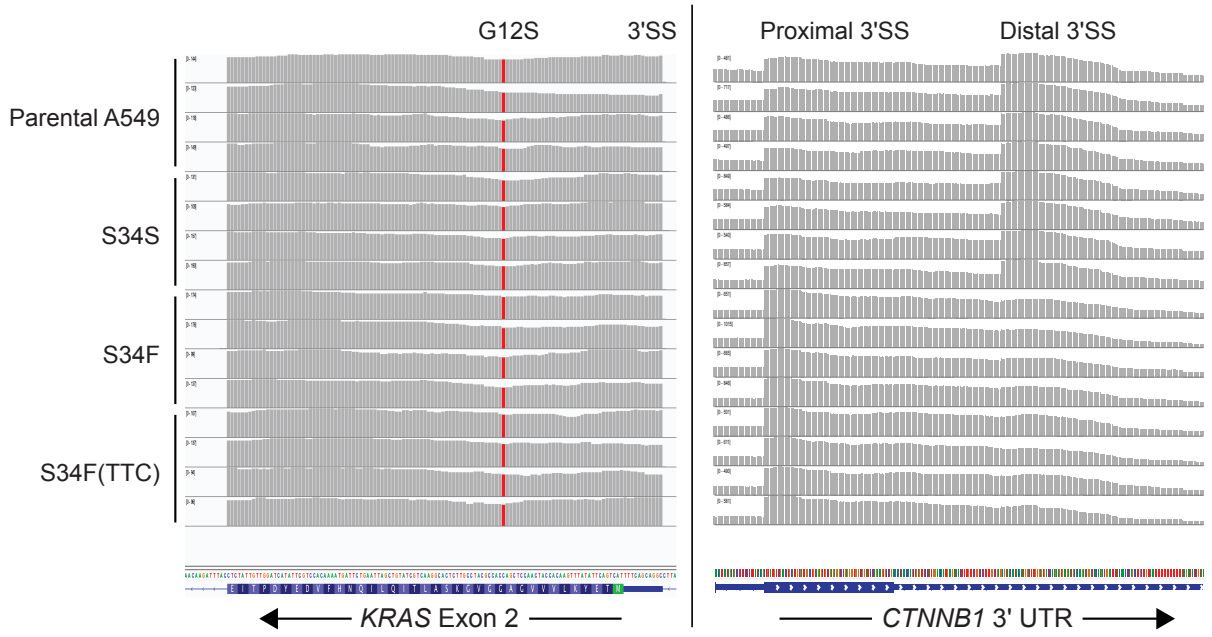
Extended Data Fig. 3: 3’ splice site preferences for all engineered U2AF1 variants. Splice site preferences were determined by examining the sequence surrounding the 3’ splice site in significantly included exons compared to the alternative skipped exon. This was performed for each U2AF1 variant compared to parental A549 cells, and WebLogos[54] were produced for each data set. The fraction of included (opaque bars) and skipped (partially translucent bars) junctions with a given base preceding the AG dinucleotide is displayed on the right.

Extended Data Fig. 4



Extended Data Fig. 4: $KRAS^{G12S}$-mutant cells undergo *KRAS* exon 2 skipping which can be reversed by $U2AF1^{S34F}$ mutations. a) Frequency of genetic alterations in $U2AF1^{S34F}$-mutant cancers compared to all lung adenocarcinomas as determined using the AACR Project GENIE dataset (v15.0)[11,26]. b) Quantification of the fraction of RNA-sequencing reads with *KRAS4A* isoform usage across parental A549 cells or those harbouring S34S, S34C, S34A, S34Y, S34F or S34F(TTC) mutations (n=4 clones for each). c) RT-PCR detection of *KRAS* exon 2 skipping in A549 cells with $U2AF1^{S34S}$ or $U2AF1^{S34F}$ mutations. d) Quantification of the fraction of *KRAS* transcripts with exon 2 inclusion from c). e) Diagram of *KRAS* gene structure showing exons used in *KRAS4A*, *KRAS4B*, and *KRAS4B* with exon 2 skipping. White segments represent untranslated regions. When exon 2 skipping occurs, this results in skipping of the translation start site and predicted use of an internal translation start site in exon 3. Diagram modified from Raso *et al.*[62] f) Diagram of KRAS protein structure with and without exon 2 skipping. When exon 2 skipping occurs, translation is predicted to begin at an internal translation site at amino acid 66. Diagram modified from Kwan *et al.*[63]. g) Additional analysis of CCLE data from Fig. 1f, comparing *KRAS* exon 2 skipping in samples with wildtype KRAS, and with $KRAS^{G12S}$ in the presence or absence of $U2AF1^{S34F}$. h) Additional analysis of TCGA data from Fig. 1g, comparing *KRAS* exon 2 skipping in samples with wildtype KRAS in the presence or absence of $U2AF1^{S34F}$, and with $KRAS^{G12S}$ in the presence or absence of $U2AF1^{S34F}$. i) Correlation between the fraction of *KRAS* exon 2 inclusion and variant allele frequency of $KRAS^{G12S}$ in pan-cancer patients from The Cancer Genome Atlas dataset[32]. Red sample indicates a $KRAS^{G12S}$-mutant sample which also contains a $U2AF1^{S34F}$ mutation, and was excluded from statistical analyses. j) Quantification of the fraction of RNA-Sequencing reads with *KRAS4A* isoform usage for cell lines from (CCLE)[31] with wildtype KRAS (n=836), $KRAS^{G12A}$ (n=8), $KRAS^{G12C}$ (n=23), $KRAS^{G12D}$ (n=53), $KRAS^{G12R}$ (n=6), $KRAS^{G12S}$ (n=3), $KRAS^{G12V}$ (n=34), $KRAS^{G13}$ (n=17), or $KRAS^{Q61}$ (n=14) mutations. k) Quantification of the fraction of RNA-Sequencing reads with *KRAS4A* isoform usage for pan-cancer patient samples from (TCGA)[32] with wildtype KRAS (n=7022), $KRAS^{G12A}$ (n=31), $KRAS^{G12C}$ (n=79), $KRAS^{G12D}$ (n=131), $KRAS^{G12R}$ (n=32), $KRAS^{G12S}$ (n=17), $KRAS^{G12V}$ (n=121), $KRAS^{G13}$ (n=56), or $KRAS^{Q61}$ (n=29) mutations.

Extended Data Fig. 5



Extended Data Fig. 5: *KRAS* undergoes an exon 2 skipping event, distinct from alternative 3' splice site selection. Read count distribution across *KRAS* exon 2 or *CTNNB1* 3'UTR in parental A549 cells or those with U2AF1^{S34S}, U2AF1^{S34F} or U2AF1$^{S34F(TTC)}$ mutations. Reads were visualized using Integrative Genomics Viewer[52]. Red nucleotide in *KRAS* exon 2 indicates the homozygous KRAS^{G12S} mutation present in A549 cells. Arrow indicates the directionality of the gene.

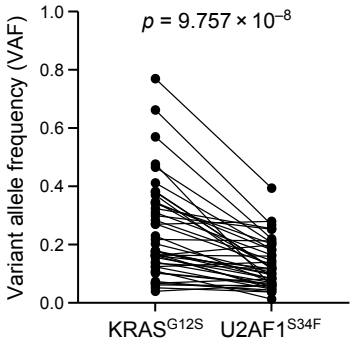
Extended Data Fig. 6

a

	KRAS^{G12S} +	KRAS^{G12S} −
U2AF1^{S34F} +	71 (13.15%)	1530 (2.00%)
U2AF1^{S34F} −	469 (86.85%)	74,847 (98.00%)

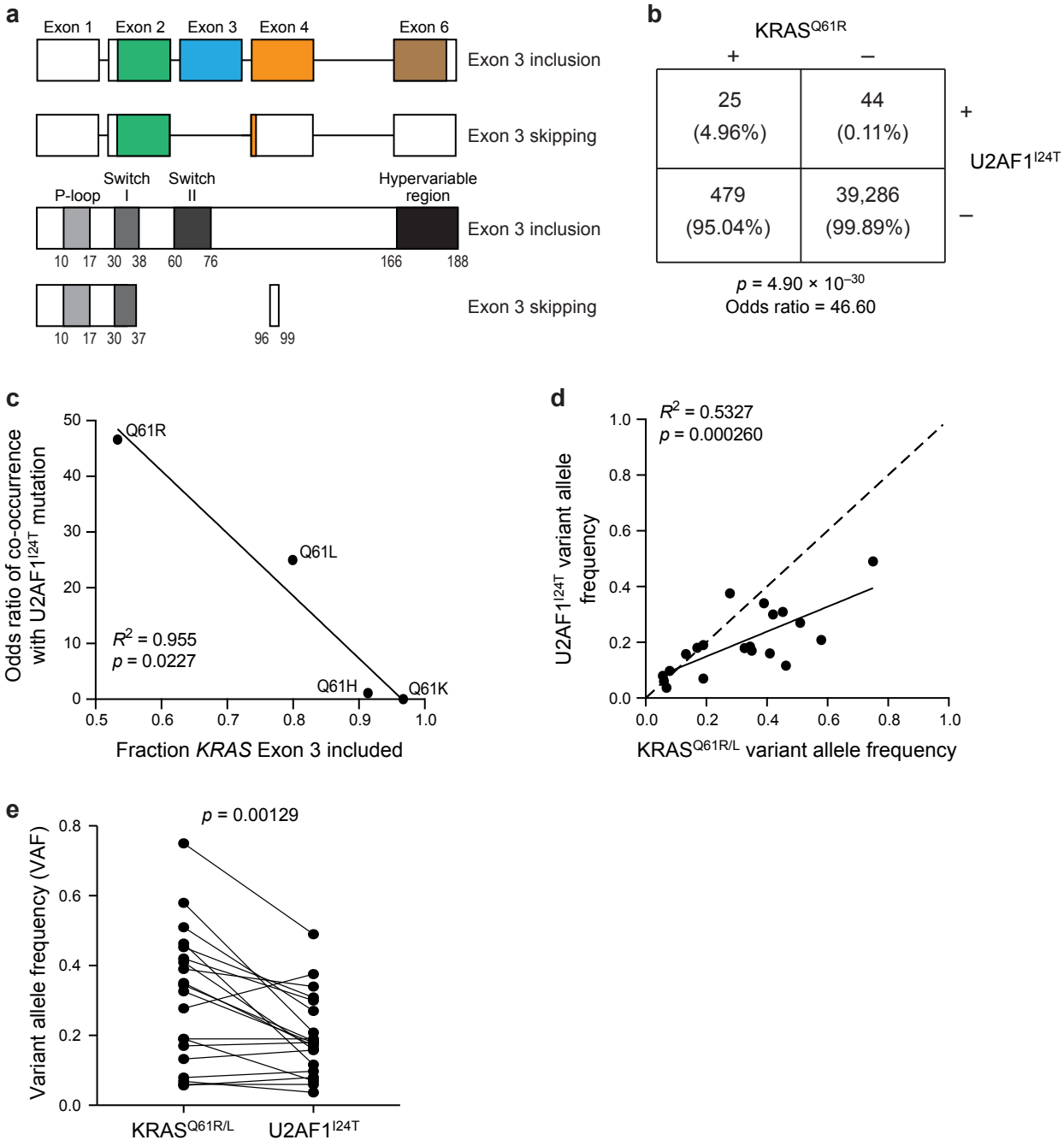
$p = 4.78 \times 10^{-73}$
Odds ratio = 7.406

b



Extended Data Fig. 6: U2AF1^{S34F} mutations are enriched as secondary mutations in KRAS^{G12S}-mutant lung adenocarcinomas. a) Contingency plot analysis of U2AF1^{S34F} and KRAS^{G12S} mutations in lung adenocarcinoma patients from Foundation Medicine Inc. (n=62009 patients) and AACR Project GENIE (v14.0, n=14908 patients)[11], including column percentages. b) Paired analysis of the variant allele frequency of KRAS^{G12S} and U2AF1^{S34F} across 43 patient samples from the Foundation Medicine Inc. dataset containing both mutations and lacking CNVs at either locus.

Extended Data Fig. 7



Extended Data Fig. 7: U2AF1^{I24T} mutations are enriched as secondary mutations in KRAS$^{Q61R/L}$-mutant pancreatic cancers. a) Diagram of *KRAS* gene and protein structure. Top shows exons used in *KRAS4B* with and without exon 3 skipping. White segments represent untranslated regions. When exon 3 skipping occurs, this results in early translation termination in exon 4. Bottom shows the KRAS protein structure with and without exon 3 skipping. When exon 3 skipping occurs, translation is predicted to terminate after amino acid 37, producing 3 incorrect amino acids before terminating. b) Contingency plot analysis of U2AF1^{I24T} and KRAS^{Q61R} mutations in pancreatic cancer patients from Foundation Medicine Inc. (n=31,530 patients) and AACR Project GENIE (v16.1, n=8,304 patients)[11], including column percentages. c) Odds ratio of co-occurrence of various KRAS^{Q61} mutations with U2AF1^{I24T} from Foundation Medicine Inc. (n=31,530 patients) and AACR Project GENIE (v16.1, n=8,304 patients)[11] compared to the mean fraction of *KRAS* exon 3 inclusion for each Q61 mutation as quantified from TCGA[32]. d) Correlation between the variant allele frequencies for KRAS$^{Q61R/L}$ and U2AF1^{I24T} in 20 patient samples from the Foundation Medicine Inc. and AACR Project GENIE (v16.1) datasets containing both mutations and lacking CNVs at either locus. Dashed line indicates a 1:1 ratio of KRAS^{G12S} and U2AF1^{S34F} allele frequencies. e) Paired analysis of the variant allele frequency of KRAS$^{Q61R/L}$ and U2AF1^{I24T} across 20 patient samples from the Foundation Medicine Inc. and AACR Project GENIE (v16.1) datasets containing both mutations and lacking CNVs at either locus.