U2AF1 mutations rescue deleterious exon skipping induced by KRAS mutations

David M. Walter[1,2,3], Katherine Cho[1,2], Smruthy Sivakumar[4], Daniel Denney[1,2], Iris T.-H. Lee[1,2], Anders B. Dohlman[1,2,3], Jakob M. Heinz[1,2,5], 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
5. Department of Biomedical Informatics, Harvard Medical School, 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}. 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]. The splicing factor U2AF1 binds to an AG dinucleotide at the 3' splice site of AG-dependent introns and recruits the U2 snRNP to these sites[2-4]. Mutations in *U2AF1* are found in acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS)[1], lung adenocarcinoma (LUAD)[5] and pancreatic ductal adenocarcinoma (PDAC)[6], and 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 advantages conferred by U2AF1^{S34F} and U2AF1^{I24T} mutations remain poorly understood.

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}. As homozygous U2AF1 mutations are lethal to cells[22,23], we used TwinPE to generate 6 heterozygous U2AF1 mutations in A549 cells. These consisted of 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 analyzed the alternative splicing events associated with each amino acid substitution and found that these were highest in cells harboring 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[24] (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)[11,25,26]. 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 tumor suppressors[27,28]. We identified splicing alterations in multiple genes of interest including *AXL, KRAS* and *NCOR2* (Fig. 1a). Combining mutation data from GENIE[11,29] 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[30,31], though both isoforms are capable of transforming cells[32]. 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,e). Interestingly, exon 2 of *KRAS* harbors the protein's translation start site as exon 1 is untranslated (Extended Data Fig. 4f). 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. 4g). Using long-read RNA-sequencing, we observed *KRAS* exon 2 skipping in both *KRAS4A* and *KRAS4B* transcripts, as well as in combination with previously described alternative 5' and 3'UTRs[33] (Extended Data Fig. 5).

To further probe the relationship between *KRAS* exon 2 skipping and U2AF1^{S34F}, we "backedited" 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[34], we examined the *KRAS* locus of the two cell lines, and noticed that while A549 cells harbor 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)[35]. 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. 6a). The same held true in 6008 cancers with reported *KRAS* splicing data from The Cancer Genome Atlas (TCGA)[24], where *KRAS* exon 2 skipping was minimal except in cancers with KRAS^{G12S} mutations (Fig. 1g, Extended Data Fig. 6b). Furthermore, we found a negative correlation between variant allele frequency (VAF) of KRAS^{G12S} and exon 2 inclusion, suggesting that tumors with a higher fraction of KRAS^{G12S}-mutant cells have a higher degree of exon 2 skipping (Extended Data Fig. 6c). The TCGA dataset also contained a KRAS^{G12S}-mutant case which had 100% *KRAS* exon 2 inclusion (Fig. 1g, Extended Data Fig. 6c, 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. 6d). 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. 6e). These data support previous findings that *KRAS* mutations in LUAD are associated with increased levels of *KRAS4A* usage[36].

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

153 We observed that the RNA sequence of the KRAS^{G12S} mutation resembles a U2AF1
154 binding site, creating a UAG trinucleotide predicted to be bound by wildtype U2AF1 protein but
155 not U2AF1^{S34F} (Fig. 2a, Extended Data Fig. 7a)[8,37]. Previous work has reported that expression
156 of KRAS^{G12V} or KRAS^{Q61H} leads to decreased phosphorylation of SR splicing factors and global
157 changes in alternative splicing[38]. Therefore, we wondered if the *KRAS* exon 2 skipping we
158 observed was due to the KRAS^{G12S} amino acid substitution itself, or due to the creation of a
159 cryptic U2AF1 binding site in the RNA. To test this, we performed prime editing on A549
160 U2AF1^{S34S} cells to mutate the homozygous KRAS^{G12S} sequence to another codon for serine
161 (TCT) that does not resemble a U2AF1 binding site (Fig. 2a). RT-PCR and RNA-sequencing
162 analysis of exon 2 skipping found that exon skipping was completely abolished in KRAS^{G12S}-
163 mutant cells with a TCT codon (Fig. 2b,c). This demonstrates that the creation of a novel cryptic
164 U2AF1 binding site is required for *KRAS* exon 2 skipping, as opposed to the G12S amino acid
165 substitution itself.
166 As it seemed surprising that a cryptic U2AF1 binding site would cause exon skipping
167 rather than alternative 3' splice site usage, we performed a thorough analysis of the RNA-Seq
168 reads using Integrative Genomics Viewer (IGV 2.16.0)[39]. Alternative 3' splice site usage would
169 correlate with a stepwise increase in read coverage near the G12 RNA sequence, similar to the
170 3'UTR of *CTNNB1*, where U2AF1^{S34F} leads to decreased usage of the distal 3' splice site[8]
171 (Extended Data Fig. 7b). However, we observed relatively constant read coverage across *KRAS*
172 exon 2 regardless of U2AF1 mutational status, confirming that KRAS^{G12S} leads to exon 2
173 skipping as opposed to alternative 3' splice site usage (Extended Data Fig. 7b).
174 As the 3' splice site upstream of *KRAS* exon 2 is only 45 bp from the cryptic binding site
175 at G12S, we hypothesized that exon 2 skipping might arise from steric hindrance caused by
176 wildtype U2AF1 binding and recruiting the spliceosome complex to this alternative site. To test
177 whether such a process was possible, we targeted a catalytically inactive version of the RNA
178 binding CasRX protein (dCasRx)[40] to various locations spanning *KRAS* exon 2 in NCI-H2023
179 cells with wildtype *KRAS* (Extended Data Table 3). Indeed, binding of dCasRx to the canonical
180 3' splice site leads to *KRAS* exon 2 skipping, likely due to interference with normal U2AF1
181 binding (Fig. 2d,e). Remarkably, directing dCasRx to the G12 sequence also resulted in
182 significant levels of exon 2 skipping, while binding ~60 bases downstream lessened the impact
183 on exon 2 skipping, likely due to decreased steric hindrance further from the upstream 3' splice
184 site (Fig. 2d,e). This suggests that aberrant U2AF1 binding to the G12S codon could lead to
185 *KRAS* exon 2 skipping due to steric hindrance with the proximal 3' splice site.
186 To directly assess the impact of *KRAS* sequence variation on splicing, we chose to use
187 a simplified genomic system: a splicing reporter minigene using the pSpliceExpress system[41,42].
188 Here we incorporated exon 2 of *KRAS* and 500 bp of each flanking intron between 2
189 constitutively expressed exons of the rat insulin gene. We found that the introduction of G12S
190 mutations resulted in significant skipping of *KRAS* exon 2, while other mutations at G12 had
191 minimal to no effect, demonstrating that KRAS^{G12S} mutations directly lead to exon skipping (Fig.
192 2f, Extended Data Fig. 7c). As observed in our edited A549 cells, KRAS^{G12S} using a TCT codon
193 resulted in no exon skipping (Fig. 2f, Extended Data Fig. 7c).
194 One mechanistic hypothesis for *KRAS* exon 2 skipping driven by KRAS^{G12S} mutation is
195 the introduction of a novel U2AF1 binding site containing an AG dinucleotide. If this hypothesis
196 is correct, then introducing other AG-containing U2AF1 binding sites could similarly cause exon
197 skipping. Using our splicing reporter minigene, we generated a series of nucleotide substitutions
198 upstream of or near the G12 codon of *KRAS*. We observed significant levels of *KRAS* exon 2
199 skipping for each variant that introduces an AG dinucleotide and less exon 2 skipping for any
200 variant that did not introduce an AG dinucleotide (Fig. 2g, Extended Data Fig. 7d). Additionally,
201 exon skipping was less pronounced for a U2AF1 binding site created further downstream from
202 the 3'SS (Fig. 2g, Extended Data Fig. 7d). Taken together, these results suggest that KRAS^{G12S}
203 mutation promotes exon skipping via the creation of a new AG-containing U2AF1 binding site.

204
205 ***KRAS* exon 2 skipping leads to reduced MAPK signaling and cell growth**
206
207 While the effects of *KRAS* exon 2 skipping have not been studied, a similar event has
208 been identified in *HRAS* whereby exon 2 skipping is thought to lead to decreased signaling in
209 the context of Costello syndrome[34]. Therefore, we hypothesized that *KRAS* exon 2 skipping
210 might lead to decreased MAPK signaling, while acquisition of U2AF1^{S34F} mutations would
211 enhance KRAS signaling by restoring exon 2 inclusion. We used flow cytometry to measure
212 phosphorylation of Erk as a readout for MAPK signaling and found that that the acquisition of
213 U2AF1^{S34F} mutations increased p-Erk levels by ~37% in KRAS^{G12S}-mutant A549 cells (Fig.
214 3a,b). Similarly, A549 cells in which the homozygous KRAS^{G12S} mutation was mutated to a TCT
215 codon had a ~64% increase in p-Erk levels (Extended Data Fig. 8a,b). The higher levels of p-
216 Erk in cells harboring the KRAS$^{G12S(TCT)}$ mutation compared to those with U2AF1^{S34F} may be in
217 part due to "locking in" *KRAS* to a state of 100% exon 2 inclusion through the removal of the
218 cryptic U2AF1 binding site.
219 To further test the impact of *KRAS* exon 2 inclusion on cell growth and signaling, we
220 employed splice switching oligonucleotides (SSOs) targeting *KRAS* exon 2 (Extended Data
221 Table 4)[34,43]. Treatment of NCI-H23 cells (KRAS^{G12C}) with increasing concentrations of exon-
222 skipping SSOs led to increased levels of exon 2 skipping (Fig. 3c,d). While canonical KRAS
223 signaling proceeds through a RAF/MEK/ERK axis, there is evidence that KRAS-mediated
224 PI3K/AKT/mTOR signaling is also important for tumor initiation and maintenance[44]. We found
225 that higher levels of *KRAS* exon 2 skipping led to decreased signaling through both pathways as
226 indicated by reduced levels of p-Erk and p-Akt, as well as full length KRAS protein (Fig. 3e). We
227 also treated cells with increasing levels of exon-skipping SSOs and grew them on ultra-low
228 attachment plates, where we observed even small increases in *KRAS* exon 2 skipping led to
229 decreased cell growth (Fig. 3f). These findings indicate that *KRAS* exon 2 skipping limits KRAS
230 signaling and is detrimental to cell proliferation.
231
232 **There is positive selection for U2AF1^{S34F} mutations in KRAS^{G12S}-mutant cancers**
233
234 Our data show that KRAS^{G12S} mutations result in exon 2 skipping which can be reversed
235 by U2AF1^{S34F} mutations, and that exon 2 skipping leads to reduced MAPK signaling and cell
236 growth. Therefore, we hypothesized that U2AF1^{S34F} mutations would be favored in KRAS^{G12S}-
237 mutant cancers as a means of driving KRAS signaling and tumor growth. To determine whether
238 this was the case, we examined a set of 76,917 LUAD cases including 62,009 from Foundation
239 Medicine and 14,908 from AACR Project GENIE and calculated the odds ratio of co-occurrence
240 between U2AF1^{S34F} and varying *KRAS* mutations[11,26]. Remarkably, we found that KRAS^{G12S}
241 mutation had an odds ratio of 7.4 for U2AF1^{S34F} mutations, far higher than any other *KRAS*
242 mutation, with a p value of 4.8×10^{-73} (Fig. 3g, Extended Data Fig. 8c). Interestingly, KRAS^{G12C}
243 and KRAS^{G12V} mutations were also significantly enriched, though to a lesser degree, and
244 displayed intermediate levels of exon 2 skipping in our splicing reporter assay. In addition, we
245 found significant co-occurrence of U2AF1^{S34F} mutations and KRAS^{G12S} mutations in pancreatic
246 adenocarcinoma cases (Extended Data Fig. 8d).
247 U2AF1^{S34F} has been described as a truncal mutation in both hematological malignancies
248 and LUAD[15,45,46]. However, as U2AF1^{S34F} rescues cryptic exon 2 skipping in KRAS^{G12S}-mutant
249 cancers, we could imagine that U2AF1^{S34F} mutations might occur as secondary events after the
250 acquisition of a KRAS^{G12S} mutation. To test this possibility, we analyzed the variant allele
251 frequencies (VAF) of KRAS^{G12S} and U2AF1^{S34F} in 43 LUAD cases from the Foundation Medicine
252 Inc. dataset which harbored both mutations and lacked copy number variations (CNVs) at either
253 locus. We found a strong positive correlation between the two mutations, however the VAF for
254 KRAS^{G12S} was significantly higher than the VAF for U2AF1^{S34F} across the tumors, suggesting

that U2AF1^{S34F} occurs as a secondary mutation in KRAS^{G12S}-mutant LUADs (Fig. 3h, Extended Data Fig. 8e).

Spliceosome inhibitors have gained recent interest as potential clinical targets, with a particular focus on compounds targeting the splicing factor SF3B1. It is possible that inhibition of SF3B1 may disrupt normal splicing of *KRAS* and limit MAPK signaling. To test this possibility, we treated NCI-H2023 cells with increasing concentrations of SF3B1 inhibitor Pladienolide B[47,48] and tested its impact on *KRAS* exon 2 skipping and MAPK signaling as measured by phosphorylation of Erk. We found that treatment with Pladienolide B led to increased *KRAS* exon 2 skipping (Extended Data Fig. 8f) and decreased p-Erk after 20 hours of treatment (Extended Data Fig. 8g,h).

KRAS$^{Q61R/Q61L}$ mutations lead to *KRAS* exon 3 skipping and are 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. 4a). Interestingly, the strongest relationship was between another pair of *KRAS* and *U2AF1* mutations: KRAS^{Q61R} and U2AF1^{I24T} (Fig. 4a).

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. 9a). 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[24]. 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. 4b). Interestingly, work from Kobayashi *et al*. found that KRAS^{Q61K} mutations lead to either complete exon 3 skipping or skipping of 112 bp of the exon[49]. However, this is rescued by the acquisition of adjacent silent mutations, explaining the absence of exon 3 skipping in Q61K-mutant samples.

To further interrogate the impact of Q61 mutations on *KRAS* exon 3 inclusion, we again employed the pSpliceExpress splicing reporter[41,42]. By incorporating exon 3 of *KRAS* and 500 bp each of flanking intronic sequence into the vector, we found that Q61R and Q61L mutations led to dramatic exon 3 skipping, while Q61P and Q61H mutations resembled wildtype Exon 3 (Fig. 4c and Extended Data Fig. 9b). Interestingly, when we introduced a Q61K mutation, close to 100% of transcripts skipped 112 bp of exon 3, a splicing event previously identified by Kobayashi *et al.*[49]. However, as described[49], a silent mutation in G60 restored exon 3 inclusion to similar levels as the wildtype reporter.

I24 of U2AF1 is located between its first and second RNA-binding zinc finger domains[50], consistent with an impact on RNA binding specificity. Indeed, a study of rare spliceosomal mutations found that U2AF1^{I24T} leads to preferential inclusion of exons containing CAGG sequences at the 3' splice site[51]. This CAGG motif is the exact sequence of the 3' splice site of *KRAS* exon 3, suggesting that the acquisition of U2AF1^{I24T} mutations may restore inclusion of *KRAS* exon 3 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[52], 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. 4d,e). Interestingly, we also observed that over-expression of wildtype U2AF1 resulted in a small but statistically significant increase in *KRAS* exon 3 inclusion (Fig. 4d,e).

306
307 ***KRAS* exon 3 skipping is detrimental to cell growth, while the acquisition of secondary**
308 **U2AF1^{I24T} mutations is selected for in KRAS$^{Q61R/L}$-mutant cases**
309
310 We previously observed that *KRAS* exon 2 skipping restricts cell growth and
311 hypothesized that the same may be true for exon 3. To test whether *KRAS* exon 3 skipping
312 hindered cell growth as expected, we employed SSOs targeting exon 3. As before, delivery of
313 increasing amounts of exon 3 SSO led to increased exon skipping as quantified by RT-PCR
314 (Fig. 5a,b). Furthermore, we found that KRAS^{G12C}-mutant NCI-H23 cells were sensitive to *KRAS*
315 exon 3 skipping, with relatively small increases in skipping leading to significant reductions in
316 cell growth (Fig. 5c).
317 As a result, we would expect enrichment of secondary U2AF1^{I24T} mutations in
318 KRAS$^{Q61R/L}$-mutant cancers. Patient data from AACR Project GENIE reveals that U2AF1^{I24T}
319 mutations are ~4 times more likely to occur in pancreatic cancer compared to other cancer
320 types[11]. Therefore, we interrogated genomic data of 31,530 pancreatic cancer cases from
321 Foundation Medicine Inc. and 8,304 cases from AACR Project GENIE (v16.1) to determine the
322 odds ratio of co-occurrence for U2AF1^{I24T} and various mutations at Q61. Remarkably, U2AF1^{I24T}
323 and KRAS^{Q61R} had an odds ratio for co-occurrence of 46.6, while an odds ratio of 25 was
324 calculated for the co-occurrence of U2AF1^{I24T} and KRAS^{Q61L} (Fig. 5d, Extended Data Fig. 10a).
325 These values were dramatically higher than for the co-occurrence of any other Q61 mutation
326 with either U2AF1^{I24T} or U2AF1^{S34F}. Additionally, there was a strong correlation between the
327 magnitude of exon 3 skipping caused by each Q61 mutation and the likelihood of co-occurrence
328 with U2AF1^{I24T}, suggesting that U2AF1^{I24T} mutations are selected for as a means of rescuing
329 normal *KRAS* splicing (Extended Data Fig. 10b).
330 To further determine whether U2AF1^{I24T} mutations occur secondary to KRAS^{Q61}
331 mutations, we analyzed the VAF of KRAS$^{Q61R/L}$ and U2AF1^{I24T} mutations in 20 pancreatic cancer
332 samples from Foundation Medicine Inc. and AACR Project GENIE which harbored both
333 mutations and had no copy number alterations at the *KRAS* or *U2AF1* loci. Similar to KRAS^{G12S}
334 and U2AF1^{S34F}, there was a positive correlation between the two mutations, and the VAF for
335 KRAS$^{Q61R/L}$ was significantly higher than the VAF for U2AF1^{I24T}, suggesting that U2AF1^{I24T}
336 occurs as a secondary mutation in KRAS^{Q61}-mutant pancreatic cancer (Extended Data Fig.
337 10c,d).
338 In total, these results provide a second example of splicing factor mutations that occur
339 as a means of promoting KRAS signaling.
340
341
342 **DISCUSSION**
343
344 The work presented here describes a mechanism for splicing factor mutation in cancer:
345 certain oncogenic mutations lead to splicing defects that limit their impact on pathogenesis,
346 leading to cascading selection whereby secondary mutations in splicing factors that restore
347 normal splicing are positively selected.
348 Although U2AF1 mutations have been implicated in several biological pathways that are
349 important in cancer[7,15,18,19], it has proved difficult to discern the specific selective advantage that
350 these mutations confer in solid tumors. By using a combination of cell modelling and
351 computational investigations of somatic mutations that frequently occurred with mutations in
352 *U2AF1*, we discovered that mutant U2AF1 can regulate the alternative splicing of the *KRAS*
353 oncogene. Our findings suggest that as cancer cells acquire oncogenic KRAS^{G12S} or KRAS$^{Q61R/L}$
354 mutations, inadvertent exon skipping in *KRAS* leads to reduced KRAS signaling and cell growth.
355 However, the acquisition of U2AF1^{S34F} or U2AF1^{I24T} mutations respectively, reduces exon
356 skipping, driving increased expression of the full-length *KRAS* transcript and increased MAPK

signaling and 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.

Our findings describe a novel mechanism by which mutant U2AF1 may act in cancer: 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.

We observe a clear enrichment of U2AF1^{S34F} and U2AF1^{I24T} mutations in KRAS^{G12S} and KRAS$^{Q61R/L}$-mutant cancers respectively. Additionally, a recent study has found that U2AF1^{S34F} mutations are able to promote cell proliferation in response to overexpression of KRAS^{G12V}, perhaps in part by suppressing the expression of inflammatory cytokines[53]. This is consistent with the modest (roughly 2-fold) but statistically significant enrichment of U2AF1^{S34F} mutations in KRAS^{G12V}-mutant cancers. However, 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 almost complete *KRAS* exon 3 skipping and early protein termination, resulting in 100% of KRAS^{Q61K}-mutant cancers acquiring silent mutations in codon A59 or G60 which restore *KRAS* exon 3 inclusion[49]. We found limited *KRAS* exon 3 skipping in KRAS^{Q61K}-mutant TCGA samples, which also show secondary silent mutations[49]. 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 complete exon 3 skipping and always leads to acquisition of a simultaneous silent mutations[49]. In contrast, other KRAS^{Q61} mutations result in less exon 3 skipping, allowing these mutations to be better tolerated by cancer cells before acquiring secondary U2AF1^{I24T} mutations to drive KRAS signaling.

Our understanding of why certain amino acid substitutions in KRAS are more favored 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[54]. 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 favorable 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 signaling. 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 [34,43,55]. 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 signaling in a perfect range to induce oncogenesis[55].

While a ~15% variation in *KRAS* exon 2 inclusion associated with G12S mutation may seem small, this change was associated with a ~37% difference in p-Erk levels. Over the course of a developing tumor, this could be sufficient to make significant differences in tumor growth and progression. Additionally, even though a modest decrease in exon 2 inclusion may attenuate the impact of $KRAS^{G12S}$ mutation on cancer cells, we hypothesize that the KRAS mutations that cause exon skipping are nevertheless subject to positive selection as they lock a significant fraction of the protein in a constitutively active state. While genetically engineered mouse models for $KRAS^{G12S}$ are currently unavailable, their development would allow scientists to test these hypotheses and study the impact of combined KRAS and U2AF1 mutations on tumor growth and progression *in vivo*.

Our study also raises the possibility of additional avenues to target KRAS clinically by modulating its splicing. Previous work from the laboratory of Brage Andresen[34,43] as well as our findings here demonstrate that targeting exon 2 or exon 3 of *KRAS* via SSOs can limit cell growth significantly. Furthermore, we have observed that pharmacological inhibition of the splicing factor SF3B1 also drives *KRAS* exon 2 skipping and suppresses MAPK signaling. As a result, further studies are warranted to test the effectiveness of these therapies in cancers with splice-disrupting KRAS mutations, such as G12S, Q61R and Q61L, as well as in combination with targeted KRAS inhibitors.

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[56]. Previous research on *U2AF1* mutations has suggested these occur early in tumor evolution[15,45,46]. However, the variant allele frequencies of $KRAS^{G12S}$ or $KRAS^{Q61R/L}$ are significantly higher than $U2AF1^{S34F}$ or $U2AF1^{I24T}$ in tumors harboring both mutations. Sequential clinical sequencing data are required to provide direct evidence for the secondary nature of U2AF1 mutations. However, our findings are consistent with a model in which a first round of selection events occurs during tumor 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% fetal 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[57,58]. 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

489 pU6-tevopreq1-U2AF1-S34C-Reverse
490 pU6-tevopreq1-U2AF1-S34A-Forward
491 pU6-tevopreq1-U2AF1-S34A-Reverse
492
493 pU6-tevopreq1-U2AF1-S34F-to-WT-Forward
494 pU6-tevopreq1-U2AF1-S34F-to-WT-Reverse
495 pU6-tevopreq1-U2AF1-S34F-to-S-Forward
496 pU6-tevopreq1-U2AF1-S34F-to-S-Reverse
497 pU6-tevopreq1-U2AF1-S34F-to-F(TTC)-Forward
498 pU6-tevopreq1-U2AF1-S34F-to-F(TTC)-Reverse
499 pU6-tevopreq1-U2AF1-S34F-to-Y-Forward
500 pU6-tevopreq1-U2AF1-S34F-to-Y-Reverse
501 pU6-tevopreq1-U2AF1-S34F-to-C-Forward
502 pU6-tevopreq1-U2AF1-S34F-to-C-Reverse
503 pU6-tevopreq1-U2AF1-S34F-to-A-Forward
504 pU6-tevopreq1-U2AF1-S34F-to-A-Reverse
505
506 pU6-tevopreq1-KRAS-G12S(TCT)-Forward
507 pU6-tevopreq1-KRAS-G12S(TCT)-Reverse
508
509
510 Twin prime editing (TwinPE) was carried out by transfecting human cell lines with pCMV-
511 PEMax-P2A-BSD (Addgene plasmid #174821) alongside 2 complementary epegRNAs cloned
512 into the pU6-tevopreq1-GG-acceptor backbone (Addgene plasmid #174038)[21]. Cells were
513 plated at 2×10^5 cells per well in a 6-well plate, and 24 hours after plating transfections were
514 conducted using the Mirus Bio TransIT-X2 Dynamic Delivery System (Mirus Bio LLC, MIR6004)
515 according to the manufacturer's recommended protocol for 6-well plates. A549 cells were
516 treated with 5 μl of TransIT-X2 reagent in 6 well plates, while NCI-H441 cells were treated with
517 7.5 μl of TransIT-X2 reagent. 500 ng each of the forward and reverse epegRNA vectors were
518 used, along with 1.5 μg of pCMV-PEMax-P2A-BSD. Cells were then treated with 10 mg/ml
519 Blasticidin S HCl (Gibco, A1113903) 24-hours post-transfection and changed to normal media
520 72 hours post-transfection.
521
522 **Single Cell Cloning**
523
524 Cells were counted 1 week after prime editing, and 10,000 cells were plated in well A1 of a 96-
525 well plate before being subjected to array dilution[59]. After 2-4 weeks, wells with single colonies
526 were trypsinized and half of the cell volume was passaged in 24-well plates. 1 mL PBS was
527 added to the remaining half of cells before being centrifuged for 5 minutes at 340 × *g*. The PBS
528 was aspirated, and the cells were resuspended in 10 μl Elution Buffer (QIAGEN) and boiled at
529 100°C for 15 minutes. Samples were then placed on ice for 3 minutes before being pelleted by
530 centrifugation at 20,000 × *g*. for 1 minute. PCR amplification of the desired locus was performed
531 with 12.5 μl Q5 High-Fidelity 2X Master Mix (New England BioLabs, M0494S) along with 7.5 μl
532 of the isolated supernatant as template, and 2.5 μl each of 5 μM forward and 5 μM reverse PCR
533 primers (Extended Data Table 5). Samples were then submitted to Quintara Biosciences for
534 Sanger sequencing.
535
536 **RNA-Sequencing**
537

538 RNA for each cell clone was extracted using the RNeasy Plus Mini Kit (QIAGEN, 74134) as per
539 the manufacturer's instructions. RNA concentration was quantified using the Qubit RNA Broad
540 Range Quantification Assay Kit (Invitrogen, Q10210) according to the manufacturer's protocol.
541 RNA was then submitted to Novogene for mRNA-Sequencing including the following steps
542 described in this paragraph. Sample quality was confirmed to have a RIN score >9 by
543 BioAnalyzer. Messenger RNA was purified from total RNA using poly-T oligo-attached magnetic
544 beads. After fragmentation, the first strand cDNA was synthesized using random hexamer
545 primers, followed by the second strand cDNA synthesis using dTTP. The samples were then
546 processed by end repair, A-tailing, adapter ligation, size selection, amplification, and
547 purification. The library quality was checked with Qubit and real-time PCR for quantification and
548 BioAnalyzer for size distribution detection. Paired-end clean reads were aligned to the reference
549 genome GRCh38[60] using Hisat2 (v2.0.5)[61]. FeatureCounts (v1.5.0-p3) was used to count the
550 reads numbers mapped to each gene[62].
551
552 Alternative splicing events were identified using rMATS (4.1.0)[63] and defined as significant
553 according to an FDR < 0.05. Sashimi plots were viewed and generated using Integrative
554 Genomics Viewer (v2.16.10)[64]. Filtering for known oncogenes and tumor suppressors was
555 performed by examining genes that were defined as cancer genes by at least 4 sources in
556 OncoKB[27,28].
557
558 *KRAS* exon 2 read coverage and fraction of mutant reads for *U2AF1* were analyzed using
559 Integrative Genomics Viewer (v2.16.10)[64]. The overlap of significant alternative splicing events
560 between $U2AF1^{S34F}$-mutant A549 cells and human lung adenocarcinomas was quantified by
561 performing a Fisher's exact text on alternative splicing events ($p<0.05$ by rank-sum test) present
562 or absence in each dataset as determined by rMATS (A549 cells) or SplAdder (TCGA)[24].
563
564 To identify mutant U2AF1 3' splice site sequence preferences, significant exon splicing events
565 were first determined using rMATS. The regions of interest were set as 10 bp preceding and
566 following the first base of each skipped exon, as well as 10 bp preceding and following the first
567 base of each included exon. The coordinates of these genomic regions were recorded in BED
568 file format. Sequences corresponding to these regions were extracted from the Hg38 human
569 reference genome fasta file using the getfasta function from bedtools (v2.26.0)[65], and graphics
570 were generated using Weblogo (v.3.7.11)[66].
571
572 For long-read RNA-Sequencing, RNA was isolated as above and submitted to Broad Clinical
573 Labs (Broad Institute). An aliquot of 300ng of total RNA was used as the input for Kinnex™ full
574 length cDNA synthesis and amplification (PacBio Iso-Seq express 2.0 kit, 103-071-500). The
575 amplified cDNA was then quantified by Qubit and Tapestation (Qubit dsDNA HiSens,
576 QUBDSDNA500KT, and HiSense D5000 ScreenTape, 50675592) and barcoded cDNA was
577 then pooled before proceeding to 8-fold Kinnex PCR (PacBio Kinnex PCR 8-fold kit, 103-071-
578 600). Samples then were programmatically concatenated into single molecules optimal for long-
579 read sequencing (Kinnex full-length RNA kit, 103-072-000). Success of array formation was
580 evaluated through qubit quantification and Agilent TapeStation fragment size QC (Qubit dsDNA
581 HiSens, QUBDSDNA500KT, and GenomicDNAScreenTape, 50675366). Each pooled Kinnex
582 library underwent sequencing preparation using the Revio polymerase kit (Pacific Biosciences,
583 102-739-100). Using the Sample Setup page in SMRTLink, the appropriate volumes of each
584 reagent was calculated by factoring in the insert size, sample concentration, and target plate
585 loading concentration of 150pM. The run design was set up with the Application Type "Kinnex
586 full-length RNA" and Library Type "Kinnex." Samples were loaded onto the sequencer within 24
587 hours of completed sequencing preparation and were sequenced using the 24-hour movie
588 setting.

589
590 Following sequencing, unaligned BAM files were first converted to FASTQ format with SAMtools
591 fastq[67]. Samples were then aligned to the primary GRCh38.p14[60] assembly with Minimap2[68]
592 using the "-ax splice:hq -uf" parameters. The alignments were then sorted and indexed with
593 samtools index and samtools sort, respectively, to create aligned BAM files. For transcript
594 quantification and discovery, Isoquant[69] was run on the aligned BAM files individually with
595 transcript discovery enabled using the Gencode.v48 primary annotation[70]. Transcript discovery
596 was performed on each sample individually, rather than through a joint transcript model of all
597 samples or a subset of samples, to provide us with high sensitivity to novel transcripts. Stringtie
598 –merge[71] was then used to extend the Gencode.v48.primary annotation with the unique set of
599 novel transcripts that were identified. Two novel KRAS transcripts (MSTRG14141.1 and
600 MSTRG14141.13) were identified, and then Isoquant was re-run without transcript discovery to
601 quantify all samples using the merged annotation.
602
603 **RT-PCR quantification of *KRAS* exon 2 or exon 3 skipping**
604
605 RT-PCR primers to detect *KRAS* exon 2 or exon 3 skipping were designed using Primer-
606 BLAST[72], with the forward and reverse primers being present on exons flanking the exon of
607 interest (Extended Data Table 5). Cells were trypsinized and centrifuged at 340 × *g* for 5 min,
608 followed by RNA extraction using RNeasy Plus Mini Kit (QIAGEN, 74134) as per manufacturer's
609 protocol. RNA concentrations were quantified with the Qubit RNA BR Assay Kit (Invitrogen,
610 Q10210), and 100 ng of RNA was used to make cDNA using the SuperScript IV VILO Master
611 Mix (Invitrogen, 11756050). The samples were PCR amplified using Phusion Plus PCR Master
612 Mix (Thermo Scientific, F631S), using a 60°C annealing temperature per the manufacturer's 3-
613 step protocol. The PCR products were run on a 2% agarose gel and imaged with a ChemiDoc
614 MP (Bio-Rad Laboratories, 12003154). Band intensities were quantified using the Analyze Gels
615 feature in Fiji (ImageJ). Briefly, horizontal boxes were drawn around the top (exon included) and
616 bottom (exon skipped) bands, and the intensity of each band was quantified. The sum of the
617 band intensities was then calculated, and each band intensity was divided by the total intensity
618 to give the fraction of exon inclusion.
619
620 **qRT-PCR quantification of *KRAS* exon 2 skipping**
621 qRT-PCR primers to detect *KRAS* exon 2 skipping and inclusion were designed using Primer-
622 BLAST[72], with the forward primer spanning either the exon 1-exon 3 junction (for skipping) or
623 the exon 1-exon 2 junction (for inclusion) (Extended Data Table 5). qRT-PCR primers to detect
624 Beta-Actin expression for an internal control were also designed using Primer-BLAST[72]
625 (Extended Data Table 5). Cells were trypsinized and centrifuged at 340 × *g* for 5 min, followed
626 by RNA extraction using the RNeasy Plus Mini Kit (QIAGEN, 74134) as per the manufacturer's
627 instructions. RNA concentrations were quantified with the Qubit RNA BR Assay Kit (Invitrogen,
628 Q10210), and 100 ng of RNA was used to make cDNA using the SuperScript IV VILO Master
629 Mix (Invitrogen, 11756050). cDNA was diluted 1:25 and 1.5 μl was used as input along with 0.5
630 μl each of forward and reverse primers at a concentration of 100 nM each, as well as 2.5 μl of
631 Power SYBR Green PCR Master Mix (2X) (Applied Biosciences, 4368577) for qRT-PCR
632 following the manufacturer's two-step protocol. The fold change of *KRAS* exon 2 skipping or
633 inclusion was quantified using the $2^{-\Delta\Delta Ct}$ method by first normalizing expression to Beta-actin,
634 then normalizing all samples to U2AF1^{S34S} sample 1, before quantifying the fold change.
635
636 **Modeling U2AF1 binding to *KRAS* RNA**
637
638 To model U2AF1 binding to *KRAS* RNA, RBPSuite (v1.0)[37] was used to generate predicted
639 U2AF1 binding scores. Human was used as the species, linear RNA was selected as RNA type,

640 and U2AF1 was selected as the specific model. A 46 bp RNA sequence centered at the 3'
641 splice site of *KRAS* exon 2, or a 34 bp RNA sequence centered at wildtype *KRAS* amino acid
642 G12, or mutant *KRAS* G12S was used as the input sequence.
643
644 Exon 2 3' splice site RNA sequence input:
645 CATTTTCATTATTTTTATTATAAGGCCTGCTGAAAATGACTGAATA
646
647 Wildtype G12 RNA sequence input:
648 GTGGTAGTTGGAGCTGGTGGCGTAGGCAAGAGTG
649
650 G12S RNA sequence input:
651 GTGGTAGTTGGAGCTAGTGGCGTAGGCAAGAGTG
652
653
654 **Modeling steric hindrance and *KRAS* exon 2 skipping using dCasRx**
655
656 pXR002: EF1a-dCasRx-2A-EGFP and pXR003: CasRx gRNA cloning backbone were gifts from
657 Patrick Hsu (Addgene plasmid # 109050 and 109053)[40]. dCasRx sgRNAs were designed using
658 the cas13design tool (https://cas13design.nygenome.org)[73,74], ordered from IDT and
659 resuspended in QIAGEN elution buffer (Extended Data Table 3). Each oligonucleotide pair was
660 annealed by combining 1 μl of top (forward) oligonucleotide, 1 μl of bottom (reverse)
661 oligonucleotide, and 23 μl of annealing buffer (water supplemented with 10 mM Tris-HCl pH 8.5
662 and 50 mM NaCl), heating at 95 °C for 3 minutes then cooling to 22 °C at 0.1 °C per second.
663 Annealed oligonucleotide pairs were then cloned into the CasRX gRNA cloning backbone via
664 golden gate assembly, using BbsI-HF (NEB R3539S)[20]. Cloned plasmids were propagated in
665 NEB Stable Competent *E. coli* (New England BioLabs, C3040) and grown on 50 μg/ml
666 carbenicillin plates (Corning, MT46100RG). Plasmids were extracted using the QIAprep Spin
667 Miniprep Kit (QIAGEN, 27106) and plasmid concentrations were quantified using a Nanodrop
668 ND-1000 Spectrophotometer (Thermo Scientific). Plasmids were then submitted to Quintara
669 Biosciences for Sanger sequencing, or Plasmidsaurus Inc. for whole plasmid long-read
670 sequencing.
671
672 Plasmids:
673 pXR003-KRAS-Exon2-3'SS-1
674 pXR003-KRAS-Exon2-3'SS-2
675 pXR003-KRAS-G12-1
676 pXR003-KRAS-G12-2
677 pXR003-KRAS-Downstream-4
678 pXR003-KRAS- Downstream-5
679
680
681 To test the impact of dCasRx binding to different locations on the *KRAS* exon 2 RNA, NCI-
682 H2023 with wildtype *KRAS* were plated at 200,000 cells per well on 6-well plates. Cells were
683 transfected 24 hours later with 5 μg EF1a-dCasRx-2A-EGFP and 5 μg dCasRx sgRNA in 500 μl
684 Opti-MEM with 20 μl TransIT-X2. 48 hours post-transfection, cells were trypsinized and
685 centrifuged at 340 × *g* for 5 min, followed by RNA extraction using RNeasy Plus Mini Kit
686 (QIAGEN, 74134). 100 ng of RNA was used to make cDNA using the SuperScript IV VILO
687 Master Mix (Invitrogen, 11756050). Exon 2 skipping was then quantified by RT-PCR, and the
688 PCRs were run on a 2% agarose gel.
689

690 **Splicing Reporter Minigene**
691
692 The pSpliceExpress system used for splicing reporter minigene experiments was a gift from
693 Stefan Stamm (Addgene plasmid #32485)[41]. Wildtype *KRAS* Exon 2 and Exon 3 DNA
694 (containing 500 bp of intronic sequence upstream and downstream of each exon) was
695 synthesized as gBlocks by Integrated DNA Technologies (IDT) and then cloned into the
696 pSpliceExpress via Gibson assembly. Simply, the gBlock was amplified using 2X Phusion Plus
697 PCR Master Mix (Thermo Fisher, F631S) and primers with overlapping sequences to the
698 pSpliceExpress backbone to insert the fragment between rat insulin exons 2 and 3, following the
699 manufacturer's recommended 3-step protocol. The pSpliceExpress vector was also linearized
700 by PCR using 2X Phusion Plus PCR Master Mix (Thermo Fisher, F631S) following the
701 manufacturer's recommended 3-step protocol. Gibson assembly primers were designed using
702 the NEBuilder Assembly Tool (NEB). 10% of each PCR product was run on a 1% agarose gel to
703 confirm the correct size, and the remaining product was PCR purified using the QIAquick PCR &
704 Gel Cleanup Kit (QIAGEN, 28506). Gibson assembly was then carried out using PCR products
705 for pSpliceExpress and *KRAS* exon 2 or exon 3, and NEBuilder Hifi DNA Assembly master mix
706 (NEB, E2621S) using the manufacturer's recommended protocol. The cloned plasmids were
707 propagated in NEB Stable Competent E. coli (New England BioLabs, C3040) and grown on 50
708 µg/ml carbenicillin plates (Corning, MT46100RG). Plasmid extraction was completed with the
709 QIAprep Spin Miniprep Kit (QIAGEN, 27106) and plasmid concentrations were measured using
710 a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific) and submitted to Quintara
711 Biosciences for whole plasmid sequencing.
712
713 *KRAS* exon 2 and exon 3 mutations were introduced using a modified site-directed mutagenesis
714 protocol[75]. Site directed mutagenesis primers were designed using PrimerX
715 (https://www.bioinformatics.org/primerx/) and site-directed mutagenesis was performed by PCR
716 amplification with Platinum SuperFi II Master Mix (Thermo Fisher, 12368010) following the
717 manufacturer's protocol. The reaction was then treated with DpnI (NEB, R0176S) for 1 hour at
718 37°C before bacterial transformation.
719
720 To study levels of KRAS exon 2 and exon 3 skipping cells, 293T cells were seeded at 100,000
721 cells per well in a 24-well plate. After 24 hours, cells were transfected with 1 µg pSpliceExpress
722 plasmid in 100 uL of Opti-MEM I (Gibco, 31985062) with 3 µl of Mirus Bio TransIT-X2 Dynamic
723 Delivery System (Mirus Bio LLC, MIR6004). 24 hours after transfection, cells were trypsinized
724 and centrifuged at 340 x g for 5 min, followed by RNA extraction using RNeasy Plus Mini Kit
725 (QIAGEN, 74134). 200 ng of RNA was used for cDNA production using the SuperScript IV VILO
726 Master Mix (Invitrogen, 11756050). Exon skipping was then quantified by RT-PCR using
727 Phusion Plus PCR Master Mix (ThermoFisher, F632S) and primers specific to the
728 pSpliceExpress system[76] (Extended Data Table 5). PCRs were run on a 2% agarose gel and
729 band intensities were quantified using the Analyze Gels feature in Fiji (ImageJ) as described
730 above.
731
732 Plasmids:
733 pSplice-Express-KRAS-Exon2
734 pSplice-Express-KRAS-G12S
735 pSplice-Express-KRAS-G12S(TCT)
736 pSplice-Express-KRAS-G12A
737 pSplice-Express-KRAS-G12C
738 pSplice-Express-KRAS-G12D
739 pSplice-Express-KRAS-G12V
740

741 pSplice-Express-KRAS-G10R(AGA)
742 pSplice-Express-KRAS-G10R(CGA)
743 pSplice-Express-KRAS-A11A(GCA)
744 pSplice-Express-KRAS-A11A(GCC)
745 pSplice-Express-KRAS-G13S(AGC)
746 pSplice-Express-KRAS-G13S(TCC)
747 pSplice-Express-KRAS-F28L(TTA)
748 pSplice-Express-KRAS-F28L(TTG)
749
750 pSplice-Express-KRAS-Exon3
751 pSplice-Express-KRAS-Q61R
752 pSplice-Express-KRAS-Q61L
753 pSplice-Express-KRAS-Q61H
754 pSplice-Express-KRAS-Q61P
755 pSplice-Express-KRAS-Q61K
756 pSplice-Express-KRAS-GQ60GK
757
758 **Flow Cytometry**
759
760 For flow cytometry experiments to measure p-Erk levels, 500,000 cells were first plated in 6 well
761 plates. 24 hours after plating, cells were deprived of serum and then were collected 16 hours
762 later by trypsinization using 500 µl of trypsin. An equal volume of pre-warmed BD Phosflow Fix
763 Buffer I (BD Biosciences, 557870) was added directly to the plate, cells were incubated at 37C
764 for 10 minutes before being collected and centrifuged at 340 × *g* for 5 min. Cells were washed
765 with PBS and permeabilized with 1 ml ice cold BD Phosflow Perm Buffer III (BD Biosciences,
766 558050) and incubated on ice for 30 minutes. Cells were counted, and equal numbers of cells
767 were washed twice with staining buffer (1X PBS, 1% FBS and 0.09% sodium azide) before
768 being resuspended in 50 µl of staining buffer with BD Phosflow PE-Cy7 anti-Erk1/2
769 (pT202/pY204) antibody (BD Biosciences, 560116) at a 1:5 dilution. Cells were stained for 30
770 minutes at room temperature in the dark, then washed with PBS and resuspended in PBS at a
771 final concentration of 1 million cells per ml. Samples were then analyzed on a BD LSRFortessa
772 Cell Analyzer (BD Biosciences) at the Dana-Farber Cancer Institute Flow Cytometry Core.
773
774
775 **Pladienolide B Treatment**
776
777 To determine the impact of SF3B1 inhibition on *KRAS* exon 2 skipping and MAPK signaling,
778 500,000 NCI-H2023 cells were first plated in 6 well plates. 24 hours later, cells were treated with
779 DMSO or Pladienolide B at concentrations of 100 nM, 50 nM, 10 nM, 5 nM or 1 nM, being sure
780 that all samples received equal total amounts of DMSO to control for potential DMSO-mediated
781 effects. For flow cytometry measurement of p-Erk, cells were also changed to serum free media.
782 20 hours after Pladienolide B treatment, cells were collected for RNA isolation or flow cytometry.
783 *KRAS* exon 2 skipping and p-Erk levels were measured by RT-PCR and flow cytometry
784 respectively, as described above.
785
786
787 **Splice Switching Oligonucleotide Treatment**
788
789 For *KRAS* exon 2 skipping, SSOs designed and validated by the laboratory of Brage Andresen
790 were used[43]. For *KRAS* exon 3 skipping, SSOs were designed using the eSkip-Finder
791 application[77], using 2'OMe chemistry and a length of 25 bp. SSO sequences are available in

792 Extended Data Table 4. NCI-H23 cells were plated at 2×10^6 cells per 10 cm plate. 24 hours
793 after plating, cells were transfected using Transit-X2 transfection reagent according to the
794 manufacturer's recommended protocol for 10 cm plates, with 45 μl of TransIT-X2 reagent. Each
795 well was treated with 120 nM total SSO, comprising of a range of concentrations of either
796 scramble SSO control, or *KRAS* exon 2 or exon 3 skipping SSO (120 nM scramble SSO alone,
797 6 nM *KRAS* SSO with 114 nM scramble SSO, 12 nM *KRAS* SSO with 108 nM scramble SSO,
798 30 nM *KRAS* SSO with 90 nM scramble SSO, 60 nM *KRAS* SSO with 60 nM scramble SSO, or
799 120 nM *KRAS* SSO alone) to keep the total amount of SSO delivery the same between
800 conditions[34,43]. Cells were plated for cell growth assays 24 hours after transfection, while cells
801 were collected for RNA or protein 72 hours following transfection.
802
803
804 **Immunoblot Analysis**
805
806 For immunoblot analyses, cells were deprived of serum for 16 hours before cell collection to
807 reduce the impact of extracellular growth factor stimulation on MAPK signaling. Cells were lysed
808 with RIPA Buffer (Millipore Sigma, R0278-50ML) containing Halt Protease and Phosphatase
809 Inhibitor Cocktail (100X) (Thermo Scientific, 78440) and quantified with the Pierce BCA Protein
810 Assay Kit (Thermo Scientific, 23225). Following addition of NuPAGE 4X LDS Sample Buffer
811 (Invitrogen, NP0008) and NuPAGE 10X Sample Reducing Agent (Invitrogen, NP0009), the
812 samples were run on a NuPAGE 4-12% Bis-Tris, 1.5 MM, Mini Protein Gel (Invitrogen,
813 NP0335BOX) in NuPAGE 20X MOPS SDS Running Buffer (Invitrogen, NP0001) at 125V for
814 105 min on ice. The gel was transferred on a nitrocellulose membrane in NuPAGE 20X Transfer
815 Buffer (Invitrogen, NP00061) at 70V for 2 hours on ice. Blocking and Immunostaining for HSP90
816 (BD Biosciences, 610418, 1:10,000), p-Erk (Cell Signaling, 4370S, 1:2000), Total-Erk (Cell
817 Signaling, 9102S, 1:2000), p-Akt (Cell Signaling, 4060S, 1:1000), Total-Akt (Cell Signaling,
818 58295S, 1:500) and KRAS (Invitrogen, 11H35L14, 1:2500) were done in Intercept (TBS)
819 Blocking Buffer (LI-COR Biosciences, 927-60001), with 0.1% Tween-20 for primary and
820 secondary antibody blockings. LI-COR Odyssey Classic Imaging System (LI-COR Biosciences,
821 ODY-9120) was used to image the blots.
822
823
824 **Cell Proliferation Assays**
825
826 Cells were plated on ultra-low attachment plates (Corning, 3474) at 1,000 cells per well in 100 μl
827 of media for each desired time point to measure growth. At each time point, cells were moved to
828 white-walled 96-well plates (Corning, 3903). 100 μl of CellTiter-Glo Luminescent Cell Viability
829 Assay (Promega, G7572) diluted 1:1 in PBS was then added to each well. After a 10-minute
830 incubation at room temperature while rocking, the plates were analyzed using a SpectraMax M5
831 (Molecular Devices) plate reader using the included CellTiter-Glo protocol with a luminescence
832 integration time of 500 ms. Luminescence values were normalized to data points from the
833 earliest time point.
834
835
836 **Cloning of lentiviral vectors**
837
838 $U2AF1^{S34F}$ cDNA was synthesized by IDT and subsequently cloned into the pLX301 backbone
839 (Addgene #25895) via Gibson assembly. In brief, the $U2AF1^{S34F}$ cDNA was amplified using 2X
840 Phusion Plus PCR Master Mix (Thermo Fisher, F631S) and primers with overlapping sequences
841 to the pLX301 backbone following the CMV promoter. Gibson assembly primers were designed
842 using the NEBuilder Assembly Tool (NEB) and contained a FLAG tag sequence added to the C-

843 terminal end of U2AF1. The PCR product was run on a 1% agarose gel and gel purified using
844 the QIAquick PCR & Gel Cleanup Kit (QIAGEN, 28506). The pLX301 backbone was linearized
845 via restriction digested with BsrGI-HF (NEB, R3575S) for 1 hour at 37°C and then run on a 1%
846 agarose gel and gel purified. Gibson assembly was then carried out using linearized pLX301,
847 U2AF1 PCR product, and NEBuilder Hifi DNA Assembly master mix (NEB, E2621S) using the
848 manufacturer's recommended protocol.
849
850 The vector was then mutated to express wildtype U2AF1 and U2AF1^{I24T} via sequential site-
851 directed mutagenesis. First, pLX301-U2AF1^{S34F}-FLAG was mutated to pLX301-U2AF1-FLAG,
852 before being mutated again to form pLX301-U2AF1^{I24T}-FLAG. Site directed mutagenesis
853 primers were designed using PrimerX (https://www.bioinformatics.org/primerx/) and site-directed
854 mutagenesis was performed by PCR amplification with Platinum SuperFi II Master Mix (Thermo
855 Fisher, 12368010) followed by restriction digestion with DpnI (NEB, R0176S) for 1 hour at 37°C
856 before bacterial transformation.
857
858 Plasmids:
859 pLX301-U2AF1-WT-FLAG
860 pLX301-U2AF1-S34F-FLAG
861 pLX301-U2AF1-I24T-FLAG
862
863
864 **Lentivirus Production and Transduction**
865
866 Lentivirus was produced as described previously[78]. Briefly, 8 × 10^6 293T cells were plated on 15
867 cm plates coated with 0.1% gelatin. 24 hours after plating, cells were transfected with 10 μg
868 pLX301 lentiviral vector, 7.5 μg delta 8.9 plasmid, and 2.5 μg pCMV-VSV-G along with 80 μl
869 PEIMax (Polysciences, 24765) in 1 ml of Opti-MEM I (Gibco, 31985062). 24 hours after
870 transfection, the media was replaced with fresh DMEM supplemented with 25 mM HEPES
871 (Gibco, 15630-080) and 3 mM caffeine (Sigma, C0750). Lentivirus-containing supernatant was
872 collected from the cells at 48 and 72 hours following transfection and filtered through 0.45-μm
873 filters (Thermo Scientific 723-2545).
874
875 Lentivirus was then concentrated by centrifugation through a sucrose cushion[79]. A 10% sucrose
876 solution (50 mM Tris-HCl,100 mM NaCl, 0.5 mM EDTA) was carefully pipetted underneath the
877 lentivirus-containing media at a 4:1 ratio of lentivirus-containing media to sucrose solution. The
878 lentivirus was then centrifuged at 10,000 × g for 4 hours at 4°C. Media was aspirated and the
879 pellet was resuspended in 100 μl of PBS.
880
881 For Panc 02.13 cell transduction, 35,000 cells were plated in 300 μl insulin-containing RPMI in
882 48-well plates. 24 hours after plating, cells were transduced via spinfection. Briefly, fresh RPMI
883 containing 8 μg/ml polybrene (Santa Cruz Biotechnology, sc-134220) was added to each well.
884 The titer of each lentivirus was estimated using Lenti-X GoStix Plus (Takara, 631280), and
885 equal amounts of lentivirus was delivered to each well. The 48-well plate was then centrifuged
886 at 1,000 × g for 2 hours at 30°C. 24 hours after spinfection, fresh RPMI containing 2 μg/ml
887 puromcycin (Gibco, A1113803) was added to each well. 48 hours after puromycin treatment,
888 cells were trypsinized and collected for RNA isolation and RT-PCR.
889
890 **Genomic Data**
891

All data from the AACR Project GENIE Consortium[11] were obtained through the dedicated cBioPortal website[29]. 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[26]. 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*.[35], and *KRAS* and *U2AF1* mutational data was obtained via cBioPortal[29]. Alternative splicing data from TCGA was obtained from Kahles *et al.*[24], and *KRAS* and *U2AF1* mutation data was obtained via cBioPortal[29].

Statistical Analysis

All statistical analyses were performed in GraphPad Prism 10 except where noted. Means and error bars (standard deviation) are plotted for all analyses. Measurements were taken from distinct samples (biological replicates) unless noted. Alternative splicing events were analyzed, and FDR q values were calculated using rMATS[63] for short-read RNA-sequencing, and rMATS-long[63] for long-read RNA-Sequencing. A simple linear regression was performed for 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. A nonlinear (semi-log) fit was performed to analyze the correlation between the percentage of patients with $U2AF1^{S34F}$ mutations and the odds ratio of a smoker developing the disease, and the correlation between the number alternatively spliced genes and the relative frequency of each U2AF1 mutation across all cancers.

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 two-sided Student's t-test was used to analyze the fraction of RNA-sequencing reads showing *KRAS* exon 2 skipping in backedited A549 cells and engineered NCI-H441 cells, the fraction of *KRAS* exon 2 skipping in $U2AF1^{S34F}$-mutant A549 cells by RT-PCR and qRT-PCR, and p-Erk levels in A549 cells with $U2AF1^{S34S}$, $U2AF1^{S34F}$ and $KRAS^{G12S(TCT)}$ mutations. For the comparison of the variant allele frequency of $KRAS^{G12S}$ and $U2AF1^{S34F}$, and $KRAS^{Q61R/L}$ and $U2AF1^{I24T}$, a two-sided 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.

Data Availability

Short- and long-read 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 analyzed 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, Owen Hirschi, and all the members of the Meyerson laboratory for their advice and technical assistance; Mitchell suggested the steric hindrance experiment and Owen suggested the use of splicing reporter minigenes. 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., K.C. and D.D. designed and performed most of the experiments and analyzed most of the data. S.S. and G.F. performed and analyzed the genomics data from Foundation Medicine Inc. I.T.-H.L., A.B.D., J.M.H., E.S., 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; holds equity in Isabl and Karyoverse; 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.

1035 14. Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y.,
1036 Fujimoto, A., Nakagawa, H., Shibata, T., et al. (2016). Mutational signatures associated
1037 with tobacco smoking in human cancer. Science *354*, 618-622. 10.1126/science.aag0299.
1038 15. Esfahani, M.S., Lee, L.J., Jeon, Y.J., Flynn, R.A., Stehr, H., Hui, A.B., Ishisoko, N., Kildebeck,
1039 E., Newman, A.M., Bratman, S.V., et al. (2019). Functional significance of U2AF1 S34F
1040 mutations in lung adenocarcinomas. Nat Commun *10*, 5712. 10.1038/s41467-019-
1041 13392-y.
1042 16. Zhao, Y., Cai, W., Hua, Y., Yang, X., and Zhou, J. (2022). The Biological and Clinical
1043 Consequences of RNA Splicing Factor U2AF1 Mutation in Myeloid Malignancies. Cancers
1044 (Basel) *14*. 10.3390/cancers14184406.
1045 17. Fei, D.L., Zhen, T., Durham, B., Ferrarone, J., Zhang, T., Garrett, L., Yoshimi, A., Abdel-
1046 Wahab, O., Bradley, R.K., Liu, P., and Varmus, H. (2018). Impaired hematopoiesis and
1047 leukemia development in mice with a conditional knock-in allele of a mutant splicing
1048 factor gene U2af1. Proc Natl Acad Sci U S A *115*, E10437-e10446.
1049 10.1073/pnas.1812669115.
1050 18. Wheeler, E.C., Vora, S., Mayer, D., Kotini, A.G., Olszewska, M., Park, S.S., Guccione, E.,
1051 Teruya-Feldstein, J., Silverman, L., Sunahara, R.K., et al. (2022). Integrative RNA-omics
1052 Discovers GNAS Alternative Splicing as a Phenotypic Driver of Splicing Factor–Mutant
1053 Neoplasms. Cancer Discovery *12*, 836-855. 10.1158/2159-8290.Cd-21-0508.
1054 19. Palangat, M., Anastasakis, D.G., Fei, D.L., Lindblad, K.E., Bradley, R., Hourigan, C.S.,
1055 Hafner, M., and Larson, D.R. (2019). The splicing factor U2AF1 contributes to cancer
1056 progression through a noncanonical role in translation regulation. Genes Dev *33*, 482-
1057 497. 10.1101/gad.319590.118.
1058 20. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J.,
1059 Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome
1060 editing without double-strand breaks or donor DNA. Nature *576*, 149-157.
1061 10.1038/s41586-019-1711-4.
1062 21. Anzalone, A.V., Gao, X.D., Podracky, C.J., Nelson, A.T., Koblan, L.W., Raguram, A., Levy,
1063 J.M., Mercer, J.A.M., and Liu, D.R. (2022). Programmable deletion, replacement,
1064 integration and inversion of large DNA sequences with twin prime editing. Nature
1065 Biotechnology *40*, 731-740. 10.1038/s41587-021-01133-w.
1066 22. Fei, D.L., Motowski, H., Chatrikhi, R., Prasad, S., Yu, J., Gao, S., Kielkopf, C.L., Bradley, R.K.,
1067 and Varmus, H. (2016). Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic
1068 of U2AF1-Mutant Tumors and Is Required for Cell Survival. PLoS Genet *12*, e1006384.
1069 10.1371/journal.pgen.1006384.
1070 23. Wadugu, B.A., Nonavinkere Srivatsan, S., Heard, A., Alberti, M.O., Ndonwi, M., Liu, J.,
1071 Grieb, S., Bradley, J., Shao, J., Ahmed, T., et al. (2021). U2af1 is a haplo-essential gene
1072 required for hematopoietic cancer cell survival in mice. J Clin Invest *131*.
1073 10.1172/jci141401.
1074 24. Kahles, A., Lehmann, K.V., Toussaint, N.C., Hüser, M., Stark, S.G., Sachsenberg, T., Stegle,
1075 O., Kohlbacher, O., Sander, C., and Rätsch, G. (2018). Comprehensive Analysis of
1076 Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell *34*, 211-224.e216.
1077 10.1016/j.ccell.2018.07.001.

1078 25. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M.,
1079 Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the
1080 Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291-
1081 304.e296. 10.1016/j.cell.2018.03.022.
1082 26. Milbury, C.A., Creeden, J., Yip, W.-K., Smith, D.L., Pattani, V., Maxwell, K., Sawchyn, B.,
1083 Gjoerup, O., Meng, W., Skoletsky, J., et al. (2022). Clinical and analytical validation of
1084 FoundationOne®CDx, a comprehensive genomic profiling assay for solid tumors. PLoS
1085 One *17*, e0264138. 10.1371/journal.pone.0264138.
1086 27. Chakravarty, D., Gao, J., Phillips, S., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger,
1087 R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge
1088 Base. JCO Precision Oncology, 1-16. 10.1200/PO.17.00011.
1089 28. Suehnholz, S.P., Nissan, M.H., Zhang, H., Kundra, R., Nandakumar, S., Lu, C., Carrero, S.,
1090 Dhaneshwar, A., Fernandez, N., Xu, B.W., et al. (2024). Quantifying the Expanding
1091 Landscape of Clinical Actionability for Patients with Cancer. Cancer Discovery *14*, 49-65.
1092 10.1158/2159-8290.CD-23-0467.
1093 29. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A.,
1094 Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an
1095 open platform for exploring multidimensional cancer genomics data. Cancer Discov *2*,
1096 401-404. 10.1158/2159-8290.CD-12-0095.
1097 30. Nuevo-Tapioles, C., and Philips, M.R. (2022). The role of KRAS splice variants in cancer
1098 biology. Front Cell Dev Biol *10*, 1033348. 10.3389/fcell.2022.1033348.
1099 31. Whitley, M.J., Tran, T.H., Rigby, M., Yi, M., Dharmaiah, S., Waybright, T.J., Ramakrishnan,
1100 N., Perkins, S., Taylor, T., Messing, S., et al. Comparative analysis of KRAS4a and KRAS4b
1101 splice variants reveals distinctive structural and functional properties. Science Advances
1102 *10*, eadj4137. 10.1126/sciadv.adj4137.
1103 32. Voice, J.K., Klemke, R.L., Le, A., and Jackson, J.H. (1999). Four human ras homologs differ
1104 in their abilities to activate Raf-1, induce transformation, and stimulate cell motility. J
1105 Biol Chem *274*, 17164-17170. 10.1074/jbc.274.24.17164.
1106 33. Adamopoulos, P.G., Tsiakanikas, P., Boti, M.A., and Scorilas, A. (2021). Targeted Long-
1107 Read Sequencing Decodes the Transcriptional Atlas of the Founding RAS Gene Family
1108 Members. Int J Mol Sci *22*. 10.3390/ijms222413298.
1109 34. Hartung, A.-M., Swensen, J., Uriz, I.E., Lapin, M., Kristjansdottir, K., Petersen, U.S.S.,
1110 Bang, J.M.V., Guerra, B., Andersen, H.S., Dobrowolski, S.F., et al. (2016). The Splicing
1111 Efficiency of Activating HRAS Mutations Can Determine Costello Syndrome Phenotype
1112 and Frequency in Cancer. PLOS Genetics *12*, e1006039. 10.1371/journal.pgen.1006039.
1113 35. Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd,
1114 Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation
1115 characterization of the Cancer Cell Line Encyclopedia. Nature *569*, 503-508.
1116 10.1038/s41586-019-1186-3.
1117 36. Yang, I.S., and Kim, S. (2018). Isoform specific gene expression analysis of KRAS in the
1118 prognosis of lung adenocarcinoma patients. BMC Bioinformatics *19*, 40.
1119 10.1186/s12859-018-2011-y.

1120 37. Pan, X., Fang, Y., Li, X., Yang, Y., and Shen, H.-B. (2020). RBPsuite: RNA-protein binding
1121 sites prediction suite based on deep learning. BMC Genomics *21*, 884. 10.1186/s12864-
1122 020-07291-6.
1123 38. Lo, A., McSharry, M., and Berger, A.H. (2022). Oncogenic KRAS alters splicing factor
1124 phosphorylation and alternative splicing in lung cancer. BMC Cancer *22*, 1315.
1125 10.1186/s12885-022-10311-1.
1126 39. Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer
1127 (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*,
1128 178-192. 10.1093/bib/bbs017.
1129 40. Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N., and Hsu, P.D. (2018).
1130 Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell *173*, 665-
1131 676.e614. 10.1016/j.cell.2018.02.033.
1132 41. Kishore, S., Khanna, A., and Stamm, S. (2008). Rapid generation of splicing reporters with
1133 pSpliceExpress. Gene *427*, 104-110. 10.1016/j.gene.2008.09.021.
1134 42. Hirschi, O.R., Felker, S.A., Rednam, S.P., Vallance, K.L., Parsons, D.W., Roy, A., Cooper,
1135 G.M., and Plon, S.E. (2024). Combined bioinformatic and splicing analysis of likely benign
1136 intronic and synonymous variants reveals evidence for pathogenicity. Genetics in
1137 Medicine Open *2*, 101850.
1138 43. Andresen, B. (2019). RAS exon 2 skipping for cancer treatment. Google Patents.
1139 44. Krygowska, A.A., and Castellano, E. (2018). PI3K: A Crucial Piece in the RAS Signaling
1140 Puzzle. Cold Spring Harb Perspect Med *8*. 10.1101/cshperspect.a031450.
1141 45. Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris,
1142 C.C., Koboldt, D.C., Larson, D.E., et al. (2011). Recurrent mutations in the U2AF1 splicing
1143 factor in myelodysplastic syndromes. Nat Genet *44*, 53-57. 10.1038/ng.1031.
1144 46. Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon,
1145 C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al. (2013). Clinical and biological implications
1146 of driver mutations in myelodysplastic syndromes. Blood *122*, 3616-3627; quiz 3699.
1147 10.1182/blood-2013-08-518886.
1148 47. Sakai, T., Sameshima, T., Matsufuji, M., Kawamura, N., Dobashi, K., and Mizui, Y. (2004).
1149 Pladienolides, new substances from culture of Streptomyces platensis Mer-11107. I.
1150 Taxonomy, fermentation, isolation and screening. J Antibiot (Tokyo) *57*, 173-179.
1151 10.7164/antibiotics.57.173.
1152 48. Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y.,
1153 Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural
1154 product pladienolide. Nature Chemical Biology *3*, 570-575. 10.1038/nchembio.2007.16.
1155 49. Kobayashi, Y., Chhoeu, C., Li, J., Price, K.S., Kiedrowski, L.A., Hutchins, J.L., Hardin, A.I.,
1156 Wei, Z., Hong, F., Bahcall, M., et al. (2022). Silent mutations reveal therapeutic
1157 vulnerability in RAS Q61 cancers. Nature *603*, 335-342. 10.1038/s41586-022-04451-4.
1158 50. Yoshida, H., Park, S.Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., Tsuda, K., Kuwasako,
1159 K., Unzai, S., Muto, Y., et al. (2015). A novel 3' splice site recognition by the two zinc
1160 fingers in the U2AF small subunit. Genes Dev *29*, 1649-1660. 10.1101/gad.267104.115.
1161 51. Pangallo, J., Kiladjian, J.J., Cassinat, B., Renneville, A., Taylor, J., Polaski, J.T., North, K.,
1162 Abdel-Wahab, O., and Bradley, R.K. (2020). Rare and private spliceosomal gene

1163 mutations drive partial, complete, and dual phenocopies of hotspot alterations. Blood
1164 *135*, 1032-1043. 10.1182/blood.2019002894.
1165 52. Yang, X., Boehm, J.S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas,
1166 S.R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression
1167 library of human ORFs. Nat Methods *8*, 659-661. 10.1038/nmeth.1638.
1168 53. Liang, C.E., Hrabeta-Robinson, E., Behera, A., Arevalo, C., Fetter, I.J., Soulette, C.M.,
1169 Thornton, A.M., Sikandar, S.S., and Brooks, A.N. (2024). U2AF1 S34F enhances
1170 tumorigenic potential of lung cells by exhibiting synergy with KRAS mutation and altering
1171 response to environmental stress. bioRxiv. 10.1101/2024.09.11.612492.
1172 54. Huynh, M.V., Hobbs, G.A., Schaefer, A., Pierobon, M., Carey, L.M., Diehl, J.N., DeLiberty,
1173 J.M., Thurman, R.D., Cooke, A.R., Goodwin, C.M., et al. (2022). Functional and biological
1174 heterogeneity of KRAS(Q61) mutations. Sci Signal *15*, eabn2694.
1175 10.1126/scisignal.abn2694.
1176 55. Lampson, B.L., Pershing, N.L., Prinz, J.A., Lacsina, J.R., Marzluff, W.F., Nicchitta, C.V.,
1177 MacAlpine, D.M., and Counter, C.M. (2013). Rare codons regulate KRas oncogenesis.
1178 Curr Biol *23*, 70-75. 10.1016/j.cub.2012.11.031.
1179 56. Mina, M., Iyer, A., Tavernari, D., Raynaud, F., and Ciriello, G. (2020). Discovering
1180 functional evolutionary dependencies in human cancers. Nature Genetics *52*, 1198-
1181 1207. 10.1038/s41588-020-0703-5.
1182 57. Chow, R.D., Chen, J.S., Shen, J., and Chen, S. (2021). A web tool for the design of prime-
1183 editing guide RNAs. Nature Biomedical Engineering *5*, 190-194. 10.1038/s41551-020-
1184 00622-8.
1185 58. Mathis, N., Allam, A., Kissling, L., Marquart, K.F., Schmidheini, L., Solari, C., Balázs, Z.,
1186 Krauthammer, M., and Schwank, G. (2023). Predicting prime editing efficiency and
1187 product purity by deep learning. Nature Biotechnology *41*, 1151-1159. 10.1038/s41587-
1188 022-01613-7.
1189 59. Ryan, J.A. (2008). Cell cloning by serial dilution in 96 well plates. Corning.
1190 60. Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P.A., Murphy,
1191 T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D., et al. (2017). Evaluation of GRCh38 and
1192 de novo haploid genome assemblies demonstrates the enduring quality of the reference
1193 assembly. Genome Research *27*, 849-864. 10.1101/gr.213611.116.
1194 61. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome
1195 alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology *37*,
1196 907-915. 10.1038/s41587-019-0201-4.
1197 62. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose
1198 program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.
1199 10.1093/bioinformatics/btt656.
1200 63. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014).
1201 rMATS: robust and flexible detection of differential alternative splicing from replicate
1202 RNA-Seq data. Proc Natl Acad Sci U S A *111*, E5593-5601. 10.1073/pnas.1419161111.
1203 64. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and
1204 Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology *29*, 24-26.
1205 10.1038/nbt.1754.

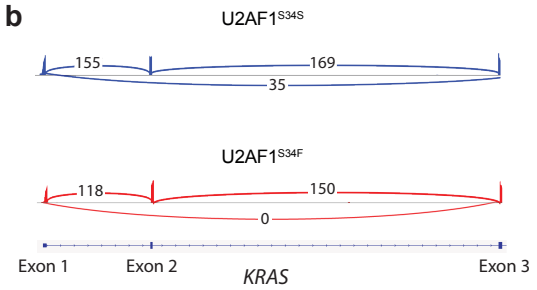
1206 65. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing
1207 genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.
1208 66. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence
1209 logo generator. Genome Res *14*, 1188-1190. 10.1101/gr.849004.
1210 67. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1211 Durbin, R., and Subgroup, G.P.D.P. (2009). The Sequence Alignment/Map format and
1212 SAMtools. Bioinformatics *25*, 2078-2079. 10.1093/bioinformatics/btp352.
1213 68. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*,
1214 3094-3100. 10.1093/bioinformatics/bty191.
1215 69. Prjibelski, A.D., Mikheenko, A., Joglekar, A., Smetanin, A., Jarroux, J., Lapidus, A.L., and
1216 Tilgner, H.U. (2023). Accurate isoform discovery with IsoQuant using long reads. Nature
1217 Biotechnology *41*, 915-918. 10.1038/s41587-022-01565-y.
1218 70. Mudge, Jonathan M., Carbonell-Sala, S., Diekhans, M., Martinez, Jose G., Hunt, T.,
1219 Jungreis, I., Loveland, Jane E., Arnan, C., Barnes, I., Bennett, R., et al. (2024). GENCODE
1220 2025: reference gene annotation for human and mouse. Nucleic Acids Research *53*,
1221 D966-D975. 10.1093/nar/gkae1078.
1222 71. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L.
1223 (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq
1224 reads. Nature Biotechnology *33*, 290-295. 10.1038/nbt.3122.
1225 72. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012).
1226 Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.
1227 BMC Bioinformatics *13*, 134. 10.1186/1471-2105-13-134.
1228 73. Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z., and Sanjana, N.E.
1229 (2020). Massively parallel Cas13 screens reveal principles for guide RNA design. Nature
1230 Biotechnology *38*, 722-727. 10.1038/s41587-020-0456-9.
1231 74. Guo, X., Rahman, J.A., Wessels, H.-H., Méndez-Mancilla, A., Haro, D., Chen, X., and
1232 Sanjana, N.E. (2021). Transcriptome-wide Cas13 guide RNA design for model organisms
1233 and viral RNA pathogens. Cell Genomics *1*, 100001.
1234 https://doi.org/10.1016/j.xgen.2021.100001.
1235 75. Laible, M., and Boonrod, K. (2009). Homemade Site Directed Mutagenesis of Whole
1236 Plasmids. JoVE, e1135. doi:10.3791/1135.
1237 76. Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., Sun, H.Y., Li, A., Ping, X.L.,
1238 Lai, W.Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol Cell
1239 *61*, 507-519. 10.1016/j.molcel.2016.01.012.
1240 77. Chiba, S., Lim, Kenji Rowel Q., Sheri, N., Anwar, S., Erkut, E., Shah, Md Nur A., Aslesh, T.,
1241 Woo, S., Sheikh, O., Maruyama, R., et al. (2021). eSkip-Finder: a machine learning-based
1242 web application and database to identify the optimal sequences of antisense
1243 oligonucleotides for exon skipping. Nucleic Acids Research *49*, W193-W198.
1244 10.1093/nar/gkab442.
1245 78. Walter, D.M., Venancio, O.S., Buza, E.L., Tobias, J.W., Deshpande, C., Gudiel, A.A., Kim-
1246 Kiselak, C., Cicchini, M., Yates, T.J., and Feldser, D.M. (2017). Systematic In Vivo
1247 Inactivation of Chromatin-Regulating Enzymes Identifies Setd2 as a Potent Tumor
1248 Suppressor in Lung Adenocarcinoma. Cancer Res *77*, 1719-1729. 10.1158/0008-
1249 5472.CAN-16-2159.

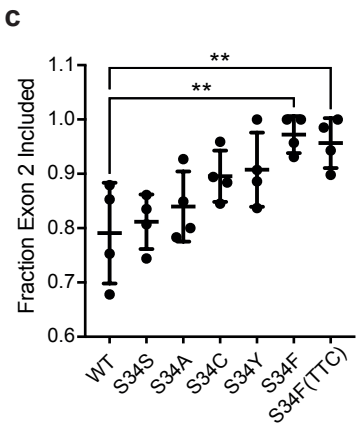
1250 79. Jiang, W., Hua, R., Wei, M., Li, C., Qiu, Z., Yang, X., and Zhang, C. (2015). An optimized
1251 method for high-titer lentivirus preparations without ultracentrifugation. Sci Rep *5*,
1252 13875. 10.1038/srep13875.
1253 80. Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdóttir, H., Robinson, J.T.,
1254 Mesirov, J.P., Airoldi, E.M., and Burge, C.B. (2015). Quantitative visualization of
1255 alternative exon expression from RNA-seq data. Bioinformatics *31*, 2400-2402.
1256 10.1093/bioinformatics/btv034.
1257 81. Rásó, E. (2020). Splice variants of RAS—translational significance. Cancer and Metastasis
1258 Reviews *39*, 1039-1049. 10.1007/s10555-020-09920-8.
1259 82. Kwan, A.K., Piazza, G.A., Keeton, A.B., and Leite, C.A. (2022). The path to the clinic: a
1260 comprehensive review on direct KRASG12C inhibitors. Journal of Experimental & Clinical
1261 Cancer Research *41*, 27. 10.1186/s13046-021-02225-w.
1262

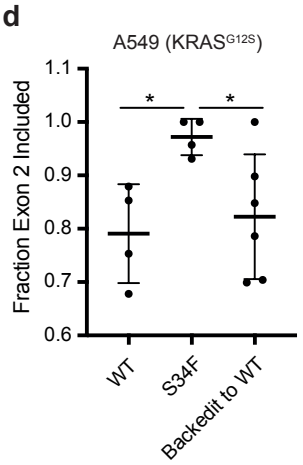
Figure 1

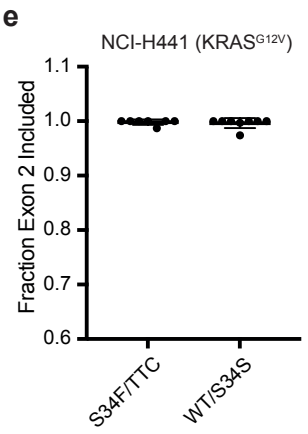
a

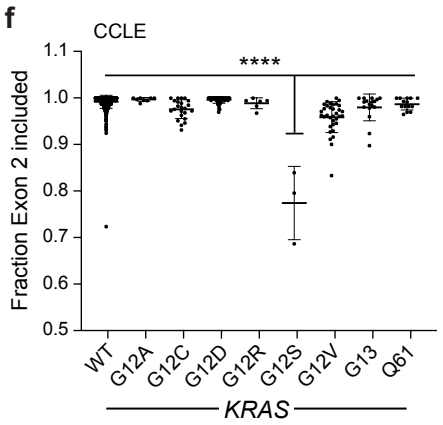
Gene	AS event	FDR	Delta mean PSI
AXL	Skipped Exon	7.38×10^{-8}	0.244
KRAS	Skipped Exon	5.85×10^{-6}	0.181
NCOR2	Skipped Exon	1.57×10^{-3}	0.146
KDM5A	Skipped Exon	2.86×10^{-3}	-0.143
DIS3	Skipped Exon	7.47×10^{-3}	0.182

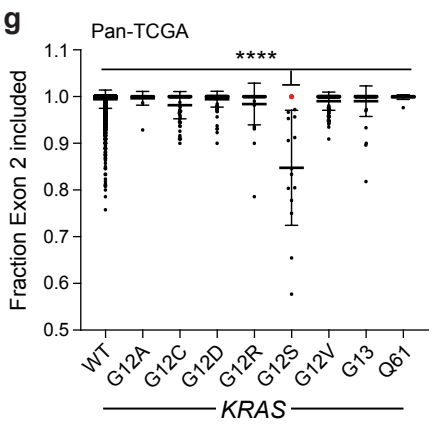






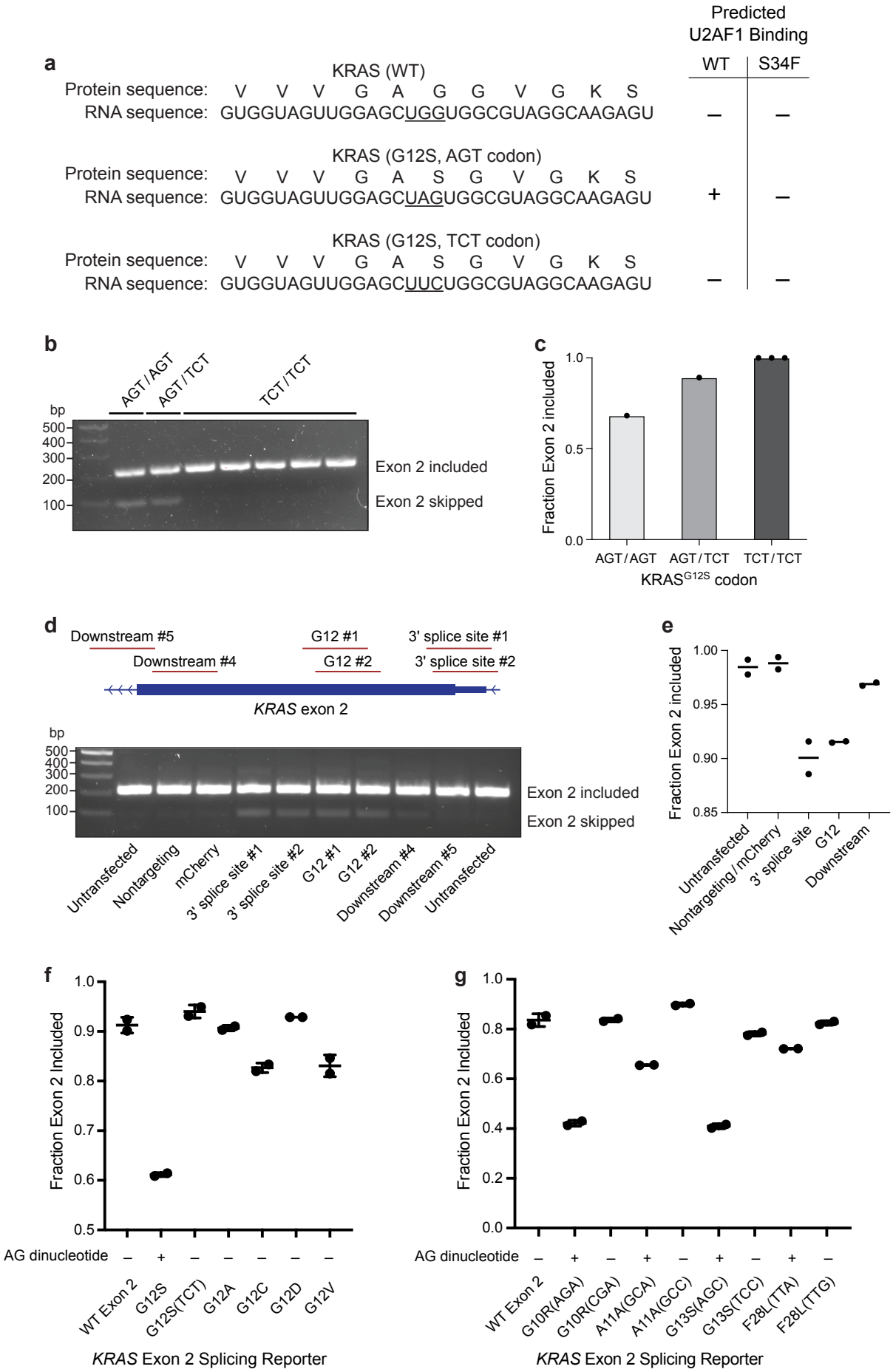






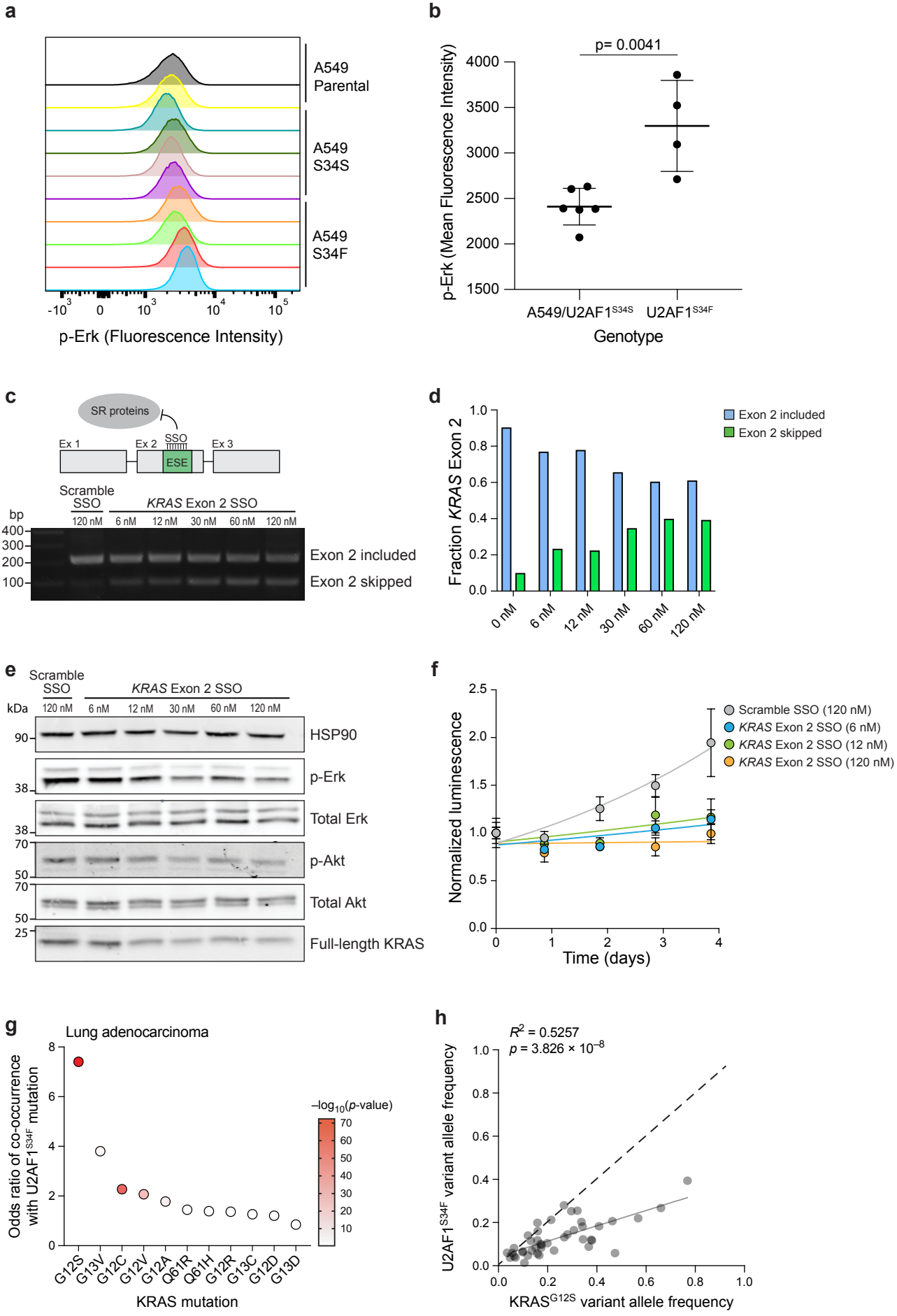
1263 **Fig. 1:** Cancers with KRAS^{G12S} mutations undergo *KRAS* exon 2 skipping which can be
1264 reversed by acquisition of U2AF1^{S34F} mutations. a) Table of the top alternative splicing events
1265 unique to U2AF1^{S34F} and U2AF1$^{S34F(TTC)}$ -mutant cells in known oncogenes and tumor
1266 suppressors as defined by their presence in OncoKB[27,28]. Table shows the gene, type of
1267 alternative splicing event, false discovery rate-corrected p value, and difference in percent
1268 spliced in (PSI) between U2AF1^{S34F}-mutant and parental A549 cells. b) Representative sashimi
1269 plots[80] for the alternative splicing of *KRAS* exon 2 in A549 cells with U2AF1^{S34S} or U2AF1^{S34F}
1270 mutations. Sashimi plot was generated using Integrative Genomics Viewer[64]. c) Quantification of
1271 the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion across parental A549 cells or
1272 those harboring S34S, S34C, S34A, S34Y, S34F or S34F(TTC) mutations (n=4 clones for
1273 each). Significance shown for S34F (q ratio=4.25, DF=21, p=0.0019) and S34F(TTC) (q
1274 ratio=3.89, DF=21, p=0.0043). d) Quantification of fraction of RNA-sequencing reads with *KRAS*
1275 exon 2 inclusion for parental A549 cells (n= 4 clones), U2AF1^{S34F}-mutant cells (n= 4 clones), or
1276 cells whose U2AF1^{S34F} mutations have been reverted to wildtype by prime editing (n= 6 clones).
1277 Statistical comparison between parental A549 cells and U2AF1^{S34F}-mutant cells (t=3.67, df=6,
1278 95% confidence interval=0.060 to 0.302, p=0.010) and between U2AF1^{S34F}-mutant cells and
1279 backedited cells (t=2.44, df=8, 95% confidence interval= -0.291 to -0.0083, p=0.010) are shown.
1280 e) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion for NCI-
1281 H441 cells with U2AF1^{S34F} or U2AF1$^{S34F(TTC)}$ mutations (n= 7 clones) or wildtype U2AF1 or
1282 U2AF1^{S34S} mutations (n=8 clones) (t=0.46, df=13, 95% confidence interval= -0.010 to 0.007,
1283 p=0.66). f) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2 inclusion
1284 for cell lines from the Cancer Cell Line Encyclopedia (CCLE)[35] with wildtype KRAS (n=835),
1285 KRAS^{G12A} (n=8), KRAS^{G12C} (n=23), KRAS^{G12D} (n=53), KRAS^{G12R} (n=6), KRAS^{G12S} (n=3),
1286 KRAS^{G12V} (n=34), KRAS^{G13} (n=17), or KRAS^{Q61} (n=14) mutations. Statistical analysis comparing
1287 KRAS^{G12S} to all other groups is shown (q ratio>19.27, DF=984, p=3.31x10^{-13} for all
1288 comparisons). g) Quantification of the fraction of RNA-sequencing reads with *KRAS* exon 2
1289 inclusion for pan-cancer patient samples from The Cancer Genome Atlas (TCGA)[24] with
1290 wildtype KRAS (n=5570), KRAS^{G12A} (n=24), KRAS^{G12C} (n=66), KRAS^{G12D} (n=120), KRAS^{G12R}
1291 (n=30), KRAS^{G12S} (n=15), KRAS^{G12V} (n=109), KRAS^{G13} (n=52), or KRAS^{Q61} (n=22) mutations.
1292 Statistical analysis comparing KRAS^{G12S} to all other groups is shown (q ratio>20.76, DF=5999,
1293 p=1.71x10^{-12} for all comparisons). Patient sample containing both a U2AF1^{S34F} mutation and
1294 concurrent KRAS^{G12S} mutation is marked in red.

Figure 2


a
Predicted U2AF1 Binding
WT
S34F
KRAS (WT)
Protein sequence: V V V G A G G V G K S
RNA sequence: GUGGUAGUUGGAGCUGGUGGCGUAGGCAAGAGU
– –
KRAS (G12S, AGT codon)
Protein sequence: V V V G A S G V G K S
RNA sequence: GUGGUAGUUGGAGCUAGUGGCGUAGGCAAGAGU
+ –
KRAS (G12S, TCT codon)
Protein sequence: V V V G A S G V G K S
RNA sequence: GUGGUAGUUGGAGCUUCUGGCGUAGGCAAGAGU
– –
b
AGT/AGT
AGT/TCT
TCT/TCT
bp
500
400
300
200
100
Exon 2 included
Exon 2 skipped
c
Fraction Exon 2 included
1.0
0.5
0.0
AGT/AGT
AGT/TCT
TCT/TCT
KRAS G12S codon
d
Downstream #5
Downstream #4
G12 #1
G12 #2
3' splice site #1
3' splice site #2
KRAS exon 2
bp
500
400
300
200
100
Exon 2 included
Exon 2 skipped
Untransfected
Nontargeting
mCherry
3' splice site #1
3' splice site #2
G12 #1
G12 #2
Downstream #4
Downstream #5
Untransfected
e
Fraction Exon 2 included
1.00
0.95
0.90
0.85
Untransfected
Nontargeting/mCherry
3' splice site
G12
Downstream
f
Fraction Exon 2 Included
1.0
0.9
0.8
0.7
0.6
0.5
AG dinucleotide – + – – – – –
WT Exon 2
G12S
G12S(TCT)
G12A
G12C
G12D
G12V
KRAS Exon 2 Splicing Reporter
g
Fraction Exon 2 Included
1.0
0.8
0.6
0.4
0.2
0.0
AG dinucleotide – + – + – + – + –
WT Exon 2
G10R(AGA)
G10R(CGA)
A11A(GCA)
A11A(GCC)
G13S(AGC)
G13S(TCC)
F28L(TTA)
F28L(TTG)
KRAS Exon 2 Splicing Reporter

1295 **Fig. 2:** The RNA sequence encoding KRAS^{G12S} creates a cryptic U2AF1 binding site which
1296 disrupts normal exon 2 inclusion. a) RNA and protein sequence for wildtype KRAS and
1297 KRAS^{G12S} indicating a cryptic U2AF1 binding site which is predicted to be bound by wildtype
1298 U2AF1 but not U2AF1^{S34F}. KRAS^{G12S} using an alternative codon for serine (TCT) does not
1299 resemble a U2AF1 binding site. b) RT-PCR detection of *KRAS* exon 2 skipping for A549 cells
1300 with KRAS^{G12S} mutations using the codons AGT/AGT (n=1 clone), AGT/TCT (n=1 clone) or
1301 TCT/TCT (n=3 clones). c) Quantification of the fraction of RNA-sequencing reads with *KRAS*
1302 exon 2 inclusion in A549 cells with AGT/AGT, AGT/TCT or TCT/TCT (n = 3 clones) codons. d)
1303 Schematic of dCasRx sgRNA locations and agarose gel for RT-PCR detection of *KRAS* exon 2
1304 skipping in NCI-H2023 cells. Cells were transfected with dCasRx in addition to nontargeting
1305 sgRNA (n=1) or sgRNAs targeting mCherry RNA (n=1), the 3' splice site of *KRAS* RNA (n=2),
1306 the G12 sequence of *KRAS* RNA (n=2), or downstream regions of *KRAS* RNA (n=2). e)
1307 Quantification of the fraction of *KRAS* exon 2 inclusion from d) in untransfected NCI-H2023 cells
1308 compared to cells following treatment with dCasRx alongside nontargeting or mCherry targeting
1309 sgRNA (n=2), 3' splice site sgRNA (n=2), G12 sgRNA (n=2) or downstream sgRNA (n=2). f)
1310 Quantification of the fraction of *KRAS* Exon 2 inclusion using a *KRAS* exon 2 splicing reporter
1311 and measured by RT-PCR from Extended Data Fig. 7c. 293T cells were transfected in biological
1312 duplicate with splicing reporter DNA containing wildtype *KRAS* exon 2 (n=2), or *KRAS* exon 2
1313 with common mutations including G12S (n=2), G12S(TCT) (n=2), G12A (n=2), G12C (n=2),
1314 G12D (n=2) or G12V (n=2). g) Quantification of the fraction of *KRAS* Exon 2 inclusion using a
1315 *KRAS* exon 2 splicing reporter and measured by RT-PCR from Extended Data Fig. 7d. 293T
1316 cells were transfected in biological duplicate with splicing reporter DNA containing wildtype
1317 *KRAS* exon 2 (n=2), or *KRAS* exon 2 with mutations introducing novel U2AF1 binding sites (AG
1318 dinucleotides) or synonymous mutations without U2AF1 binding sites. Mutations that generated
1319 novel AG dinucleotides include G10R(AGA) (n=2), A11A(GCA) (n=2), G13S(AGC) (n=2) and
1320 F28L(TTA) (n=2), while those that do not contain AG dinucleotides include G10R(CGA) (n=2),
1321 A11A(GCC) (n=2), G13S(TCC) (n=2), and F28L(TTG) (n=2).

Figure 3

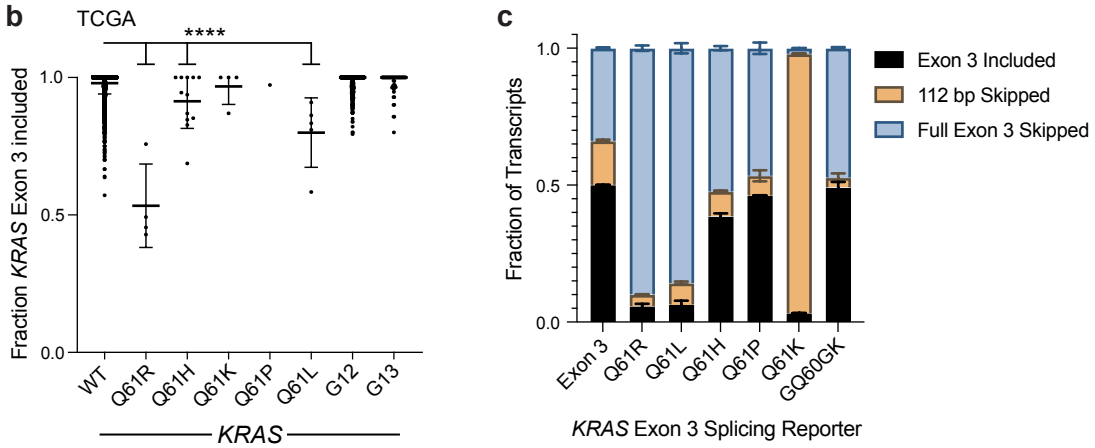

a
A549 Parental
A549 S34S
A549 S34F
p-Erk (Fluorescence Intensity)
b
p= 0.0041
p-Erk (Mean Fluorescence Intensity)
A549/U2AF1S34S
U2AF1S34F
Genotype
c
SR proteins
Ex 1
Ex 2
SSO
Ex 3
ESE
Scramble SSO
KRAS Exon 2 SSO
bp
120 nM
6 nM
12 nM
30 nM
60 nM
120 nM
Exon 2 included
Exon 2 skipped
d
Fraction KRAS Exon 2
Exon 2 included
Exon 2 skipped
0 nM
6 nM
12 nM
30 nM
60 nM
120 nM
e
Scramble SSO
KRAS Exon 2 SSO
kDa
HSP90
p-Erk
Total Erk
p-Akt
Total Akt
Full-length KRAS
f
Normalized luminescence
Time (days)
Scramble SSO (120 nM)
KRAS Exon 2 SSO (6 nM)
KRAS Exon 2 SSO (12 nM)
KRAS Exon 2 SSO (120 nM)
g
Lung adenocarcinoma
Odds ratio of co-occurrence with U2AF1S34F mutation
–log10(p-value)
G12S
G13V
G12C
G12V
G12A
Q61R
Q61H
G12R
G13C
G12D
G13D
KRAS mutation
h
R2 = 0.5257
p = 3.826 × 10–8
U2AF1S34F variant allele frequency
KRASG12S variant allele frequency

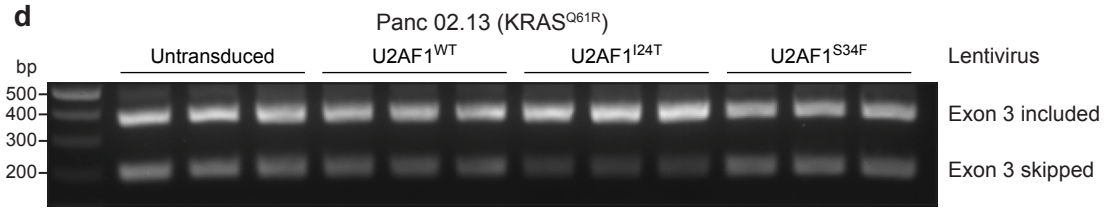
1322 **Fig. 3:** *KRAS* exon 2 skipping limits MAPK signaling and cell growth, leading to the positive
1323 selection of U2AF1^{S34F} mutations in KRAS^{G12S}-mutant lung adenocarcinomas. a) Fluorescence
1324 intensity of p-Erk as measured by flow cytometry for parental A549 cells (n= 2 clones),
1325 U2AF1^{S34S}-mutant A549 cells (n= 4 clones), or U2AF1^{S34F}-mutant A549 cells (n= 4 clones). b)
1326 Quantification of the mean fluorescence intensity of p-Erk staining in parental A549 or
1327 U2AF1^{S34S} cells (n= 6 clones) or U2AF1^{S34F} cells (n= 4 clones) (t=3.98, df=8, 95% confidence
1328 interval = 373 to 1401, p=0.0041). c) Schematic indicates the binding of splice switching
1329 oligonucleotides (SSOs) to exonic splicing enhancers (ESEs) in *KRAS* exon 2, blocking SR
1330 protein binding and inhibiting exon 2 inclusion. RT-PCR gel demonstrates the detection of
1331 *KRAS* exon 2 skipping for NCI-H23 cells treated with scramble SSO or increasing
1332 concentrations of *KRAS* exon 2 SSO. d) Quantification of the fraction of transcripts with *KRAS*
1333 exon 2 skipping or inclusion from c). e) Immunoblot analysis of HSP90, p-Erk, total Erk, p-Akt,
1334 total Akt, and full length KRAS protein in NCI-H23 cells treated with increasing concentrations of
1335 *KRAS* exon 2 skipping SSOs 3 days after treatment. f) Cell viability of NCI-H23 cells grown on
1336 ultra-low attachment plates and treated with increasing concentrations of *KRAS* exon 2 skipping
1337 SSOs over a period of 5 days as measured by CellTiter-Glo (n= 4 technical replicates at each
1338 time point). g) Odds ratio of co-occurrence for U2AF1^{S34F} mutations and indicated KRAS
1339 mutations in lung adenocarcinoma patients using data from Foundation Medicine Inc. (n=62009
1340 patients) and AACR Project GENIE (v14.0, n=14908 patients)[11]. h) Correlation between the
1341 variant allele frequencies for KRAS^{G12S} and U2AF1^{S34F} in 43 patient samples from the
1342 Foundation Medicine Inc. dataset containing both mutations and lacking CNVs at either locus
1343 (F=45.4, DFn=1, DFd=41, R^2=0.526, p=3.83x10^{-8}). Dashed line indicates a 1:1 ratio of
1344 KRAS^{G12S} and U2AF1^{S34F} allele frequencies.

Figure 4

a

Mutation in TCGA	Co-occurring splicing 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





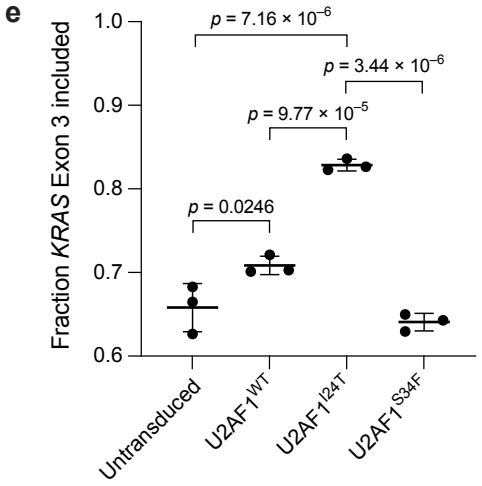
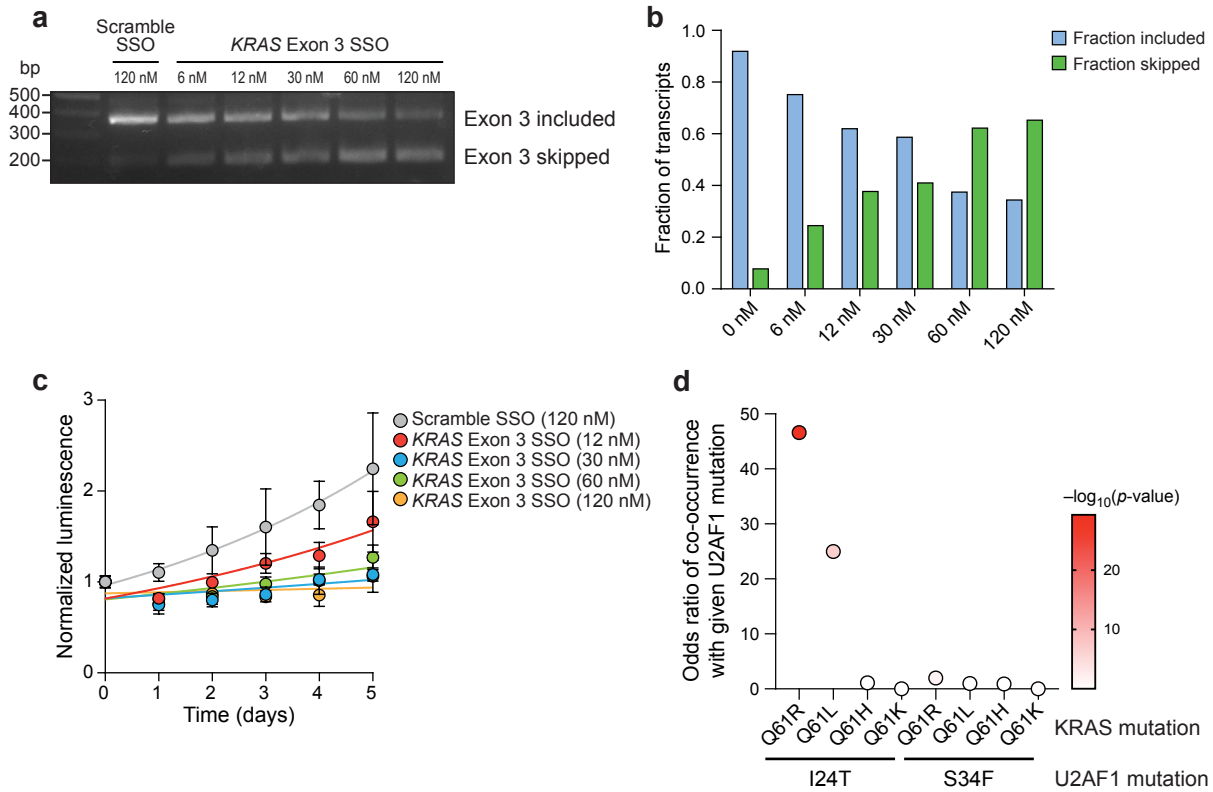


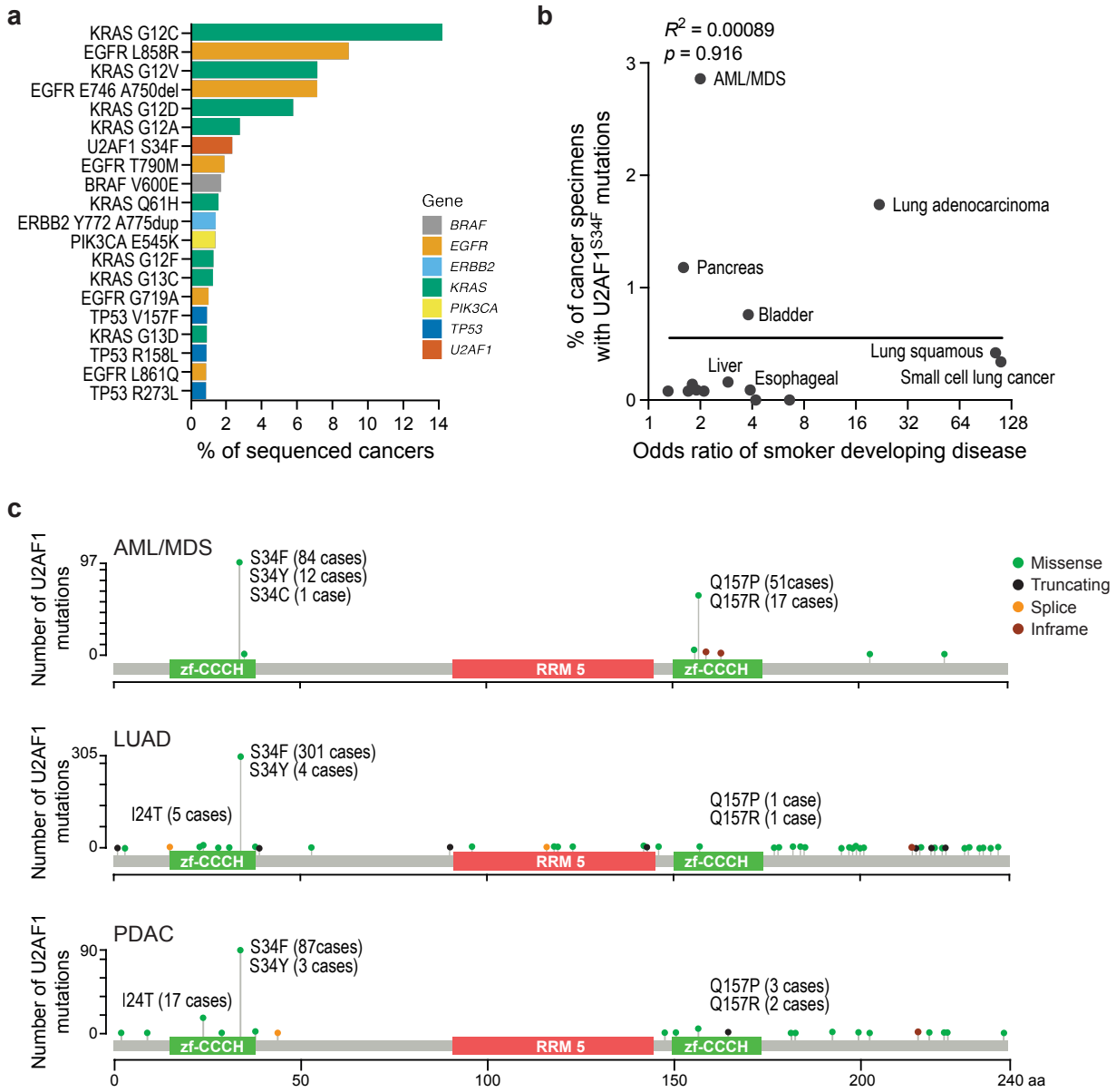
Fig. 4: $KRAS^{Q61R/L}$-mutant pancreatic cancers acquire secondary $U2AF1^{I24T}$ mutations that rescue inadvertent skipping of *KRAS* exon 3. a) Table of mutations associated with exon skipping events in TCGA data[24], and which co-occur with splicing factor mutations (data from AACR Project GENIE[11]). 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[24] 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. Statistical comparisons between wildtype KRAS and $KRAS^{Q61R}$ (q ratio=22.3, DF=6524, $p<1x10^{-15}$), $KRAS^{Q61H}$ (q ratio=5.71, DF=6524, $p=8.2x10^{-8}$), and $KRAS^{Q61L}$ (q ratio=10.08, DF=6524, $p<1x10^{-15}$) are shown. c) Quantification of the fraction of *KRAS* Exon 3 inclusion, skipping of a 112 bp fragment of exon 3, or skipping the entirety of exon 3 using a *KRAS* exon 3 splicing reporter. Splicing was measured by RT-PCR from Extended Data Fig. 9b. 293T cells were transfected in biological duplicate with splicing reporter DNA containing wildtype *KRAS* exon 3 (n=2), or *KRAS* exon 3 with common mutations including Q61R (n=2), Q61L (n=2), Q61H (n=2), Q61P (n=2), Q61K (n=2) or GQ60GK (n=2) mutation. d) RT-PCR detection of *KRAS* exon 3 skipping for untransduced $KRAS^{Q61R}$-mutant Panc 02.13 cells (n= 3 biological replicates) or after delivery of lentivirus expressing wildtype U2AF1 (n=3 biological replicates), $U2AF1^{I24T}$ (n=3 biological replicates) or $U2AF1^{S34F}$ (n=3 biological replicates). e) Quantification of the fraction of *KRAS* exon 3 skipping in Fig. 4d. Statistical comparisons for untransduced vs. $U2AF1^{WT}$ (q ratio=5.252, DF=8, p=0.0246), untransduced vs. $U2AF1^{I24T}$ (q ratio=17.76, DF=8, $p=7.16x10^{-6}$), $U2AF1^{WT}$ vs. $U2AF1^{I24T}$ (q ratio=12.51, DF=8, $p=9.77x10^{-5}$), and $U2AF1^{I24T}$ vs. $U2AF1^{S34F}$ (q ratio=19.57, DF=8, $p=3.44x10^{-6}$) are shown.

Figure 5


a
Scramble SSO
KRAS Exon 3 SSO
bp
120 nM
6 nM
12 nM
30 nM
60 nM
120 nM
500
400
300
200
Exon 3 included
Exon 3 skipped
b
Fraction of transcripts
Fraction included
Fraction skipped
0 nM
6 nM
12 nM
30 nM
60 nM
120 nM
c
Normalized luminescence
Time (days)
Scramble SSO (120 nM)
KRAS Exon 3 SSO (12 nM)
KRAS Exon 3 SSO (30 nM)
KRAS Exon 3 SSO (60 nM)
KRAS Exon 3 SSO (120 nM)
d
Odds ratio of co-occurrence with given U2AF1 mutation
–log10(p-value)
Q61R
Q61L
Q61H
Q61K
KRAS mutation
I24T
S34F
U2AF1 mutation

1368 **Fig. 5:** *KRAS* exon 3 skipping leads to reduced cell growth, while U2AF1^{I24T} mutations are
1369 enriched in KRAS$^{Q61R/L}$-mutant pancreatic cancers. a) RT-PCR detection of *KRAS* exon 3
1370 skipping for NCI-H23 cells treated with scramble SSO or increasing concentrations of *KRAS*
1371 exon 3 SSO. b) Quantification of the fraction of transcripts with *KRAS* exon 3 skipping or
1372 inclusion from a). c) Cell viability of NCI-H23 cells grown on ultra-low attachment plates and
1373 treated with increasing concentrations of *KRAS* exon 3 skipping SSOs over a period of 5 days
1374 as measured by CellTiter-Glo (n= 4 technical replicates at each time point). d) Odds ratio of co-
1375 occurrence for U2AF1^{I24T} or U2AF1^{S34F} mutations with KRAS^{Q61R}, KRAS^{Q61L}, KRAS^{Q61H} or
1376 KRAS^{Q61K} mutations in pancreatic cancers using data from Foundation Medicine Inc. (n=31,530
1377 patients) and AACR Project GENIE (v16.1, n=8,304 patients)[11].

Extended Data Fig. 1



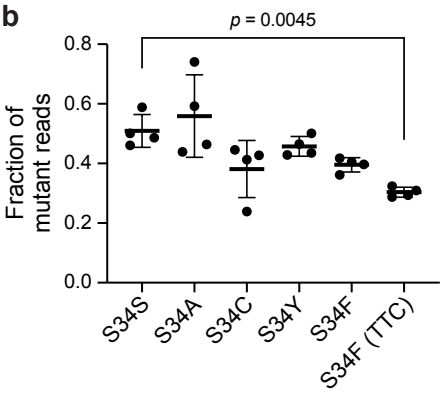
1378 **Extended Data Fig. 1:** U2AF1^{S34F} mutations are over-represented in human lung
1379 adenocarcinomas. a) Top 20 most frequent hotspot mutations in human lung adenocarcinomas
1380 based on an analysis of data from AACR Project GENIE (v15.0)[11]. b) Comparison of the
1381 percentage of cancer specimens with U2AF1^{S34F} mutations across cancer types (AACR Project
1382 GENIE v15.0)[11] compared to the odds ratio of a smoker developing that disease[14] (F=0.012,
1383 DFn=1, DFd=13,R^2=0.00089, p=0.92). c) Lollipop plots of mutations identified in the *U2AF1*
1384 gene in acute myeloid leukemia/myelodysplastic syndrome (n=181 patients), pancreatic ductal
1385 adenocarcinoma (n= 131 patients) and lung adenocarcinoma (n=380 patients) based on data
1386 from AACR Project GENIE (v16.0)[11,29].

Extended Data Fig. 2

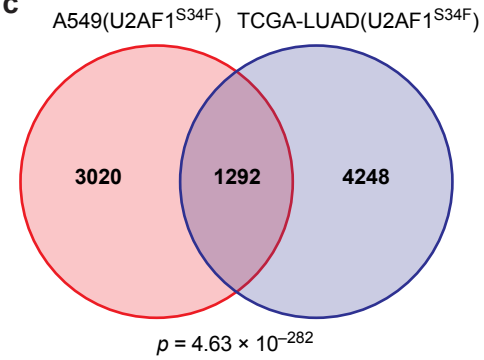
a

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

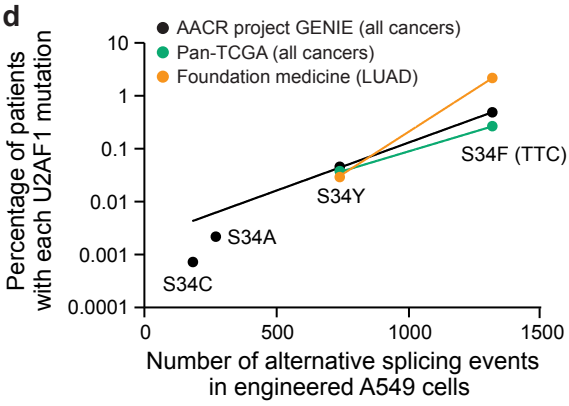
b



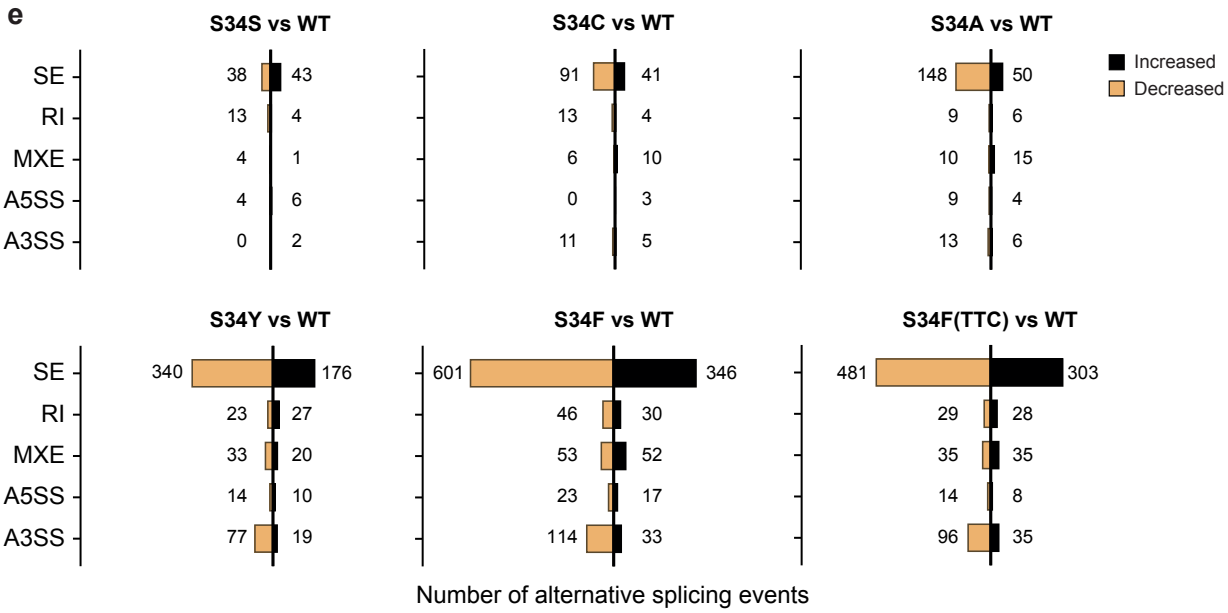
c



d

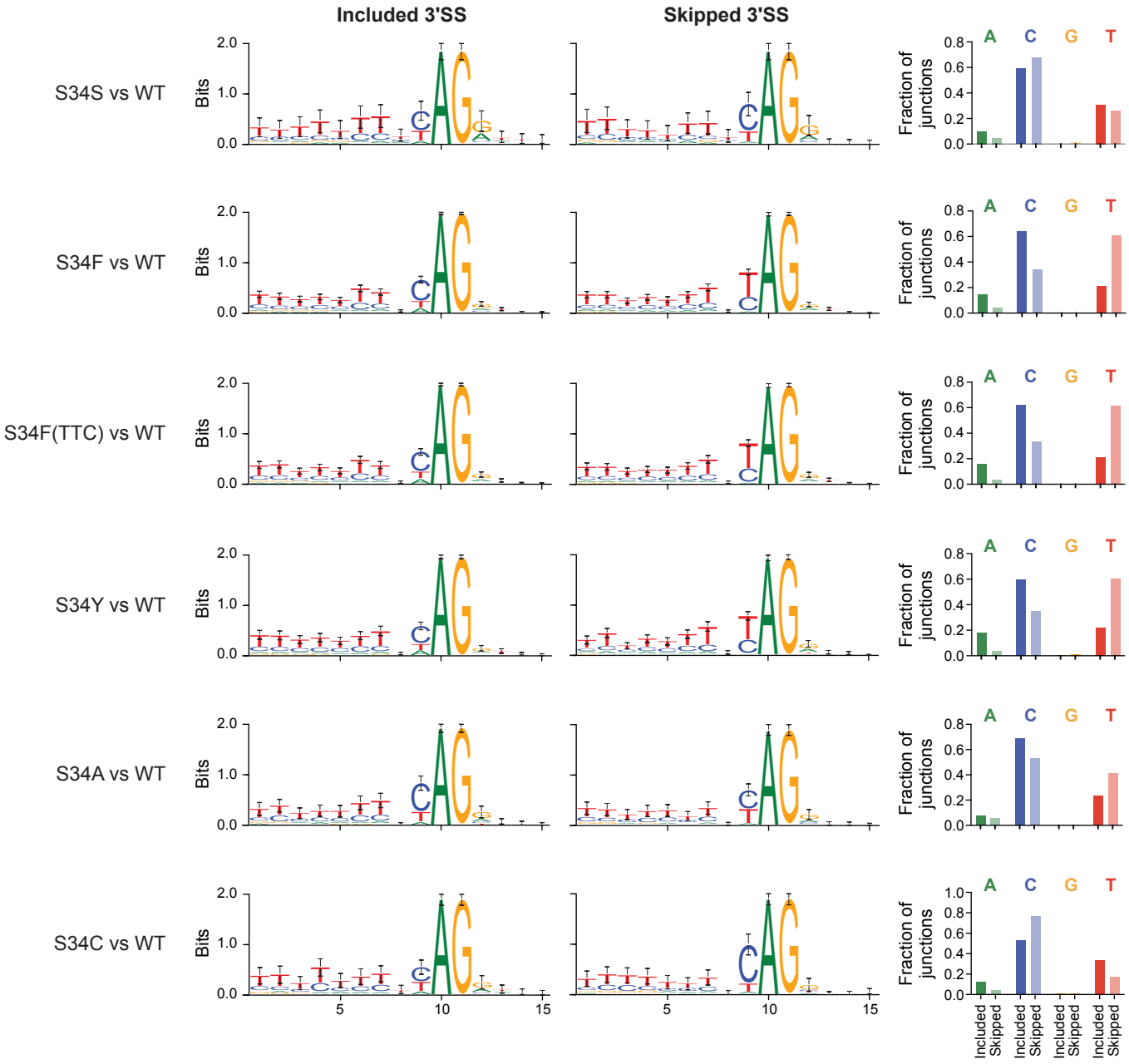


e



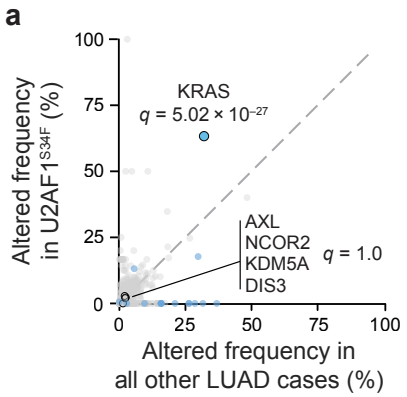
1387 **Extended Data Fig. 2:** Alternative splicing analysis of A549 cells with varying U2AF1 amino
1388 acid substitutions. a) Number of significant alternative splicing events identified in the context of
1389 each engineered *U2AF1* mutation in A549 cells (n=4 clones each). Significant alternative
1390 splicing events were identified by comparing each variant to transcripts from parental A549 cells
1391 using rMATS (v4.1.0) and defined by an FDR q value < 0.05. b) Fraction of RNA-sequencing
1392 reads for each engineered *U2AF1* variant sequence relative to wildtype *U2AF1* in A549 cells
1393 (n=4 clones each, q ratio=3.90, DF=18, p=0.0045 for S34S vs. S34F(TTC)). c) Overlap of
1394 significant alternative splicing events between mRNA from A549 cells with an engineered
1395 $U2AF1^{S34F}$ mutation (n=4 clones compared to 4 parental A549 cell controls) and mRNA from
1396 $U2AF1^{S34F}$-mutant lung adenocarcinoma patient samples from TCGA[24] (n=7 cases compared to
1397 100 randomly sampled controls with wildtype *U2AF1*) (odds ratio=4.18, $p=4.63x10^{-282}$). d)
1398 Correlation between the number of alternative splicing events that occur in the presence of each
1399 amino acid substitution in A549 cells and the frequency of each substitution across all human
1400 cancers as determined by the AACR Project GENIE dataset (v15.1), TCGA Pan-Cancer Atlas,
1401 or Foundation Medicine Inc. LUAD data[11,25,26]. e) Number of significant alternative splicing
1402 events for A549 cells with S34S (n=4 clones), S34C (n=4 clones), S34A (n=4 clones), S34Y
1403 (n=4 clones), S34F (n=4 clones) or S34F(TTC) (n=4 clones) amino acid substitutions in U2AF1,
1404 compared to parental A549 cells (n=4 clones) as determined by rMATS (v4.1.0)[63]. Detected
1405 splicing events consist of 5 categories: skipped exons (SE), retained introns (RI), mutually
1406 exclusive exons (MXE), alternative 5' splice site usage (A5SS) and alternative 3' splice site
1407 usage (A3SS).

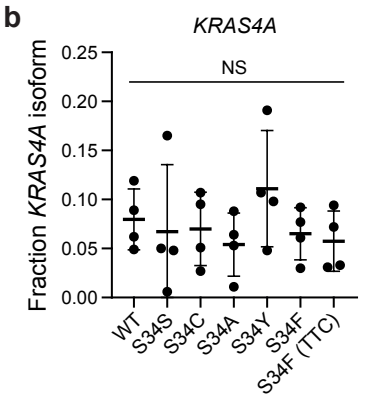
Extended Data Fig. 3

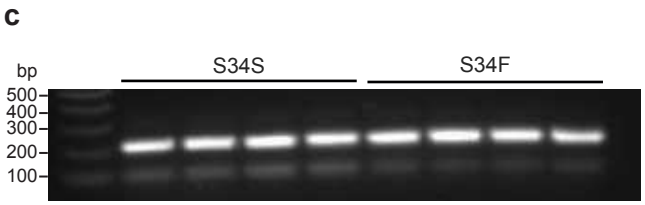


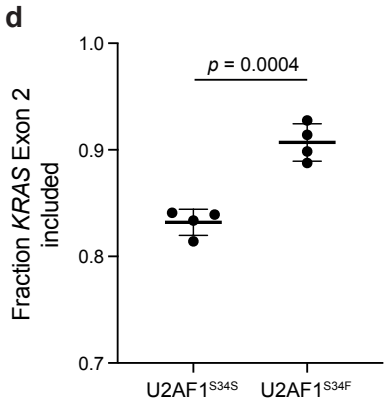
1408 **Extended Data Fig. 3:** 3' splice site preferences for all engineered U2AF1 variants. Splice site
1409 preferences were determined by examining the sequence surrounding the 3' splice site in
1410 significantly included exons compared to the alternative skipped exon. This was performed for
1411 S34S (n=81 splicing events), S34F (n=947 splicing events), S34F(TTC) (n=784 splicing events),
1412 S34Y (n=516 splicing events), S34A (n=198 splicing events), and S34C (n=132 splicing events)
1413 cells compared to parental A549 cells, and WebLogos[66] were produced for each data set. The
1414 fraction of included (opaque bars) and skipped (partially translucent bars) junctions with a given
1415 base preceding the AG dinucleotide is displayed on the right.

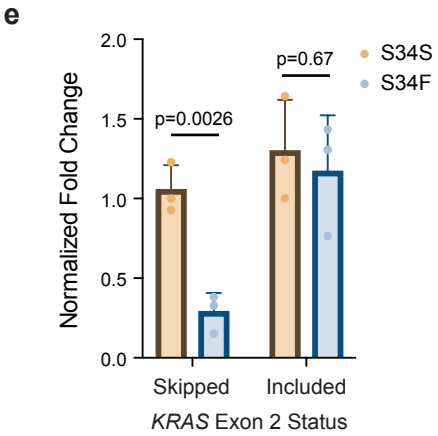
Extended Data Fig. 4

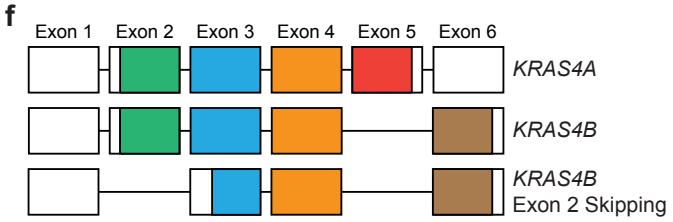


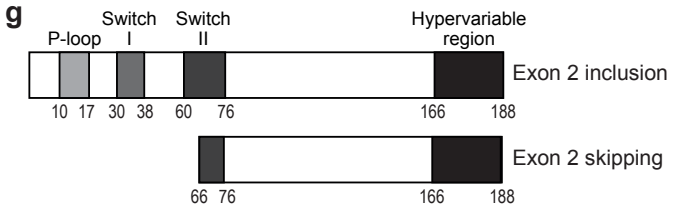






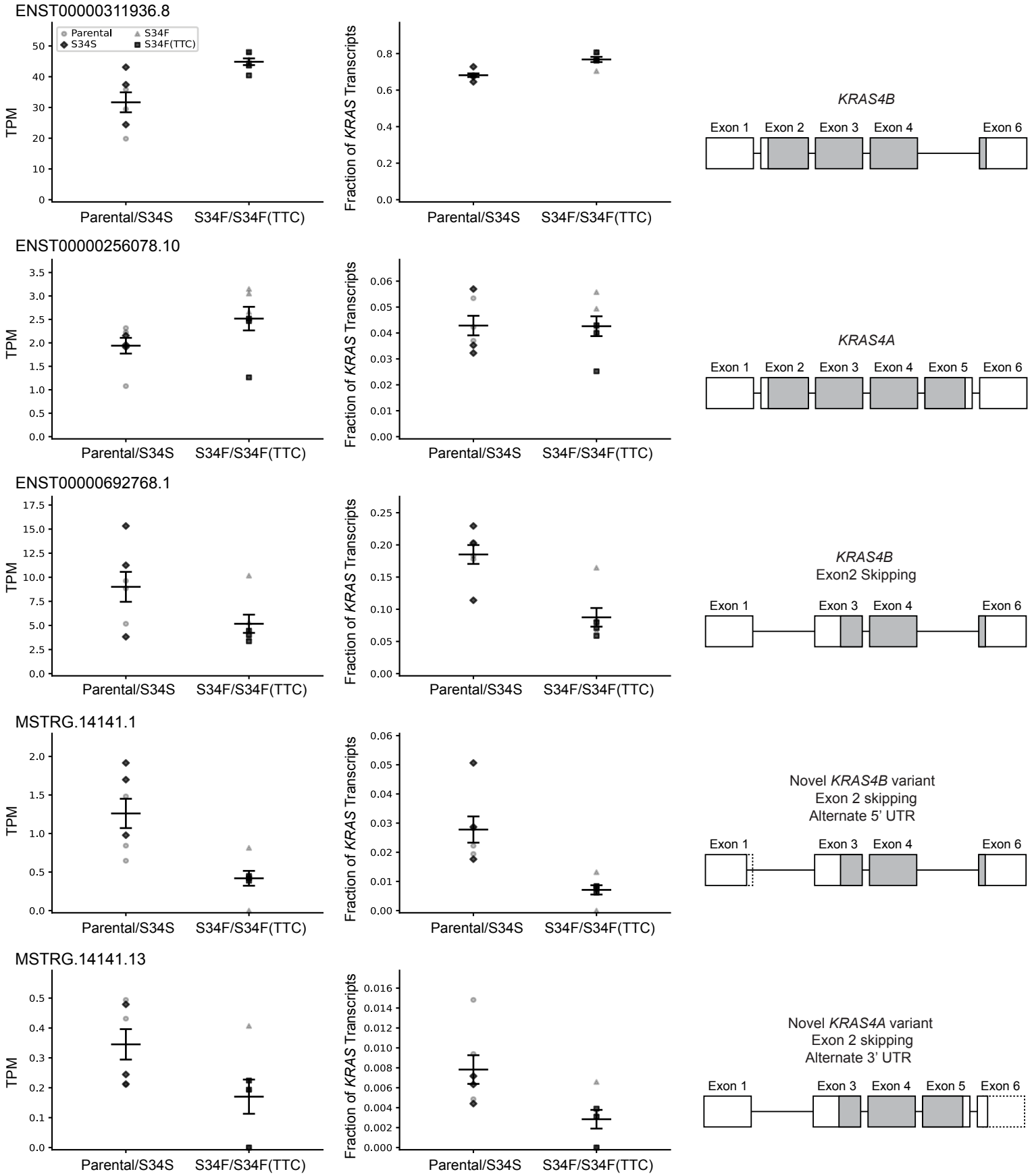






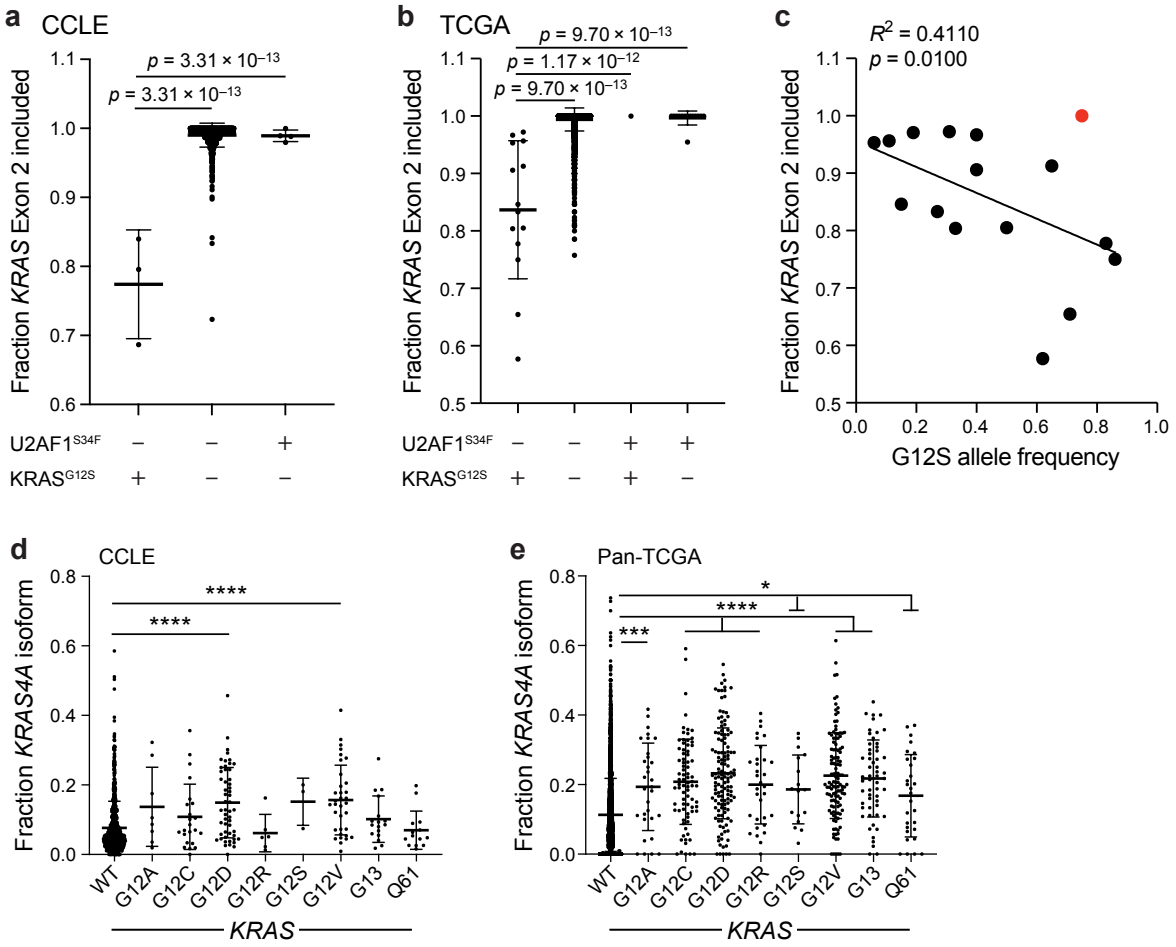
1416 **Extended Data Fig. 4:** KRAS^{G12S}-mutant cells undergo *KRAS* exon 2 skipping which can be
1417 reversed by U2AF1^{S34F} mutations. a) Frequency of genetic alterations in U2AF1^{S34F}-mutant
1418 cancers compared to all lung adenocarcinomas as determined using the AACR Project GENIE
1419 dataset (v15.0)[11,29]. b) Quantification of the fraction of RNA-sequencing reads with *KRAS4A*
1420 isoform usage across parental A549 cells or those harboring S34S, S34C, S34A, S34Y, S34F
1421 or S34F(TTC) mutations (n=4 clones for each). Statistical analysis comparing each variant to
1422 parental A549 cells is shown (q ratio <1.02, DF=21, p>0.81 for all variants). c) RT-PCR
1423 detection of *KRAS* exon 2 skipping in A549 cells with U2AF1^{S34S} (n=4 clones) or U2AF1^{S34F} (n=4
1424 clones) mutations. d) Quantification of the fraction of *KRAS* transcripts with exon 2 inclusion
1425 from c) (t=6.99, df=6, 95% confidence interval = 0.0487 to 0.101, p=0.0004). e) qRT-PCR
1426 quantification of *KRAS* exon 2 skipping or inclusion in A549 cells with U2AF1^{S34S} (n=3 clones) or
1427 U2AF1^{S34F} (n=3 clones) mutations. Fold change is normalized to Beta-actin expression.
1428 Statistical analysis comparing the amount of *KRAS* exon 2 skipping or inclusion in A549 cells
1429 with U2AF1^{S34S} or U2AF1^{S34F} mutations is shown (exon skipping: t=6.69, df=4, 95% confidence
1430 interval = -1.082 to -0.448, p=0.0026, exon inclusion: t=0.461, df=4, 95% confidence interval = -
1431 0.900 to 0.644, p=0.67). f) Diagram of *KRAS* gene structure showing exons used in *KRAS4A*,
1432 *KRAS4B*, and *KRAS4B* with exon 2 skipping. White segments represent untranslated regions.
1433 When exon 2 skipping occurs, this results in skipping of the translation start site and predicted
1434 use of an internal translation start site in exon 3. Diagram modified from Raso *et al.*[81] g)
1435 Diagram of KRAS protein structure with and without exon 2 skipping. When exon 2 skipping
1436 occurs, translation is predicted to begin at an internal translation site at amino acid 66. Diagram
1437 modified from Kwan *et al.*[82].

Extended Data Fig. 5



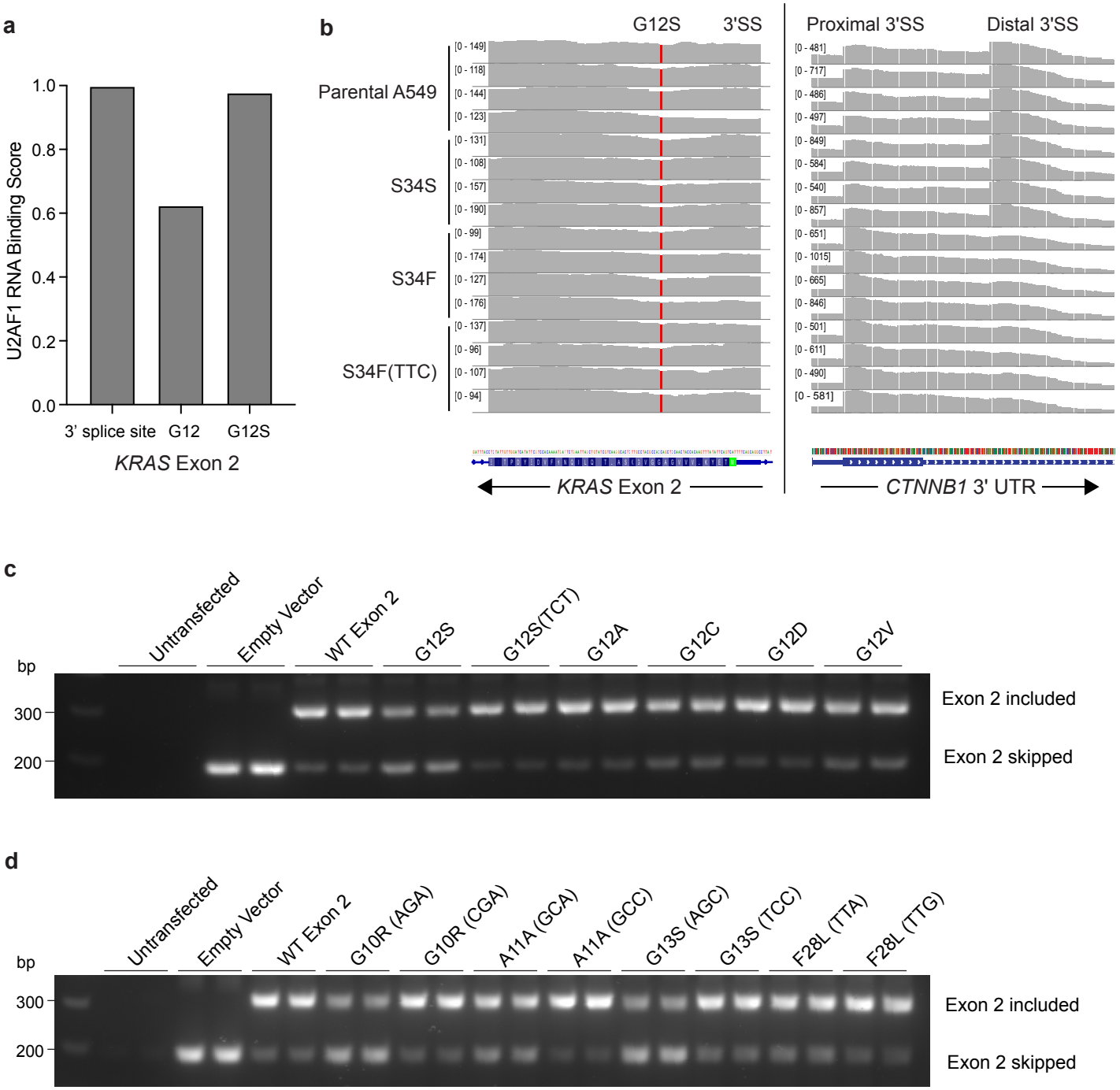
1438 **Extended Data Fig. 5:** Long-read transcript usage for *KRAS4A, KRAS4B* and annotated *KRAS*
1439 exon 2 skipping transcripts. MSTRG14141.1 is a *KRAS4B* exon 2 skipping transcript observed
1440 using an alternative 5' UTR. MSTRG14141.13 is a *KRAS4A* exon 2 skipping transcript observed
1441 using an alternative 3' UTR. Transcripts per million (TPM) and fraction of *KRAS* transcripts are
1442 shown for each transcript. For each splicing diagram, boxes represent included exons, grey
1443 shading represents the predicted translation product, and dashed lines indicate alternative UTR
1444 usage.

Extended Data Fig. 6



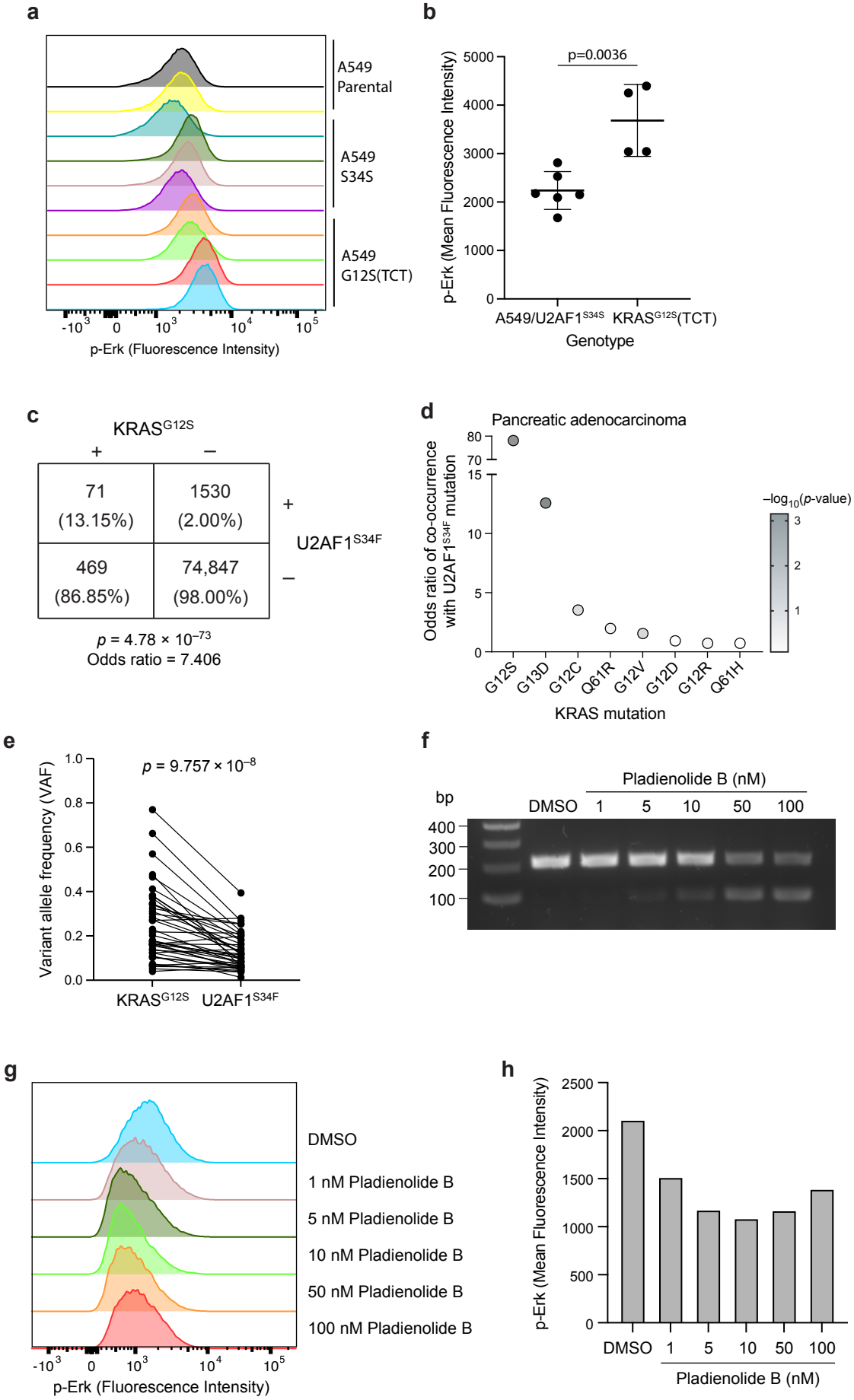
1445 **Extended Data Fig. 6:** $KRAS^{G12S}$-mutant cancer cell lines and patient tumors exhibit *KRAS*
1446 exon 2 skipping. a) Additional analysis of CCLE data from Fig. 1f, comparing *KRAS* exon 2
1447 skipping in samples with $KRAS^{G12S}$ and wildtype U2AF1 (n=3 cases), wildtype KRAS and
1448 wildtype U2AF1 (n=1001 cases), and wildtype KRAS and $U2AF1^{S34F}$ (n=4 cases). Statistical
1449 comparison of wildtype U2AF1 + $KRAS^{G12S}$ samples vs. wildtype U2AF1 + wildtype KRAS (q
1450 ratio=29.76, DF=1005, $p=3.31x10^{-13}$) and comparison of wildtype U2AF1 + $KRAS^{G12S}$ vs.
1451 $U2AF1^{S34F}$ + wildtype KRAS (q ratio=22.44, DF=1005, $p=3.31x10^{-13}$) are shown. b) Additional
1452 analysis of TCGA data from Fig. 1g, comparing *KRAS* exon 2 skipping in samples with
1453 $KRAS^{G12S}$ and wildtype U2AF1 (n=14 cases), wildtype KRAS and wildtype U2AF1 (n=6028
1454 cases), $KRAS^{G12S}$ and $U2AF1^{S34F}$ (n=1 case), and wildtype KRAS and $U2AF1^{S34F}$ (n=14 cases).
1455 Statistical comparisons between wildtype U2AF1 + $KRAS^{G12S}$ samples and wildtype U2AF1 +
1456 wildtype KRAS samples (q ratio=40.19, DF=6053, $p=9.70x10^{-13}$), $U2AF1^{S34F}$ + $KRAS^{G12S}$
1457 samples (q ratio=10.77, DF=6053, $p=1.17x10^{-12}$) and $U2AF1^{S34F}$ + wildtype KRAS samples (q
1458 ratio=28.92, DF=6053, $p=9.70x10^{-13}$) are shown. c) Correlation between the fraction of *KRAS*
1459 exon 2 inclusion and variant allele frequency of $KRAS^{G12S}$ in pan-cancer patients from The
1460 Cancer Genome Atlas dataset (n=16 cases, F=9.07, DFn=1, DFd=13, $R^2=0.411$, p=0.0100)[24].
1461 Red sample indicates a $KRAS^{G12S}$-mutant sample which also contains a $U2AF1^{S34F}$ mutation,
1462 and was excluded from statistical analyses. d) Quantification of the fraction of RNA-Sequencing
1463 reads with *KRAS4A* isoform usage for cell lines from (CCLE)[35] with wildtype KRAS (n=836),
1464 $KRAS^{G12A}$ (n=8), $KRAS^{G12C}$ (n=23), $KRAS^{G12D}$ (n=53), $KRAS^{G12R}$ (n=6), $KRAS^{G12S}$ (n=3),
1465 $KRAS^{G12V}$ (n=34), $KRAS^{G13}$ (n=17), or $KRAS^{Q61}$ (n=14) mutations. Statistical comparisons
1466 between wildtype KRAS and $KRAS^{G12D}$ (q ratio=6.48, DF=985, $p=1.14x10^{-9}$), and between
1467 wildtype KRAS and $KRAS^{G12V}$ (q ratio=5.79, DF=985, $p=7.54x10^{-8}$) are shown. e) Quantification
1468 of the fraction of RNA-Sequencing reads with *KRAS4A* isoform usage for pan-cancer patient
1469 samples from (TCGA)[24] comparing wildtype KRAS (n=7022) to $KRAS^{G12A}$ (n=31, q ratio=4.21,
1470 DF=7509, p=0.0002), $KRAS^{G12C}$ (n=79, q ratio=7.91, DF=7509, $p<1x10^{-15}$), $KRAS^{G12D}$ (n=131, q
1471 ratio=12.73, DF=7509, $p<1x10^{-15}$), $KRAS^{G12R}$ (n=32, q ratio=4.60, DF=7509, $p=3.41x10^{-5}$),
1472 $KRAS^{G12S}$ (n=17, q ratio=2.83, DF=7509, p=0.0364), $KRAS^{G12V}$ (n=121, q ratio=11.58,
1473 DF=7509, $p<1x10^{-15}$), $KRAS^{G13}$ (n=56, q ratio=7.32, DF=7509, $p<1x10^{-15}$), or $KRAS^{Q61}$ (n=29, q
1474 ratio=2.77, DF=7509, p=0.0435) mutations.

Extended Data Fig. 7



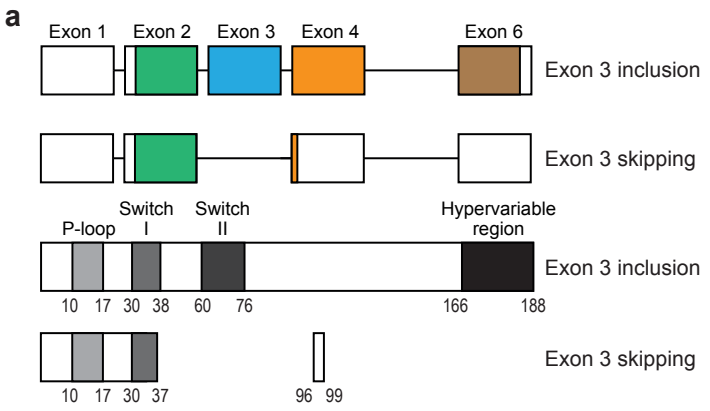
1475 **Extended Data Fig. 7:** *KRAS* undergoes an exon 2 skipping event caused by the creation of a
1476 U2AF1 binding site upon $KRAS^{G12S}$ mutation. a) U2AF1 binding score for an RNA sequence
1477 centered on the *KRAS* exon 2 3' splice site, wildtype *KRAS* amino acid G12, or mutant *KRAS*
1478 G12S. Binding score was determined by RBPsuite (v1.0)[37]. b) Read count distribution across
1479 *KRAS* exon 2 or *CTNNB1* 3'UTR in parental A549 cells (n=4 clones) or those with $U2AF1^{S34S}$
1480 (n=4 clones), $U2AF1^{S34F}$ (n=4 clones) or $U2AF1^{S34F(TTC)}$ (n=4 clones) mutations. Reads were
1481 visualized using Integrative Genomics Viewer[64]. Red nucleotide in *KRAS* exon 2 indicates the
1482 homozygous $KRAS^{G12S}$ mutation present in A549 cells. Arrow indicates the directionality of the
1483 gene. c) RT-PCR detection of *KRAS* exon 2 skipping for 293T cells transfected with *KRAS* exon
1484 2 splicing reporter constructs. Cells were either untransfected (n=2), transfected with an empty
1485 vector without *KRAS* exon 2 (n=2), transfected with a wildtype *KRAS* exon 2 splicing reporter
1486 (n=2), or *KRAS* exon 2 with G12S (n=2), G12S(TCT) (n=2), G12A (n=2), G12C (n=2), G12D
1487 (n=2) or G12V (n=2) mutations. d) RT-PCR detection of *KRAS* exon 2 skipping for 293T cells
1488 transfected with *KRAS* exon 2 splicing reporter constructs. Cells were either untransfected
1489 (n=2), transfected with an empty vector without *KRAS* exon 2 (n=2), transfected with a wildtype
1490 *KRAS* exon 2 splicing reporter (n=2), or *KRAS* exon 2 with mutations introducing novel U2AF1
1491 binding sites (AG dinucleotides) or synonymous mutations without U2AF1 binding sites.
1492 Mutations that generated novel AG dinucleotides include G10R(AGA) (n=2), A11A(GCA) (n=2),
1493 G13S(AGC) (n=2) and F28L(TTA) (n=2), while those that do not contain AG dinucleotides
1494 include G10R(CGA) (n=2), A11A(GCC) (n=2), G13S(TCC) (n=2), and F28L(TTG) (n=2).

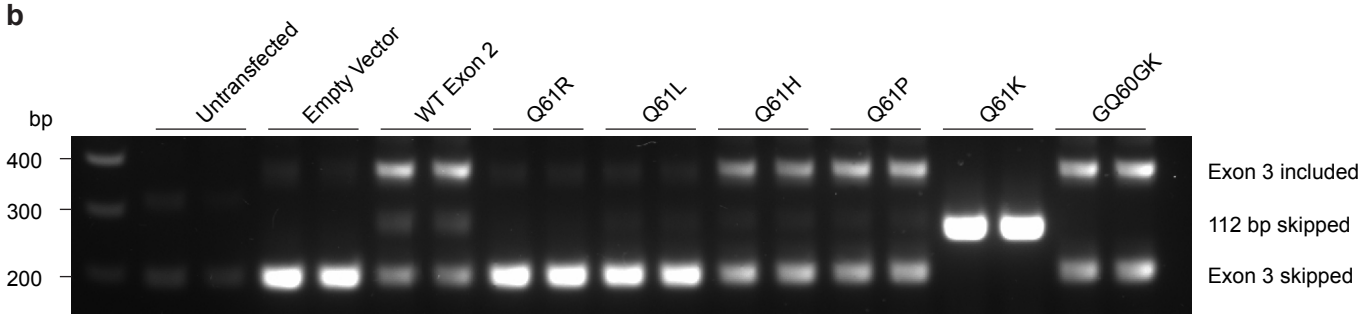
Extended Data Fig. 8



1495 **Extended Data Fig. 8:** U2AF1^{S34F} mutations are enriched as secondary mutations in KRAS^{G12S}-
1496 mutant cancers. a) Fluorescence intensity of p-Erk as measured by flow cytometry for parental
1497 A549 cells (n= 2 clones), U2AF1^{S34S}-mutant A549 cells (n= 4 clones), or KRAS$^{G12S(TCT)}$-mutant
1498 A549 cells (n= 4 clones). b) Quantification of the mean fluorescence intensity of p-Erk staining
1499 in parental A549 or U2AF1^{S34S} cells (n= 6 clones) or U2AF1^{S34F} cells (n= 4 clones) (t=4.07, df=8,
1500 95% confidence interval = 627 to 2261, p=0.0036). c) Contingency plot analysis of U2AF1^{S34F}
1501 and KRAS^{G12S} mutations in lung adenocarcinoma patients from Foundation Medicine Inc.
1502 (n=62009 patients) and AACR Project GENIE (v14.0, n=14908 patients)[11], including column
1503 percentages (odds ratio=7.406, p=4.78x10^{-73}). d) Odds ratio of co-occurrence for U2AF1^{S34F}
1504 mutations and indicated KRAS mutations in pancreatic adenocarcinoma patients using data
1505 from AACR Project GENIE (v15.0, n=6528 patients)[11]. e) Paired analysis of the variant allele
1506 frequency of KRAS^{G12S} and U2AF1^{S34F} across 43 patient samples from the Foundation Medicine
1507 Inc. dataset containing both mutations and lacking CNVs at either locus (t=6.42, df=42, 95%
1508 confidence interval= -0.155 to -0.081, p= 9.757x10^{-8}). f) RT-PCR detection of *KRAS* exon 2
1509 skipping in NCI-H2023 cells treated with DMSO or 1 nM, 5 nM, 10 nM, 50 nM or 100 nM of the
1510 SF3B1 inhibitor Pladienolide B. g) Fluorescence intensity of p-Erk as measured by flow
1511 cytometry for NCI-H2023 cells treated with DMSO or 1 nM, 5 nM, 10 nM, 50 nM or 100 nM of
1512 Pladienolide B. h) Quantification of the mean fluorescence intensity of p-Erk staining in NCI-
1513 H2023 cells treated with DMSO or 1 nM, 5 nM, 10 nM, 50 nM or 100 nM of Pladienolide B.

Extended Data Fig. 9





1514 **Extended Data Fig. 9:** $KRAS^{Q61R/L}$ mutations lead to *KRAS* exon 3 skipping. a) Diagram of
1515 *KRAS* gene and protein structure. Top shows exons used in *KRAS4B* with and without exon 3
1516 skipping. White segments represent untranslated regions. When exon 3 skipping occurs, this
1517 results in early translation termination in exon 4. Bottom shows the KRAS protein structure with
1518 and without exon 3 skipping. When exon 3 skipping occurs, translation is predicted to terminate
1519 after amino acid 37, producing 3 incorrect amino acids before terminating. b) RT-PCR detection
1520 of *KRAS* exon 3 skipping for 293T cells transfected with *KRAS* exon 3 splicing reporter
1521 constructs. Cells were either untransfected (n=2), transfected with an empty vector without
1522 *KRAS* exon 3 (n=2), transfected with a wildtype *KRAS* exon 3 splicing reporter (n=2), or *KRAS*
1523 exon 3 with Q61R (n=2), Q61L (n=2), Q61H (n=2), Q61P (n=2), Q61K (n=2) or GQ60GK (n=2)
1524 mutations.

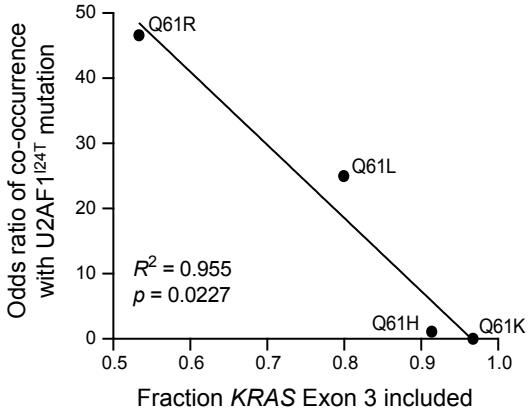
Extended Data Fig. 10

a

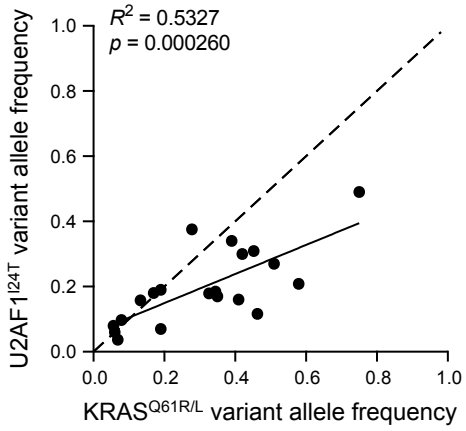
	KRAS^{Q61R} +	KRAS^{Q61R} –
U2AF1^{I24T} +	25 (4.96%)	44 (0.11%)
U2AF1^{I24T} –	479 (95.04%)	39,286 (99.89%)

$p = 4.90 \times 10^{-30}$
Odds ratio = 46.60

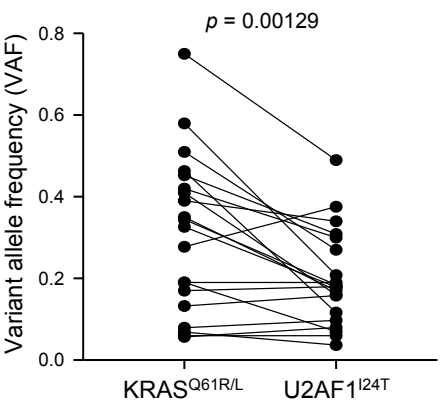
b



c



d



1525 **Extended Data Fig. 10:** U2AF1^{I24T} mutations are enriched as secondary mutations in
1526 KRAS$^{Q61R/L}$-mutant pancreatic cancers. a) Contingency plot analysis of U2AF1^{I24T} and KRAS^{Q61R}
1527 mutations in pancreatic cancer patients from Foundation Medicine Inc. (n=31,530 patients) and
1528 AACR Project GENIE (v16.1, n=8,304 patients)[11], including column percentages (odds
1529 ratio=46.60, p=4.90x10^{-30}). b) Odds ratio of co-occurrence of various KRAS^{Q61} mutations with
1530 U2AF1^{I24T} from Foundation Medicine Inc. (n=31,530 patients) and AACR Project GENIE (v16.1,
1531 n=8,304 patients)[11] compared to the mean fraction of *KRAS* exon 3 inclusion for each Q61
1532 mutation as quantified from TCGA[24] (F=42.48, DFn=1, DFd=2, R^2=0.955, p=0.0227). c)
1533 Correlation between the variant allele frequencies for KRAS$^{Q61R/L}$ and U2AF1^{I24T} in 20 patient
1534 samples from the Foundation Medicine Inc. and AACR Project GENIE (v16.1) datasets
1535 containing both mutations and lacking CNVs at either locus (F=20.52, DFn=1, DFd=18,
1536 R^2=0.5327, p=0.000260). Dashed line indicates a 1:1 ratio of KRAS^{G12S} and U2AF1^{S34F} allele
1537 frequencies. d) Paired analysis of the variant allele frequency of KRAS$^{Q61R/L}$ and U2AF1^{I24T}
1538 across 20 patient samples from the Foundation Medicine Inc. and AACR Project GENIE (v16.1)
1539 datasets containing both mutations and lacking CNVs at either locus (t=3.77, df=19, 95%
1540 confidence interval=-0.175 to -0.050, p=0.00129).