

Reviewed Preprint

v1 • June 2, 2026

Not revised

✉ For correspondence:

kumarp@illinois.eduthomas.gonatopoulos@nih.govashish.lal@nih.gov

Competing interests: No competing interests declared

Funding: See [page 21](#)

Reviewing editor: Arun Kumar Ganesan, University of New Mexico, United States

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0 public domain dedication](#).

Intron Retention Controls Localization of lncRNAs *PURPL* and *MALAT1* to Promote Cell Proliferation and Migration

Ioannis Grammatikakis¹, Chosita Norkaew¹, You Jin Song², Amit K Behera³, Erica C Pehrsson^{4,5}, Corrine Corrina R Hartford¹, Shreya Kordale³, Rishabh Prasanth², Yongmei Zhao⁶, Biraj Shrethsa⁶, Xiao Ling Li¹, Ravi Kumar¹, Ragini Singh¹, Tayvia Brownmiller⁷, Xinyu Wen^{4,5}, Natasha J Caplen⁷, Pablo Perez-Pinera⁸, Kannanganattu V Prasanth²✉, Thomas Gonatopoulos-Pournatzis³✉, Ashish Lal¹✉

¹Regulatory RNAs and Cancer Section, Genetics Branch, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, United States • ²Department of Cell and Developmental Biology, Cancer Center at Illinois, University of Illinois at Urbana-Champaign, Urbana, United States • ³Functional Transcriptomics Section, RNA Biology Laboratory, NCI, NIH, Frederick, United States • ⁴OMICS Technology Facility, Bioinformatics, Genetics Branch, CCR, NCI, NIH, Bethesda, United States • ⁵Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research, Frederick, United States • ⁶Sequencing Facility Bioinformatics Group, Bioinformatics and Computational Science Directorate, Frederick National Laboratory for Cancer Research, NCI, NIH, Frederick, United States • ⁷Functional Genetics Section, Genetics Branch, CCR, NCI, NIH, Bethesda, United States • ⁸Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, United States

eLife Assessment

This manuscript provides **important** insights into how U2AF2-dependent intron retention regulates the localization and function of long noncoding RNAs, with evidence supported by multiple complementary approaches. The work is notable for linking intron retention to nuclear speckle localization and cellular phenotypes, including proliferation and migration, although the mechanistic basis remains **incompletely** resolved. Overall, the study presents a **compelling** dataset with clear biological implications but would benefit from additional analyses to strengthen mechanistic interpretation and generality.

<https://doi.org/10.7554/eLife.111394.1.sa3>

Abstract

Intron retention (IR) is increasingly recognized as a feature of long noncoding RNAs (lncRNAs), yet the mechanisms that shape IR in lncRNAs and the functional consequences of this process remain largely unexplored. To investigate how IR contributes to lncRNA regulation, we performed a genome-wide screen to identify factors controlling IR in the lncRNA *PURPL*. This approach uncovered a prominent role for U2AF2, which promotes retention of a specific intron in *PURPL* through a weak polypyrimidine tract. IR of this intron drives nuclear enrichment of *PURPL* and enhances cell proliferation, revealing biological relevance. Transcriptome-wide analyses showed that although U2AF2 broadly supports canonical splicing consistent with its well-established function in promoting splicing, it also facilitates IR within a distinct subset of RNAs, including the nuclear speckle-associated lncRNA *MALAT1*. Loss of U2AF2 disrupts *MALAT1* speckle localization and using *MALAT1* knockout cells reconstituted with wild-type or intron deleted variants, we identified a single intron critical for *MALAT1*'s speckle localization. Deletion of this intron from endogenous *MALAT1* impaired speckle localization and reduced cell migration, phenocopying the

loss of *MALAT1*. Together, these findings reveal IR as a key regulatory mechanism governing lncRNA localization and function and uncover an unexpected role for U2AF2 in promoting IR within specific lncRNA contexts.

Introduction

Long non-coding RNAs (lncRNAs) are RNA molecules >200 nucleotides (nt) long that lack coding potential. Similar to protein-coding genes, lncRNAs are transcribed by RNA Polymerase II, are 5'-capped, spliced and are often polyadenylated (1). Several characteristics distinguish lncRNAs from protein-coding genes, including higher tissue-specific expression, lower abundance and stability, predominant nuclear localization, and having fewer but longer exons (2–6). One of the most distinguished features of lncRNAs is that they are less efficiently spliced compared to mRNAs (4,7–9). Interestingly, lncRNAs contain introns with much slower intron-excision kinetics compared to protein-coding genes (10). However, the factors that regulate lncRNA splicing and their impact on lncRNA biology remain largely unclear. We previously reported that the lncRNA *PURPL* exhibits prominent splicing defects, including frequent retention of its second intron. Transcriptomic analyses using both long-read and short-read sequencing revealed multiple *PURPL* isoforms, several of which retain intron 2 (11). While *PURPL* is also known to be transcriptionally induced by p53 and to contribute to cancer cell growth through multiple mechanisms, the regulatory factors governing intron 2 splicing and the functional significance of its retention have not been defined. This gap provides an opportunity to uncover broader principles of IR regulation in lncRNAs.

Splicing is a highly regulated process where *cis*- and *trans*-acting factors coordinate proper and efficient spliceosomal recruitment (12). Several consensus sequences in the RNA are crucial for splicing such as the polypyrimidine (Py)-tract, a stretch of pyrimidines close to the 3' splice site (ss) (13). The splicing factor U2AF1 (U2-auxiliary factor 1) binds to the 3' ss and U2AF2 binds to the Py-tract. The U2AF heterodimer is recruited to the branch point, and this complex serves as a critical intermediate for the subsequent assembly of the spliceosome and splicing catalysis (14–18). The vast majority of human genes encode pre-mRNAs that contain introns that are spliced out and most nascent transcripts undergo alternative splicing, which leads to the combinatorial joining of exons. Alternative splicing is a major contributor to protein isoform diversity, enabling the generation of multiple transcript variants from a single gene and frequently shaping measurable cellular phenotypes (19–22). One of the least studied forms of alternative splicing is intron retention (IR). Lately, it has drawn more attention since it has an important role in post-transcriptional gene expression regulation, and it is believed that more than half of human multi-exonic genes are affected by IR (23). IR frequently leads to nuclear detention of transcripts and eventually degradation which may serve as a defense mechanism of the cell against deleterious products. The extra sequence added by the retained intron may change the open reading frame (ORF) of protein-coding genes and lead to nonsense-mediated decay (NMD) (24–26) or initiate the use of an upstream ORF (27,28). In addition, they can contain binding sites for microRNAs (miRNAs) or RNA-binding proteins (RBPs), or they may form structural changes that impact the function of the transcript (29). Interestingly, IR has been associated with several diseases such as cancer where tumor cells show higher levels of IR compared to normal tissues (30). Studies on IR have shown that retained introns have been associated with weaker (non-consensus) splice sites, higher GC content, various lengths of introns, and higher intronic sequence conservation (23,31). lncRNAs have been shown to participate in splicing regulation, with *MALAT1* as the most prominent example. *MALAT1* is a highly abundant nuclear lncRNA which has oncogenic activity and is one of the first lncRNAs associated with cancer (32–34). It also has the ability to regulate alternative splicing because it resides in the nuclear speckles and can interact with several SR (Serine/Arginine rich) proteins to regulate their phosphorylation state and distribution between transcription sites and speckles (33,35–37). Nuclear speckles are sub-nuclear compartments that act as hubs in coordinating transcription, RNA processing and mRNA export (38). Several RBPs reside in nuclear speckles such as the protein SON which is a key component and marker of

nuclear speckles (39). However, *MALAT1* itself is not essential for the structural integrity of nuclear speckles (40,41). Nevertheless, it remains unclear whether *MALAT1* undergoes alternative splicing and how this process impacts its localization to nuclear speckles.

Here, we show that the intron 2-retaining *PURPL* isoform, which localizes to the nucleus, has a shorter half-life than the spliced isoform and promotes cell proliferation. To identify regulatory factors controlling the splicing of this functionally relevant intron, we utilized our recently developed CRISPR-based genome-wide screening strategy using a pooled guide RNA library with a *PURPL* intron 2 minigene reporter containing its flanking exons. Unexpectedly, this approach revealed a role of the splicing activator U2AF2 in promoting *PURPL* IR. U2AF2 was further found to promote a broader intron retention program that includes two introns within *MALAT1*. Notably, depletion of U2AF2 resulted in *MALAT1* exclusion from nuclear speckles. *MALAT1* knockout cells exogenously expressing *MALAT1* truncation mutants lacking specific introns identified one intron essential for the nuclear speckle association of the *MALAT1* transcript. Breast cancer cells where this intron is deleted showed reduced cell migration demonstrating a functional role of the intronic sequences. These findings provide insights into the splicing regulation of *PURPL* and *MALAT1*, uncovering a previously unrecognized functional role of intron retention in lncRNAs.

Results

A CRISPR-based screen identifies U2AF2 as an activator of *PURPL* intron 2 retention

We previously identified *PURPL* as a p53-regulated lncRNA that promotes the growth of colorectal and liver cancer cells (11,42). In those studies, we reported that *PURPL* undergoes alternative splicing, with the ~1.9 kb intron 2 either retained or spliced in the mature transcript (11). To identify regulators of *PURPL* intron 2 retention, we applied our recently developed screening platform, CRISPR-based identification of Regulators of Alternative Splicing with Phenotypic Sequencing (CRASP-seq) (43). This approach utilizes a genome-wide knockout lentiviral library targeting ~18,500 protein-coding genes using a dual hybrid guide (hg)RNA system that combines *Streptococcus pyogenes* (*Sp*)Cas9 and optimized *Acidaminococcus sp.* (*op*)Cas12a nucleases. (22,44–46). The library also includes ~2,400 pairs and ~500 curated genetic interaction pairs, allowing combinatorial perturbations from a single construct. Each gene or gene pair is targeted by four independent hgRNAs alongside ~500 intergenic and non-targeting controls, totaling ~91,000 hgRNAs. The lentiviral library incorporates a doxycycline-inducible *PURPL* IR reporter comprised of the full native intron 2 sequences (~1.9 kb) flanked by the adjacent exons 2 (161 bp) and 3 (163 bp) (Figure 1A). (47).

We performed the CRASP-screen in two cell lines: HAP1 and RPE1. Following puromycin selection for 48 h and doxycycline induction for an additional 24 h, RNA was extracted, and poly-A RNA was purified using oligo(dT) beads to exclude transcripts still undergoing processing. cDNA was synthesized using customized primers containing unique molecular identifiers (UMIs), facilitating the removal of PCR duplicates during downstream analysis. The synthesized cDNA was then utilized to construct Illumina sequencing libraries for paired-end sequencing. This sequencing approach enables simultaneous quantification of transcripts that either retain or splice out an intron of interest from one read, while the corresponding hgRNA sequences were captured in the other read to pinpoint specific genetic perturbations. By directly linking *PURPL* splicing outcomes with genetic modifications introduced by each hgRNA in the library, this platform allowed the genome-wide identification of *PURPL* splicing regulators (Figure 1B). (48).

We employed an analytical pipeline that prioritized candidate genes based on multiple hgRNAs showing significant deviations from intergenic controls, with consistent identification across two technical replicates. This analysis identified 62 genes that influence *PURPL* intron 2 splicing, the majority of which act as suppressors of intron 2 splicing (Figure 1C and Table S1). (49). Gene ontology analysis of biological processes revealed a strong enrichment for terms related to mRNA processing and RNA splicing. Additionally, analysis of cellular components indicated that the

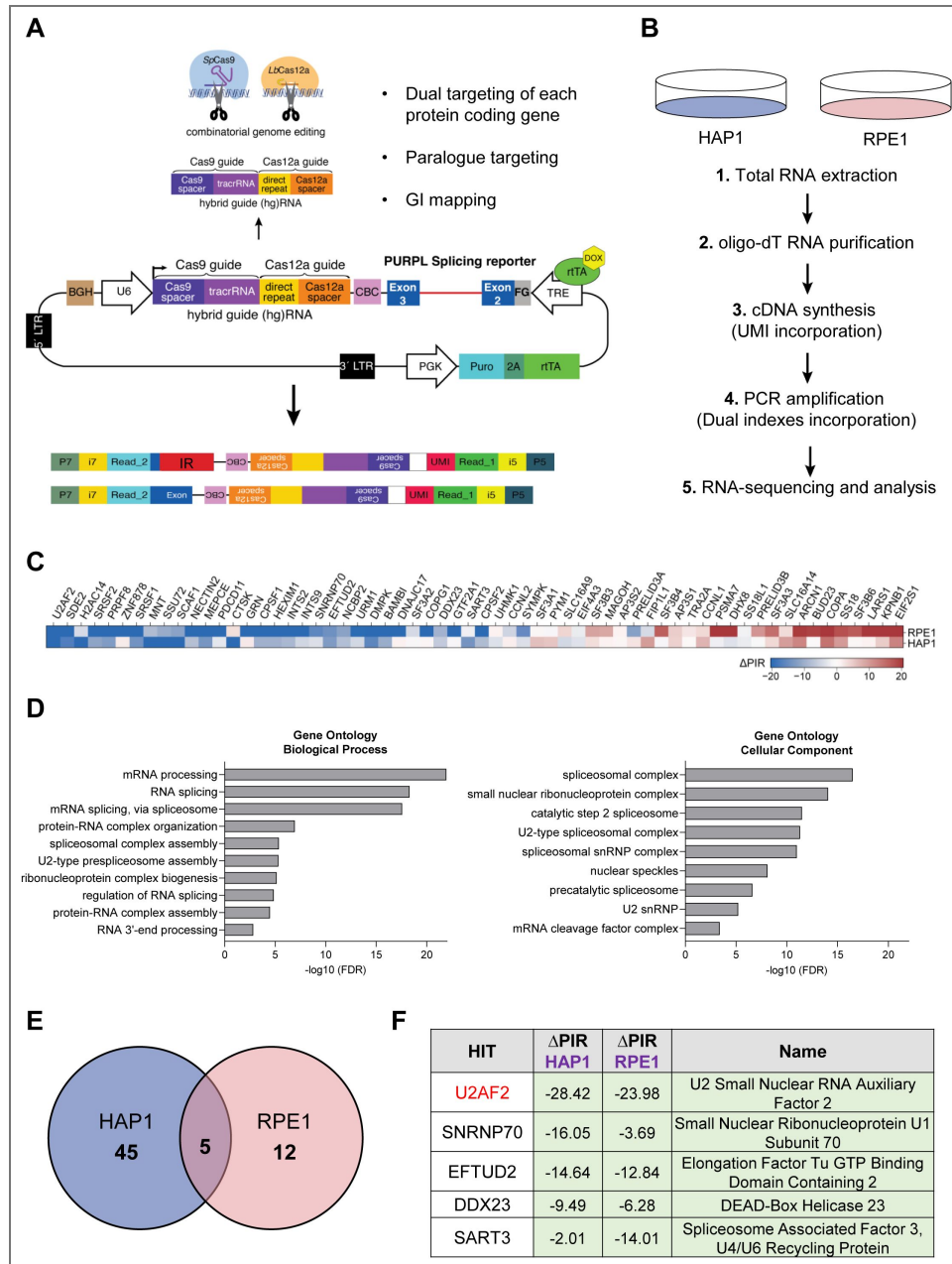


Figure 1. CRISPR-based screen identifies U2AF2 as a promoter of intron retention.

(A) Schematic of the *PURPL* IR reporter used in the screen to identify IR regulators. The pLCHKO vector contains the puromycin resistance gene used for selection. The transgene also contains the genomic sequences of exon 2 (161 bp), intron 2 (1,943 bp), and exon 3 (163 bp) of the *PURPL* gene under a Doxycycline-inducible promoter. A hybrid guide (hg)RNA for guiding *SpCas9* and *LbCas12a* is expressed under a U6 promoter. Top arrow: The hgRNA is cleaved by Cas12a to produce 2 guide RNAs for dual targeting of each protein coding gene, paralog targeting, and genetic interaction mapping. Bottom arrow: Upon intron splicing or intron retention, two different reads are generated. **(B)** Two different cell lines, stably expressing Cas9 and Cas12a, are used: HAP1, and RPE1. The CRISPR-seq pipeline until Illumina paired-end sequencing of CRISPR libraries is depicted. **(C)** Heatmap showing single genes only (after removing gene pairs), whether identified individually or as part of a gene pair. The hits are ordered by the average Δ PIR values across HAP1 and RPE1, from lowest to highest. **(D)** Left: Gene ontology enrichment analysis of biological processes regulated by the hits of the screen. The most significant pathways involve mRNA processing and RNA splicing. Right: Gene ontology enrichment analysis of cellular components regulated by the hits of the screen. These components include mainly Spliceosomal Complex, small nuclear ribonucleoprotein complex and the Catalytic step 2 Spliceosome. **(E)** Venn diagram showing that there are 5 common hits which cause a reduction in the percentage of intron retention (Δ PIR) of the splicing reporter in the two cell lines. **(F)** Table showing the top 5 hits of the screen with a Δ PIR for each hit in each cell line and their full names.

identified hits are predominantly associated with Spliceosomal Complex, including the U2-type Spliceosomal Complex and the Catalytic step 2 Spliceosome (Figure 1D). These findings validate the effectiveness of our screening approach.

Consistent with the endogenous *PURPL* transcript isoforms, the basal levels of IR of the reporter were very high, with the majority of amplified transcripts corresponding to intron-containing isoforms, as expected based on the intensity of the Illumina libraries before sequencing (Figure S1A). Thus, the detection of genes that suppress IR might be below the detection levels of this screen. Therefore, we focused on the genes that promote intron retention. By calculating the change in Percent Intron Retention ($\Delta\text{PIR} = \text{PIR}_{\text{guide}} - \text{PIR}_{\text{intergenic}}$) across the two cell lines, we identified 5 unique genes as regulators of *PURPL* intron 2 retention, all of which act to suppress intron 2 splicing (Figure 1E and Table S1). This highly significant overlap between the two cell lines (Figure 1E; odds ratio = 204.4, $p=2.81 \times 10^{-10}$; Fisher's exact test) indicated a shared regulatory mechanism governing *PURPL* IR, although cell-type-specific effects may also contribute.

Surprisingly, the gene with the strongest impact based on the average ΔPIR metric and identified as a hit in both cell lines, was *U2AF2* (Figure 1F). This was followed by Small Nuclear Ribonucleoprotein U1 Subunit 70 (SNRNP70), a core component of the U1 snRNP, and Elongation Factor Tu GTP Binding Domain, which is associated with U5 snRNP (Figure 1F). Our finding that *U2AF2* promotes IR of *PURPL* intron 2 was unexpected because *U2AF2* is a component of the spliceosome machinery, and its canonical function is to promote intron removal (14,18).

***U2AF2* binds to the *PURPL* intron 2 transcript and promotes intron retention**

U2AF2, as a heterodimer with *U2AF1*, binds to the Py-tract and the 3' ss, respectively, to promote splicing (14,17). To elucidate its role in *PURPL* splicing, we investigated whether *U2AF2* binds to intron 2. Enhanced Crosslinking and Immunoprecipitation sequencing (eCLIP-seq) (47), from the ENCODE project (48,49) from HepG2 cells revealed multiple *U2AF2* and *U2AF1* binding sites not only proximal to the 3' ss of *PURPL* intron 2 but also broadly distributed across intron 2 (Figure 2A). Compared to eCLIP data for other RBPs, *U2AF1* and *U2AF2* showed the highest density of binding sites and the most prominent peaks. Specific binding of *U2AF2* to intron 2 was confirmed by performing RNA immunoprecipitation (RNA-IP) using a *U2AF2*-specific antibody and primer pairs designed to detect *PURPL* transcripts. *U2AF2* was found to interact with transcripts containing *PURPL* intron 2, but not with the spliced *PURPL* transcripts (Figure 2B and S1B).

To functionally validate the role of *U2AF2* in *PURPL* intron 2 retention, we performed RT-PCR assays from HAP1 and RPE1 cells transduced with *PURPL* minigene reporter constructs and control or *U2AF2*-targeting sgRNAs. In both cell lines, we observed a striking decrease in intron 2 retention upon *U2AF2* depletion as compared to control cells expressing sgRNAs targeting intergenic regions (Figure 2C). As an alternative approach, we knocked down *U2AF2* using two independent siRNAs and conducted semi-quantitative RT-PCR assays from SKHEP1, HepG2, and normal human diploid fibroblasts W138 cells, using a primer triplet (Figures 2D-2F; Figures S2A and S2B). In all tested cell lines, *U2AF2* depletion increased *PURPL* intron 2 splicing, from ~20% to ~40-60%, more than doubling the abundance of the spliced isoform. Given that *U2AF2* functions as a heterodimer with *U2AF1*, we also knocked down *U2AF1* in SKHEP1 cells and observed an effect similar to *U2AF2* knockdown on *PURPL* intron 2 retention regulation (Figure S2C). We expected that *U2AF1* would also come as one of the top hits in the screen but *U2AF1* sgRNAs were not included in the library due to high off-target scores of the guides. These results demonstrate that the *U2AF* complex promotes *PURPL* intron 2 retention and that both *U2AF1* and *U2AF2* promote intron retention in *PURPL*.

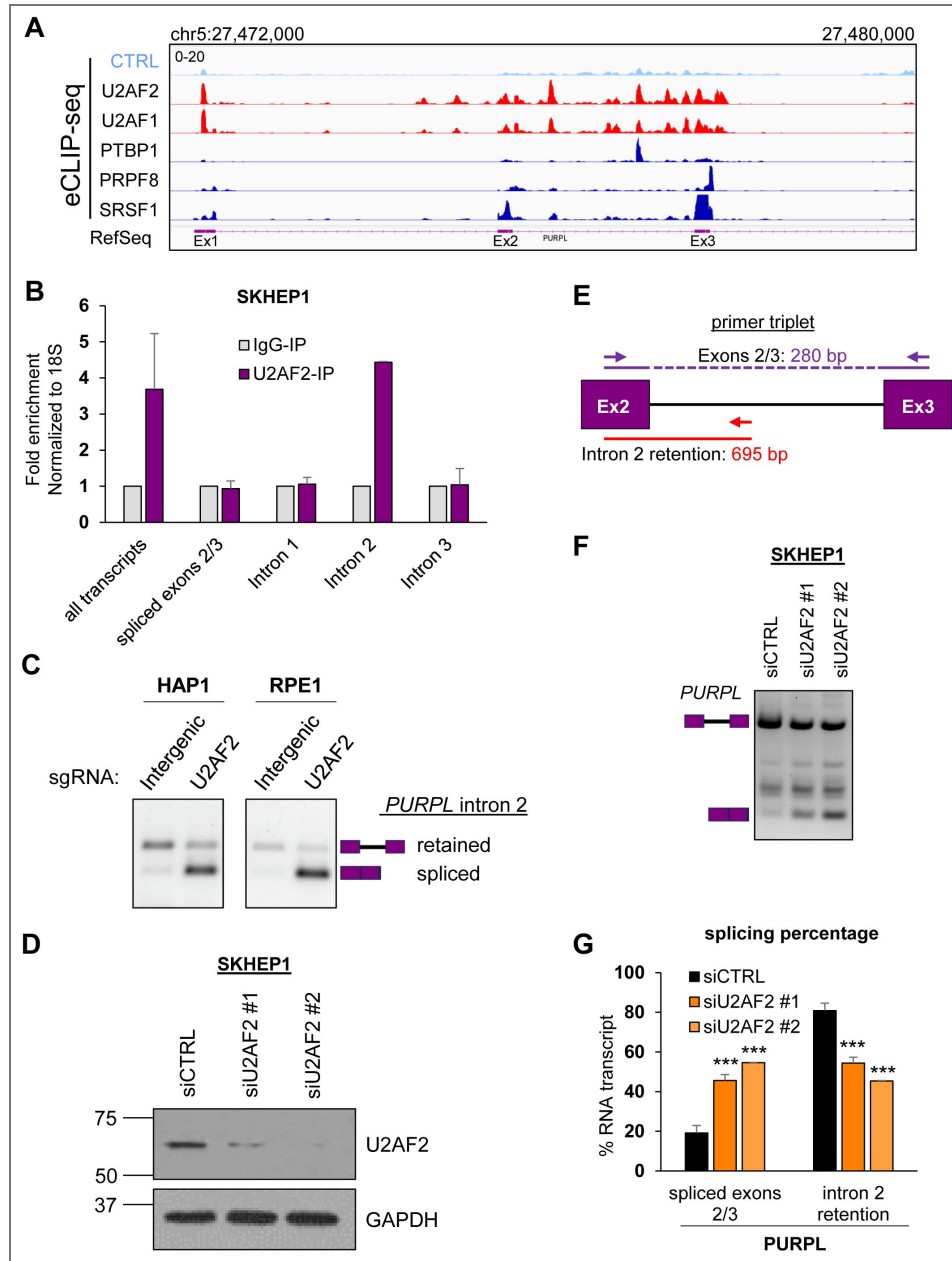


Figure 2. U2AF2 binds to intron 2-containing *PURPL* transcripts and promotes intron 2 retention.

(A) IGV snapshot of eCLIP-seq data showing binding sites and enrichment of U2AF2 and U2AF1 on the *PURPL* transcripts around intron 2. eCLIP-seq data for PTBP1, PRPF8, and SRSF1 are indicated. The annotated locus by RefSeq is also indicated. eCLIP data-seq was downloaded from encodeproject.org. (B) RT-qPCR after RNA-IPs using a U2AF2 antibody with primer pairs specifically detecting *PURPL* transcripts as indicated in Figure S1B. U2AF2 binds to transcripts containing intron 2 but not the ones with intron 1, intron 3, or spliced exons 2 and 3. Samples were normalized to IgG-IP. 18S was used as a loading control. (C) Validation of hgRNAs using the *PURPL* minigene in HAP1 and RPE1 cells. The gels show RT-PCR products for *PURPL* transcripts with specific primers for the minigene. The hgRNAs used were targeting intergenic region, or U2AF2. The schematics next to the gel indicate the expected products of the intron-retained and spliced isoforms. (D) Western blot for U2AF2 showing successful knockdown of U2AF2 protein in SKHEP1 cells. GAPDH was used as a loading control (lower panel). (E) Schematic of the PCR primer triplet used to detect intron 2 retention (red) or splicing (purple). The length for each PCR product is indicated. (F) Gel with RT-PCR products for *PURPL* upon knockdown of U2AF2 with 2 different siRNAs in SKHEP1 cells. The schematics next to the gel indicate the expected products of the intron-retained and spliced isoforms. Between the two expected PCR products, we observed an extra band corresponding to the inclusion of an alternative exon inside *PURPL* intron 2 as observed in RefSeq, the inclusion of which is not affected by U2AF2. (G) Graph with quantitation of the gel bands in (F). Error bars represent standard deviations from 3 (G) experiments. *** $p < 0.001$.

The intron-retained *PURPL* transcript is unstable and localizes to the nucleus

We previously observed an enrichment of the intron 2-retaining *PURPL* transcript in the nucleus in liver cancer cells (11). To validate these findings, we conducted single molecule RNA-FISH in HCT116 cells using probes hybridizing to *PURPL* intron 2 (Figure 3A). Because *PURPL* is upregulated in response to DNA damage (42), we treated the cells with hydroxyurea for 24 h to induce *PURPL* expression. RNA-FISH revealed that the *PURPL* intron 2 containing transcript was exclusively localized to the nucleus. To assay for subnuclear compartmentalization, we also performed RNA-FISH to detect *MALAT1*, which resides in nuclear speckles (36). Intron 2-retained *PURPL* transcripts did not colocalize with *MALAT1* indicating that *PURPL* is excluded from nuclear speckles (Figure 3A). We further validated this result by conducting nuclear/cytoplasmic fractionation assays followed by RT-qPCR and found that *PURPL* intron 2-containing transcripts were almost exclusively nuclear (Figure S3A).

Intron retention can affect transcript stability, with longer RNA species often being more vulnerable to degradation. To test if *PURPL* intron 2 effects transcript turnover, we conducted RNA stability assays using the transcription inhibitor Actinomycin D (ActD) and observed that the intron 2-retained *PURPL* transcript was less stable and had a significantly lower half-life than the *PURPL* transcript in which intron 2 was spliced out (Figure 3B). To determine the functional impact of *PURPL* intron 2 retention, we used CRISPRi to generate *PURPL*-depleted SKHEP1 cells by targeting 3 independent sgRNAs to the p53-response element of the *PURPL* promoter that we had used in our previous study (11). We observed >80% reduction in the levels of endogenous *PURPL* in these experiments (Figure 3C). We next stably transduced these *PURPL*-depleted cells with a lentivirus to express exogenous *PURPL* transcript specifically containing intron 2 under the control of a doxycycline (doxy)-inducible promoter and observed ~2-3-fold induction of the exogenous *PURPL* intron 2 containing transcript (Figure 3D). Because we and others have previously shown that *PURPL* is overexpressed in some cancers and *PURPL* promotes proliferation (11,42,50,51), we next performed cell proliferation assays from the doxy-inducible cells. Our results showed that cells expressing intron 2-containing *PURPL* proliferate faster (~2-fold) as compared to the no-doxy cells after 6 days of doxy treatment (Figure 2E). Induction of intron 2-containing *PURPL* transcripts did not have any effect on live cell percentage which remained at ~100% (Figure S3B). Importantly, as a negative control, there was no effect of doxy treatment on the proliferation of parental SKHEP1 cells (Figure S3C). To confirm that proliferation increase induced by exogenous *PURPL* transcript containing intron 2 was not due to the spliced isoform, we performed semi-quantitative RT-PCR. As shown in Figure 3F, doxy treatment resulted exclusively in the detection of the intron-retained *PURPL* transcript. These data suggest that intron 2 retention modulates *PURPL* function by sequestering the transcript in the nucleus and reducing its stability, while simultaneously promoting cell proliferation, highlighting a critical regulatory role for intron 2.

A weak polypyrimidine tract is critical for *PURPL* intron 2 retention

U2AF2 binds to the Py-tract to initiate spliceosomal assembly leading to splicing and intron removal (14). Suboptimal splicing efficiency may be the result of multiple factors, including splice site strength and the Py-tract sequence composition. The Py-tract is a stretch of 15-20 pyrimidines (mostly Us) immediately upstream of the 3' ss. We analyzed the Py-tract of *PURPL* intron 2 and noticed that it deviates from the consensus, containing numerous purines (Figure S4A). To investigate whether U2AF2 regulation on intron 2 retention relies on the Py-tract sequence, we used the *PURPL* intron 2 minigene reporter system (Figure 1A). We mutated the Py-tract, replacing it either with a strong Py-tract from *PURPL* intron 1 or with a stretch of Us, simulating a strong Py-tract. U2AF2 was knocked down in SKHEP1 cells stably expressing the mutated Py-tract constructs and RT-PCR was conducted to monitor intron 2 splicing. Interestingly, the basal levels of intron retention were very different with different mutants; ~90% in intron 2 WT (wild-type), ~30% in *PURPL* intron 1 and ~20% in U-rich construct (Figure S4B and S4C). This data

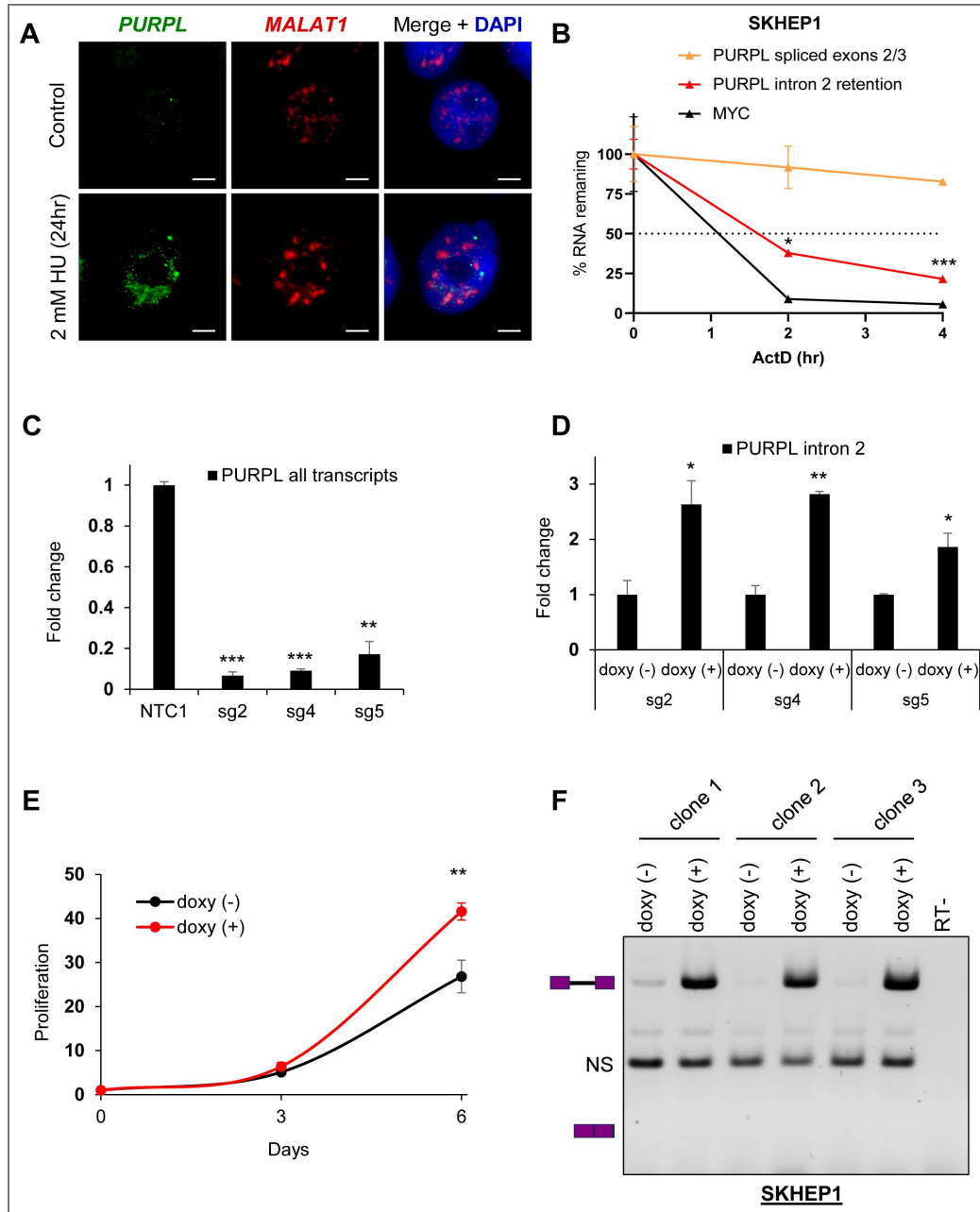


Figure 3. Expression of Intron 2 Retaining *PURPL* leads to high proliferation.

(A) RNA-FISH images for *PURPL* with intron 2 retention and *MALAT1* in HCT116 cells without treatment or after 24 h of 2 mM of Hydroxyurea (HU) to induce *PURPL* expression. Scale bar is 10 μm. (B) RNA stability assays were performed for *PURPL* transcripts by measuring their levels by RT-qPCR following 0, 2, and 4 h of ActD treatment in SKHEP1 cells. *MYC* was used as a positive control for unstable RNA. (C) RT-qPCR showing *PURPL* depletion in SKHEP1 cells using 3 different gRNAs (sg2, sg4, sg5) compared to a Non-Targeting Control gRNA. (D) RT-qPCR for intron 2-containing *PURPL* transcript after 48 h of 1 μg/mL doxycycline treatment in comparison to no treatment in SKHEP1 *PURPL*-CRISPRi populations using the 3 gRNAs as in (C). (E) Proliferation assay showing the effect of overexpression of intron 2-containing *PURPL* transcript in the proliferation of SKHEP1 cells where the endogenous *PURPL* is knocked down with CRISPRi. Error bars represent standard deviation from 3 populations with different gRNAs. The cells were treated with 1 μg/mL doxycycline to induce intron 2-retained *PURPL* expression and cell proliferation was monitored at 3 and 6 days. (F) Gel with RT-PCR products for transcripts using primers as indicated in Fig. 2E upon doxycycline treatment of SKHEP1 CRISPRi cells. The three repeats represent 3 different clones of cells. The last lane is an RT- control. The schematics next to the gel indicate the expected products of the intron-retained and spliced isoforms. NS indicates a non-specific band. Error bars in (B), (C), and (D) represent standard deviations from 2 experiments. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

demonstrates that the weak Py-tract is responsible for the inefficient intron 2 removal in *PURPL* (Figure S4C). *U2AF2* knockdown resulted in modest increase in splicing of the mutant constructs as well. Because the minigene reporter contained the mutant Py-tract along with the remaining endogenous sequences from *PURPL* intron 2, the modest increase in splicing observed for the mutants after *U2AF2* knockdown may reflect *U2AF2* inhibitory binding at additional sites within *PURPL* intron 2 beyond the Py-tract (Figure 2A). These data indicate that the endogenous intron 2 Py-tract in *PURPL* is a weak Py-tract that controls the high intron retention observed in *PURPL*.

U2AF2 directly regulates multiple IR events including *MALAT1*

To determine if *U2AF2* promotes IR in additional genes, we next performed total RNA-seq and PacBio Iso-seq from SKHEP1 cells transfected with siCTRL or si*U2AF2*. Consistent with our data, *U2AF2* knockdown resulted in a marked increase in the splicing efficiency of *PURPL* intron 2 (Figures S5A and S5B). Additionally, *U2AF2* depletion increased overall *PURPL* expression, with increased read coverage across all exons (Figures S5A and S5B; Tables S3 and S4). To identify IR events regulated by *U2AF2* at a transcriptome-wide level, we utilized IRFinder, an algorithm specifically designed to calculate changes in IR (52). This analysis revealed ~1000 statistically significant IR events regulated by *U2AF2*. For majority of these events (732), there was an increase in the IR ratio (Intronic abundance/(Intronic abundance + exonic abundance)) after *U2AF2* depletion, indicating that *U2AF2* predominantly promotes splicing, as expected. However, in 269 events, *U2AF2* knockdown resulted in decreased IR, indicating that *U2AF2* promotes IR in a subset of genes (Figure 4A and Table S2). We next analyzed ENCODE *U2AF2* eCLIP-seq data from HepG2 cells to determine if *U2AF2* directly binds to the regulated transcripts. *U2AF2* binding was detected in 96 out of the 269 IR events supporting a direct role for *U2AF2* in promoting IR (Figure S5C and Table S2).

To validate these findings, we selected three of the identified introns from *MDM1*, *RETREG2*, and *TMEM41B* (Figure S6A). We observed multiple *U2AF2* binding peaks within the regulated introns from *MDM1* (Figure S6B). We next assessed their splicing in SKHEP1 cells following *U2AF2* knockdown using two independent siRNAs. In all three cases, *U2AF2* depletion significantly enhanced intron removal (Figures S6C-S6E). IRFinder validated that *PURPL* intron 2 was one of the events showing a significant decrease (~2-fold) in IR ratio upon *U2AF2* depletion (Figure 4B).

Among the lncRNAs in which IR is promoted by *U2AF2*, was the lncRNA *MALAT1*. IRFinder identified two introns in *MALAT1* that have weak Py-tract sequences similar to *PURPL* intron 2 (Figure S7A-S7C). For both introns, *U2AF2* knockdown resulted in a significant decrease (~2-fold) in the IR ratio compared to siCTRL (Figures 4A and 4B). This result was confirmed by RT-PCR from HCT116 and SKHEP1 cells upon *U2AF2* knockdown (Figures 4C-4F).

Intron 2 of *MALAT1* drives its localization to nuclear speckles

MALAT1 regulates splicing by interacting with splicing factors in nuclear speckles.(36) To determine the impact of IR on *MALAT1* localization to nuclear speckles, we performed RNA-FISH for *MALAT1* upon *U2AF2* depletion in SKHEP1 and HCT116 cells. Remarkably, *U2AF2* knockdown resulted in a significant decrease in the ratio of *MALAT1* signal intensity in speckles relative to nucleoplasm (Figures 5A and 5B; Figures S8A and S8B). *U2AF2* depletion did not affect total *MALAT1* levels as observed in our RNA-seq (Table S3). This reduction in *MALAT1* levels in nuclear speckles did not compromise the structural integrity of nuclear speckles as indicated by immunostaining for the RNA-binding protein SON, a well-established nuclear speckle marker (39) (Figures 5A and S8A) and is consistent with previous reports showing that *MALAT1* is not essential for the integrity of nuclear speckles (40,41). These data suggest that *U2AF2* plays an important role in promoting *MALAT1* localization to nuclear speckles.

To determine whether it is the intron retention in the transcript *per se* which directs *MALAT1* localization or if *U2AF2* indirectly regulates its localization, we assessed the nuclear speckle localization of wild-type (WT) and intron-deleted *MALAT1* transcripts. To do this, we used CRISPR/Cas9 to generate two *MALAT1* knockout (KO) clones in HCT116 by inserting ~100 bp

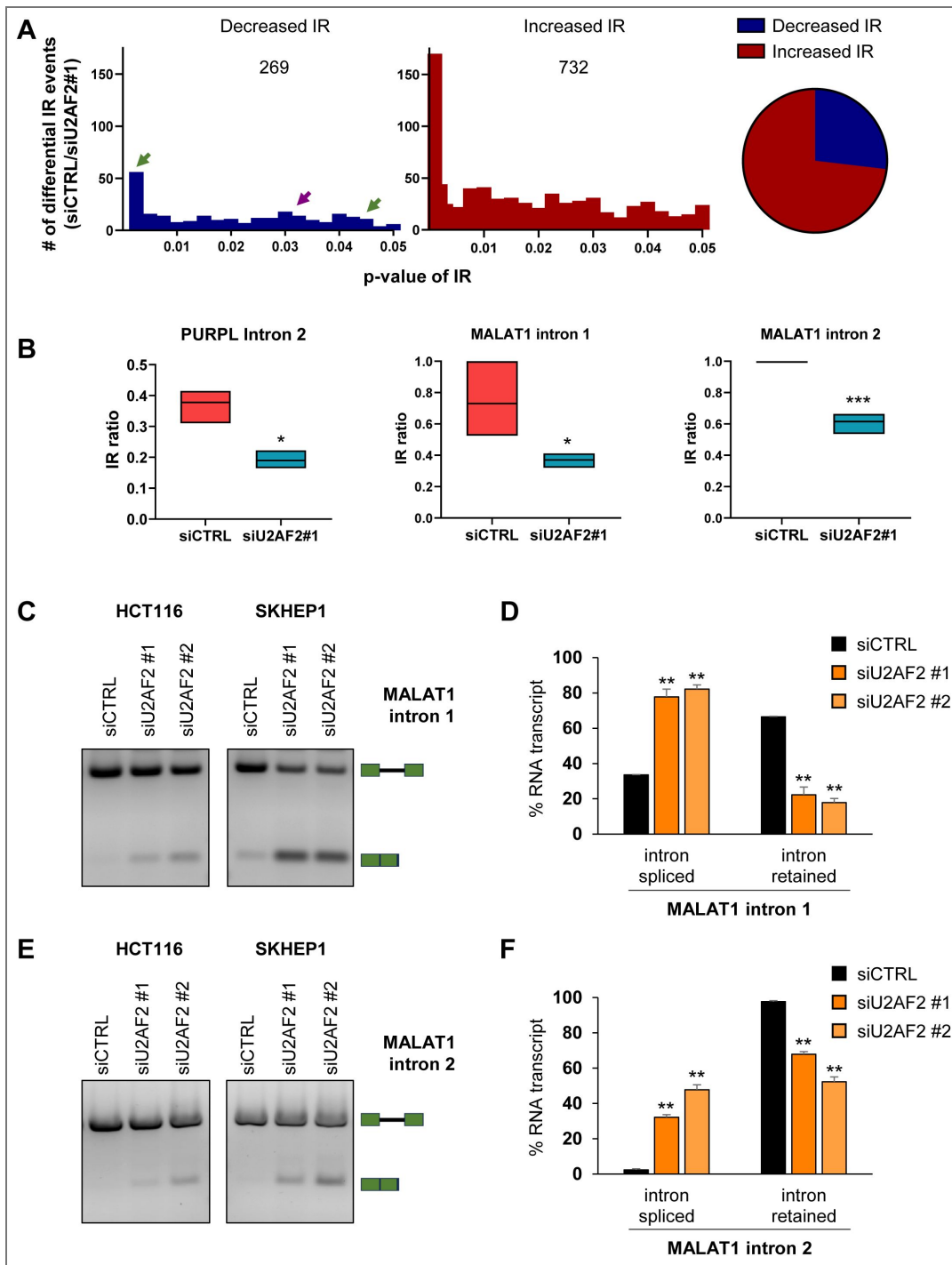


Figure 4. U2AF2 directly promotes Intron Retention of *MALAT1*.

(A) U2AF2 was knocked down in SKHEP1 cells and 72 h later, RNA was extracted and RNA-seq was performed. *Left*: Number of decreased (blue) and increased (red) IR events at various p-values after U2AF2 knockdown as analyzed with the IR Finder algorithm. The purple arrow indicates the *PURPL* IR event and green arrows indicate *MALAT1* IR events *Right*: Pie chart of the numbers of increased and decreased IR events upon U2AF2 knockdown. (B) Floating bar plots showing the IR ratio of *PURPL* intron 2 and the IR ratio of intron 1 (middle) and intron 2 (right) of *MALAT1* in siCTRL and siU2AF2#1 samples as analyzed with the IRFinder algorithm. (C) and (E) RT-PCR for *MALAT1* using a primer pair flanking the regulated intron 1 (C) or intron 2 (E) upon knockdown of U2AF2 with 2 different siRNAs in HCT116 and SKHEP1 cells. The schematics next to the gel indicate the expected products of the intron-retained and spliced isoforms. (D) and (F) Bar graph with quantitation of the gel bands from (C) and (E) in SKHEP1 cells. Error bars represent standard deviations from 2 independent experiments. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

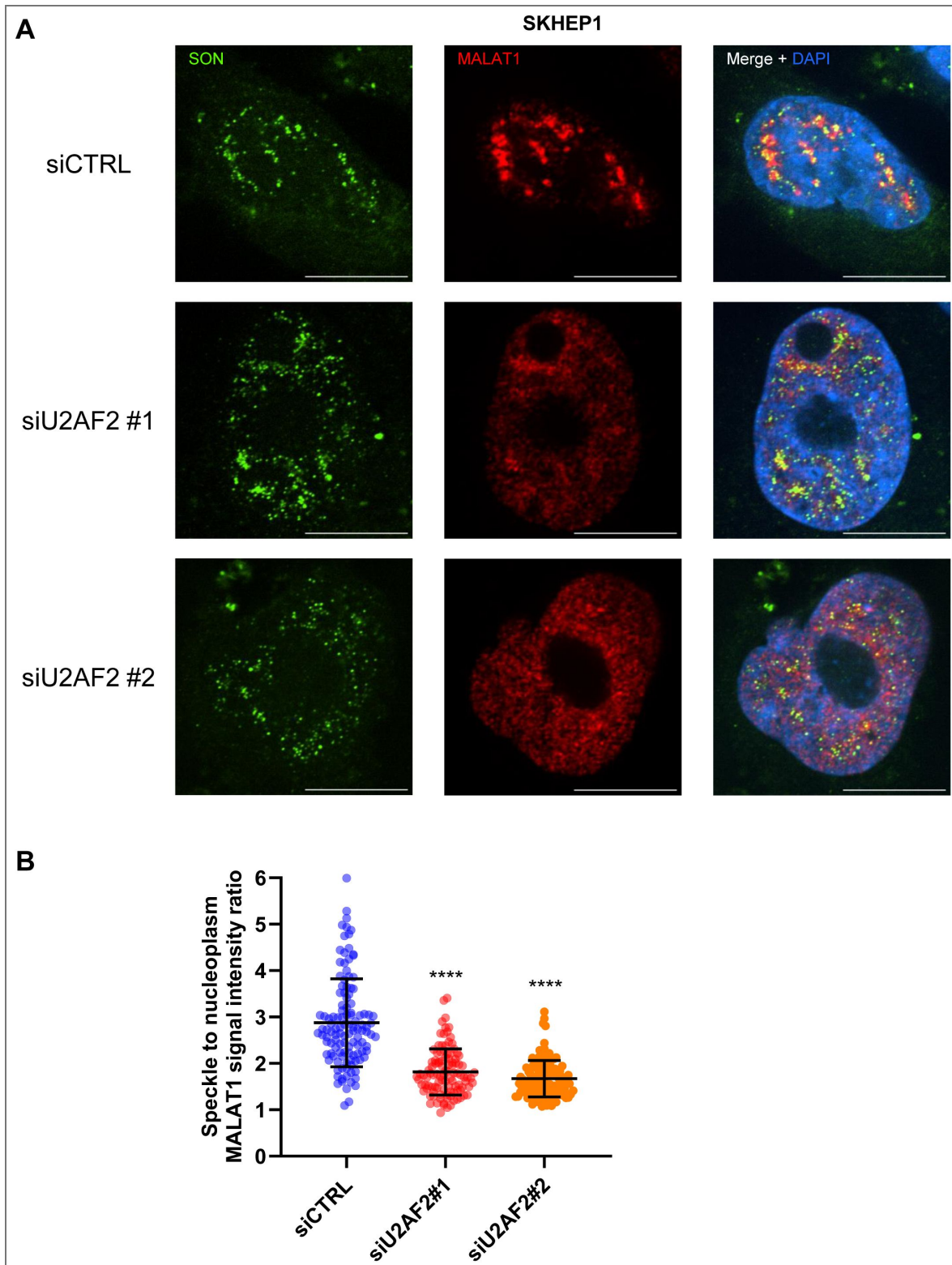


Figure 5. U2AF2 promotes localization of MALAT1 to nuclear speckles.

(A) RNA-FISH images for MALAT1 and Immunofluorescence images for SON is shown upon transfection of SKHEP1 cells with siCTRL or siU2AF2 with two different siRNAs. MALAT1 is enriched in nuclear speckles in the siCTRL but not upon U2AF2 knockdown. (B) Quantitation of the speckle to nuclear plasma MALAT1 signal ratio in the three replicates in panel (A). **** $p < 0.0001$.

downstream of the transcription start site, a puromycin resistance gene in one allele and a hygromycin resistance gene in the other allele. In both KO clones, >95% reduction in *MALAT1* RNA levels was observed as compared to a wild-type clone (Figure S9A [↗](#)).

Next, we transduced the *MALAT1* KO clones with an empty vector or doxycycline-inducible lentiviral constructs that exogenously express full-length *MALAT1*-WT in which both introns are intact, intron 1 deleted *MALAT1* (del-1), intron 2 deleted *MALAT1* (del-2) or both introns were deleted (del-1+2). After 48 h of doxy-treatment to induce *MALAT1* expression, which was confirmed with RT-qPCR (Figure S9B [↗](#)), RNA-FISH combined with immunofluorescence for the nuclear speckle marker SON, revealed distinct localization patterns. As expected, exogenously expressed *MALAT1*-WT predominantly (~80%) localized to nuclear speckles (Figures 6A [↗](#) and S10A). Similarly, *MALAT1*-del-1 also predominantly (~80%) localized to nuclear speckles. In contrast, *MALAT1* lacking intron 2 (del-2) or both introns (del-1+2) exhibited a marked reduction (<15%) in nuclear speckle localization and instead displayed diffused localization in the nucleus (Figures 6A [↗](#) and S10A). Importantly, this diffused localization upon intron 2 deletion was observed in both *MALAT1* knockout clones and was further confirmed through quantitative analyses (Figures 6B [↗](#) and S10B).

Deletion of *MALAT1* Intron 2 using CRISPR/Cas9 impairs cell migration

MALAT1 has been shown to promote cell migration and metastasis ([32,53](#)) and its expression has been correlated with high tumor progression and metastasis in various tumors ([54](#)). We therefore tested whether the intronic sequences that regulate *MALAT1* nuclear speckle localization are related to the ability of *MALAT1* to promote cell migration. To investigate this, we used the metastatic breast cancer cell line MDA-MB-231 and first performed RNA-FISH for *MALAT1* after knocking down *U2AF2*. Similar to SKHEP1 and HCT116 cells, *U2AF2* knockdown resulted in decreased localization of *MALAT1* to speckles in MDA-MB-231 cells (Figures 7A [↗](#) and 7B [↗](#)). We next utilized CRISPR/Cas9 to generate KO clones in MDA-MB-231 cells where either the whole *MALAT1* locus or intron 2 of *MALAT1* were deleted. *MALAT1* depletion in the *MALAT1* KO clone was confirmed with RT-qPCR (Figure S11A [↗](#)). RNA-FISH in 2 separate clones showed that intron 2 deletion reduced *MALAT1* localization to nuclear speckles compared to the WT clone (Figures 7C [↗](#)-7E [↗](#)). In the intron-2 deletion clones, the transcripts containing intron 2 were, as expected, undetected with primers specifically amplifying intron 2. Interestingly, there was a modest increase in total *MALAT1* expression suggesting a compensatory mechanism for the loss of intron 2 (Figure S11B [↗](#)). Functionally, when we tested the migration potential of these cells in transwell cell migration assays, we found significantly reduced cell migration in the intron 2-deleted clones, phenocopying the complete loss of *MALAT1* locus (Figures 7F [↗](#) and S11C). These data suggest that the retention of intron 2 is essential for *MALAT1* localization to nuclear speckles and it plays a key role in promoting cell migration. Together, our data uncovers a *U2AF2*-driven intron retention program in lncRNAs that regulates cell proliferation and migration.

Discussion

The goal of this study was to examine the causes and consequences of IR in lncRNAs, focusing on *PURPL* and *MALAT1*. Intron retention is a regulated process that shapes gene expression networks through mechanisms such as NMD and influences alternative protein isoform generation during differentiation, development, and disease ([25,55](#)). We previously identified a p53-induced lncRNA, *PURPL*, which exhibits growth-promoting functions, and that intron 2 of *PURPL* is retained in multiple cell lines ([11](#)). To accomplish our goal, we employed an innovative RNA-seq coupled genome-wide CRISPR screen on a reporter containing *PURPL* intron 2 and unexpectedly identified *U2AF2* as a major inhibitor of intron 2 removal in *PURPL*. *U2AF2* forms a heterodimer with *U2AF1*, binding to the Py-tract and the 3' ss, respectively, to promote splicing by recruiting core spliceosomal components ([17,18,56](#)). Therefore, given the established role of *U2AF2* as a critical positive regulator of splicing catalysis, we were surprised that *U2AF2* inhibits the removal of

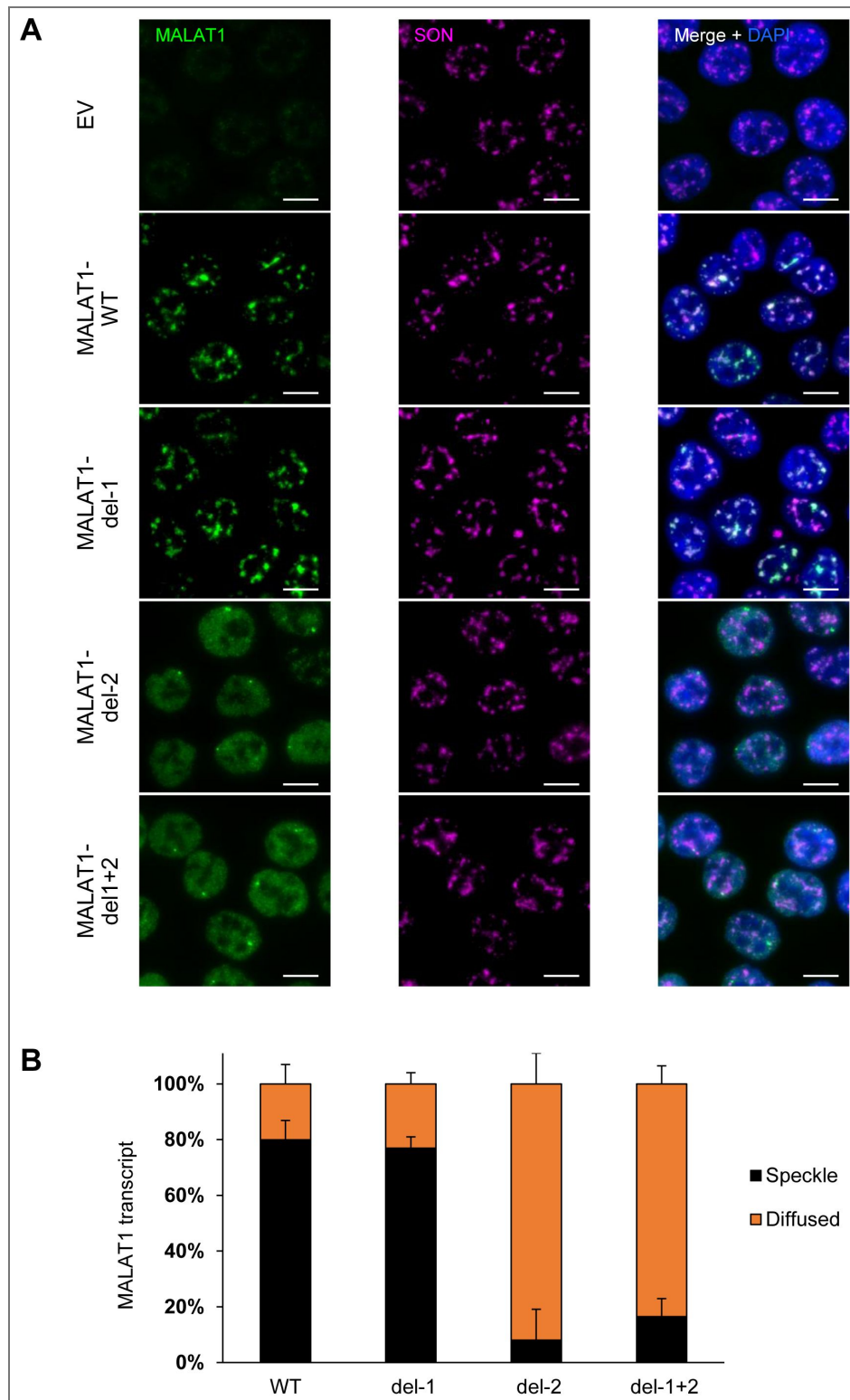
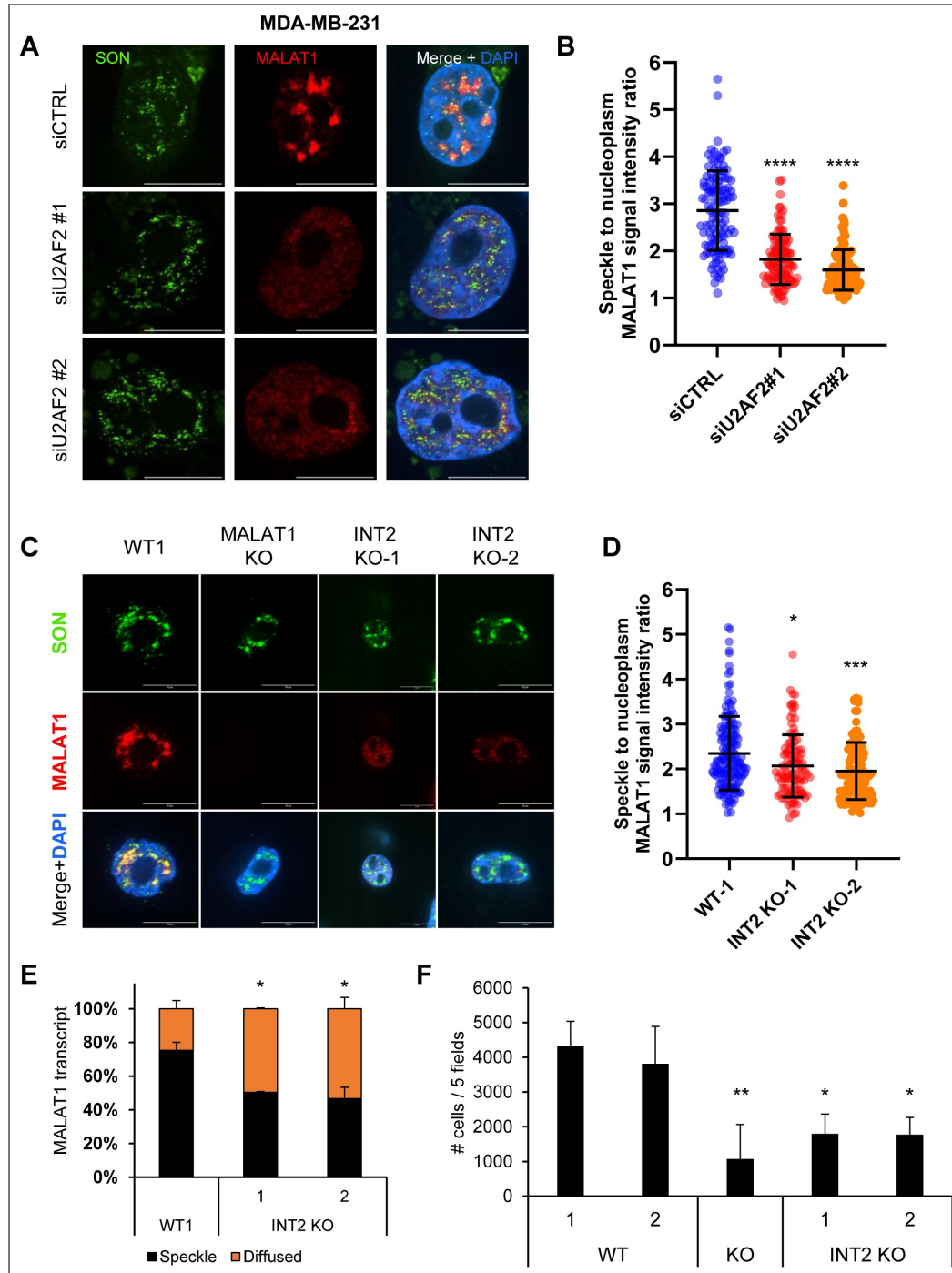


Figure 6. Intron 2 of *MALAT1* dictates its localization to nuclear speckles.

(A) RNA-FISH images for *MALAT1* and Immunofluorescence images for SON in a HCT116 *MALAT1*-KO clone (clone 1) where Empty Vector (EV) or constructs that exogenously express *MALAT1* transcripts were re-introduced using a doxy-inducible lentivirus. Expressed *MALAT1* transcripts were full-length (WT) or intron 1 deleted (del-1) or intron 2 deleted (del-2) or both introns deleted simultaneously (del-1+2). The cells were first treated with 1 $\mu\text{g}/\text{mL}$ doxycycline for 48 h to induce *MALAT1* expression. *MALAT1* without intron 2 does not enrich in speckles as it is shown with SON protein. Scale bar is 10 μm . **(B)** Graph showing quantification of the percentage of cells where *MALAT1* transcripts are enriched in speckles or being diffused in HCT116 cells clone 1. Error bars represent standard deviations from 2 independent experiments.



PURPL intron 2. While U2AF1 was not identified as a hit in our screen, follow-up experiments revealed that U2AF1 also suppresses *PURPL* intron 2 removal, suggesting a previously unrecognized role for the U2AF complex as a negative regulator *PURPL* IR.

Beyond IR in *PURPL*, we sought to determine whether U2AF2 promotes IR in other transcripts. Using IR Finder (52), we identified other genes where U2AF2 promoted IR. One of the genes that drew our attention was the lncRNA *MALAT1* which showed two intron retention events regulated and bound by U2AF2. These introns contain weak splice sites suggesting that this may be the reason that they are regulated, and they are not constitutively spliced. Because *MALAT1* interacts with splicing factors within nuclear speckles (36), we examined whether U2AF2 affects *MALAT1* localization to speckles. Indeed, our results establish *MALAT1* intron 2 as a critical determinant of its nuclear speckle localization. Future studies involving mapping of the intronic sequences and targeted mutagenesis of the splice sites within *MALAT1* intron 2 could help distinguish whether speckle localization of *MALAT1* is inhibited by the splicing process itself or depends on a discrete element within the intron.

For *PURPL*, previous work established its context-dependent oncogenic activity (11,42), but those studies depleted all transcripts from the *PURPL* locus despite the presence of multiple alternatively spliced isoforms. In the current study, overexpression of the intron 2-retaining *PURPL* isoform in CRISPRi-mediated *PURPL*-depleted cells increased proliferation, demonstrating isoform-specific, growth-promoting function. Our findings reveal that intron retention in both *MALAT1* and *PURPL* is a regulatory mechanism with direct consequences on lncRNA localization and cancer-related phenotypes.

The surprising finding of U2AF2 promoted IR promoted us to determine the underlying mechanism. We first asked whether U2AF2 influences transcription initiation or elongation at the *PURPL* locus, given the tight coupling between transcription and splicing (57,58). Transcription kinetics can modulate IR (23,59) and lncRNAs often lack proximal RNA-pol II phosphorylation over 5' ss (10). We therefore assayed for PolII-pSer5 (transcription initiation) and PolII-pSer2 (transcription elongation) occupancy after U2AF2 knockdown using CUT&RUN-Seq but there was no significant difference around intron 2 or elsewhere across the *PURPL* locus (data not shown). We also examined the chromatin mark H3K36me3 which has been linked to IR (60) and again observed no effect after U2AF2 depletion (data not shown).

U2AF2 functions by binding to the Py-tract, and the Py-tract of *PURPL* intron 2 and *MALAT1* introns 1 and 2 deviate from the consensus, containing a mix of pyrimidines and purines. These atypical sequences likely contribute to splicing inhibition as revealed by our data from the mini-gene reporter for *PURPL*. Given that U2AF1 and U2AF2 bind to multiple intronic sites, a plausible model is that their docking obstructs subsequent spliceosomal assembly. lncRNAs generally undergo less efficient splicing and consequently, they have higher levels of IR compared to protein coding genes (8,10). A study showed that lncRNA splicing strength correlates with 5' ss strength and lncRNA Py-tract has higher thymidine content compared to protein coding genes (61). The reduced evolutionary pressure to maintain efficient splicing as they are not used as templates for protein synthesis, likely contributes to this pattern. This is also supported by their predominant nuclear localization (3,4).

Our data shows that *PURPL* transcripts can be either nuclear or cytoplasmic, and U2AF2 regulation of IR offers a mechanism of nuclear detention. Similar to this notion, the lncRNA *TUG1* remains in the nucleus due to intron retention (62), and the mouse lncRNA *pCharme* undergoes an IR event which renders it chromatin-bound through its association with MATR3 and PTBP1 (63). Another aspect of the effects of IR lies in the turnover of the transcripts where additional sequences added by the introns may cause lower transcript stability. In support of this, nuclear lncRNAs have been shown to be less stable than cytoplasmic ones (5), and *PURPL* also follows this rule. Notably, the human lncRNA *hFAST* is cytoplasmic and promotes pluripotency through WNT signaling, but the mouse homolog does not because of its nuclear localization (64).

Introns have been proposed to act as docking elements that tether lncRNAs to chromatin (4,63), and a recent study showed that nuclear speckles are associated with distinct classes of retained introns (65). This raises the question of what makes intron 2 important for speckle localization of *MALAT1*. One possibility is that this region recruits specific RBPs or adopts a structural configuration required for speckle association. Our data further indicates that intron 2 retention contributes to *MALAT1*'s role in cell migration, potentially because speckle localization is necessary to sustain this phenotype. Future studies are warranted to dig deeper into the underlying mechanism and to determine how intron retention shapes the function of *PURPL* and *MALAT1*.

Collectively, our findings reveal an expected function of U2AF2 in promoting intron retention that governs the subcellular localization and function of the lncRNAs *PURPL* and *MALAT1*. As intron retention continues to emerge as a regulated form of alternative splicing, the evidence presented here expands the framework for the mechanisms and biological consequences of intron retention.

Materials and Methods

Cell culture and treatment

HAP1, HCT116, HEK293T, HepG2, RPE1, SKHEP1, U2OS, WI38, and MDA-MB-231 cells were purchased from ATCC and maintained in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% (v/v) fetal bovine serum (Gibco) and 1% (v/v) PenStrep (Gibco) in a 5% CO₂ atmosphere at 37°C. All cell lines were routinely confirmed to be free of mycoplasma using Venor GeM Mycoplasma Detection Kit (Millipore Sigma-Aldrich).

CRASP-Seq *PURPL* screen

The *PURPL* minigene reporter, composed of exons 2 and 3 with the full native intronic sequence, was cloned into the CRASP-Seq genome-wide library as described previously (43). Reporter fragments were synthesized by TWIST Biosciences and assembled into the Eco32I-digested CRASP-Seq knockout library using NEBuilder (46). Library coverage (>250-fold) was maintained through parallel cloning reactions and electroporation into Endura competent cells followed by plasmid amplification and extraction.

Lentivirus was produced following established protocols (46). The pooled library was used to infect HAP1 and RPE1 cells expressing *SpCas9* and *opCas12a* nucleases (46), at a multiplicity of infection (MOI) of ~0.2. After puromycin selection (2 µg/mL for 48 hours), cells were pooled, induced with doxycycline to activate reporter expression and harvested 24 hours later for RNA analysis.

Total RNA was extracted, poly(A)⁺ RNA isolated, and cDNA synthesized using a U6-annealing primer. Spliced and intron-retaining reporter transcripts were amplified using exon- or intron-specific primers incorporating Illumina barcodes. Libraries were sequenced on the NovaSeq 6000 platform with a 25% PhiX spike-in and a 300-cycle kit. The sequencing strategy consisted of the following configuration: Read 1 (210 cycles), Index Read 1 (8 cycles), Index Read 2 (8 cycles) and Read 2 (85 cycles). Each sample yielded over 500 million total reads.

CRASP-seq data analysis

CRASP-Seq data were processed using a custom pipeline (43). Guides, cell barcodes, and UMIs were retrieved using STAR, and alignment were managed with SAMTools. Reads containing complete hybrid guide, barcodes, UMI, and splicing outcomes were deduplicated prior to analysis. Splicing outcomes were quantified as percent intron retention (PIR): $PIR = [IE / (IE + EE)] \times 100$, where IE represents intron retention reads, and EE represents exon-exon junction reads. Differential PIR (ΔPIR) was calculated as: $\Delta PIR = PIR_{\text{guide}} - PIR_{\text{intergenic}}$

The ΔPIR for each gene and for each guide separately is shown in Table S1 (43). Guides with fewer than 10 reads were excluded. A ΔPIR threshold corresponding to 10% FDR (based on intergenic controls) defined significant hits. At the gene level, hits required FPKM ≥ 0.1 , at least two

qualifying guides, and $\geq 50\%$ of guides exceeding the Δ PIR threshold in both technical replicates. Genes identified through combinatorial targeting that met these criteria were included as hits (Table S1). The heatmap and Venn diagram in Figure 1C and 1E display single-gene hits as well as collapsed gene hits resulting from combinatorial targeting.

eCLIP-seq data and functional enrichment analysis

eCLIP-seq data for U2AF2, U2AF1, PTBP1, PRPF8, and SRSF1 were obtained from the ENCODE database (encodeproject.org) with the following identifiers: ENCFF989JBA (CTRL), ENCFF566CFJ (CTRL), ENCFF536AFD (U2AF2), ENCFF913WRH (U2AF2), ENCFF159SPZ (U2AF2), ENCFF368XEI (U2AF2), ENCFF542NOA (U2AF1), ENCFF651VRQ (U2AF1), ENCFF363UDO (PTBP1), ENCFF044RKN (PRPF8), and ENCFF327NVE (SRSF1). For visualization of RNA-seq data, the Integrated Genome Viewer (IGV) version 17 was used. The functional enrichment analysis was performed using g:Profiler (66) and the results are shown in the 'gProfiler_Results' tab in Table S1.

siRNA transfections

For siRNA transfections, we used AllStars Negative Control siRNA (QIAGEN). U2AF2 siRNAs were purchased from Dharmacon (J-012380-18 and J-012380-19). U2AF1 siRNA was also purchased from Dharmacon (J-012325-10). Cells were reverse transfected with siRNA at a final concentration of 20 nM using Lipofectamine RNAiMAX (Invitrogen) in Opti-MEM I Reduced Serum Medium (Gibco) according to the manufacturer's protocol. SKHEP1 cells were transfected for 72 hr before RNA harvesting. U2OS, HepG2 and WI38 cells were transfected for two rounds of 48 h + 72 h or 72 h + 72 h. For RNA FISH, SKHEP1, HCT116, and MDA-MB-231 cells were transfected for 72 hr. SKHEP1 expressing the *PURPL* Py-tract mutants were transfected for two rounds of 48 h and 24 h before harvesting, the cells were treated with 1 μ g/mL doxycycline to induce the transgene expression. For RNA-seq, the cells were transfected for 72 h before harvested. For RNA stability assay, the cells were reseeded in 12-well plates 48 h after siRNA transfection and 24 h later they were treated with Act D for the indicated time points at 5 μ g/mL before RNA extraction and subsequent assays. For RNA-IP, the cells were transfected for 72 h before harvested in 10 cm-plates.

RNA extraction, RT-qPCR, semi-quantitative RT-PCR, and splicing gels

Total RNA was extracted using TRIzol (Invitrogen) according to the manufacturer's protocol. cDNA was then made using iScript[™] Reverse Transcription Supermix (Bio-Rad). For real-time qPCR, all reactions were carried out on StepOnePlus real-time PCR System (Applied Biosystems) using FastStart SYBR Green Master Mix (Millipore Sigma). *GAPDH* mRNA was used to normalize expression except for RNA-IPs in Figure 2B where *18S* was used for normalization, and the relative expression/enrichment of RNAs was calculated using the $2^{-\Delta\Delta C_t}$ method. RT-qPCR primer sequences for each gene are indicated in Table S3. For splicing gels, cDNA was used to amplify the region flanking the retained intron with primers in the flanking exons with the addition of one primer inside the intron as indicated in Figure S1B. In the case of *MALAT1*, primers in the flanking exons were utilized due to the short length of introns. For the assay of the effect of U2AF2 on the Py-tract mutants of *PURPL*, primers on the transgene were used instead of the endogenous *PURPL* exon 2 and 3. The sequences of the primers can be found in Table S5.

RNA-IPs

For RNA-IP, two 10 cm-plates of SKHEP1 cells were lysed with RIPA buffer (Thermo Scientific) for 10 min on ice before spin down for 10 min at 4°C. The lysate was incubated with 1 μ g of anti-U2AF2 (CST) or rabbit IgG (CST) in 500 μ L total volume overnight at 4°C. The next day, 50 μ L of Protein A/G Dynabeads (Pierce) were added to the mix and incubated for 2 h at 4°C. Beads were then washed 4 times with RIPA buffer and RNA was isolated using TRIzol. Equal volumes of extracted RNA were used for cDNA synthesis and RT-qPCR as described above.

Immunoblotting

For immunoblotting, total cell lysate was prepared using RIPA buffer (Thermo Scientific) containing protease inhibitor cocktail (Roche). Protein concentration was determined using the BCA protein quantitation kit (Thermo Scientific). Cell lysate was loaded onto SDS-PAGE gel and transferred to a PVDF membrane (Thermo Scientific) using a semi-dry apparatus (Bio-Rad). The membrane was blocked in 5% skim milk (Millipore) in TBS containing 0.05% Tween 20. The following antibodies were used: anti-U2AF2 (1:1000, rabbit) (CST, 70471S), anti-GAPDH (1:10,000, rabbit) (CST, 2118L) and secondary anti-rabbit (1:5000) (CST, 7074S).

Nuclear cytoplasmic fractionation

For nuclear cytoplasmic fractionation, SKHEP1 cells were harvested after trypsinization at 80-90% confluence. Cells were pelleted and resuspended in 500 μ L RSB buffer (10 mM Tris-HCl, 100 mM NaCl, and 2.5 mM MgCl₂) containing 40 μ g/mL digitonin (Invitrogen), and 100 U of RNase Out. After incubation on ice for 10 min, cell viability was tested with trypan blue to make sure that the cells were permeabilized. The cytoplasmic fraction was harvested as supernatant after centrifugation at $1,500 \times g$ for 5 min at 4°C. The nuclear pellet was resuspended in 100 μ L RSB buffer. 100 μ L of the cytoplasmic fraction were transferred to a fresh tube and 900 μ L of TRIzol was added as well as to the nuclear extracts for RNA extraction and subsequent cDNA synthesis and RT-qPCR.

Cell proliferation assay

The cell proliferation assay was conducted as previously described (67). To measure proliferation, parental SKHEP1 cells and *PURPL*-CRISPRi cells overexpressing *PURPL* containing intron 2 were seeded at 1×10^5 cells per well in 6-well plates in the presence of 1 μ g/mL doxycycline. After 3 days, the cell numbers and viability were measured. The cells were then passaged at 1×10^5 and measured again at 6 days, and cumulative cell numbers were plotted.

RNA-seq

The TruSeq Stranded mRNA libraries were pooled and sequenced on one NovaSeq6000 S1 flowcell using 2×101 cycles for pair-end run. The Real Time Analysis software (RTA v3.4.4) was used for processing raw data files, the Illumina bcl2fastq v2.20 was used to demultiplex and convert binary base calls and quality scores to fastq format. The sequencing reads were trimmed of adapters and low-quality bases using Cutadapt (v1.18). The trimmed reads were mapped to human reference genome (hg38) and Gencode annotation GENCODE_v30 GTF using STAR aligner (version 2.7.0f) with two-pass alignment option. RSEM (v1.3.1) was used for gene and transcript quantification based on GENCODE annotation file. Palantir Foundry was used for analysis of Differential Expression of Genes inside the secure NIH Integrated Data Analysis Platform (NIDAP). The DEG Analysis results are shown in Table S3 [↗](#).

Iso-Seq

The libraries were prepared using PacBio standard Iso-seq library prep using SMRTbell prep kit 3.0. The libraries were sequenced on PacBio Sequel II platform using PacBio v2.0 chemistry.

PacBio SMRTlink raw subreads were converted into HiFi circular consensus sequences (CCS) with minimum 3 passes. PacBio IsoSeq v3 pipeline was used to demultiplex the barcodes and remove primers. Additional refine steps included trimming poly-A tails and removing concatemers to generate Full Length Non-Concatemer (FLNC) reads. The FLNC reads were used to map to human reference (hg38) using Minimap2 software to generate alignment bam files. Hierarchical clustering and merging steps were performed to obtain consensus isoforms and the full-length (FL) consensus sequences. The high-quality FL transcripts were mapped to the reference genome by using the minimap2 software. Isoform classification and quality control were done using SQANTI3 software and Illumina short-read gene expression data was combined with PacBio Iso-

seq for transcript quantification. The Iso-Seq Analysis results are shown in [Table S4](#). The resulting bam files from the Iso-seq analysis were used for plotting and visualization with the R package Gviz (<https://bioconductor.org/packages/release/bioc/html/Gviz.html>).

IRFinder

IRFinder ([52](#)) was used to identify intron retention events that changed significantly on U2AF2 knockdown. A docker image was downloaded from Docker Hub (`docker://cloxd/irfinder:2.0.1`). The reference was built using pre-downloaded resources (IRFinder BuildRefProcess), using Ensembl release 92 and the 100 bp mappability file provided by IRFinder. The Build-BED-refs.sh script was adjusted so that transcript biotype was not restricted to “processed_transcript” or “protein_coding”. Intron retention was quantified using IRFinder in FastQ mode with adapter trimming. Differential intron retention events between siU2AF2 vs. siCTRL were identified with IRFinder Diff using DESeq2 with the default IR ratio ≥ 0.05 in at least one sample and no warning filters, which profiled 16,859 unique introns.

Construct generation

For Py-tract mutants, gene fragments were purchased from Twist Biosciences and cloned into the pLCHKO-puro vector containing the WT sequences of *PURPL* ([Figure 1A](#)), between sites for NseI and AfeI. The *PURPL* transcript containing intron 2 was cloned into the pCW57 vector (Hygro) (Addgene #80922) in Sall and MluI sites. *MALAT1* transcripts were cloned into the pCW57 vector (BSD) (Addgene #80921) in Sall and MluI sites. The *PURPL* insert was generated by Twist Biosciences and the *MALAT1* sequence was amplified using Phusion DNA polymerase (NEB). For deletion mutants, we used the Q5® Site-Directed Mutagenesis Kit (NEB). The primers used for cloning are listed in [Table S5](#).

Generation of *MALAT1* KO cell lines

For HCT116 *MALAT1* knockout, we used integration by non-homologous end joining, which was accomplished by introducing a simultaneous double-strand break in genomic DNA and in the targeting vectors (pCMV-Puro and pCMV-Hygro) ([68,69](#)). Plasmids encoding spCas9 and sgRNAs were obtained from Addgene (Plasmids #41815 and #47108). Oligonucleotides for construction of sgRNAs were obtained from Integrated DNA Technologies, hybridized, phosphorylated and cloned into the sgRNA plasmid or targeting vector using BbsI sites ([70](#)). Target sequences for sgRNAs are provided in [Table S5](#). These vectors were transfected into HCT116 using Lipofectamine 2000 following recommendations from the manufacturer. Three days after transfection, the cells were selected with Puromycin (0.5 $\mu\text{g}/\text{mL}$) and Hygromycin (100 $\mu\text{g}/\text{mL}$) to generate clonal populations. Genomic DNA from each clone was isolated using DNEasy Blood and Tissue Kit (Qiagen). PCRs to detect integration of the targeting vector at the target site were performed using KAPA2G Robust PCR kits (Kapa Biosystems) according to the manufacturer’s instructions. A typical reaction contained 20-100 ng of genomic DNA in Buffer A (5 μL), Enhancer (5 μL), dNTPs (0.5 μL), primers forward (PINCR Det FP, 1.25 μL) and reverse (Targeting vector Det RP, 1.25 μL) and KAPA2G Robust DNA Polymerase (0.5 U). The DNA sequences of the primers for each target are provided in [Table S5](#). PCR products were visualized in 2% agarose gels and images were captured using a ChemiDoc-It2 (UVP).

To generate MDA-MB-231 whole-locus and intron 2 *MALAT1* knockout clones, we used the CRISPR/Cas9 system by transfecting the cells with gRNAs and purified Cas9. 2 μL of each 30 μM gRNA was mixed with 1 μL (~5 μg) of purified Cas9 and 100 μL of 4D-Nucleofector® X SE buffer SE from the Cell Line Kit (Lonza) supplemented with 18.5 μL of SE supplement per 100 μL total buffer volume. The mixture was incubated at room temperature for 10 min and 1×10^6 cells were resuspended in nucleofection cuvettes. Nucleofections were carried out using the 4D-Nucleofector® System (Lonza). The cells were transferred to a 6-well plate, and they were single-cell sorted 48 h later. The clones were grown for 3-4 weeks before genomic DNA extraction and genotyping with PCR using the primers in [Table S5](#).

PURPL and MALAT1 overexpression and transduction

PURPL intron 2 constructs were transduced in previously generated *PURPL*-CRISPRi SKHEP1 cells (11). *MALAT1*-WT and deletion mutants were transduced in the two HCT116-*MALAT1* KO clones. For lentivirus generation, HEK293T cells were seeded at 2×10^5 cells/well onto 6-well plates, and 1,200 ng of each construct was transfected with lentiviral package vectors using Lipofectamine 2000 (Life Technologies, Invitrogen) according to the manufacturer's protocol. Virus-containing media were harvested at 48 and 72 h post-transfection and added to cells for 24 h. Transduced cells were selected with 500 $\mu\text{g}/\text{mL}$ hygromycin for 2 weeks (SKHEP1) or 10 $\mu\text{g}/\text{mL}$ Blasticidin (HCT116).

RNA-FISH, IF and Quantification of FISH

The *PURPL* intron probe set was custom-designed using the Stellaris Probe Designer, tagged with Quasar 570 dye and purchased from Stellaris. The *MALAT1* smFISH probe set with Quasar 670 dye was purchased from Stellaris (Cat# VSMF-2211-5). For *PURPL* intron smRNA-FISH, SKHEP1 cells were seeded on coverslips and cultured for 24 hours at 37°C with 5% CO₂ before inducing DNA damage by incubating the cells with 2 mM hydroxyurea for 24 hours (or regular growth media for control). For *MALAT1* smRNA-FISH, HCT116 and MDA-MB-231 *MALAT1* KO cells were seeded on coverslips and cultured for 24 hours at 37°C with 5% CO₂. *MALAT1* WT, *MALAT1* del-1, *MALAT1* del-2, *MALAT1* del1+2, or empty vector expression was induced by incubating the cells with 1 $\mu\text{g}/\text{mL}$ doxycycline for 48 hours. The cells that were transfected with siRNA for U2AF2 were seeded on coverslips at the time of transfection. Cells were fixed with 4% PFA for 15 min at room temperature, permeabilized with 0.5% Triton X-100, and washed with washing buffer (10% formamide, 2XSSC) for 5 min. Probe was added to hybridization buffer (10% formamide in Stellaris hybridization buffer Cat# SMF-HB1-10) at a final concentration of 125 nM. Hybridization was done in a humidified chamber in the dark overnight at 37°C. After hybridization, the coverslips were washed twice with wash buffer for 30 min at 37°C and then post-fixed with 4% PFA for 10 min at room temperature. The coverslips were blocked with 5% normal goat serum and then incubated with SON antibody (Sigma Cat# HPA023535). Primary and secondary antibodies were diluted in 1% normal goat serum. Antibody incubations were done at room temperature for 1 h and washed with PBS. DNA was counterstained by DAPI. The coverslips were then washed with 4XSSC for 5 min at room temperature and mounted in VectaShield Antifade Mounting Medium (Vector Laboratories). Images were taken using DeltaVision microscope (GE Healthcare) equipped with 60 X/1.42 NA oil immersion objective (Olympus) and CoolSNAP-HQ2 camera. To quantify the speckle to nucleoplasm signal intensity ratio of *MALAT1* in Figures 5B, 7B, 7D, and S8B ImageJ was used to create nucleus masks based on DAPI and speckles mask based on SON. The signal intensity of *MALAT1* in the speckles mask was divided by the *MALAT1* signal intensity in the nucleus-minus-speckles area. For the quantification of Figures 6B, 7E and S10B, the cells with *MALAT1* FISH signal showing prominent speckle patterns similar to SON were classified as "Speckle" pattern, and the cells with a diffused *MALAT1* FISH signal were classified as "Diffused". Only *MALAT1*-positive cells were counted for quantification.

Transwell cell migration assays

For the transwell assay, WT and KO MDA-MB-231 clones were seeded at 10×10^5 cells per well in 12-well plates. 24 h later, the media were removed and replaced with media containing 1% FBS. After 5 h, the cells were trypsinized and seeded in the top chamber of the transwell plates (Corning, Catalog#354578) at 25,000 cells per well at a volume of 100 μL . The lower chamber of the transwell is filled with 700 μL of media containing 10% FBS. The cells were incubated for 24 h and the transwell were removed from the plates. The non-migrated cells were scraped with a cotton swab, and the migrated cells were fixed with ice-cold methanol for 5 min and stained with 0.05% crystal violet blue. The migrated cells were counted under a light microscope.

Statistics

Statistics were performed using the Student's t-test.

Data availability

The Iso-Seq data files have been deposited at GEO with accession number GSE288684 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE288684>). The RNA-Seq data files have been deposited at GEO with accession number GSE288685 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE288685>).

Acknowledgements

This research was supported by the Intramural Research Program (A.L.) of the National Cancer Institute (NCI), Center for Cancer Research (CCR) (project numbers ZIA BC011646, ZIA BC012019), NIH. KVP laboratory is supported by grants from NIH (R01-GM132458), Cancer center at Illinois seed grant, ARPA-H, and National Science Foundation (NSF) center for quantitative cell biology. YJS is supported by NIH T32EB019944. We thank the CCR Genomics Core, CCR, NCI, Bethesda, MD for its valuable assistance with Sanger and Nanopore sequencing. We also thank the Sequencing Facility, CCR, NCI, Frederick, MD for performing the RNA sequencing assays. The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Additional information

Author Contributions

IG, CN, YJS, CCH, SK, RP, XLL, RK, RS, and PPZ conducted experiments. IG, AKB, EP, YZ, BS, TB, XW, TGP, KP, and AL did data analysis. IG and AL wrote the manuscript. NC, KP, and TGP edited the manuscript.

Funding

Funder	Grant reference number	Author
HHS NIH National Cancer Institute (NCI)	ZIA BC011646	Ashish Lal

Author ORCID iDs

Ioannis Grammatikakis: <https://orcid.org/0000-0002-8455-1584>

Thomas Gonatopoulos-Pournatzis: <https://orcid.org/0000-0003-2669-2570>

Ashish Lal: <https://orcid.org/0000-0002-4299-8177>

Additional files

[Supplemental Figures](#)

[Supplemental Table 1](#)

[Supplemental Table 2](#)

[Supplemental Table 3](#)

[Supplemental Table 4](#)

[Supplemental Table 5](#)

References

1. Quinn J.J., Chang H.Y (2016) Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* **17**:47-62 <https://doi.org/10.1038/nrg.2015.10> | [PubMed](#)

2. Ulitsky I., Bartel D.P (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**:26-46 <https://doi.org/10.1016/j.cell.2013.06.020> | PubMed
3. Guo C.J., Xu G., Chen L.L (2020) Mechanisms of Long Noncoding RNA Nuclear Retention. *Trends Biochem Sci* **45**:947-960 <https://doi.org/10.1016/j.tibs.2020.07.001> | PubMed
4. Zuckerman B., Ulitsky I (2019) Predictive models of subcellular localization of long RNAs. *RNA* **25**:557-572 <https://doi.org/10.1261/rna.068288.118> | PubMed
5. Clark M.B., Johnston R.L., Inostroza-Ponta M., Fox A.H., Fortini E., Moscato P., Dinger M.E., Mattick J.S (2012) Genome-wide analysis of long noncoding RNA stability. *Genome Res* **22**:885-898 <https://doi.org/10.1101/gr.131037.111> | PubMed
6. Cabili M.N., Dunagin M.C., McClanahan P.D., Biaisch A., Padovan-Merhar O., Regev A., Rinn J.L., Raj A (2015) Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16**:20 <https://doi.org/10.1186/s13059-015-0586-4> | PubMed
7. Tilgner H., Knowles D.G., Johnson R., Davis C.A., Chakraborty S., Djebali S., Curado J., Snyder M., Gingeras T.R., Guigo R (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22**:1616-1625 <https://doi.org/10.1101/gr.134445.111> | PubMed
8. Mele M., Mattioli K., Mallard W., Shechner D.M., Gerhardinger C., Rinn J.L (2017) Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res* **27**:27-37 <https://doi.org/10.1101/gr.214205.116> | PubMed
9. Kumar M., Grammatikakis I (2025) Causes, Consequences and Challenges of Intron Retention in lncRNAs. *Mol Cell Biol* **45**:576-584 <https://doi.org/10.1080/10985549.2025.2539755> | PubMed
10. Mukherjee N., Calviello L., Hirsekorn A., de Pretis S., Pelizzola M., Ohler U (2017) Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat Struct Mol Biol* **24**:86-96 <https://doi.org/10.1038/nsmb.3325> | PubMed
11. Hartford C.C.R., Shrestha R.L., Pongor L., Zhao Y., Chen X., Fromont C., Chaudhary R., Li X.L., Pasterczyk K.R., Kumar R., *et al.* (2022) Context-Dependent Function of Long Noncoding RNA PURPL in Transcriptome Regulation during p53 Activation. *Mol Cell Biol* **42**:e0028922 <https://doi.org/10.1128/mcb.00289-22> | PubMed
12. Black D.L (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**:291-336 <https://doi.org/10.1146/annurev.biochem.72.121801.161720> | PubMed
13. Wilkinson M.E., Charenton C., Nagai K (2020) RNA Splicing by the Spliceosome. *Annu Rev Biochem* **89**:359-388 <https://doi.org/10.1146/annurev-biochem-091719-064225> | PubMed
14. Ruskin B., Zamore P.D., Green M.R (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**:207-219 [https://doi.org/10.1016/0092-8674\(88\)90509-0](https://doi.org/10.1016/0092-8674(88)90509-0) | PubMed
15. Berglund J.A., Chua K., Abovich N., Reed R., Rosbash M (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**:781-787 [https://doi.org/10.1016/s0092-8674\(00\)80261-5](https://doi.org/10.1016/s0092-8674(00)80261-5) | PubMed
16. Berglund J.A., Abovich N., Rosbash M (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* **12**:858-867 <https://doi.org/10.1101/gad.12.6.858> | PubMed
17. Wu S., Romfo C.M., Nilsen T.W., Green M.R (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**:832-835 <https://doi.org/10.1038/45590> | PubMed
18. Zorio D.A., Blumenthal T (1999) Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* **402**:835-838 <https://doi.org/10.1038/45597> | PubMed
19. Wang E.T., Sandberg R., Luo S., Khrebukova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., Burge C.B (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**:470-476 <https://doi.org/10.1038/nature07509> | PubMed
20. Nilsen T.W., Graveley B.R (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**:457-463 <https://doi.org/10.1038/nature08909> | PubMed

21. Pan Q., Shai O., Lee L.J., Frey B.J., Blencowe B.J (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**:1413-1415 <https://doi.org/10.1038/ng.259> | [PubMed](#)
22. Gier R.A., Budinich K.A., Evitt N.H., Cao Z., Freilich E.S., Chen Q., Qi J., Lan Y., Kohli R.M., Shi J (2020) High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening. *Nat Commun* **11**:3455 <https://doi.org/10.1038/s41467-020-17209-1> | [PubMed](#)
23. Braunschweig U., Barbosa-Morais N.L., Pan Q., Nachman E.N., Alipanahi B., Gontopoulos-Pournatzis T., Frey B., Irimia M., Blencowe B.J (2014) Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24**:1774-1786 <https://doi.org/10.1101/gr.177790.114> | [PubMed](#)
24. Wong J.J., Ritchie W., Ebner O.A., Selbach M., Wong J.W., Huang Y., Gao D., Pinello N., Gonzalez M., Baidya K., et al. (2013) Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154**:583-595 <https://doi.org/10.1016/j.cell.2013.06.052> | [PubMed](#)
25. Monteuis G., Wong J.J.L., Bailey C.G., Schmitz U., Rasko J.E.J (2019) The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res* **47**:11497-11513 <https://doi.org/10.1093/nar/gkz1068> | [PubMed](#)
26. Eom T., Zhang C., Wang H., Lay K., Fak J., Noebels J.L., Darnell R.B (2013) NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. *eLife* **2**:e00178 <https://doi.org/10.7554/eLife.00178> | [PubMed](#)
27. Hossain M.A., Claggett J.M., Edwards S.R., Shi A., Pennebaker S.L., Cheng M.Y., Hasty J., Johnson T.L (2016) Posttranscriptional Regulation of Gcr1 Expression and Activity Is Crucial for Metabolic Adjustment in Response to Glucose Availability. *Mol Cell* **62**:346-358 <https://doi.org/10.1016/j.molcel.2016.04.012> | [PubMed](#)
28. Cha S., Hong C.P., Kang H.A., Hahn J.S (2021) Differential activation mechanisms of two isoforms of Gcr1 transcription factor generated from spliced and un-spliced transcripts in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **49**:745-759 <https://doi.org/10.1093/nar/gkaa1221> | [PubMed](#)
29. Schmitz U., Pinello N., Jia F., Alasmari S., Ritchie W., Keightley M.C., Shini S., Lieschke G.J., Wong J.J., Rasko J.E.J (2017) Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol* **18**:216 <https://doi.org/10.1186/s13059-017-1339-3> | [PubMed](#)
30. Dvinge H., Bradley R.K (2015) Widespread intron retention diversifies most cancer transcriptomes. *Genome Med* **7**:45 <https://doi.org/10.1186/s13073-015-0168-9> | [PubMed](#)
31. Boutz P.L., Bhutkar A., Sharp P.A (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* **29**:63-80 <https://doi.org/10.1101/gad.247361.114> | [PubMed](#)
32. Gutschner T., Hammerle M., Eissmann M., Hsu J., Kim Y., Hung G., Revenko A., Arun G., Stentrup M., Gross M., et al. (2013) The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* **73**:1180-1189 <https://doi.org/10.1158/0008-5472.can-12-2850> | [PubMed](#)
33. Hutchinson J.N., Ensminger A.W., Clemson C.M., Lynch C.R., Lawrence J.B., Chess A (2007) A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**:39 <https://doi.org/10.1186/1471-2164-8-39> | [PubMed](#)
34. Ji P., Diederichs S., Wang W., Boing S., Metzger R., Schneider P.M., Tidow N., Brandt B., Buerger H., Bulk E., et al. (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**:8031-8041 <https://doi.org/10.1038/sj.onc.1206928> | [PubMed](#)
35. Bernard D., Prasanth K.V., Tripathi V., Colasse S., Nakamura T., Xuan Z., Zhang M.Q., Sedel F., Jourden L., Couplier F., et al. (2010) A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* **29**:3082-3093 <https://doi.org/10.1038/emboj.2010.199> | [PubMed](#)

36. Tripathi V., Ellis J.D., Shen Z., Song D.Y., Pan Q., Watt A.T., Freier S.M., Bennett C.F., Sharma A., Bubulya P.A., *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39**:925-938 <https://doi.org/10.1016/j.molcel.2010.08.011> | PubMed
37. Shinn M.K., Tomares D.T., Liu V., Pant A., Qiu Y., Vitalis A., Song Y.J., Ayala Y., Ruff K.M., Strout G.W., *et al.* (2026) Nuclear speckle proteins form intrinsic and MALAT1-dependent microphases. *Cell* **189**:832-852. <https://doi.org/10.1016/j.cell.2025.11.026> | PubMed
38. Spector D.L., Lamond A.I. (2011) Nuclear speckles. *Cold Spring Harb Perspect Biol* **3** <https://doi.org/10.1101/cshperspect.a000646> | PubMed
39. Sharma A., Takata H., Shibahara K., Bubulya A., Bubulya P.A. (2010) Son is essential for nuclear speckle organization and cell cycle progression. *Mol Biol Cell* **21**:650-663 <https://doi.org/10.1091/mbc.e09-02-0126> | PubMed
40. Nakagawa S., Ip J.Y., Shioi G., Tripathi V., Zong X., Hirose T., Prasanth K.V. (2012) Malat1 is not an essential component of nuclear speckles in mice. *RNA* **18**:1487-1499 <https://doi.org/10.1261/rna.033217.112> | PubMed
41. Zhang B., Arun G., Mao Y.S., Lazar Z., Hung G., Bhattacharjee G., Xiao X., Booth C.J., Wu J., Zhang C., *et al.* (2012) The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep* **2**:111-123 <https://doi.org/10.1016/j.celrep.2012.06.003> | PubMed
42. Li X.L., Subramanian M., Jones M.F., Chaudhary R., Singh D.K., Zong X., Gryder B., Sindri S., Mo M., Schetter A., *et al.* (2017) Long Noncoding RNA PURPL Suppresses Basal p53 Levels and Promotes Tumorigenicity in Colorectal Cancer. *Cell Rep* **20**:2408-2423 <https://doi.org/10.1016/j.celrep.2017.08.041> | PubMed
43. Behera A.K., Kim J.J., Kordale S., Pekovic F., Damodaran A.P., Kumari B., Vidak S., Dickson E., Xiao M.S., Duncan G., *et al.* (2026) RNA-coupled CRISPR screens reveal ZNF207 as a regulator of LMNA aberrant splicing in progeria. *Mol Cell* **86**:41-59. <https://doi.org/10.1016/j.molcel.2025.12.003> | PubMed
44. Aregger M., Xing K., Gonatopoulos-Pournatzis T. (2021) Application of CHyMERa Cas9-Cas12a combinatorial genome-editing platform for genetic interaction mapping and gene fragment deletion screening. *Nat Protoc* **16**:4722-4765 <https://doi.org/10.1038/s41596-021-00595-1> | PubMed
45. Gonatopoulos-Pournatzis T., Aregger M., Brown K.R., Farhangmehr S., Braunschweig U., Ward H.N., Ha K.C.H., Weiss A., Billmann M., Durbic T., *et al.* (2020) Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9-Cas12a platform. *Nat Biotechnol* **38**:638-648 <https://doi.org/10.1038/s41587-020-0437-z> | PubMed
46. Xiao M.S., Damodaran A.P., Kumari B., Dickson E., Xing K., On T.A., Parab N., King H.E., Perez A.R., Guiblet W.M., *et al.* (2024) Genome-scale exon perturbation screens uncover exons critical for cell fitness. *Mol Cell* **84**:2553-2572. <https://doi.org/10.1016/j.molcel.2024.05.024> | PubMed
47. Van Nostrand E.L., Pratt G.A., Shishkin A.A., Gelboin-Burkhart C., Fang M.Y., Sundararaman B., Blue S.M., Nguyen T.B., Surka C., Elkins K., *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**:508-514 <https://doi.org/10.1038/nmeth.3810> | PubMed
48. Consortium E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**:57-74 <https://doi.org/10.1038/nature11247> | PubMed
49. Luo Y., Hitz B.C., Gabdank I., Hilton J.A., Kagda M.S., Lam B., Myers Z., Sud P., Jou J., Lin K., *et al.* (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* **48**:D882-D889 <https://doi.org/10.1093/nar/gkz1062> | PubMed
50. Berhane T., Holm A., Karstensen K.T., Petri A., Ilieva M.S., Krarup H., Vyberg M., Lovendorf M.B., Kauppinen S. (2022) Knockdown of the long noncoding RNA PURPL induces apoptosis and sensitizes liver cancer cells to doxorubicin. *Sci Rep* **12**:19502 <https://doi.org/10.1038/s41598-022-23802-9> | PubMed

51. Han S., Li X., Wang K., Zhu D., Meng B., Liu J., Liang X., Jin Y., Liu X., Wen Q., *et al.* (2021) PURPL represses autophagic cell death to promote cutaneous melanoma by modulating ULK1 phosphorylation. *Cell Death Dis* **12**:1070 <https://doi.org/10.1038/s41419-021-04362-8> | PubMed
52. Middleton R., Gao D., Thomas A., Singh B., Au A., Wong J.J., Bomane A., Cosson B., Eyras E., Rasko J.E., *et al.* (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol* **18**:51 <https://doi.org/10.1186/s13059-017-1184-4> | PubMed
53. Martinez-Terroba E., Plasek-Hegde L.M., Chiotakakos I., Li V., de Miguel F.J., Robles-Oteiza C., Tyagi A., Politi K., Zamudio J.R., Dimitrova N (2024) Overexpression of Malat1 drives metastasis through inflammatory reprogramming of the tumor microenvironment. *Sci Immunol* **9**:eadh5462 <https://doi.org/10.1126/sciimmunol.adh5462> | PubMed
54. Arun G., Aggarwal D., Spector D.L (2020) MALAT1 Long Non-Coding RNA: Functional Implications. *Noncoding RNA* **6** <https://doi.org/10.3390/ncrna6020022> | PubMed
55. Jacob A.G., Smith C.W.J (2017) Intron retention as a component of regulated gene expression programs. *Hum Genet* **136**:1043-1057 <https://doi.org/10.1007/s00439-017-1791-x> | PubMed
56. Shao C., Yang B., Wu T., Huang J., Tang P., Zhou Y., Zhou J., Qiu J., Jiang L., Li H., *et al.* (2014) Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol* **21**:997-1005 <https://doi.org/10.1038/nsmb.2906> | PubMed
57. Naftelberg S., Schor I.E., Ast G., Kornblihtt A.R (2015) Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem* **84**:165-198 <https://doi.org/10.1146/annurev-biochem-060614-034242> | PubMed
58. Herzog L., Ottoz D.S.M., Alpert T., Neugebauer K.M (2017) Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**:637-650 <https://doi.org/10.1038/nrm.2017.63> | PubMed
59. Wong J.J., Gao D., Nguyen T.V., Kwok C.T., van Geldermalsen M., Middleton R., Pinello N., Thoeng A., Nagarajah R., Holst J., *et al.* (2017) Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. *Nat Commun* **8**:15134 <https://doi.org/10.1038/ncomms15134> | PubMed
60. Guo R., Zheng L., Park J.W., Lv R., Chen H., Jiao F., Xu W., Mu S., Wen H., Qiu J., *et al.* (2014) BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Mol Cell* **56**:298-310 <https://doi.org/10.1016/j.molcel.2014.08.022> | PubMed
61. Krchnakova Z., Thakur P.K., Krausova M., Bieberstein N., Haberman N., Muller-McNicoll M., Stanek D (2019) Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5' splice-site sequences due to weak interactions with SR proteins. *Nucleic Acids Res* **47**:911-928 <https://doi.org/10.1093/nar/gky1147> | PubMed
62. Dumbovic G., Braunschweig U., Langner H.K., Smallegan M., Biayna J., Hass E.P., Jastrzebska K., Blencowe B., Cech T.R., Caruthers M.H., *et al.* (2021) Nuclear compartmentalization of TERT mRNA and TUG1 lncRNA is driven by intron retention. *Nat Commun* **12**:3308 <https://doi.org/10.1038/s41467-021-23221-w> | PubMed
63. Desideri F., Cipriano A., Petrezselyova S., Buonaiuto G., Santini T., Kasperek P., Prochazka J., Janson G., Paiardini A., Calicchio A., *et al.* (2020) Intronic Determinants Coordinate lncRNA Nuclear Activity through the Interaction with MATR3 and PTBP1. *Cell Rep* **33**:108548 <https://doi.org/10.1016/j.celrep.2020.108548> | PubMed
64. Guo C.J., Ma X.K., Xing Y.H., Zheng C.C., Xu Y.F., Shan L., Zhang J., Wang S., Wang Y., Carmichael G.G., *et al.* (2020) Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell* **181**:621-636. <https://doi.org/10.1016/j.cell.2020.03.006> | PubMed
65. Barutcu A.R., Wu M., Braunschweig U., Dyakov B.J.A., Luo Z., Turner K.M., Durbic T., Lin Z.Y., Weatheritt R.J., Maass P.G., *et al.* (2022) Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol Cell* **82**:1035-1052. <https://doi.org/10.1016/j.molcel.2021.12.010> | PubMed

66. Kolberg L., Raudvere U., Kuzmin I., Adler P., Vilo J., Peterson H (2023) g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**:W207-W212 <https://doi.org/10.1093/nar/gkad347> | PubMed
67. Lee J.S., Dan T., Zhang H., Cheng Y., Rehfeld F., Brugarolas J., Mendell J.T (2025) An ultraconserved snoRNA-like element in long noncoding RNA CRNDE promotes ribosome biogenesis and cell proliferation. *Mol Cell* **85**:1543-1560. <https://doi.org/10.1016/j.molcel.2025.03.006> | PubMed
68. Brown A., Winter J., Gapinske M., Tague N., Woods W.S., Perez-Pinera P (2019) Multiplexed and tunable transcriptional activation by promoter insertion using nuclease-assisted vector integration. *Nucleic Acids Res* **47**:e67 <https://doi.org/10.1093/nar/gkz210> | PubMed
69. Brown A., Woods W.S., Perez-Pinera P (2016) Multiplexed Targeted Genome Engineering Using a Universal Nuclease-Assisted Vector Integration System. *ACS Synth Biol* **5**:582-588 <https://doi.org/10.1021/acssynbio.6b00056> | PubMed
70. Brown A., Woods W.S., Perez-Pinera P (2017) Targeted Gene Activation Using RNA-Guided Nucleases. *Methods Mol Biol* **1468**:235-250 https://doi.org/10.1007/978-1-4939-4035-6_16 | PubMed

Peer reviews

Reviewer #1 (Public review):

Intron retention is observed in many long noncoding RNAs. The authors here used a powerful genome-wide screening strategy to identify proteins controlling intron retention in the long noncoding RNA PURPL. One of the top hits across multiple cell lines surprisingly, was U2AF2, which is well known to bind the polypyrimidine tract close to the 3' splice site to promote splicing. Nonetheless, U2AF2 is working in the opposite direction here. Convincing follow-up RT-PCR experiments confirmed that knocking down U2AF2 does indeed lead to reduced intron retention of PURPL. The authors then show that this intron retention event is functionally important for both the nuclear retention of PURPL as well as its ability to enhance cell proliferation.

The authors then used transcriptome-wide analyses to look for additional intron retention events affected by U2AF2. Among the ~250 genes with decreased intron retention (more splicing) upon U2AF2 knockdown was MALAT1, a well-established long noncoding RNA that normally localizes to nuclear speckles. Depletion of U2AF2 or removal of the MALAT1 2nd intron resulted in reduced speckle localization and cell migration, revealing a critical and fascinating role for this intron retention event. Overall, the authors have used a set of complementary approaches to clearly demonstrate a very intriguing role for U2AF2 in controlling intron retention and functionality of a set of long noncoding RNAs.

I feel the current work has revealed an important role of intron retention in controlling the localization and functionality of long noncoding RNAs, which is likely broad in scope and is likely regulated by cell state.

One experimental suggestion: The authors show that expressing intron-2 containing PURPL in PURPL-depleted cells is sufficient to induce faster proliferation, but a valuable comparison would be identifying the phenotype expressing spliced PURPL transcript.

<https://doi.org/10.7554/eLife.111394.1.sa2>

Reviewer #2 (Public review):

Summary:

This study identified U2AF1/2 as a regulator of pre-mRNA splicing that either promotes or suppresses the splicing of introns on different genes. The authors then focused on two genes

PURPL and MALAT1 that U2AF1/2 can promote intron retention of specific introns, and characterized the biological implications of these introns regulated by U2AF1/2.

Strengths:

(1) The experiments in this manuscript are relatively rigorously designed and performed, often with validation checks such as verifying the knockout, verifying the treatment itself doesn't have an effect, etc.

(2) The experiments provided comprehensive support for the claims that these specific introns are important for the stability or nuclear localization of the RNA, as well as that U2AF1/2 suppresses the splicing of these introns.

(3) The writing of the manuscript is very clear and doesn't overstate the conclusions that can be drawn from the experiments.

Weaknesses:

I think one main weakness of this study is the lack of a deeper analysis of the mechanisms. Whether studying the mechanism is within the scope of this paper is probably debatable, but with the current experiment setup and data, I believe there are some analyses that can be relatively easily done to enhance the value or significance of this study. My detailed questions and suggestions are listed below:

(1) Line 194-195 and Figure 2A: How many RBPs are included in "other RBPs" in line 194? Does "other RBPs" only include PTBP1, PRPF8 and SRSF1 in Figure 2A, or do they include all the ~100 RBPs with HepG2 eCLIP data available on ENCODE? If U2AF1/2 have the highest occupancy around the intron 2 region among the ~100 RBPs, it would be nice to visualize it.

(2) Figure 2A and 2B: Why didn't U2AF2 show interaction with exon 2 and 3 in RNA-IP but showed enrichment over exon 2 and exon 3 regions in the eCLIP data?

(3) Figure 3C - 3F: Maybe I misinterpreted the experiments, but to my understanding, these experiments showed that the exogenous PURPL with intron 2 promoted cell proliferation compared to when the exogenous PURPL wasn't induced, but didn't compare to the effect of the same amount of PURPL with intron 2 removed. Wouldn't it be clearer to compare the effects of exogenous PURPL with intron 2 and exogenous PURPL without intron 2 to pinpoint whether the effect is related to intron 2? Without an intron 2 specific experiment, these current experiments don't seem to provide much added value than "PURPL promotes cell proliferation".

(4) It's not very clear what proportion of these introns are retained in the endogenous PURPL and MALAT1 in various tissues, cell types and conditions. I think it will be valuable to provide this background (either from previous research, public database or data from this study).

(5) Since U2AF1/2 have a wide range of targets as demonstrated by Figure 4A, I think it would be valuable to have some experiments that directly disrupt the interaction between U2AF1/2 and PURPL and MALAT1 and test the effect on splicing outcomes, such as by mutating the sequence that U2AF1/2 bind to. The section on the weak py-tract of PURPL touched upon this topic but focused more on how the weak py-tract causes the intron 2 retention in the background rather than how U2AF1/2 binding and action were affected by sequence mutations. I think experiments on disrupting the direct binding between U2AF1/2 on targets can provide valuable mechanistic insights.

(6) Across all the target genes of U2AF1/2, it might be feasible to do some systematic analysis to find what correlates with whether U2AF1/2 have a promoting or suppressing effect on intron splicing. For example, do genes with decreased IR after U2AF2 depletion systematically

have a weak py-tract compared to genes with increased IR? This dataset can potentially provide many hypotheses for understanding the dual role of U2AF1/2.

<https://doi.org/10.7554/eLife.111394.1.sa1>

Reviewer #3 (Public review):

Summary:

This manuscript characterized the splicing regulation of two long non-coding RNAs relevant to cancer, starting with a focus on PURPL and ending with insights into MALAT1. A CRISPR screen for the regulators of PURPL intron retention revealed a role for the U2AF heterodimer in inducing this retention, with U2AF2 as the actual hit. This is surprising, because the canonical function of U2AF is to recognize the polypyrimidine tract (PPT) and 3' splice site junction to induce splicing at the site. The brief mechanistic characterization of this phenomenon showed that this intron retention accounts for the nuclear localization and instability of the PURPL transcript, and seems to confer the enhanced cell proliferation feature. U2AF2 also induces retention of two introns in MALAT1, and one of them is essential for its nuclear speckle localization and enhanced cell migration.

Strengths:

These findings about PURPL and MALAT1 are clear and interesting.

Weaknesses:

The results are not sufficiently connected to each other, because one regulation is nuclear-speckle dependent but not the other.

Here are my specific comments:

Major comments:

The main issue is the lack of focus because of the distinct and incomplete analysis pertaining to the two long noncoding RNAs, PURPL and MALAT1. The paper starts with a very good genetic screen on the former, and immunofluorescence and functional analysis on the latter, with U2AF2 as the main link to induce intron retention. The first one does not show clear localization while the second docks to nuclear speckles, apparently because of the retained intron. Hence the two mechanisms are related yet distinct. Here are some suggestions to enhance the characterization and connection between the two cases:

- (1) As the MALAT1 intron 2 retention contributes to its speckle localization but not the retained PURPL intron, the retained introns or their 3' splice site sequences should be swapped to see if they determine the localization.
- (2) Figure 3, the rescue of the PURPL knockout by the intron-retained RNA to induce proliferation is a powerful experiment, that is lacking the rescue with the RNA without the intron as a control. This must be done and shown.
- (3) The weakness of the PPT of PURPL intron 2 appears as a clear feature of its retention dependent on U2AF2, which appears direct, as backed by CLIP data. It would be good to show direct binding by EMSA or equivalent techniques. Furthermore, the data is also consistent with other determinants. The exon and upstream intronic sequences, including the branch point, could also be involved, so mutations in these are also required.
- (4) In brief, what are the commonalities and differences between PURPL and MALAT1 with regard to their U2AF2-dependent intron retention?

<https://doi.org/10.7554/eLife.111394.1.sa0>