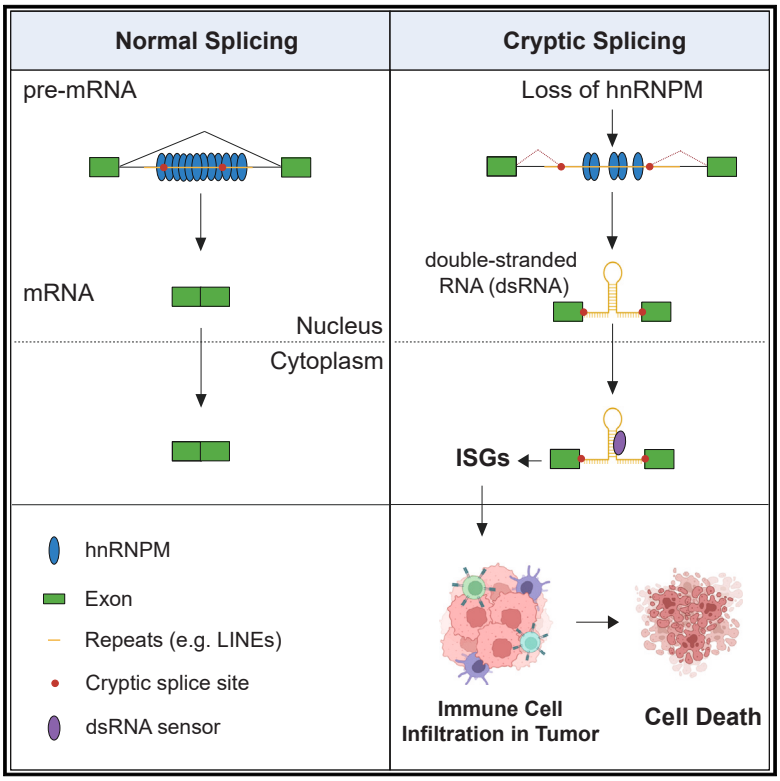
Molecular Cell

Article

hnRNPM protects against the dsRNA-mediated interferon response by repressing LINE-associated cryptic splicing

Graphical abstract



Authors

Rong Zheng, Mikayla Dunlap, Georg O.M. Bobkov, ..., Eric L. Van Nostrand, Xinshu Xiao, Chonghui Cheng

Correspondence

gxxiao@g.ucla.edu (X.X.), chonghui.cheng@bcm.edu (C.C.)

In brief

Zheng et al. reveal that depletion of hnRNPM results in global cryptic splicing. hnRNPM-repressed cryptic exons are enriched in LINEs and can generate long, stem-like double-stranded RNAs (dsRNAs), which in turn stimulate antiviral signaling, elevate tumor immune infiltration, and correlate with improved cancer patient survival.

Highlights

- hnRNPM is a major RBP that occupies deep introns to repress cryptic splicing
- hnRNPM binding is enriched in LINEs, repressing pseudo splice sites
- LINE/Alu-containing cryptic exons can form dsRNAs, triggering IFN-I responses
- hnRNPM^{low} tumors show increased cryptic splicing, IFN-I, and immune infiltration

Zheng et al., 2024, Molecular Cell *84*, 2087–2103
June 6, 2024
https://doi.org/10.1016/j.molcel.2024.05.004

CellPress

Molecular Cell

CellPress

Article

hnRNPM protects against the dsRNA-mediated interferon response by repressing LINE-associated cryptic splicing

Rong Zheng,[1] Mikayla Dunlap,[1] Georg O.M. Bobkov,[1] Carlos Gonzalez-Figueroa,[3] Khushali J. Patel,[1] Jingyi Lyu,[1] Samuel E. Harvey,[1] Tracey W. Chan,[4] Giovanni Quinones-Valdez,[3] Mudra Choudhury,[4] Charlotte A. Le Roux,[2] Mason D. Bartels,[2] Amy Vuong,[1] Ryan A. Flynn,[6,8] Howard Y. Chang,[6,7] Eric L. Van Nostrand,[2] Xinshu Xiao,[3,4,5,*] and Chonghui Cheng[1,9,*]

[1]Lester & Sue Smith Breast Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[2]Verna & Marrs McLean Department of Biochemistry & Molecular Pharmacology and Therapeutic Innovation Center, Baylor College of Medicine, Houston, TX 77030, USA
[3]Department of Integrative Biology and Physiology and the Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA
[4]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA
[5]Department of Bioengineering, University of California, Los Angeles, Los Angeles, CA 90095, USA
[6]Center for Personal Dynamic Regulome, Stanford University School of Medicine, Stanford, CA 94305, USA
[7]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
[8]Present address: Stem Cell Program, Boston Children's Hospital, Boston, MA 02115, USA
[9]Lead contact
*Correspondence: gxxiao@g.ucla.edu (X.X.), chonghui.cheng@bcm.edu (C.C.)
https://doi.org/10.1016/j.molcel.2024.05.004

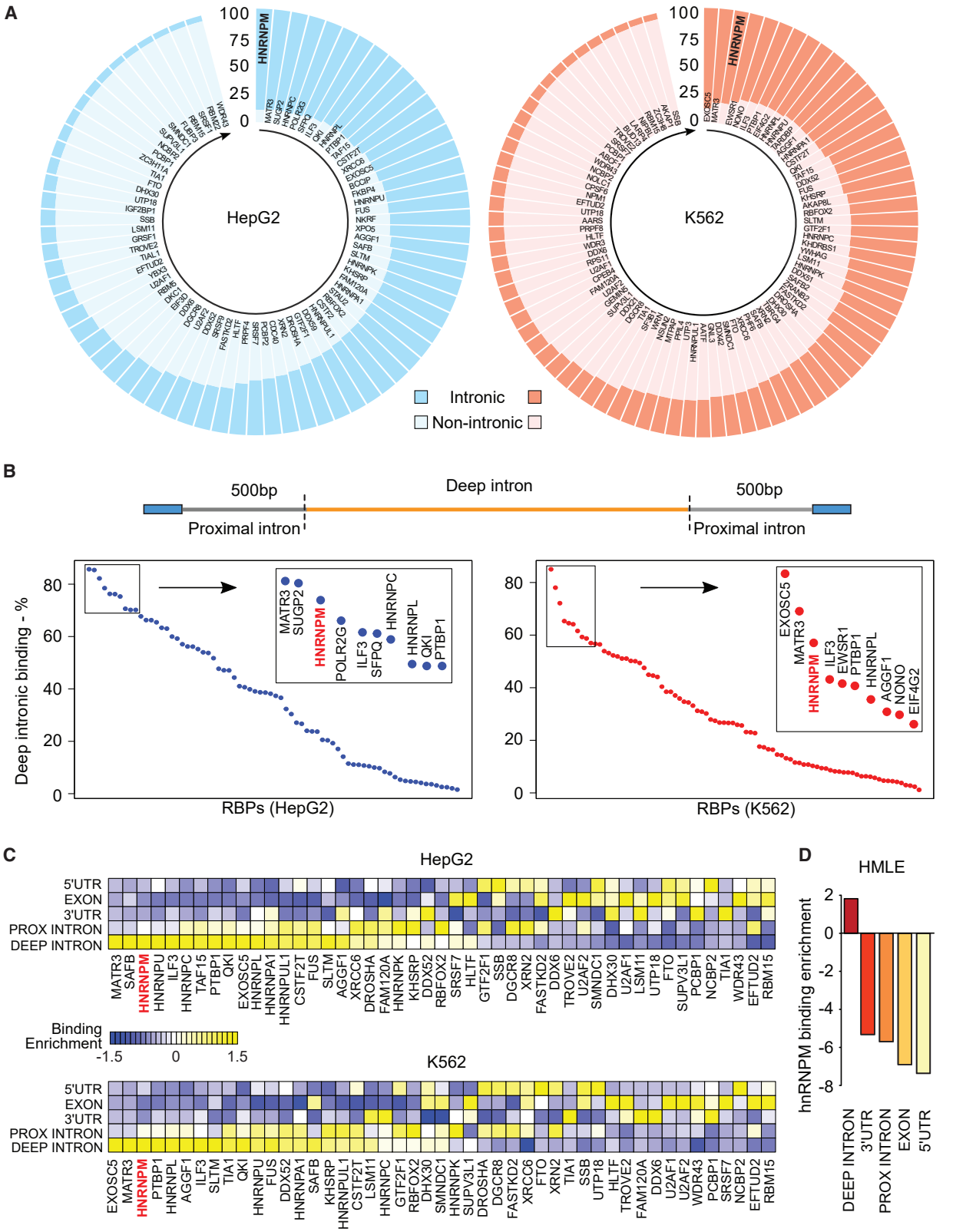
SUMMARY

RNA splicing is pivotal in post-transcriptional gene regulation, yet the exponential expansion of intron length in humans poses a challenge for accurate splicing. Here, we identify hnRNPM as an essential RNA-binding protein that suppresses cryptic splicing through binding to deep introns, maintaining human transcriptome integrity. Long interspersed nuclear elements (LINEs) in introns harbor numerous pseudo splice sites. hnRNPM preferentially binds at intronic LINEs to repress pseudo splice site usage for cryptic splicing. Remarkably, cryptic exons can generate long dsRNAs through base-pairing of inverted ALU transposable elements interspersed among LINEs and consequently trigger an interferon response, a well-known antiviral defense mechanism. Significantly, hnRNPM-deficient tumors show upregulated interferon-associated pathways and elevated immune cell infiltration. These findings unveil hnRNPM as a guardian of transcriptome integrity by repressing cryptic splicing and suggest that targeting hnRNPM in tumors may be used to trigger an inflammatory immune response, thereby boosting cancer surveillance.

INTRODUCTION

In eukaryotic evolution, the expansion of genome size increases genome complexity and biological diversity. A significant portion of genetic expansion is accounted for by intron net gain. In humans, introns constitute approximately 25% of the genome, representing an extensive expansion compared with lower eukaryotes and other mammals.[1–3] Transposition by intronic retrotransposable elements, such as long interspersed nuclear elements (LINEs), short interspersed elements (SINEs), and endogenous retroviruses (ERVs), is a major driver of intron expansion across various species,[4–6] expanding average intron length in human genes to several kilobases.[7,8] Long introns harbor numerous pseudo splice sites that are highly similar to annotated splice sites.[9–12] Misusage of pseudo splice sites results in cryptic splicing, which may be detrimental to cell viability, leading to diseases.[13–16] In spite of the presence of many pseudo splice sites, RNA splicing occurs in an accurate and precise manner. This suggests that potent protective mechanisms exist to repress cryptic splicing and ensure transcriptome integrity.

RNA-binding proteins (RBPs) regulate RNA processing, including splicing, localization, polyadenylation, and translation, through interaction with RNAs and other proteins.[17–20] RBP binding to pseudo splice sites may conceivably suppress cryptic splicing by preventing pseudo splice site usage. One such example is the function of hnRNPC to suppress aberrant ALU exonization through preventing spurious U2AF65 binding at cryptic splice sites.[21] Furthermore, expression of cryptic exons (CryEx) can be deleterious to normal gene expression, a mechanism partly responsible for the contribution of TDP-43 to



(legend on next page)

amyotrophic lateral sclerosis (ALS).[22] Conversely, CryEx may evolve into tissue-specific exons regulated by RBPs.[23] Despite these findings, the field of cryptic splicing is still in its infancy.

Here, we identify heterogeneous nuclear ribonucleoprotein M (hnRNPM) as a key RBP to repress cryptic splicing. We develop a bioinformatic pipeline that nominates CryEx from RNA sequencing (RNA-seq) datasets. We show that hnRNPM represses a large quantity of CryEx. These CryEx are enriched in LINEs, and some can form cytoplasmic double-stranded RNAs (dsRNAs) that mimic viral dsRNA known to elicit interferon (IFN) responses.[24–26] We further link LINE-associated dsRNAs derived from CryEx to IFN-response-induced tumor immunity.

RESULTS

hnRNPM preferentially binds to deep introns

The GENCODE v19 annotation of the human genome includes approximately 1.6 billion introns with a median length of 5 kb. These introns often contain numerous pseudo splice sites, resembling canonical splice sites but not involved in normal splicing. The mechanisms preventing splicing at pseudo splice sites remain incompletely understood. We hypothesized that RBPs repress pseudo splice sites by preferentially occupying intronic regions, thereby preventing spliceosome recognition and protecting cells against generating transcripts with CryEx. Through analysis of RBP footprints using ENCODE eCLIP data from HepG2 and K562 cancer cell lines,[27] we identified approximately 40% (67 proteins) of RBPs with over 50% of their binding sites within introns (Figure 1A; Table S1).

We next analyzed RBP binding in deep introns, located beyond 500 nucleotides away from the nearest exons (Figure 1B, top). Unlike proximal intron regions, which contain *cis* elements to regulate splicing of annotated exons,[28,29] deep introns are often over 5-fold longer and harbor numerous pseudo splice sites, potentially leading to CryEx formation. Of the 67 intron-binding RBPs, 43 preferentially bound to deep introns (Figure 1B), suggesting that most of the preferentially intron-bound RBPs also predominantly occupy deep introns. To account for length differences between proximal and deep introns, we assessed length-normalized binding distribution along pre-mRNA of the 47 RBPs shared by both cell lines in Figure 1A and examined their binding preference across pre-mRNA (Figure 1C). The results revealed that MATR3, hnRNPM, ILF3, PTBP1, and hnRNPL had the highest propensity to bind to deep introns in both cell lines (Figures 1B and 1C).

Among deep-intron preferentially bound RBPs, up to 82% of hnRNPM binding peaks reside in deep introns, placing hnRNPM among the top three RBPs in HepG2 and K562 cells (Figure 1B). Moreover, hnRNPM is essential for cell survival,[30–32] and its downregulation inhibits breast tumor metastasis.[33] Thus, we investigated hnRNPM binding by performing iCLIP analysis in human mammary epithelial (HMLE) cells. We identified 54,377 highly reliable hnRNPM binding sites (Figures S1A and S1B), with approximately 90% of them residing in deep introns (Figure S1C). Length-normalized binding distribution affirmed its binding preference to deep introns (Figure 1D) at UG-rich motifs (Figure S1D), highly similar to previously known hnRNPM binding motifs.[34–37] Collectively, these results demonstrated hnRNPM as a major deep-intron-binding RBP.

Loss of hnRNPM results in global cryptic splicing

If hnRNPM's binding in deep introns masks pseudo splice sites, loss of hnRNPM could result in global cryptic splicing and transcriptome instability. To test this, we developed a bioinformatic pipeline named CryEx. CryEx enables the detection of any expressed exon, both annotated and unannotated, through the *de novo* assembly of splice junction reads from RNA-seq (Figure 2A), whereas current splicing analysis tools primarily focus on identifying annotated exons. We performed RNA-seq in HMLE cells expressing non-specific shRNA (referred to as control) and hnRNPM shRNA (referred to as hnRNPM knockdown [KD]) (Figure S2A) and applied CryEx. We identified 493,975 expressed exons, including 321,415 unannotated ones. The majority of unannotated exons are located in deep introns. Percent spliced in (PSI) from CryEx showed that the inclusion rates of 716 alternatively skipped exons (ASEs) are highly consistent with those of rMATs[38] results (Figure S2B), with a majority (480 out of 716 events) being inclusion events, indicating the role of hnRNPM in promoting exon skipping (Figure S2C). Meanwhile, CryEx unveiled more than 1,000 cryptic splicing events after hnRNPM KD, and these events are essentially absent in control samples based on CryEx parameters in Figure 2A. Most of the CryEx-containing genes exhibit no differential mRNA expression (Figures 2B and S2D), indicating that the observed cryptic events are generated by abnormal splicing rather than changes in transcription. Experimental validation using semi-RT-qPCR for 68 randomly selected ASEs and 10 cryptic splicing events demonstrated the accuracy of the CryEx pipeline (Figures S2E and S2F).

Most (78%) of these CryEx reside in introns (Figure 2C, ACYP1). Among them, 68% disrupt the reading frame, potentially causing frameshifts and generating premature stop codons leading to nonsense-mediated decay (NMD) of the mRNA transcript (Figure S2G). Some CryEx can contain a stop codon and a downstream polyA signal (Figure 2C, TRAPPC10), which generates a truncated mRNA with a novel 3′ untranslated region (UTR) (Figure S2H). Cryptic splicing can occur upstream of the first exon, extending the 5′ UTR and potentially causing the

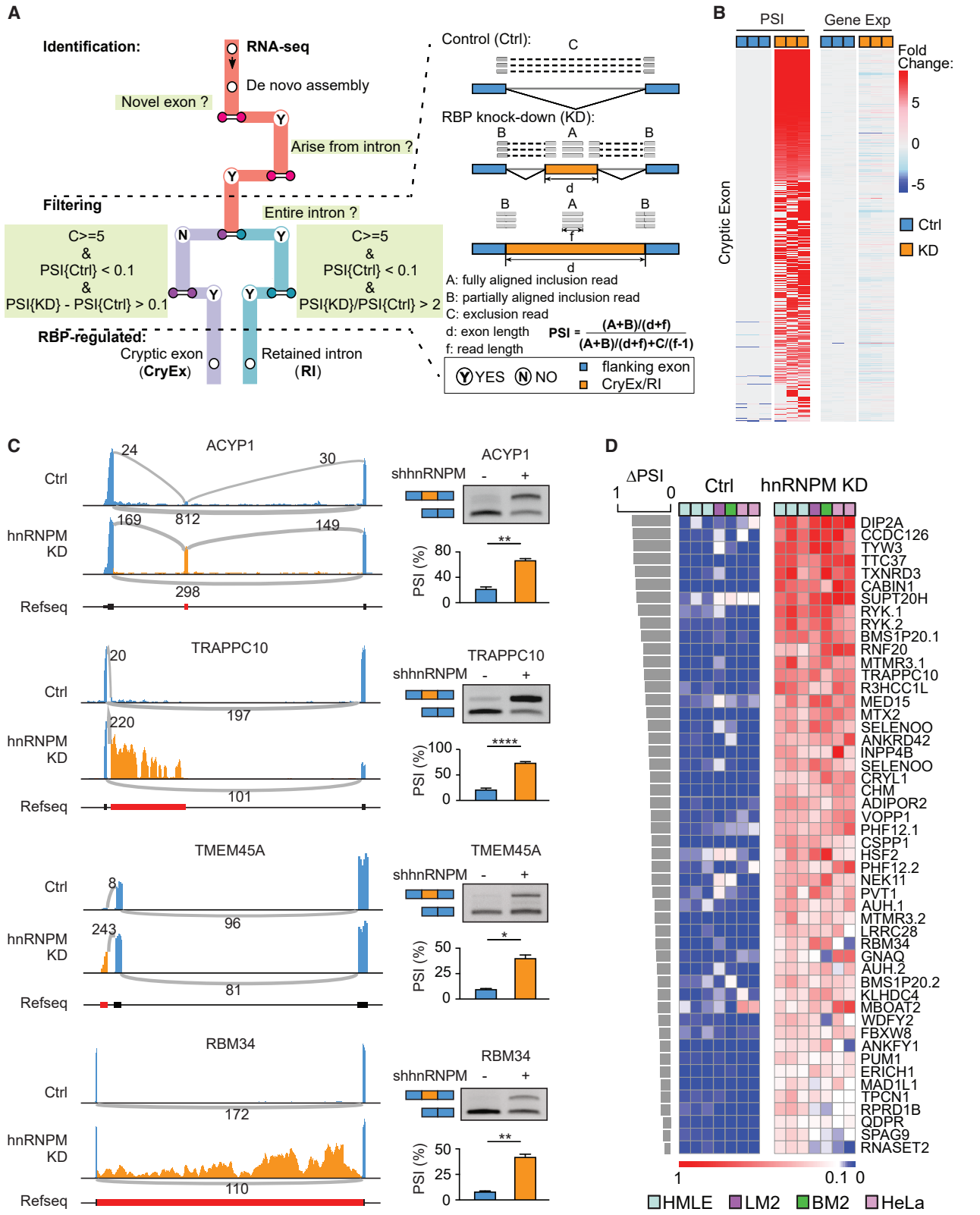
Figure 1. hnRNPM preferentially binds to deep introns

(A) RBPs binding distribution in intronic and non-intronic regions. Circular bars indicate percent of RBP binding in intronic and non-intronic regions in HepG2 (n = 72) and K562 (n = 84). Dark shades indicate intronic binding frequency; light shades indicate non-intronic binding frequency.

(B) RBP binding distribution in deep introns. Upper diagram illustrates deep and proximal introns. Dot plots indicate RBP binding frequency in deep introns. Boxes show top 10 ranked RBPs.

(C) Heatmaps indicating binding preferences of shared RBPs (n = 47) in HepG2 and K562 cell lines to 5′ UTR, exon, 3′ UTR, proximal (PROX) intron, and deep intron across pre-mRNA.

(D) Bars showing hnRNPM binding enrichment across pre-mRNA in HMLE cells. Extended data for (A)–(C) and Table S1.



(legend on next page)

usage of a new transcriptional start site (Figure 2C, TMEM45A). Cryptic splicing can also result in intron retention (IR) (Figure 2C, RBM34), possibly creating a premature stop codon and a truncated protein. Gene Ontology (GO) analysis shows that loss of hnRNPM resulted in widespread cryptic splicing affecting a broad range of biological pathways (Figure S2I). Hereafter, we will refer to these hnRNPM-repressed CryEx as CryEx.

To investigate the role of hnRNPM in repressing cryptic splicing, we constructed a splicing reporter minigene containing the ADIPOR2 CryEx and its adjacent introns flanked by two constitutive exons (Figure S3A). Splicing of the exon was gauged using semi-RT-qPCR to measure the percentage of CryEx inclusion (Figure S3B, left). Co-transfection of the minigene construct and different amount of hnRNPM cDNA in 293FT cells demonstrated that hnRNPM represses ADIPOR2 CryEx inclusion in a dose-dependent manner, resulting in a decrease of CryEx inclusion from 97% in control to 66% (Figure S3B, right). Disruption of hnRNPM binding by mutating a 154-bp GU-rich region covering putative hnRNPM binding sites downstream of the ADIPOR2 CryEx diminished the repression effect (Figure S3B). Moreover, inserting the intron sequence that is retained in MED15 into the splicing reporter minigene (Figure S3C, left) resulted in suppression of a cryptic retained intron in an hnRNPM dose-dependent manner (Figure S3C, right). These results suggest that hnRNPM directly represses cryptic splicing.

To generalize our findings, we depleted hnRNPM in the lung metastatic breast cancer line LM2, bone metastatic breast cancer line BM2, and the cervical cancer line HeLa. We found that the top 50 CryEx identified in hnRNPM KD HMLE cells were consistently upregulated (Figure 2D). This non-random cryptic splicing pattern was further confirmed in publicly available RNA-seq datasets (Figure S3D), suggesting that pseudo splice sites are repressed under normal conditions and become recognized by the splicing machinery when hnRNPM levels are low or depleted.

hnRNPM binding overlaps with cryptic splice sites and prevents spliceosome binding

Several key characteristics of CryEx became apparent from our analysis. First, they are located in significantly longer introns (Figure 3A) and genes (Figure S4A) compared with hnRNPM-regulated alternative exons or background exons. Second, 35% of CryEx show hnRNPM binding by iCLIP, markedly higher than hnRNPM-regulated alternative exons, where only 5% had detectable hnRNPM binding (Figure 3B). Third, 95.6% of hnRNPM-bound CryEx are embedded in deep introns (Figure 3C). The average distance between hnRNPM-bound CryEx and the nearest annotated exons is roughly 8 kb (Figure S4B).

Our metagene analysis of iCLIP data revealed that hnRNPM binding centers around its CryEx, similar to its binding to alternative exons, but with a significantly higher binding percentage near CryEx (Figure 3D). The motif derived from hnRNPM iCLIP binding near CryEx was GU rich (Figure S4C).[34–37,39] Metagene analysis using hnRNPM motifs further confirmed hnRNPM binding distributions (Figure S4D). Additionally, RNA pull-down experiments using RNA oligos derived from iCLIP hnRNPM peaks near the CryEx of ADIPOR2 and TRAPPC10 validated hnRNPM binding (Figure S4E).

CryEx exhibit nearly identical 5′ and 3′ splice site motifs as hnRNPM-regulated alternative exons (Figures 3E and S4F), closely resembling conserved splice site motifs.[40] These findings support the notion that pseudo splice sites bear similarity to conserved splice sites, thus providing an explanation for the non-random occurrence of CryEx (Figures 2D and S3D).

The above results led us to propose that hnRNPM represses cryptic splicing by impeding the binding of spliceosome components to pseudo splice sites (Figure 3F). To test this, we examined the occupancy of U2AF2 and PRPF8, two spliceosome proteins recognizing 3′ and 5′ splice sites, respectively. In the K562 ENCODE eCLIP dataset, no U2AF2 or PRPF8 binding was detected at the 3′ or 5′ splice sites of CryEx; however, their binding to splice sites was evident at the adjacent flanking exon (Figures S4G and S4H), indicating that spliceosome recognition of CryEx pseudo splice sites is hindered in the presence of hnRNPM. Subsequently, we conducted eCLIP of U2AF2 and PRPF8 in control and hnRNPM KD HMLE cells. In contrast to control cells where no signals were detected at the cryptic splice sites, hnRNPM KD resulted in substantial binding of U2AF2 and PRPF8 at the 3′ and 5′ splice sites of CryEx, respectively (Figures 3G and 3H). These findings suggest that hnRNPM represses cryptic splicing by preventing the spliceosome from binding to CryEx splice sites.

hnRNPM-bound cryptic splice sites are enriched in deep intronic LINEs

As introns contain extensive repeats, we used RepeatMasker[41] to assess hnRNPM binding preference at CryEx. Among hnRNPM peaks, 59% are situated in intronic repeat regions, with 53% of these regions harboring CryEx. In contrast, 41% of hnRNPM peaks are located in intronic non-repeat regions, and among these, 39% contain CryEx (Figure S4I), indicating

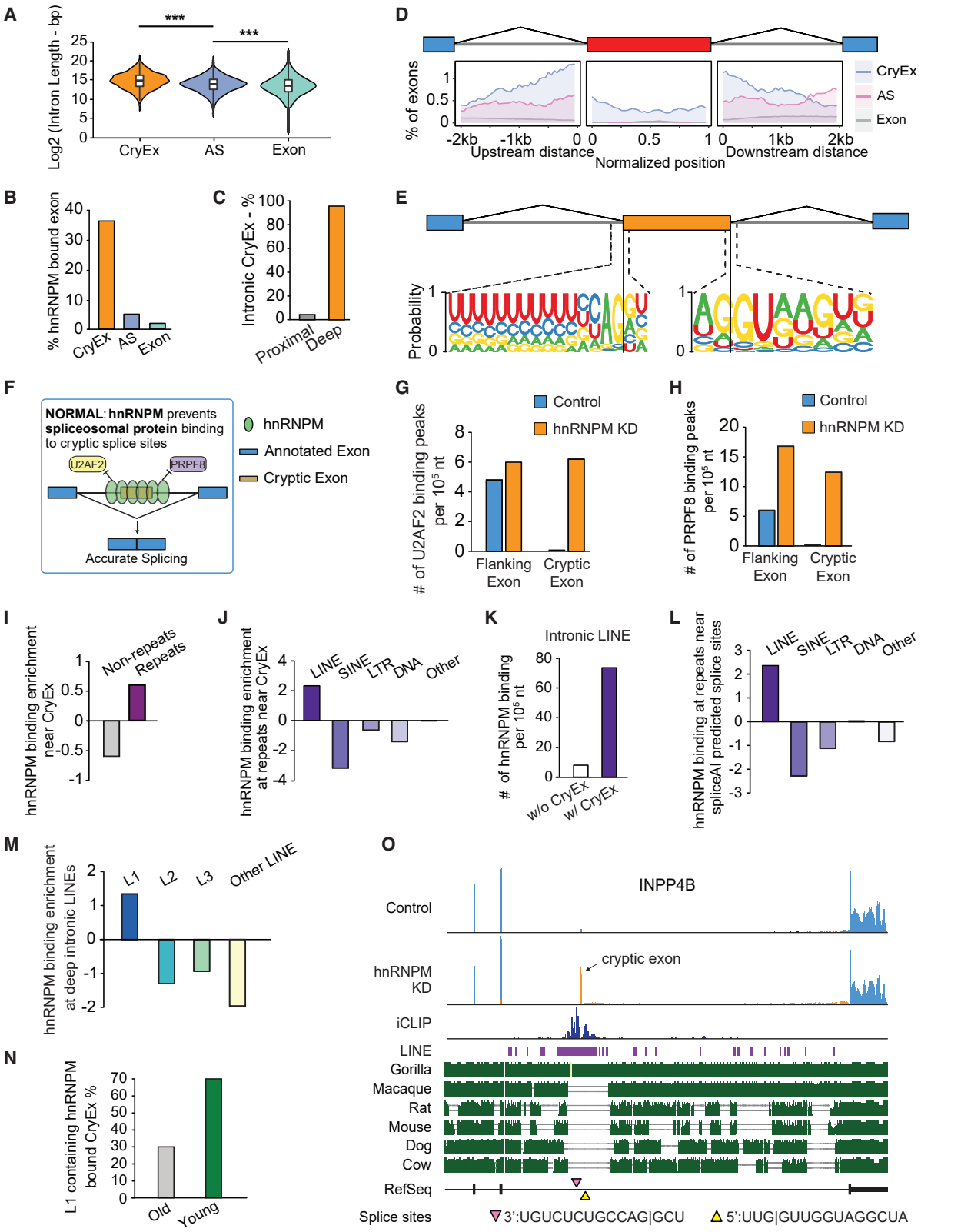
Figure 2. Loss of hnRNPM results in global cryptic splicing

(A) CryEx workflow used to identify and quantify cryptic splicing events from RNA-seq data. Shown in between dashed lines on the left are the criteria to report a cryptic exon. Reads and lengths used for percent spliced in (PSI) metric calculation are denoted on the right.

(B) Heatmaps showing that cryptic exon production is independent of gene expression changes between hnRNPM knockdown (KD, orange) and non-specific shRNA control (Ctrl, blue). Left, PSI fold change. Right, fold change of genes containing cryptic exons.

(C) Representative hnRNPM-repressed cryptic exons. Left, sashimi plots of cryptic splicing events. Numbers of splice junction reads are shown next to arches. Right, semi-qPCR validations. Showing agarose gel images (top) and PSI quantifications (bottom). Data are represented as mean ± SEM of at least three biological replicates, $*p \leq 0.05$, $**p \leq 0.01$, $****p \leq 0.0001$ (unpaired Student's t test).

(D) Heatmaps showing PSI values of top 50 representative hnRNPM-repressed cryptic exons identified from HMLE cells in four cell lines (HMLE, LM2, BM2, and HeLa) with hnRNPM KD. Cryptic events are ranked from high to low based on ΔPSI values (ΔPSI = PSI{KD} − PSI{Ctrl}) calculated from HMLE RNA-seq data. Extended data for (B) and (D) and Table S2.



(legend on next page)

that hnRNPM-bound CryEx are 2-fold more prevalent in repeat regions than non-repeat regions. Upon normalizing hnRNPM binding peaks by the genomic length, hnRNPM binding remains distinctively enriched at CryEx in repeat regions (Figure 3I). Among different intronic repeat types, hnRNPM is most specifically enriched at LINEs compared with other repeat types (Figure 3J). Moreover, hnRNPM binding density is drastically increased at regions of intronic LINEs that contain CryEx in HMLE, HepG2, and K562 cells (Figures 3K and S4J), indicating that the preferential binding of hnRNPM to CryEx in intronic LINEs is universal.

We applied SpliceAI[42] to predict potential splice sites in deep intronic repeats and found that splice sites at LINEs are highly enriched for hnRNPM binding sites (Figure 3L). To examine the association between LINEs, CryEx, and hnRNPM binding, we performed chi-squared test and found statistically significant associations between LINEs and hnRNPM binding (Figure S4K), as well as CryEx and hnRNPM binding (Figure S4L). However, the association between LINE and CryEx is not statistically significant (Figure S4M). These results suggest that hnRNPM preferentially binds CryEx that contain LINEs. Of note, hnRNPM binding is particularly enriched in LINE1 (L1), the most abundant LINE in humans (Figure 3M), consistent with previous reports.[23,43]

As the most abundant non-long terminal repeat retrotransposon found in the genome, L1 insertion is a key contributor to intron expansion during evolution.[44] We therefore assessed whether the evolutionary age of human L1 impacts CryEx repression by hnRNPM. 70% of hnRNPM-bound CryEx belong to evolutionarily young L1 elements that are conserved only in primates (Figure 3N). Using INPP4B as an example, hnRNPM binds near its CryEx and hnRNPM binding sites fall on an L1 element (L1PA5) that is only conserved between human and the closely related gorilla genomes (Figure 3O). Such binding to evolutionarily young L1s is supported by a previous study[23] and suggests a role for hnRNPM in repressing CryEx inside young L1s.

CryEx form cytoplasmic, long stem-like dsRNAs

Because of the propensity of repetitive sequences to form secondary structures,[45–47] we used RNAfold[48] to evaluate secondary structures in hnRNPM-bound CryEx. Remarkably, 24% of the repeat-containing CryEx are predicted to form extended stretches (>30 nt) of dsRNAs, compared with only 8.45% of the non-repeat-containing CryEx (Figure 4A). To assess the stability of RNA structures, we quantified absolute values of minimum free energy (|MFE|) for each exon, with larger |MFE| values indicating more stable RNA structures. Among hnRNPM-bound exons, the repeat-containing CryEx showed significantly higher maximum |MFE| values (Figure 4B), suggesting that repeat-containing CryEx produce more stable RNA structures, consistent with the notion that they form dsRNAs.

dsRNAs are the primary substrates of adenosine deaminase acting on RNA (ADAR) enzymes, which bind dsRNA and catalyze the deamination of adenosine to inosine (A-I). This process prevents the sensing of fully base-paired dsRNA in mRNA transcripts and blocks the subsequent induction of an IFN response.[49–52] ADAR editing sites thus act as indicators of dsRNA formation. To examine ADAR editing sites in control and hnRNPM KD HMLE cells, we calculated the ratio of reads with edited bases over the total reads overlapping a specific position.[53,54] In hnRNPM KD cells, we identified 1,744 editing sites in the repeat regions of CryEx, a 12-fold increase from the 150 editing sites in control cells. hnRNPM KD cells also showed a significantly higher editing ratio (Figure 4C). Conversely, non-repeat exons exhibited very few editing sites (n = 6–10), with no notable differences between control and hnRNPM KD (Figure S5A). These results further reinforce that repeat-containing CryEx form dsRNAs, which serve as substrates for ADAR editing.

To visualize dsRNA in hnRNPM KD cells directly, we performed immunofluorescence (IF) staining using a J2 antibody that binds dsRNA structures.[55] Compared with control cells, the cytoplasmic J2 signal was drastically increased in hnRNPM KD cells (Figure 4D). IF staining using an alternative dsRNA

Figure 3. hnRNPM binding overlaps with cryptic splice sites at deep intronic LINEs

(A) Violin plot showing length distributions of introns containing hnRNPM-repressed cryptic exons (CryEx, orange), hnRNPM-regulated alternative exons (AS, blue), and annotated background exons (Exon, green) separately. ***$p \leq 0.001$ (Wilcoxon rank-sum test).

(B) Bar plot showing percentage of CryEx, AS, and Exon bound by hnRNPM. Exons with 500 bp flanking regions overlapping hnRNPM binding are considered positive.

(C) Bar plot depicting the percentage of CryEx at proximal and deep introns.

(D) Metaprofile showing hnRNPM binding in a ±2-kb window flanking the splice sites of CryEx, AS, or Exon. The schematic above metaplot depicts introns (thin lines) and exons with cryptic/alternative exon marked in red and annotated exons marked in blue.

(E) Weblogo consensus motifs at splice sites of CryEx. Flanking annotated exons are marked in blue.

(F) Model of hnRNPM-mediated cryptic splicing repression.

(G and H) Bar plots showing U2AF2 and PRPF8 binding density within 200 nt window at 3′ (U2AF2, G) and 5′ (PRPF8, H) splice sites of cryptic exons and their flanking exons in control and hnRNPM KD cells.

(I and J) Bar plots showing hnRNPM binding enrichment in repeats (I) and LINEs (J) in a ±500-bp window flanking CryEx. See details in STAR Methods.

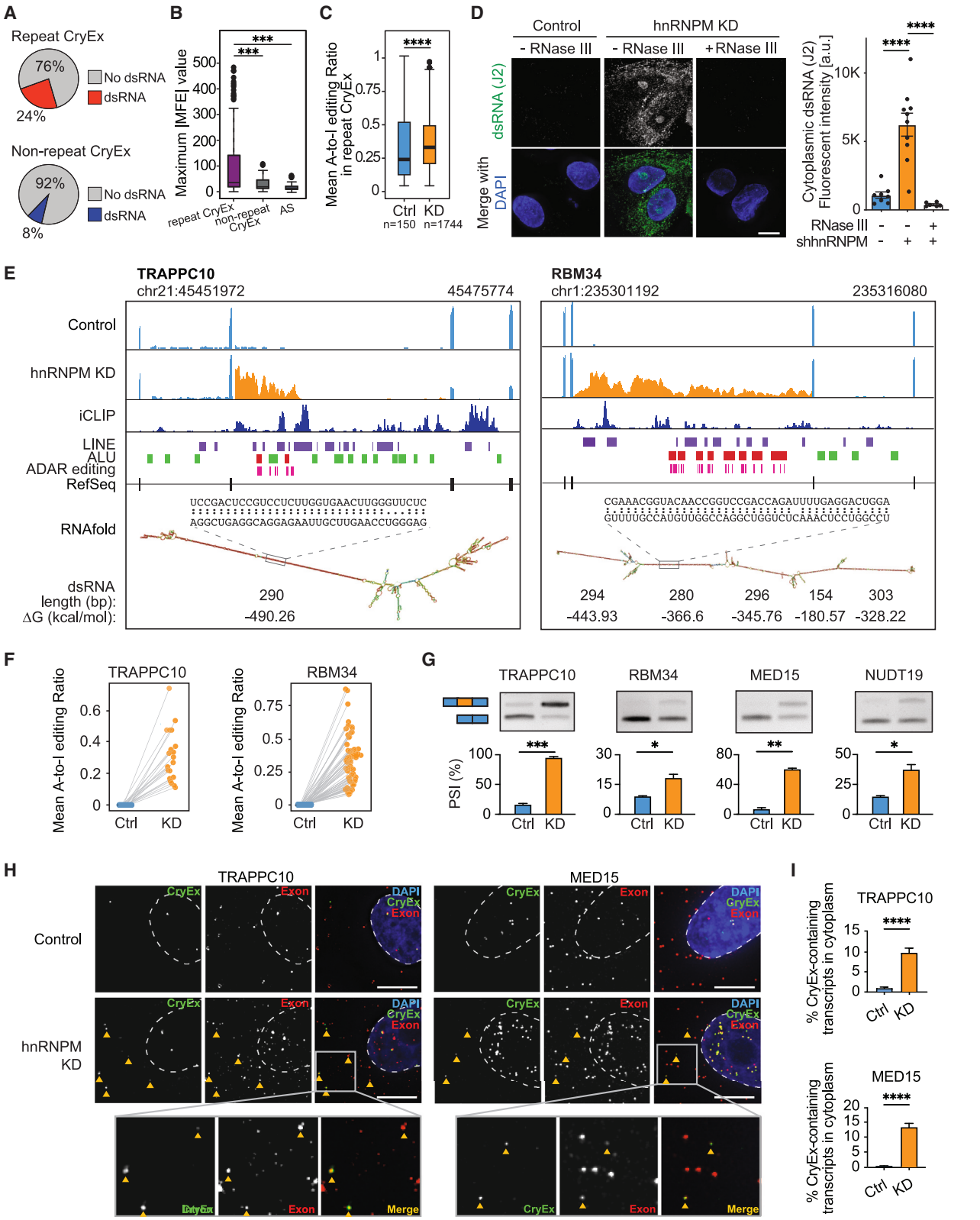
(K) Bar plot showing hnRNPM binding density at intronic LINEs without and with CryEx.

(L) Bar plot showing hnRNPM binding enrichment at repeats near spliceAI-predicted splice sites. See details in STAR Methods.

(M) Bar plot showing hnRNPM binding enrichment at different LINE subtypes.

(N) Bar plot depicting the percentage of hnRNPM-bound cryptic exons that contain evolutionarily old or young L1s.

(O) Example of an hnRNPM-repressed cryptic exon in the INPP4B gene. Top two Integrative Genomics Viewer (IGV) tracks represent RNA-seq reads aligned in control shRNA and hnRNPM KD cells. Annotated exons are represented in blue; the cryptic exon is marked in orange. hnRNPM iCLIP binding peaks (dark blue) are shown in the third track. LINEs are depicted as purple boxes below the iCLIP track. Below the LINE track, a snapshot of the Primates Multiz Alignment & Conservation tracks (green) from UCSC genome browser are shown, which include 6 species, with the first 2 being primates. RefSeq track shows annotated exons (vertical bars) and sequences near spliceAI-predicted 3′ (pink) and 5′ (yellow) splice sites indicated by triangles.



(legend on next page)

antibody (9D5)[56] revealed a similar cytoplasmic signal in hnRNPM KD cells, which highly overlapped with J2 staining (Figure S5B). RNase III treatment abolished the J2 and 9D5 signals detected in hnRNPM KD cells, suggesting that the observed cytoplasmic signal stems from dsRNAs (Figures 4D and S5B).

RNAfold analysis revealed several CryEx that form dsRNAs with continuous stretches of nearly 300 bp (Figures 4E and S5C). In addition to multiple LINEs (purple in Figures 4E and S5C), the long dsRNA regions of TRAPPC10 also contain ALU elements (red and green), a primate-specific class of SINEs.[57] The virtually perfect base pairing inside predicted dsRNAs stems from inverted ALU elements that are scattered in between LINEs (red). These predicted dsRNA regions show low Gibbs free energy (ΔG in Figures 4E and S5C) and more ADAR editing sites (magenta), indicating that they are double-stranded moieties.

Analysis of A-I editing of the predicted dsRNA regions in TRAPPC10, RBM34, MED15, and NUDT19 revealed significantly increased ADAR editing sites and editing levels in hnRNPM KD cells (Figures 4F and S5D). Experimentally, we reverse-transcribed RNA from the hnRNPM KD cells and PCR-amplified the TRAPPC10 CryEx. DNA sequencing of this PCR product showed virtually perfect base-pairing of the inverted ALU repeats (Figure S6A). We then performed J2 binding on dot blots carrying *in vitro* transcription products from the predicted TRAPPC10 dsRNA region (wild type [WT]) and a reverted mutant containing an inverted ALU arm to disrupt dsRNA formation (Figure S6B). The J2 antibody specifically recognized the WT CryEx, but not the reverted mutant, supporting our finding that the TRAPPC10 CryEx produces dsRNAs (Figure S6C, left). This notion is further validated by the MED15 dot blot (Figure S6C, right).

To determine whether the cytoplasmic dsRNA signals in hnRNPM KD cells stem from CryEx, we isolated the cytoplasmic fraction of HMLE cells (Figure S6D) and measured CryEx inclusion by semi-RT-qPCR. As expected, cytoplasmic RNAs in control cells contain fully spliced transcripts with little or no detection of CryEx. In contrast, cryptic transcripts are readily detected in the cytoplasm of hnRNPM KD HMLE (Figure 4G) and LM2 cells (Figure S6E).

To investigate the subcellular localization of individual dsRNA-containing cryptic transcripts, we performed RNA fluorescence *in situ* hybridization (FISH) for TRAPPC10 and MED15 transcripts. RNA probes are designed to specifically recognize the non-repetitive region of the CryEx (green) or their neighboring annotated exons (red, denoted as Exon, Figure S6F). RNA-FISH experiments revealed a significant increase in cytoplasmic CryEx-containing mRNAs in hnRNPM KD cells (Figures 4H and 4I). hnRNPM depletion significantly boosts CryEx-containing transcript production, promoting their cytoplasmic export (Figure 4H). Collectively, these results provide multiple lines of evidence indicating that loss of hnRNPM produces CryEx-containing transcripts capable of forming dsRNAs, which are also present in the cytoplasm.

Cryptic-splicing-produced dsRNA elevates IFN-I response

dsRNAs longer than 30 bp act as a potent trigger of antiviral immunity.[58–60] Host dsRNA sensors recognize cytoplasmic dsRNAs and trigger the activation of a type I IFN (IFN-I) response and expression of IFN-stimulated genes (ISGs) as a cellular defense mechanism against viral infection.[25,61] Because hnRNPM KD leads to the accumulation of cytoplasmic dsRNAs, we examined whether hnRNPM KD activates the IFN-I response and increases ISG production (Figure 5A).

Comparing transcriptomes through GO and gene set enrichment analysis (GSEA) revealed that hnRNPM KD cells exhibit significant positive enrichment in antiviral signaling pathways, including IFN-I responses (Figures 5B and 5C). Additionally, terms associated with epithelial cell state showed positive enrichment, while cell cycle pathways were downregulated in hnRNPM KD cells (Figure S7A), consistent with previous reports.[33,35,39,62,63]

Figure 4. hnRNPM-repressed cryptic exons form cytoplasmic long stem-like dsRNAs

(A) Pie charts depicting the percentage of hnRNPM-bound cryptic exons (CryEx) predicted to form dsRNA structures by RNAfold.

(B) Boxplots showing highest absolute minimum free energy (|MFE|) values for hnRNPM-bound CryEx and alternatively spliced exons (AS). MFE values were calculated using RNAfold.

(C) Boxplots showing mean A-to-I editing levels of repeat-containing and hnRNPM-bound CryEx in non-specific shRNA control (Ctrl) and hnRNPM knockdown (KD) cells. Numbers of ADAR editing sites are shown at the bottom. Statistics in (B) and (C): $^{****}p \leq 0.0001$ (Wilcoxon rank-sum test).

(D) Detection of cytoplasmic dsRNA upon hnRNPM KD. Control shRNA (shLuc) and shhnRNPM-expressing HMLE cells were stained for dsRNA (J2, green); DAPI (blue) served as a marker for nuclei. Representative images of control (left), hnRNPM KD (middle), and hnRNPM KD treated with RNase III (right) cells are shown. Scale bar represents 10 μm. Quantification of cytoplasmic dsRNA signal intensity are shown on the right. Data are represented as mean ± SEM, each data point represents the average cytoplasmic signal of 10–20 cells from three independent biological repeats. $^{****}p \leq 0.0001$ (ordinary one-way ANOVA).

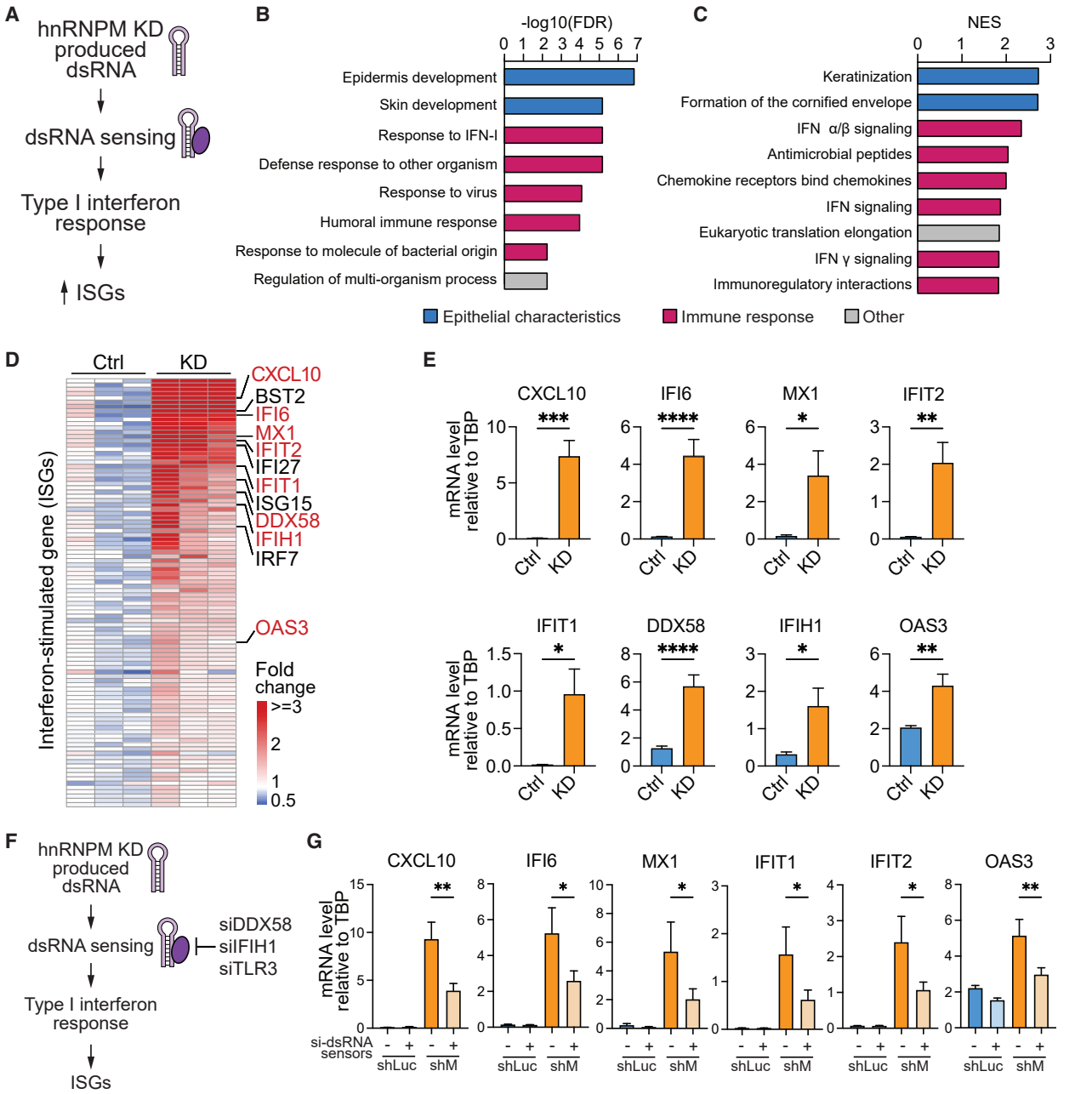
(E) Representative examples of dsRNA-forming cryptic exons. Left, TRAPPC10. Right, RBM34. Top two tracks represent RNA-seq reads aligned in control shRNA and hnRNPM KD cells. hnRNPM iCLIP binding peaks are shown in the third track. Below the iCLIP track shows colored boxes indicating LINEs (purple), ALU elements (green), ALU elements predicted to form dsRNA (red), and ADAR editing sites (pink). Seventh track indicates RefSeq annotation. Eighth track represents RNAfold-predicted secondary structure of the cryptic exon. Box indicates the position of a dsRNA fragment whose sequence is depicted above. Approximate lengths of highly base-paired dsRNA regions are indicated together with predicted folding free energies (ΔG).

(F) Dot plots depicting ADAR editing sites found in predicted dsRNA-forming regions of TRAPPC10 (left) and RBM34 (right) in control shRNA and hnRNPM KD HMLE cells.

(G) Semi-RT-qPCR validations for cytoplasmic cryptic splicing products showing agarose gel images (top) and PSI quantifications (bottom). Data are represented as mean ± SEM of at least three biological replicates. $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$ (unpaired Student's t test).

(H) RNA-FISH images of dsRNA containing cryptic exons (CryEx, green) and constitutive exons (Exon, red) for TRAPPC10 (left) and MED15 (right). Nuclei are stained with DAPI; nuclei boundaries are shown as dotted lines. Orange triangles indicate overlapped cytoplasmic signals of CryEx and Exon foci. Boxes depict the position of the 3× enlarged inset. Scale bar represents 10 μm.

(I) Quantification of images in (H). Data are represented as mean ± SEM, $n = 3$, 8–25 cells per experiment, $^{****}p \leq 0.0001$ (unpaired Student's t test).



(legend on next page)

Examination of dsRNA-specific antiviral genes[64] and the hallmark IFN-α, -β, and -γ[65] showed significant upregulation of a broad spectrum of ISGs in hnRNPM KD cells (Figure 5D), with the top 40 upregulated ISGs showing a more than 4-fold increase in gene expression (Figure S7B). RT-qPCR validated the results (Figures 5E and S7C). Moreover, elevated RNA and protein levels of IFNB1 and/or IFNL2/3 were observed in hnRNPM KD cells (Figures S7D and S7E). Collectively, these results suggest that hnRNPM depletion induces IFN-I response and ISG upregulation.

To determine whether hnRNPM-KD-induced ISG upregulation is mediated through dsRNAs produced by cryptic splicing, we depleted three dsRNA sensors—DDX58, IFIH1, and TLR3 (Figure S7F)—and assessed changes in ISG induction (Figure 5F). Control cells exhibit very low levels of ISGs and show negligible changes of ISG levels upon dsRNA sensor KD (Figure 5G). In contrast, hnRNPM KD cells show a drastic upregulation of ISGs, which is significantly suppressed upon concomitant KD of dsRNA sensors, suggesting that dsRNA sensing is required for ISG upregulation in hnRNPM KD cells (Figure 5G).

As IFN-I response signals through the JAK-STAT pathway to stimulate ISG expression,[66,67] we treated HMLE control and hnRNPM KD cells with the FDA-approved JAK inhibitor ruxolitinib and measured its effect on ISG expression. Although control cells showed no notable changes in ISG levels, ruxolitinib treatment significantly inhibited ISG upregulation in hnRNPM KD cells (Figure 5H), indicating that hnRNPM-KD-induced ISG upregulation requires JAK-STAT signaling. Overall, these findings show that CryEx can form dsRNAs, which are recognized by dsRNA sensors, subsequently triggering IFN-I-mediated ISG upregulation.

Tumors with low hnRNPM expression show upregulated IFN-I responses and better patient survival

Elevation of the IFN-I pathway is widely observed to activate anti-tumor immunity.[68–70] Given that loss of hnRNPM stimulates IFN-I response through cryptic-splicing-induced dsRNA formation, we examined whether hnRNPM expression is associated with IFN-I signatures in tumors. Using The Cancer Genome Atlas (TCGA) database, we separated tumors into hnRNPM^{high} and hnRNPM^{low} groups based on normalized hnRNPM levels and conducted GSEAs across cancers, focusing on 50 "hallmark" gene sets (Figures 6A and S8A). Strikingly, in 16 out of 33 cancer types, hnRNPM^{low} tumors exhibited significant enrichment of immune pathways, especially the IFN-I-associated responses, including IFN-α, IFN-γ, and interleukin (IL)-6/Jak/Stat3 pathways, and impaired cell proliferation (Figure 6A), mirroring our experimental findings in hnRNPM KD cells (Figures 5B, 5C, and S7A), which were further validated using the CPTAC3 dataset (Figure S8B). Additionally, considering the high heterogeneity in breast cancer, we stratified breast tumors into its four subtypes (luminal A, luminal B, HER2, and basal) and found that the most aggressive basal subtype displayed the most pronounced immune response upregulation in hnRNPM^{low} tumors (Figure 6A).

We next compared overall ISG expression in hnRNPM^{high} and hnRNPM^{low} tumors by performing single-sample GSEA (ssGSEA)[71] across various cancer types. Six cancer types, including basal subtype of breast cancer, exhibited significantly increased ISG scores in hnRNPM^{low} tumors compared with hnRNPM^{high} tumors (Figure 6B). These results were recapitulated using CPTAC3 datasets (Figure S8C), further confirming our findings that loss of hnRNPM results in ISG upregulation.

Using basal subtype breast cancer as an example, we observed elevated dsRNA-forming cryptic splicing (MED15, RBM34, LRP11, and TRAPPC10) in hnRNPM^{low} tumors (Figure 6C), congruent with our findings that hnRNPM depletion leads to cryptic-splicing-mediated dsRNA production and subsequent IFN responses (Figure 5). Upregulated IFN responses can facilitate immune cell infiltration, particularly natural killer (NK) cells and CD8$^+$ T cells, to elicit anti-tumor immunity.[72–74] Using individual immune cell gene signatures curated from xCell,[75] we assigned immune cell infiltration scores to each TCGA basal-type breast tumor through ssGSEA (see STAR Methods). We found that increased MED15 cryptic IR level (left to right, Figure 6D) is positively correlated with immune infiltration in TCGA basal subtype breast tumors (Figure 6D).

Motivated by our findings, we built a multivariate survival regression model to generate tumor risk scores by integrating hnRNPM expression and the inclusion rates of dsRNA forming CryEx in Figure 6C. Inspecting immune cell infiltration in TCGA basal subtype breast tumors revealed that many innate and adaptive immune cells known to exhibit anti-tumor functions, including dendritic cells and CD8$^+$ T cells, show significantly higher infiltration in tumors with lower risk scores (Figures 6E and S8D).

Lastly, we conducted Kaplan-Meier survival estimation using the risk scores. Our results revealed that patients with low-risk scores show better survival in Basal breast cancer (Figure 6F) and various other cancer types, including lymphoid neoplasm

Figure 5. Cryptic-splicing-produced dsRNA elevates type I interferon response

(A) Model depicting how hnRNPM KD can trigger a type I interferon response via dsRNA sensing.

(B and C) GO (B) and GSEA (C) showing top biological processes and pathways positively enriched in hnRNPM KD cells. Immune-related GO terms and pathways are indicated in red. Gene sets related to epithelial characteristics are shown in blue and other terms are shown in gray.

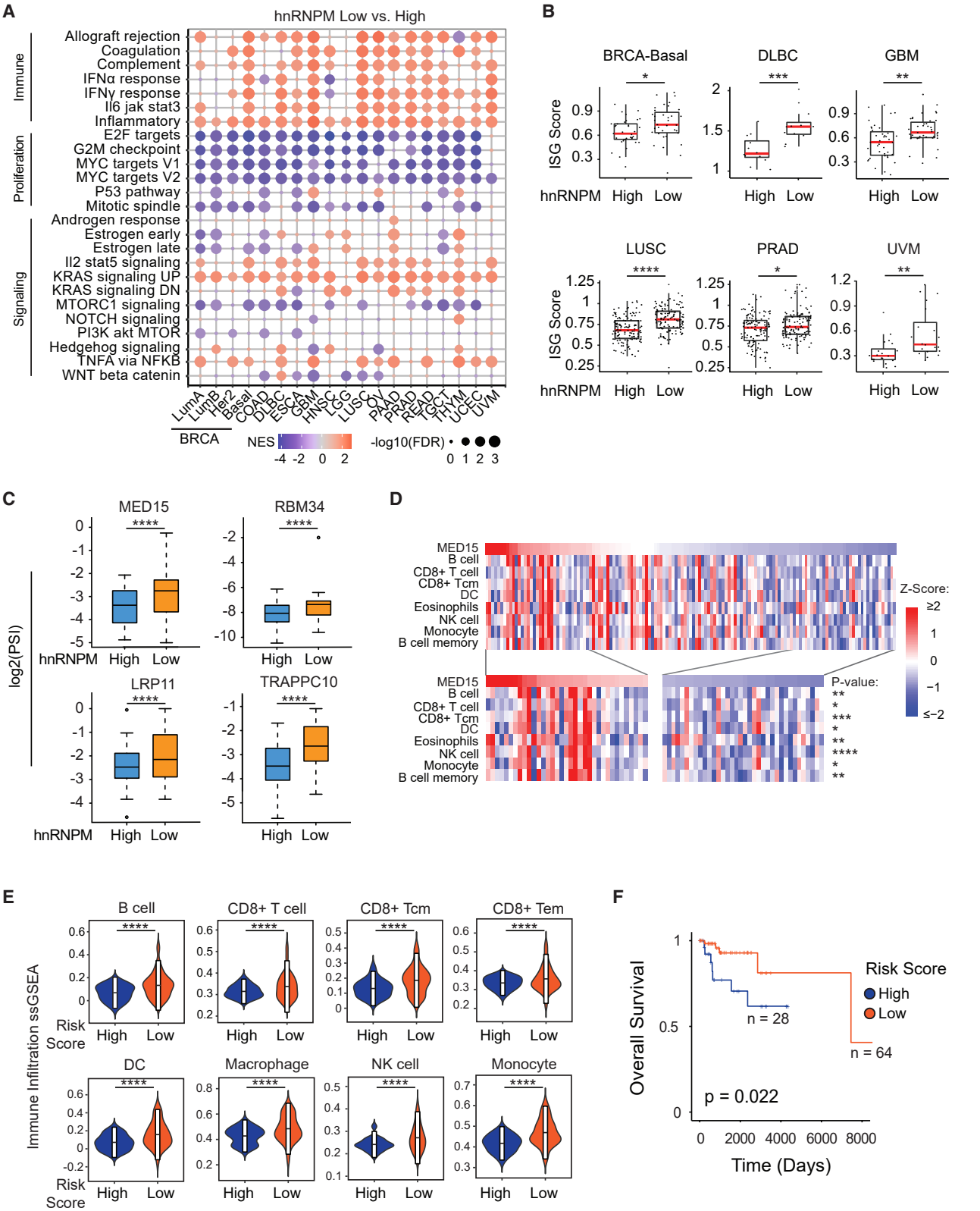
(D) Heatmaps showing ISG expression in non-specific shRNA-expressing control (Ctrl) and hnRNPM KD HMLE cells. Relative expression was calculated as mean FPKM fold change versus control. ISGs involved with virus-stimulated interferon response are highlighted in red.

(E) RT-qPCR validations for ISGs highlighted in (D). Data are normalized to TBP and represented as mean ± SEM, $n = 10$ biological replicates.

(F) Model illustrating siRNA treatment of dsRNA sensors (DDX58, IFIH1, and TLR3) inhibits type I interferon response and overall ISG expression.

(G) RT-qPCR analyses of ISG expression in control (shLuc; blue) and hnRNPM KD (shM; orange) HMLE cells. Lighter colors represent the ISG expression in cells additionally depleted of the dsRNA sensors DDX58, IFIH1, and TLR3. Data are normalized to TBP and represented as mean ± SEM, $n = 5$ biological replicates.

(H) RT-qPCR analyses of ISG expression in control (shLuc; blue) and hnRNPM KD (shM; orange) HMLE cells. Lighter colors represent the ISG expression in cells additionally treated for 48–72 h with the 2 μM of the JAK inhibitor ruxolitinib. Data relative to TBP are shown. Data are represented as mean ± SEM, $n = 7$ biological replicates, statistics in (E), (G), and (H): $^*p \leq 0.05$, $^{**}p \leq 0.01$, $^{****}p \leq 0.0001$ (two-tailed paired Student's t test).



(legend on next page)

diffuse large B cell lymphoma (DLBC), glioblastoma multiforme (GBM), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), and uveal melanoma (UVM) (Figure S8E), highlighting the clinical significance of hnRNPM expression and cryptic splicing in predicting patient survival outcome. In summary, our results reveal that hnRNPM^{low} tumors show enhanced cryptic splicing, elevated ISG expression and IFN-response, and increased immune cell infiltration, ultimately resulting in better patient survival.

DISCUSSION

Human introns harbor numerous pseudo splice sites with sequences indistinguishable from canonical splicing sites. Although normal cells effectively preserve pre-mRNA splicing accuracy, the mechanisms underlying their ability to inhibit pseudo splice site recognition and suppress cryptic splicing remain unclear. This study shows that hnRNPM preferentially binds introns at regions containing cryptic splice sites. Its binding represses the binding of spliceosomal proteins at the cryptic splice sites, preventing cryptic splicing. Our work has shed light onto an RBP-mediated mechanism for cryptic splicing suppression, which is crucial to maintaining transcriptome integrity.

hnRNPM is a ubiquitously expressed nuclear RBP and generally serves as a repressor of alternative splicing.[33,39] Loss of hnRNPM inhibits the epithelial-to-mesenchymal transition (EMT) and cancer metastasis through increased inclusion of epithelial alternative exons.[33] In this study, we found that hnRNPM can also serve as a repressor for CryEx inclusion. Most of CryEx-containing protein-coding transcripts are potential NMD targets due to a frameshift. Additionally, cryptic splicing can produce novel proteins that may impact cell fitness in normal cells and generate neo-antigens in tumors.[76–78]

Spliceosome proteins essential for 3′ and 5′ splice site recognition, such as U2AF2 and PRPF8, respectively, do not bind to CryEx splice sites according to analysis of ENCODE and our eCLIP data. Following hnRNPM KD, however, their bindings to cryptic splice sites become detectable, suggesting a mechanism by which hnRNPM prevents spliceosome proteins from binding to cryptic splice sites, thereby repressing cryptic splicing. Of note, while hnRNPM acts as a repressor at both functional and cryptic splice sites, hnRNPM-containing protein complexes between alternative and CryEx may differ. Investigating these protein complexes is crucial for elucidating the mechanisms underlying RBP-mediated cryptic splicing repression.

We and others found that hnRNPM binding is predominantly enriched at LINEs in deep introns.[23] The loss of hnRNPM binding in introns, particularly those containing LINEs, triggers cryptic splicing events, some of which generate dsRNAs, consequently activating the IFN response. Despite originating from intronic sequences, these CryEx are not sporadic and are distinguishable from intron-containing splicing intermediates.[79] Upon hnRNPM KD, CryEx consistently emerge at specific locations across different cell types. Additionally, certain CryEx we have identified occasionally appear in public databases, potentially due to reduced hnRNPM levels under those cellular conditions. Nonetheless, the consistent presence of CryEx at identical genomic locations suggests that cryptic splicing events are not random occurrences.

Cells likely employ a collective effort of RBPs to repress cryptic splicing. Using ENCODE RNA-seq data, we found that nearly a quarter of hnRNPM-repressed CryEx are also targeted by MATR3 or SUGP2. Knocking down MATR3 or SUGP2 showed similar effects on the shared CryEx as observed in hnRNPM KD cells (data not shown). The development of the CryEx pipeline opens avenues for future research into how RBPs collaboratively prevent cryptic splicing, offering a more comprehensive understanding of cryptic splicing biology.

In addition to our results, hnRNPM loss has been associated with IFN response in macrophages through hnRNPM's role in regulating IL-6 alternative splicing[80,81] and through its interaction with RIG-I-like receptors (RLRs).[82] These different modes of action suggest that hnRNPM is an essential player in repressing IFN response in various types of cells.

In recent years, there has been an abundance of treatments aimed at targeting different proteins to stimulate anti-tumor immunity.[83,84] Notably, our finding of CryEx-derived dsRNA suggests a distinct immunity stimulus. The immunity-enhancing properties of hnRNPM-repressed CryEx holds significant clinical promise, with implications extending beyond cancer therapy. As RBPs are emerging as therapeutic targets for diseases involving genomic abnormalities like cancer,[19,85,86] inhibiting hnRNPM or enhancing the splicing of dsRNA-forming CryEx could represent innovative methods to activate immunity in patients. Overall, hnRNPM emerges as a guardian of transcriptome integrity by repressing cryptic splicing, and we provide a connection between RBP-loss-induced cryptic splicing and the consequent dsRNA production and tumor immunity.

Limitations of the study

This study shows that hnRNPM plays a key role in repressing cryptic splicing in repeat elements. Other RBPs may support

Figure 6. hnRNPM deficiency in cancer patients leads to upregulated immunity and better survival

(A) Selected MsigDB hallmark gene sets enriched in hnRNPM lowly expressed tumors in TCGA cohorts. Circle size represents gene set enrichment significance ($-\log_{10}$ FDR). Color indicates normalized enrichment score (NES) for each gene set per cancer.

(B) ISG score distribution between hnRNPM-low and -high tumors across cancers. Scores from ssGSEA of the curated ISG gene signature were computed.

(C) Boxplots of $\log_2$-transformed PSI distributions of the four most representative dsRNA-forming cryptic exons between hnRNPM-low (orange) and -high (blue) tumors in TCGA basal subtype breast cancer.

(D) Heatmap showing MED15 cryptic exon inclusion levels across tumors and ssGSEA scores for xCell immune cell signatures.

(E) Violin plots showing immune infiltration ssGSEA score distributions between low- and high-risk score tumors.

(F) Kaplan-Meier survival analysis for patients with low- versus high-risk scores. p value was calculated using the logrank test. Statistics in (B)–(E): $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$ (Wilcoxon rank-sum test).

hnRNPM during this process or repress cryptic splicing independently. Although we propose a model wherein hnRNPM blocks the binding of spliceosome protein recognition at pseudo splice sites, the precise molecular mechanism remains to be elucidated. We utilized cells with stable hnRNPM KD in examining IFN-I responses, an acute depletion of hnRNPM would be better suited to avoid secondary effects. The formation of CryEx-derived dsRNAs induces an IFN response; however, other outcomes of cryptic splicing—like the creation of neo-antigens—might further contribute to the increased immune infiltration observed in hnRNPM^{low} tumors, which would be interesting to explore in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell culture
- METHOD DETAILS
  - iCLIP assay and analysis
  - Calculation of a normalized binding score for RBPs on different gene features
  - Calculation of a preferential binding metric for hnRNPM
  - Deep RNA sequencing and data analysis
  - De novo identification of hnRNPM-regulated cryptic exons and retained introns
  - RT-PCR analysis
  - Western blot
  - Splicing minigene reporter and assay
  - Cytoplasmic and nuclear fractionation for RNA isolation
  - In-vitro transcription and RNA dot blot
  - siRNA Knockdown
  - RNA pull-down
  - Identification and analysis of RNA editing sites
  - RNA fluorescent in situ hybridization for cryptic exons
  - Immunofluorescence
  - Microscopy and image analysis
  - Metaplot and motif enrichment
  - Spliceosomal protein binding density
  - RNA structure prediction
  - ISG score calculation
  - Immune infiltration score calculation
  - Clinical and molecular data of TCGA study
  - Multivariate survival model and risk score calculation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.05.004.

ACKNOWLEDGMENTS

This work is dedicated to Dr. Phillip A. Sharp for his 50-year anniversary of the Sharp Lab and for his mentorship to R.A.F. and C.C. We thank members in the Cheng lab for helpful feedback on the project. We thank Dr. Wei Hong and Dr. Chao Cheng (BCM) for discussions. We thank Dr. Charles Bou-Nader (NIH/NIDDK) for his advice on J2 antibody staining. This research was supported in part by grants from NIH R35-CA209919 (to H.Y.C.), R01CA262686 and R01AG078950 (to X.X.), and R35GM131876 (to C.C.). E.L.V.N. and C.C. are Cancer Prevention Research Institute of Texas Scholars in Cancer Research (RR200040 to E.L.V.N. and RR160009 to C.C.).

AUTHOR CONTRIBUTIONS

C.C., X.X., and R.Z. conceived of the project. R.Z. performed most of the bioinformatics analyses, with help from C.G.-F. and M.C. on ADAR analysis, S.E.H. on CLIP analysis, G.Q.-V. on CLIP metaplot analysis, and T.W.C. for TCGA bioinformatics supervision. M.D., G.O.M.B., K.J.P., J.L., and A.V. performed the experiments. C.A.L.R., M.D.B., and E.L.V.N. performed eCLIP. R.A.F. and H.Y.C. performed CLIP. E.L.V.N. helped with bioinformatics methods and interpretation. C.C. and X.X. supervised the study. R.Z. and C.C. wrote the manuscript and X.X. helped with the editing. All contributors read the manuscript and provided extensive inputs.

DECLARATION OF INTERESTS

R.A.F. is a co-founder, board of directors member, and stockholder of GanNA Bio and a board of directors member and stockholder of Chronus Health. H.Y.C. is a co-founder of Accent Therapeutics, Boundless Bio, Cartography Biosciences, and Orbital Therapeutics, an advisor of 10× Genomics, Arsenal Biosciences, Chroma Medicine, and Spring Discovery, and a member of the *Molecular Cell* advisory board. E.L.V.N. is a co-founder, member of the board of directors, on the SAB, an equity holder, and a paid consultant for Eclipse BioInnovations, on the SAB of RNAConnect, and an inventor of intellectual property owned by University of California, San Diego. E.L.V.N.'s interests have been reviewed and approved by the Baylor College of Medicine in accordance with its conflict-of-interest policies.

Received: February 24, 2023
Revised: January 9, 2024
Accepted: May 7, 2024
Published: May 29, 2024

REFERENCES

1. Jo, B.S., and Choi, S.S. (2015). Introns: the functional benefits of introns in genomes. Genomics Inform. *13*, 112–118. https://doi.org/10.5808/GI.2015.13.4.112.
2. McCoy, M.J., and Fire, A.Z. (2020). Intron and gene size expansion during nervous system evolution. BMC Genomics *21*, 360. https://doi.org/10.1186/s12864-020-6760-4.
3. Hare, M.P., and Palumbi, S.R. (2003). High intron sequence conservation across three mammalian orders suggests functional constraints. Mol. Biol. Evol. *20*, 969–978. https://doi.org/10.1093/molbev/msg111.
4. Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F.E., Singh, N.K., Abraham, L., Karisto, P., Vonlanthen, T., Mundt, C., McDonald, B.A., and Croll, D. (2021). A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. eLife *10*, e69249. https://doi.org/10.7554/eLife.69249.
5. Gozashti, L., Roy, S.W., Thornlow, B., Kramer, A., Ares, M., Jr., and Corbett-Detig, R. (2022). Transposable elements drive intron gain in diverse eukaryotes. Proc. Natl. Acad. Sci. USA *119*, e2209766119. https://doi.org/10.1073/pnas.2209766119.
6. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921. https://doi.org/10.1038/35057062.
7. Etchegaray, E., Naville, M., Volff, J.N., and Haftek-Terreau, Z. (2021). Transposable element-derived sequences in vertebrate development. Mobile DNA *12*, 1. https://doi.org/10.1186/s13100-020-00229-5.
8. Wang, D., Su, Y., Wang, X., Lei, H., and Yu, J. (2012). Transposon-derived and satellite-derived repetitive sequences play distinct functional

roles in Mammalian intron size expansion. Evol. Bioinform. Online *8*, 301–319. https://doi.org/10.4137/EBO.S9758.

9. Zhang, X.H.F., Leslie, C.S., and Chasin, L.A. (2005). Computational searches for splicing signals. Methods *37*, 292–305. https://doi.org/10.1016/j.ymeth.2005.07.011.
10. Cheng, Y., Kang, X.Z., Chan, P., Ye, Z.W., Chan, C.P., and Jin, D.Y. (2022). Aberrant splicing events caused by insertion of genes of interest into expression vectors. Int. J. Biol. Sci. *18*, 4914–4931. https://doi.org/10.7150/ijbs.72408.
11. Black, D.L. (2005). A simple answer for a splicing conundrum. Proc. Natl. Acad. Sci. USA *102*, 4927–4928. https://doi.org/10.1073/pnas.0501414102.
12. Sun, H., and Chasin, L.A. (2000). Multiple splicing defects in an intronic false exon. Mol. Cell. Biol. *20*, 6414–6425. https://doi.org/10.1128/MCB.20.17.6414-6425.2000.
13. Humphrey, J., Emmett, W., Fratta, P., Isaacs, A.M., and Plagnol, V. (2017). Quantitative analysis of cryptic splicing associated with TDP-43 depletion. BMC Med. Genomics *10*, 38. https://doi.org/10.1186/s12920-017-0274-1.
14. Aldalaqan, S., Dalgliesh, C., Luzzi, S., Siachisumo, C., Reynard, L.N., Ehrmann, I., and Elliott, D.J. (2022). Cryptic splicing: common pathological mechanisms involved in male infertility and neuronal diseases. Cell Cycle *21*, 219–227. https://doi.org/10.1080/15384101.2021.2015672.
15. Keegan, N.P., Wilton, S.D., and Fletcher, S. (2021). Analysis of pathogenic pseudoexons reveals novel mechanisms driving cryptic splicing. Front. Genet. *12*, 806946. https://doi.org/10.3389/fgene.2021.806946.
16. Metherell, L.A., Akker, S.A., Munroe, P.B., Rose, S.J., Caulfield, M., Savage, M.O., Chew, S.L., and Clark, A.J. (2001). Pseudoexon activation as a novel mechanism for disease resulting in atypical growth-hormone insensitivity. Am. J. Hum. Genet. *69*, 641–646. https://doi.org/10.1086/323266.
17. Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. *582*, 1977–1986. https://doi.org/10.1016/j.febslet.2008.03.004.
18. Brinegar, A.E., and Cooper, T.A. (2016). Roles for RNA-binding proteins in development and disease. Brain Res. *1647*, 1–8. https://doi.org/10.1016/j.brainres.2016.02.050.
19. Qin, H., Ni, H., Liu, Y., Yuan, Y., Xi, T., Li, X., and Zheng, L. (2020). RNA-binding proteins in tumor progression. J. Hematol. Oncol. *13*, 90. https://doi.org/10.1186/s13045-020-00927-w.
20. Oliveira, C., Faoro, H., Alves, L.R., and Goldenberg, S. (2017). RNA-binding proteins and their role in the regulation of gene expression in Trypanosoma cruzi and Saccharomyces cerevisiae. Genet. Mol. Biol. *40*, 22–30. https://doi.org/10.1590/1678-4685-GMB-2016-0258.
21. Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., Reyes, A., Anders, S., Luscombe, N.M., and Ule, J. (2013). Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell *152*, 453–466. https://doi.org/10.1016/j.cell.2012.12.023.
22. Ma, X.R., Prudencio, M., Koike, Y., Vatsavayai, S.C., Kim, G., Harbinski, F., Briner, A., Rodriguez, C.M., Guo, C., Akiyama, T., et al. (2022). TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. Nature *603*, 124–130. https://doi.org/10.1038/s41586-022-04424-7.
23. Attig, J., Agostini, F., Gooding, C., Chakrabarti, A.M., Singh, A., Haberman, N., Zagalak, J.A., Emmett, W., Smith, C.W.J., Luscombe, N.M., and Ule, J. (2018). Heteromeric RNP assembly at LINEs controls lineage-specific RNA processing. Cell *174*, 1067–1081.e17. https://doi.org/10.1016/j.cell.2018.07.001.
24. Chen, Y.G., and Hur, S. (2022). Cellular origins of dsRNA, their recognition and consequences. Nat. Rev. Mol. Cell Biol. *23*, 286–301. https://doi.org/10.1038/s41580-021-00430-1.
25. Hur, S. (2019). Double-stranded RNA sensors and modulators in innate immunity. Annu. Rev. Immunol. *37*, 349–375. https://doi.org/10.1146/annurev-immunol-042718-041356.
26. Bowling, E.A., Wang, J.H., Gong, F., Wu, W., Neill, N.J., Kim, I.S., Tyagi, S., Orellana, M., Kurley, S.J., Dominguez-Vidaña, R., et al. (2021). Spliceosome-targeted therapies trigger an antiviral immune response in triple-negative breast cancer. Cell *184*, 384–403.e21. https://doi.org/10.1016/j.cell.2020.12.031.
27. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature *583*, 711–719. https://doi.org/10.1038/s41586-020-2077-3.
28. Liu, S., and Cheng, C. (2013). Alternative RNA splicing and cancer. Wiley Interdiscip. Rev. RNA *4*, 547–566. https://doi.org/10.1002/wrna.1178.
29. Yeo, G.W., Van Nostrand, E.L., and Liang, T.Y. (2007). Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. PLoS Genet. *3*, e85. https://doi.org/10.1371/journal.pgen.0030085.
30. Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101. https://doi.org/10.1126/science.aac7041.
31. Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092–1096. https://doi.org/10.1126/science.aac7557.
32. Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. Cell *163*, 1515–1526. https://doi.org/10.1016/j.cell.2015.11.015.
33. Xu, Y., Gao, X.D., Lee, J.H., Huang, H., Tan, H., Ahn, J., Reinke, L.M., Peter, M.E., Feng, Y., Gius, D., et al. (2014). Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. Genes Dev. *28*, 1191–1203. https://doi.org/10.1101/gad.241968.114.
34. Ramesh, N., Kour, S., Anderson, E.N., Rajasundaram, D., and Pandey, U.B. (2020). RNA-recognition motif in Matrin-3 mediates neurodegeneration through interaction with hnRNPM. Acta Neuropathol. Commun. *8*, 138. https://doi.org/10.1186/s40478-020-01021-5.
35. Ho, J.S., Di Tullio, F., Schwarz, M., Low, D., Incarnato, D., Gay, F., Tabaglio, T., Zhang, J., Wollmann, H., Chen, L., et al. (2021). HNRNPM controls circRNA biogenesis and splicing fidelity to sustain cancer cell fitness. eLife *10*, e59654. https://doi.org/10.7554/eLife.59654.
36. Wang, X., Li, J., Bian, X., Wu, C., Hua, J., Chang, S., Yu, T., Li, H., Li, Y., Hu, S., et al. (2021). CircURI1 interacts with hnRNPM to inhibit metastasis by modulating alternative splicing in gastric cancer. Proc. Natl. Acad. Sci. USA *118*, e2012881118. https://doi.org/10.1073/pnas.2012881118.
37. Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep. *1*, 167–178. https://doi.org/10.1016/j.celrep.2012.02.001.
38. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. USA *111*, E5593–E5601. https://doi.org/10.1073/pnas.1419161111.
39. Harvey, S.E., Xu, Y., Lin, X., Gao, X.D., Qiu, Y., Ahn, J., Xiao, X., and Cheng, C. (2018). Coregulation of alternative splicing by hnRNPM and ESRP1 during EMT. RNA *24*, 1326–1338. https://doi.org/10.1261/rna.066712.118.
40. Sibley, C.R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. Nat. Rev. Genet. *17*, 407–421. https://doi.org/10.1038/nrg.2016.46.

41. Smit, A., Hubley, R., and Green, P. (1996–2004). RepeatMasker Open-3.0. http://wwwrepeatmasker.org.

42. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting splicing from primary sequence with deep learning. Cell *176*, 535–548.e24. https://doi.org/10.1016/j.cell.2018.12.015.

43. Van Nostrand, E.L., Pratt, G.A., Yee, B.A., Wheeler, E.C., Blue, S.M., Mueller, J., Park, S.S., Garcia, K.E., Gelboin-Burkhart, C., Nguyen, T.B., et al. (2020). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. Genome Biol. *21*, 90. https://doi.org/10.1186/s13059-020-01982-9.

44. Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. *10*, 691–703. https://doi.org/10.1038/nrg2640.

45. Ciesiolka, A., Jazurek, M., Drazkowska, K., and Krzyzosiak, W.J. (2017). Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. Front. Cell. Neurosci. *11*, 97. https://doi.org/10.3389/fncel.2017.00097.

46. Kaushal, S., and Freudenreich, C.H. (2019). The role of fork stalling and DNA structures in causing chromosome fragility. Genes Chromosomes Cancer *58*, 270–283. https://doi.org/10.1002/gcc.22721.

47. Alamro, H., Alzamel, M., Iliopoulos, C.S., Pissis, S.P., and Watts, S. (2021). IUPACpal: efficient identification of inverted repeats in IUPAC-encoded DNA sequences. BMC Bioinformatics *22*, 51. https://doi.org/10.1186/s12859-021-03983-2.

48. Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26. https://doi.org/10.1186/1748-7188-6-26.

49. Walkley, C.R., and Li, J.B. (2017). Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. Genome Biol. *18*, 205. https://doi.org/10.1186/s13059-017-1347-3.

50. Hong, H., Lin, J.S., and Chen, L. (2015). Regulatory factors governing adenosine-to-inosine (A-to-I) RNA editing. Biosci. Rep. *35*, e00182. https://doi.org/10.1042/BSR20140190.

51. Chung, H., Calis, J.J.A., Wu, X., Sun, T., Yu, Y., Sarbanes, S.L., Dao Thi, V.L., Shilvock, A.R., Hoffmann, H.H., Rosenberg, B.R., and Rice, C.M. (2018). Human ADAR1 prevents endogenous RNA from triggering translational shutdown. Cell *172*, 811–824.e14. https://doi.org/10.1016/j.cell.2017.12.038.

52. Lamers, M.M., van den Hoogen, B.G., and Haagmans, B.L. (2019). ADAR1: "editor-in-chief" of cytoplasmic innate immunity. Front. Immunol. *10*, 1763. https://doi.org/10.3389/fimmu.2019.01763.

53. Zhang, Q., and Xiao, X. (2015). Genome sequence-independent identification of RNA editing sites. Nat. Methods *12*, 347–350. https://doi.org/10.1038/nmeth.3314.

54. Bahn, J.H., Lee, J.H., Li, G., Greer, C., Peng, G., and Xiao, X. (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res. *22*, 142–150. https://doi.org/10.1101/gr.124107.111.

55. Schönborn, J., Oberstrass, J., Breyel, E., Tittgen, J., Schumacher, J., and Lukacs, N. (1991). Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. Nucleic Acids Res. *19*, 2993–3000. https://doi.org/10.1093/nar/19.11.2993.

56. Kitagawa, Y., Matsumoto, T., Okuhara, E., and Shikata, E. (1977). Immunogenicity of rice dwarf virus-ribonucleic acid. Tohoku J. Exp. Med. *122*, 337–343. https://doi.org/10.1620/tjem.122.337.

57. Batzer, M.A., and Deininger, P.L. (2002). Alu repeats and human genomic diversity. Nat. Rev. Genet. *3*, 370–379. https://doi.org/10.1038/nrg798.

58. Wang, Q., and Carmichael, G.G. (2004). Effects of length and location on the cellular response to double-stranded RNA. Microbiol. Mol. Biol. Rev. *68*, 432–452. https://doi.org/10.1128/MMBR.68.3.432-452.2004.

59. Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R., and Paludan, S.R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J. Virol. *80*, 5059–5064. https://doi.org/10.1128/JVI.80.10.5059-5064.2006.

60. Werner, A., Clark, J.E., Samaranayake, C., Casement, J., Zinad, H.S., Sadeq, S., Al-Hashimi, S., Smith, M., Kotaja, N., and Mattick, J.S. (2021). Widespread formation of double-stranded RNAs in testis. Genome Res. *31*, 1174–1186. https://doi.org/10.1101/gr.265603.120.

61. Onomoto, K., Onoguchi, K., and Yoneyama, M. (2021). Regulation of RIG-I-like receptor-mediated signaling: interaction between host and viral factors. Cell. Mol. Immunol. *18*, 539–555. https://doi.org/10.1038/s41423-020-00602-7.

62. Zhang, F.L., Cao, J.L., Xie, H.Y., Sun, R., Yang, L.F., Shao, Z.M., and Li, D.Q. (2018). Cancer-associated MORC2-mutant M276I regulates an hnRNPM-mediated CD44 splicing switch to promote invasion and metastasis in triple-negative breast cancer. Cancer Res. *78*, 5780–5792. https://doi.org/10.1158/0008-5472.CAN-17-1394.

63. Qiao, L., Xie, N., Li, Y., Bai, Y., Liu, N., and Wang, J. (2022). Downregulation of HNRNPM inhibits cell proliferation and migration of hepatocellular carcinoma through MAPK/AKT signaling pathway. Transl. Cancer Res. *11*, 2135–2144. https://doi.org/10.21037/tcr-21-2484.

64. Vitali, P., and Scadden, A.D. (2010). Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. Nat. Struct. Mol. Biol. *17*, 1043–1050. https://doi.org/10.1038/nsmb.1864.

65. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

66. Lazear, H.M., Schoggins, J.W., and Diamond, M.S. (2019). Shared and distinct functions of type I and Type III interferons. Immunity *50*, 907–923. https://doi.org/10.1016/j.immuni.2019.03.025.

67. Dowling, J.W., and Forero, A. (2022). Beyond good and evil: molecular mechanisms of type I and III IFN functions. J. Immunol. *208*, 247–256. https://doi.org/10.4049/jimmunol.2100707.

68. Aricò, E., Castiello, L., Capone, I., Gabriele, L., and Belardelli, F. (2019). Type I interferons and cancer: an evolving story demanding novel clinical applications. Cancers (Basel) *11*, 1943. https://doi.org/10.3390/cancers11121943.

69. Cao, X., Liang, Y., Hu, Z., Li, H., Yang, J., Hsu, E.J., Zhu, J., Zhou, J., and Fu, Y.X. (2021). Next generation of tumor-activating type I IFN enhances anti-tumor immune responses to overcome therapy resistance. Nat. Commun. *12*, 5866. https://doi.org/10.1038/s41467-021-26112-2.

70. Zitvogel, L., Galluzzi, L., Kepp, O., Smyth, M.J., and Kroemer, G. (2015). Type I interferons in anticancer immunity. Nat. Rev. Immunol. *15*, 405–414. https://doi.org/10.1038/nri3845.

71. Yi, M., Nissley, D.V., McCormick, F., and Stephens, R.M. (2020). ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. Sci. Rep. *10*, 10258. https://doi.org/10.1038/s41598-020-66986-8.

72. Cózar, B., Greppi, M., Carpentier, S., Narni-Mancinelli, E., Chiossone, L., and Vivier, E. (2021). Tumor-infiltrating natural killer cells. Cancer Discov. *11*, 34–44. https://doi.org/10.1158/2159-8290.CD-20-0655.

73. Fenton, S.E., Saleiro, D., and Platanias, L.C. (2021). Type I and II interferons in the anti-tumor immune response. Cancers (Basel) *13*, 1037. https://doi.org/10.3390/cancers13051037.

74. Crouse, J., Kalinke, U., and Oxenius, A. (2015). Regulation of antiviral T cell responses by type I interferons. Nat. Rev. Immunol. *15*, 231–242. https://doi.org/10.1038/nri3806.

75. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220. https://doi.org/10.1186/s13059-017-1349-1.

76. Biernacki, M.A., and Bleakley, M. (2020). Neoantigens in hematologic malignancies. Front. Immunol. *11*, 121. https://doi.org/10.3389/fimmu.2020.00121.

77. Seddighi, S., Qi, Y.A., Brown, A.L., Wilkins, O.G., Bereda, C., Belair, C., Zhang, Y., Prudencio, M., Keuss, M.J., Khandeshi, A., et al. (2023). Mis-spliced transcripts generate de novo proteins in TDP-43-related ALS/FTD. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.23.525149.

78. Diederichs, S., Bartsch, L., Berkmann, J.C., Fröse, K., Heitmann, J., Hoppe, C., Iggena, D., Jazmati, D., Karschnia, P., Linsenmeier, M., et al. (2016). The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. EMBO Mol. Med. *8*, 442–457. https://doi.org/10.15252/emmm.201506055.

79. Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., and Larson, D.R. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. Cell *184*, 2878–2895.e20. https://doi.org/10.1016/j.cell.2021.04.012.

80. Wagner, A.R., Scott, H.M., West, K.O., Vail, K.J., Fitzsimons, T.C., Coleman, A.K., Carter, K.E., Watson, R.O., and Patrick, K.L. (2021). Global transcriptomics uncovers distinct contributions from splicing regulatory proteins to the macrophage innate immune response. Front. Immunol. *12*, 656885. https://doi.org/10.3389/fimmu.2021.656885.

81. West, K.O., Scott, H.M., Torres-Odio, S., West, A.P., Patrick, K.L., and Watson, R.O. (2019). The splicing factor hnRNP M is a critical regulator of innate immune gene expression in macrophages. Cell Rep. *29*, 1594–1609.e5. https://doi.org/10.1016/j.celrep.2019.09.078.

82. Cao, P., Luo, W.W., Li, C., Tong, Z., Zheng, Z.Q., Zhou, L., Xiong, Y., and Li, S. (2019). The heterogeneous nuclear ribonucleoprotein hnRNPM inhibits RNA virus-triggered innate immunity by antagonizing RNA sensing of RIG-I-like receptors. PLoS Pathog. *15*, e1007983. https://doi.org/10.1371/journal.ppat.1007983.

83. Zou, S., Tong, Q., Liu, B., Huang, W., Tian, Y., and Fu, X. (2020). Targeting STAT3 in cancer immunotherapy. Mol. Cancer *19*, 145. https://doi.org/10.1186/s12943-020-01258-7.

84. Shi, C., Wang, Y., Wu, M., Chen, Y., Liu, F., Shen, Z., Wang, Y., Xie, S., Shen, Y., Sang, L., et al. (2022). Promoting anti-tumor immunity by targeting TMUB1 to modulate PD-L1 polyubiquitination and glycosylation. Nat. Commun. *13*, 6951. https://doi.org/10.1038/s41467-022-34346-x.

85. Hong, S. (2017). RNA binding protein as an emerging therapeutic target for cancer prevention and treatment. J. Cancer Prev. *22*, 203–210. https://doi.org/10.15430/JCP.2017.22.4.203.

86. Kang, D., Lee, Y., and Lee, J.S. (2020). RNA-binding proteins in cancer: functional and therapeutic perspectives. Cancers (Basel) *12*, 2699. https://doi.org/10.3390/cancers12092699.

87. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173*, 400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.

88. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

89. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

90. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

91. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

92. Shah, A., Qian, Y., Weyn-Vanhentenryck, S.M., and Zhang, C. (2017). CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. Bioinformatics *33*, 566–567. https://doi.org/10.1093/bioinformatics/btw653.

93. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

94. Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., Li, B., and Liu, X.S. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. *48*, W509–W514. https://doi.org/10.1093/nar/gkaa407.

95. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

96. Flynn, R.A., Martin, L., Spitale, R.C., Do, B.T., Sagan, S.M., Zarnegar, B., Qu, K., Khavari, P.A., Quake, S.R., Sarnow, P., and Chang, H.Y. (2015). Dissecting noncoding and pathogen RNA-protein interactomes. RNA *21*, 135–143. https://doi.org/10.1261/rna.047803.114.

97. Damianov, A., Ying, Y., Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A., and Black, D.L. (2016). Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. Cell *165*, 606–619. https://doi.org/10.1016/j.cell.2016.03.040.

98. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

99. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205. https://doi.org/10.1093/nar/gkz401.

100. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408. https://doi.org/10.1006/meth.2001.1262.

101. Harvey, S.E., and Cheng, C. (2016). Methods for characterization of alternative RNA splicing. Methods Mol. Biol. *1402*, 229–241. https://doi.org/10.1007/978-1-4939-3378-5_18.

102. Quinones-Valdez, G., Tran, S.S., Jun, H.I., Bahn, J.H., Yang, E.W., Zhan, L., Brümmer, A., Wei, X., Van Nostrand, E.L., Pratt, G.A., et al. (2019). Regulation of RNA editing by RNA-binding proteins in human cells. Commun. Biol. *2*, 19. https://doi.org/10.1038/s42003-018-0271-8.

103. Yang, E.W., Bahn, J.H., Hsiao, E.Y.H., Tan, B.X., Sun, Y., Fu, T., Zhou, B., Van Nostrand, E.L., Pratt, G.A., Freese, P., et al. (2019). Allele-specific binding of RNA-binding proteins reveals functional genetic variants in the RNA. Nat. Commun. *10*, 1338. https://doi.org/10.1038/s41467-019-09292-w.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-β-actin (AC-15)	Sigma-Aldrich	Cat# A1978; RRID: AB_476692
Mouse anti-GAPDH	EMD Millipore	Cat# MAB374; RRID: AB_2107445
Mouse anti-hnRNPM (2A6)	OriGene	Cat# TA301557
Donkey anti-rabbit AlexaFluor 488	Invitrogen	Cat# A-21206; RRID: AB_2535792
Anti-mouse AlexaFluor 488	Abcam	Cat# AB150113; RRID: AB_2576208
Sheep Anti-Mouse IgG, Whole Ab ECL Antibody, HRP Conjugated	Cytiva	Cat# NXA931; RRID: AB_772209
Rabbit anti-IFN Beta Polyclonal Antibody	Proteintech	Cat# 27506-1-AP; RRID: AB_2880893
Mouse anti-dsRNA (J2)	Jena Bioscience	Cat# RNT-SCI-10010200; RRID: AB_2922431
Rabbit anti-dsRNA (9D5)	Absolute Antibody	Cat# Ab00458-23.0; RRID: AB_2920603
Chemicals, peptides, and recombinant proteins		
TRIzol Reagent	Invitrogen	Cat# 15596026
Ruxolitinib	Selleckchem	Cat# S1378
Proteinase K	ThermoFisher	Cat# EO0491
RNAse III	ThermoFisher	Cat# AM2290
DAPI	Sigma-Aldrich	Cat# 10236276001
Hoechst	ThermoFisher	Cat# 62249
Acridine Orange	Invitrogen	Cat# A1301
Critical commercial assays		
HotStarTaq Plus DNA Polymerase	Qiagen	Cat# 203605
GoScript Reverse Transcriptase	Promega	Cat# A5004
Phusion High-Fidelity DNA Polymerase	NEB	Cat# M0530
High-Speed Plasmid Mini Kit	IBI Scientific	Cat# IB47102
Genopure Plasmid Midi Kit	Roche	Cat# 3143414001
E.Z.N.A. Total RNA Kit I	Omega Bio-Tek	Cat# R6834-02
Gel/PCR DNA Fragments Extraction Kit	IBI Scientific	Cat# IB47020
GoTaq qPCR Master Mix	Promega	Cat# A6002
Lipofectamine RNAiMAX Transfection Reagent	Invitrogen	Cat# 13778150
Lipofectamine 2000 Transfection Reagent	Invitrogen	Cat# 11668019
T4 DNA Ligase	NEB	Cat# M0202
Rapid DNA Ligation Kit	Roche	Cat# 11635379001
Riboprobe® Systems T3	Promega	Cat# P1430
Riboprobe® Systems T7	Promega	Cat# P1440
Amersham Hybond-N+ Membrane	Cytiva	Cat# RPN303B
Deposited data		
RNA-seq of HeLa cells depleted of hnRNPM	This paper	GEO: GSE227047
RNA-seq of LM2 cells depleted of hnRNPM	This paper	GEO: GSE227047
RNA-seq of BM2 cells depleted of hnRNPM	This paper	GEO: GSE227047
RNA-seq of HMLE cells depleted of hnRNPM	This paper	GEO: GSE227047
iCLIP of hnRNPM in HMLE cells	This paper	GEO: GSE227048
Unprocessed microscopic images, western blots, and agarose gel images	This paper	Mendeley link https://doi.org/10.17632/7p7dpzcc6x.1
Processed gene expression data from TCGA cohort	NCI Genomic Data Commons	https://portal.gdc.cancer.gov

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNA-seq data from TCGA cohort	NCI Genomic Data Commons	https://portal.gdc.cancer.gov
Clinical outcome endpoints data from TCGA cohort	Liu et al.[87]	N/A
All eCLIP data	ENCODE Consortium	Datasets used are listed in Table S1. Accesible via https://www.encodeproject.org/search/?type=Experiment&assay_title=eCLIP
CryEx Pipeline and other codes and scripts	This paper; Zenodo	Zenodo: https://doi.org/10.5281/zenodo.10543478
Human reference genome primary assembly (GRCh37)	GENCODE Consortium	https://www.gencodegenes.org/human/release_37lift37.html
Human transcriptome GENCODE version 24 backmap 37 comprehensive assembly	GENCODE Consortium	https://www.gencodegenes.org/human/release_24lift37.html
Repeat Masker genome annotation	Smit et al.[41]	RRID: SCR_012954
Experimental models: Cell lines		
HMLE	Jing Yang, UCSD	N/A
LM2	Yibing Kang, Princeton	N/A
BM2	Yibing Kang, Princeton	N/A
HeLa	ATCC	CCL-2™
293FT	ATCC	CRL-3216™
Oligonucleotides		
qRT-PCR and semi qRT-PCR primers	This paper	See Table S3
siRNAs and shRNAs	This paper	See Table S3
Cloning primers	This paper	See Table S3
Sequencing primers	This paper	See Table S3
MED15 and TRAPPC10 dsRNA and reverted sequences for in vitro transcription	This paper	See Table S3
ADIPOR2 ctrl and mutant sequences for minigene	This paper	See Table S3
ADIPOR2 Oligo 1	This paper	UUUCUGUGGGAUUGGUGGUA
ADIPOR2 Oligo 2	This paper	UACUUUGUAUUUCUGUGGGA
TRAPPC10 Oligo	This paper	CUUCUGCUUGUUUGUGACCC
RNA FISH probes	Stellaris	See Table S4
Recombinant DNA		
pLKO.1-shluc	Xu et al.[33]	N/A
pLKO.1-shhnRNPM	Xu et al.[33]	N/A
pcDNA3	Xu et al.[33]	N/A
pcDNA3-hnRNPM	Xu et al.[33]	N/A
pET	Xu et al.[33]	N/A
pET-ADIPOR2	This paper	N/A
pET-ADIPOR2-mut	This paper	N/A
pET-MED15	This paper	N/A
pUC57-ADIPOR2-ctrl	This paper	N/A
pUC57-ADIPOR2-mut	This paper	N/A
Software and algorithms		
GraphPad Prism (version 10.1.1)	Dotmatics	https://www.graphpad.com
STAR (version 2.7.9a)	Dobin et al.[88]	https://github.com/alexdobin/STAR
featureCounts (version 1.5.0)	Liao et al.[89]	https://subread.sourceforge.net
DESeq2	Love et al.[90]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
rMATS (version 4.0.2)	Shen et al.[38]	https://rmats.sourceforge.io

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
StringTie	Pertea et al.[91]	https://ccb.jhu.edu/software/stringtie/
Clip Tool Kit (version 1.0.3)	Shah et al.[92]	https://zhanglab.c2b2.columbia.edu/index.php/CTK_Documentation#Download
HOMER (version 4.10)	Heinz et al.[93]	http://homer.ucsd.edu/homer/
ViennaRNA Package 2.0	Lorenz et al.[48]	https://www.tbi.univie.ac.at/RNA/
TIMER2.0	Li et al.[94]	http://timer.cistrome.org
R	R Project for Statistical Computing	https://www.r-project.org/
ImageJ (FIJI, version 2.14.0/1.54f)	Schindelin et al.[95]	https://imagej.net/software/fiji/
BioRender	BioRender	https://app.biorender.com/
Other		
Nanodrop 2000	ThermoFisher	N/A
CFX Connect Real-Time PCR system	BioRad	N/A
ChemiDoc™ Touch Imaging Systems	BioRad	N/A
QIAxcel Advanced System	Qiagen	N/A
GE Healthcare DeltaVision LIVE High Resolution Deconvolution Microscope	GE Healthcare	N/A
Zeiss LSM 780 Confocal Microscope	Zeiss	LSM 780
Echo Revolve Fluorescence Microscope	ECHO	N/A
Olympus IX83 (epifluorescence deconvolution microscope)	Olympus Lifescience	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chonghui Cheng (chonghui.cheng@bcm.edu).

Materials availability

All unique reagents generated in this study are available from the lead contact upon request.

Data and code availability

- RNA-seq data of HeLa, LM2, BM2 and HMLE cells depleted of hnRNPM and iCLIP data of hnRNPM in HMLE cells have been deposited at GEO and are publicly available as of the date of publication. Publicly available datasets including eCLIP data from ENCODE Consortium and RNA-seq data from TCGA cohort were utilized in this study. Accession numbers and accessible links are listed in the key resources table. Unprocessed western blots and agarose gel images have been deposited at Mendeley and are publicly available. Additional microscopy data reported in this paper will be shared by the lead contact upon request.
- All original code has been deposited at GitHub (https://github.com/CC-Cheng-Splicing-lab-BCM/hnRNPM_CryEx_dsRNA) and Zenodo (https://doi.org/10.5281/zenodo.10543478) and is publicly available. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell culture

Human embryonic kidney cell line 293FT (CRL-3216™) and cervical cancer cell line HeLa (CCL-2™) were previously obtained from ATCC. Breast carcinoma cell line MDA-MB-231-derived LM2 and BM2 cells were previously provided by Yibin Kang (Princeton), while the immortalized human mammary epithelial cells (HMLE) were previously provided by Jing Yang (UCSD). 293FT, HeLa, LM2 and BM2 cells were grown in DMEM supplemented with 10% FBS, L-glutamine and Pen/Strep at 37℃. HMLE cells were grown in Mammary Epithelial Cell Growth Medium (Lonza, CC-3150) as described previously.[33] All cells were passaged every 2 – 3 days. 2μM JAK inhibitor Ruxolitinib (Selleckchem, S1378) was used for inhibition analysis.

METHOD DETAILS

iCLIP assay and analysis

Two biological replicates of HMLE cells were UV crosslinked and iCLIP was performed as previously described using anti-hnRNPM antibody (OriGene, TA301557).[96] Data analysis was conducted using the Clip Tool Kit (CTK) v1.0.3.[92] After processing and mapping iCLIP reads in CTK, replicates were pooled and single nucleotide resolution binding sites for hnRNPM were determined using the following strategy: 1) Statistically significant iCLIP peaks were called and delimited by one half the peak height (adjusted $p < 0.05$); 2) Significant crosslinking induced truncations sites (CITS) were called (adjusted $p < 0.05$) using the CTK CITS.pl script; 3) CITS were further filtered by retaining only those that overlapped with a significant peak; 4) CITS were extended upstream and downstream by 10 nucleotides to identify local binding sites for further analysis.

De novo motif analysis was performed using HOMER v4.10[93] findMotifsGenome.pl script with the following non-default parameters -p 4 -rna -S 10 -len 4,5,6 -size 100 -chopify. De novo motifs were computed compared to a background of shuffled human introns.

For the GU-rich hnRNPM binding motif RNA map analysis, 14 GU-rich 5-mer motifs used in a previous study to identify hnRNPM binding motifs were used to screen for motif enrichment.[97] The motifs were = ['UGUGU', 'GUGUG', 'UUGUG', 'GUGUU', 'UGUUG', 'UGUGG', 'GUUGU', 'GGUGU', 'UGGUU', 'UUGGU', 'UGGUG', 'GUGGU', 'GUUGG', 'GGUUG'].

Calculation of a normalized binding score for RBPs on different gene features

To compare binding preferences of RBPs across different gene features, a normalized binding score for each RBP on each gene feature was calculated as follows. RBPs with more than 5% of their reproducible and significantly enriched IDR peaks located in introns in both HepG2 and K562 cells were selected for further analysis.[27] Number of each RBP's binding sites at each gene feature was normalized against each RBP's overall number of binding sites across pre-mRNA, generating a binding metric at different gene features in each RBP. The metric was then normalized by the genomic length of its associated gene feature, including 5'UTR, exon, 3'UTR, proximal intron (within 500 nt from the nearest splice site), and deep intron (500 nt away from the splice sites) using Gencode version 24 backmap 37 comprehensive gene annotation. Enrichment on a specific feature is defined by normalizing to the mean between the features.

Calculation of a preferential binding metric for hnRNPM

Significant hnRNPM iCLIP binding peaks in HMLE cells were calculated using the equation:

$$\textit{hnRNPM preferential binding metric} = \log 2\left(\frac{\frac{\%\ \textit{of peaks at each feature}}{\textit{genomic length of each feature}}}{\frac{\sum \%\ \textit{of peaks in the rest features}}{\sum \textit{genomic length of the rest features}}}\right)$$

The percentage of hnRNPM iCLIP peak counts at each feature was calculated and then divided by their corresponding genomic length. The percentage of hnRNPM iCLIP peak counts in the rest of features was divided by the total genomic length of the rest features. hnRNPM preferential binding metrics were calculated by dividing the two divisions above (each feature/the rest features), then $\log_2$-transformed.

To determine the enrichment of hnRNPM binding across pre-mRNA in Figure 1D, the features used in the equation were deep intron, 3'UTR, proximal intron, exon, and 5'UTR.

To assess binding enrichment for hnRNPM at repeats vs non-repeats within cryptic exons encompassing flanking 500 nt upstream and 500 nt downstream regions in Figure 3I, the features in the above equation were repeats and non-repeats.

To assess binding enrichment for hnRNPM across different intronic repeat types within cryptic exons encompassing flanking 500 nt upstream and 500 nt downstream regions in Figure 3J, different repeat types were classified based on their repetitive DNA distribution and sequence complexity, and the features in the above equation were intronic repeat types, including LINE, SINE, LTR, DNA transposons (DNA), and other repeat elements (Other).

To assess binding enrichment for hnRNPM across intronic repeats within spliceAI version 1.3.1 predicted splice sites encompassing flanking 100 nt upstream and 100 nt downstream regions in Figure 3L, the features in the above equation were intronic repeat types, including LINE, SINE, LTR, DNA, and Other.

For enrichment at different LINE subtypes in Figure 3M, the features in the above equation were L1, L2, L3, and other LINE.

Deep RNA sequencing and data analysis

Generation of HMLE and LM2 shLuc and shhnRNPM (shM2) cell lines using pLKO.1-shLuc and pLKO.1-shhnRNPM was described previously,[33] BM2 and HeLa shLuc and shM2 cell lines were generated with the same method. Three biological replicates for control shRNA and hnRNPM knockdown HMLE cells were collected using TRIzol Reagent (Invitrogen, 15596026) according to the manufacturer's instructions. Purified RNA samples were submitted to the Genomic Facility at University of Chicago for RNA quality validation, RNA-seq library generation and paired-end sequencing on HiSeq 4000. The RNA-seq was performed using 100 bp paired-end with depths of ~80-180M reads. RNA-seq reads were aligned to the human genome (GRCh37, primary assembly) and

transcriptome (Gencode version 24 backmap 37 comprehensive gene annotation) using STAR v2.7.9a[88] with the following non-standard parameters –outSAMstrandField intronMotif –outFilterType BySJout –outFilterMultimapNmax 1 –alignSJoverhangMin 8 –alignSJDBoverhangMin 3 –alignEndsType EndToEnd. Only uniquely aligned reads were retained for downstream analysis.

Differential gene expression analysis was performed by counting reads over genes from the same annotation as alignment using featureCounts version 1.5.0[89] with the following non-default parameters -s 0 -a. Differential gene expression analysis was conducted using DESeq2[90] performed on genes with at least 5 counts present in at least half of the samples. Significantly deregulated genes were defined as genes with an $|\log_2 FC| > 2$ and FDR < 0.05.

Differential alternative splicing was quantified using rMATS version 4.0.2 with the following non-default parameters –readLength 100 –cstat 0.01 –libType fr-secondstrand. To identify significant differential splicing events, we set up the following cutoffs: FDR<0.05, |ΔPSI|>=0.1, and average junction reads per event per replicate >=20.

Gene set enrichment analysis (GSEA) was conducted through our in-house script that uses fgsea[98] on genes ranked by log2FC with 1000 gene set level permutations considering only gene sets of 15 to 500 genes. Significantly enriched pathways were selected by FDR < 0.05. Gene ontology (GO) biological process analysis was conducted using http://webgestalt.org with default setting.[99] By ranking log2FC of hnRNPM knockdown vs control shRNA, the top 300 upregulated and downregulated genes were uploaded for the analysis separately.

De novo identification of hnRNPM-regulated cryptic exons and retained introns

To predict exons from RNA-seq data, STAR v2.7.9a was used to generate coordinate sorted binary alignment map (BAM) files and Stringtie v2.1.5[91] was used to annotate any expressed exons for each sample. All exons that were not identical with exons annotated in GENCODE v19 were referred to as 'cryptic'. To narrow down to adequately expressed cryptic exons that arise from introns, 'cryptic' exons which are annotated exons with read alignments extending across the 5′ss, 3′ss, or both ends and have at least 5 reads aligned to their spliced junctions were kept. hnRNPM-regulated cryptic exons are defined as events with a PSI value less than 0.1 in non-specific shRNA-expressing control cells and ΔPSI values (ΔPSI = PSI{KD} – PSI{Ctrl}) larger than 0.1. Full-length intron retention events are analyzed separately. For hnRNPM-regulated retained introns, PSI value in control shRNA cells were required to be less than 0.1 and the PSI ratio (PSI ratio = PSI{KD} / PSI{Ctrl}) larger than 2. Scripts of the pipeline for identifying cryptic exons and intron retentions as well as their exon inclusion quantification can be found at https://github.com/CC-Cheng-Splicing-lab-BCM/hnRNPM_CryEx_dsRNA.

RT-PCR analysis

RT-PCR and semi qRT-PCR were conducted using RNA extracted from cells processed with the E.Z.N.A. Total RNA Kit I (Omega Bio-Tek, R6834-02). RNA concentration was measured using a Nanodrop 2000 (ThermoFisher). cDNA was generated via reverse transcription using the GoScript Reverse Transcription System (Promega, A5004) with 1 ul GoScript RT and 250 ng of RNA in a total volume of 20ul followed by incubation at 25 ℃ for 5min, 42 ℃ for 30min, and 70 ℃ for 15min.

qRT-PCR was performed using GoTaq qPCR Master Mix (Promega, A6002) per the manufacturer's in- structions on a CFX Connect Real-Time PCR system (BioRad) using a two-step protocol and supplied software. For every qPCR sample, two technical replicates were performed per biological replicate, and Ct counts were averaged. Normalization and quantification of qPCR data was done using the $2^{-\Delta Ct}$ method relative to TATA-binding protein (TBP) expression.[100] Primer specificity was verified with melt-curve.

Primers for semi-qRT-PCR analysis were designed on constitutive exons flanking each variable exon.[101] HotStarTaq Plus DNA Polymerase (Qiagen, 203605) was used to amplify the desired products. Semi-qRT-PCR generates both exon inclusion and skipping products in one PCR reaction, which were separated through agarose gel electrophoresis. PCR product intensity was quantified on a QIAxcel Advanced System (Qiagen). Primers for qPCR and semi-qRT-PCR are included in Table S3.

Western blot

Cell lysates from HMLE shLuc and shM2 were separated by 4-20% SDS-PAGE (GenScript, M42010) and transferred to a PVDF membrane (GE, 10600023). Membranes were blocked with 5% milk in 1x TBST (1x TBS, 0.05% Tween-20) for 30 minutes and then incubated in primary antibodies in 5% milk in TBST overnight at 4℃. Primary antibodies used were hnRNPM (Origene technologies, TA301557, 1:50,000), GAPDH (EMD Millipore, MAB374, 1:3,000), and β-actin (Sigma, A5441, 1:3,000). Membranes were then washed in TBST and incubated in the anti-mouse HRP-conjugated secondary antibody (Cytiva, NXA931) for 1 hour at room temperature. After final TBST washing, membranes were imaged using a ChemiDoc™ Touch Imaging Systems (BioRad) with Immobilon Western Chemiluminescent HRP Substrate (EMD Millipore, WBKLS0500).

Splicing minigene reporter and assay

MED15 cryptic exon and 500 bp of its up- and downstream intronic regions were amplified from genomic DNA of HMLE cells using Phusion High-Fidelity DNA Polymerase (NEB, M0530). Cloning primers for amplifying MED15 cryptic exon and proximal introns are listed in Table S3. PCR products were purified with Gel/PCR DNA Fragment Extraction Kit (IBI Scientific, IB47020). The pET backbone[33] and PCR products were digested with BamHI restriction enzyme (NEB, R3136S) and gel purified. The cut backbone and PCR products were ligated together using T4 DNA Ligase (NEB, M0202). Plasmids were purified using High-Speed Plasmid Mini Kit (IBI Scientific, IB47102) or GenoPure Plasmid MIDI kit (Roche, 3143414001). Plasmid sequences were confirmed by DNA sequencing.

ADIPOR2 ctrl and mutant sequences were synthesized by GenScript in the pUC57 backbone between XmaI and XbaI restriction enzyme sites. Sequence details are provided in Table S3. pET backbone, pUC57-ADIPOR2-ctrl, and pUC57-ADIPOR2-mut were digested with XmaI (NEB, R0180S) and XbaI (NEB, R0145S). Fragments of the appropriate size were extracted by agarose gel purification and ligated with a Rapid DNA ligation kit (Roche, 11635379001). Plasmid sequences were confirmed by DNA sequencing.

2.5x10∧5 293FT cells were seeded in 24-well plate 20–24 hours prior to transfection. 100 ng splicing minigene and increasing amount of pcDNA3-control or pcDNA3-hnRNPM plasmid[33] were cotransfected using Lipofectamine 2000 Transfection Reagent (Invitrogen, 11668019). Cells were collected 24 hours after transfection for RNA extraction and analysis.

Cytoplasmic and nuclear fractionation for RNA isolation

293FT cells growing in a 15cm tissue culture dish were collected in PBS using a tissue culture scraper. After two washes with PBS, cells were centrifuged and resuspended in 1X hypotonic buffer (20 mM Tris-HCL (pH 7.4), 10 mM NaCl, 3 mM MgCl2). After 15 minutes incubation on ice, 10% NP-40 was added, and the vial vortexed for 10 seconds. The fractions were separated by centrifugation for 10 minutes at 3,000 rpm at 4℃. The supernatant containing the cytoplasmic fraction was collected, and the RNA was extracted using TRIzol Reagent (Invitrogen, 15596026). The pellet containing the nuclear fraction was used for RNA pull-down.

In-vitro transcription and RNA dot blot

Plasmids that acted as templates for the in-vitro transcription were produced as follows: dsRNA region for MED15 was amplified from cDNA of hnRNPM KD HMLE cells using Phusion High-Fidelity DNA Polymerase (NEB, M0530). The PCR product was run through an agarose gel and purified with Gel/PCR DNA Fragment Extraction Kit (IBI Scientific, IB47020). A second PCR was performed using the PCR product and restriction enzyme flanking primers: PCR products were inserted into the pBRIT plasmid backbone using BamHI (NEB, R3136S) and XhoI (NEB, R0146S) and T4 DNA ligase (NEB, M0202). To generate the MED15 reverted mutant, a PCR was performed on the pBRIT-MED15 dsRNA plasmid with opposite restriction enzyme flanking primers. The PCR product was inserted backwards into the MED15 dsRNA plasmid by subcloning with BamHI and BsmI and T4 DNA Ligase. All 3 MED15 cloning primer sets are listed in Table S3. The dsRNA region for TRAPPC10 and its reverted mutant were synthesized and subcloned by GenScript into the pcDNA3.1 backbone using restriction enzymes BamHI and XhoI. Plasmid sequences were confirmed by DNA sequencing. Sequence details are provided in Table S3.

Plasmids were linearized using the NEB restriction enzyme XhoI for 2 hours and precipitated using ethanol. In-vitro RNA transcripts were produced using Riboprobe® Systems T3 (Promega, P1430; MED15) or T7 (Promega, P1440; TRAPPC10) kits according to the manufacturers manual. After DNAse digest, transcripts were extracted using TRIzol reagent (Invitrogen, 15596026), heated to 85°C for 30min as dsRNA structures recognizable by J2 also formed in the reverted configuration at room temperature, and dot blotted onto Amersham Hybond-N+ positively charged nylon membranes (Cytiva, RPN303B). RNA was UV cross-linked using 125mJoule/cm2 at 254nm and unbound RNA washed off using TBST (1x TBS, 0.1% Tween-20). Following a blocking of 30 minutes in TBST/5% non-fat dry milk powder (Bio-Rad 1706404), membranes were incubated for 1.5 hours with J2 antibody (Jena Bioscience, RNT-SCI-10010200; 1:1000) in TBST/5% Milk. After three washes in TBST, membranes were incubated in anti-mouse HRP antibody (Cytiva, NXA931) for 45 minutes. Following three more washes in TBST, membranes were imaged. For loading controls, membranes were washed and incubated for 30min in methylene blue staining buffer (0.2% methylene blue in 0.4M sodium acetate and 0.4M acetic acid) and imaged again. ChemiDoc™ Touch Imaging Systems (BioRad) was used for imaging.

siRNA Knockdown

Cells were plated on a 24-well tissue culture treated plate at 5-7x10∧4 cells per well. The following day, 20-60 nM siRNA was transfected using Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, 13778150). Media was replaced after 24 hours. RNA was collected 48 hours post-transfection using E.Z.N.A. Total RNA Kit I (Omega Bio-Tek, R6834-02). siRNA sequences are provided in Table S3.

RNA pull-down

RNA oligonucleotides labeled with biotin at the 5'-end were synthesized by Integrated DNA Technologies. The RNA sequences used in this study are listed as follows: ADIPOR2 Oligo 1: UUUCUGUGGGAUUGGUGGUA, ADIPOR2 Oligo 2: UACUUUGUAUUUCUGUGGGA, TRAPPC10 Oligo: CUUCUGCUUGUUUGUGACCC. 400 pmol Biotinylated RNA oligos were conjugated with 50 μl of streptavidin beads (50% slurry; ThermoFisher, 20347) in a total volume of 300 μl of RNA-binding buffer (20 mM Tris, 200 mM NaCl, 6 mM EDTA, 5 mM sodium fluoride and 5 mM β-glycerophosphate, PH 7.5) at 4 °C on a rotating shaker for 2 hours. After washing three times with RNA-binding buffer, RNA-beads conjugates were incubated with 100 μg of nuclear extracts in 500 μl RNA-binding buffer at 4 °C on a rotating shaker overnight. Beads were then washed with RNA-binding buffer for three times and the RNA pull-down samples were eluted with 2 × SDS loading buffer for western blot analysis.

Identification and analysis of RNA editing sites

Identification of RNA editing sites from RNA-seq reads was performed according to our previously published work.[102] In brief, RNA-DNA differences (RDDs) were identified from RNA-seq reads.[54] The identified RDDs were required to be covered by 5 or more reads, at least 2 edited reads, and 10% editing ratio. Specifically, RDDs were excluded if they are located in homopolymers,

splice sites, dbSNP database, or simple repeats.[103] The GIREMI tool was used to further select editing sites based on their mutual information with genetic variants.[53] To identify differentially edited sites between the two conditions, we used the parametric test for allelic bias as previously described.[103] Site-specific normal distributions were parameterized by the mean editing level between replicates and the expected variance calculated from the mean coverage. The FDR was calculated using the Benjamini-Hochberg method. Differentially edited sites were called by requiring FDR $\leq$ 10% and the absolute change in the editing level between knockdown and control shRNA $\geq$5.

RNA fluorescent in situ hybridization for cryptic exons

The Stellaris RNA FISH probe designer was used to generate 45-48 probes labelling constitutive exons and cryptic exons of TRAPPC10 and MED15 (Table S4). For visualization, exon probes were coupled to Quasar670 and cryptic exon probes to Quasar570. HMLE control shRNA and hnRNPM KD cells were fixed in PBS/4% PFA for 30 min at room temperature and RNA staining was performed according to the Stellaris RNA FISH protocol for adherent cells.

Immunofluorescence

For J2 (Jena Bioscience, RNT-SCI-10010200) and 9D5 (Absolute Antibody, Ab00458-23.0) staining, cells were grown for two days on coverslips after the second round of selection following shLuc or shhnRNPM transfection and fixed for 30 minutes in PBS/4%PFA. Fixed cells were permeabilized for 10 minutes using PBS/0.5% Triton and treated for 30 minutes with 15U/ml Proteinase K (ThermoFisher, EO0491) at 37°C. Where indicated, samples were treated for 1h with 10U of RNAse III (ThermoFisher, AM2290) in reaction buffer at 37°C for 3 hours. Following a 30-minute blocking step in PBS/2%BSA, primary antibody incubation was performed o/n at 4°C and the respective secondary antibody for 45 minutes at room temperature. For IFN beta (Proteintech, 27506-1-AP) staining, proteinase K treatment step was omitted. The full list of immunofluorescence antibodies used in this study can be found in the key resources table.

Microscopy and image analysis

High magnification images were acquired using either a GE Healthcare DeltaVision LIVE High Resolution Deconvolution Microscope or an Olympus IX83 and were deconvolved using softWoRx Explorer. Images of fixed cells were taken as 50–65z stacks of 0.2 μM increments using a 100× oil immersion objective. Quantification of signal intensities was performed using Fiji.[95] Enrichment of cytoplasmic J2 and 9D5 dsRNA signal in hnRNPM KD cells was quantified from pictures containing 10-20 cells. To achieve average cytoplasmic signal intensities, the signal from the DAPI-labelled nuclear regions was subtracted from the total signal of the whole image and divided by the number of cells present in the picture. Other immunofluorescence images were acquired using either a Zeiss LSM 780 Confocal Microscope or an Echo Revolve Fluorescence Microscope.

Metaplot and motif enrichment

For the CLIP binding metagenes, the 2 kB up- and downstream intronic sequences flanking the cryptic exons were also obtained. For hnRNPM-regulated cryptic exons (CryEx), alternatively spliced exons, and annotated background exons, we padded their coordinates with 2 kB up- and downstream. We then intersected them with the hnRNPM iCLIP binding sites using bedtools. The "% exon" represents the number of exons containing an overlapping binding site at a given position over the total number of exons. The equation was as follows:

$$\%\ exons_{position\ =\ i}\ =\ \frac{\#\ of\ exons\ with\ clip\ at\ position\ =\ i}{total\ \#\ of\ exons} \times 100\%$$

The binding probability (% exons) was further smoothed by using a moving average of 500-bases.

The motif score is defined as the mean percentage of bases containing any of the hnRNPM motifs in a 50-bases window.[97] The motifs were = ['UGUGU', 'GUGUG', 'UUGUG', 'GUGUU', 'UGUUG', 'UGUGG', 'GUUGU', 'GGUGU', 'UGGUU', 'UUGGU', 'UGGUG', 'GUGGU', 'GUUGG', 'GGUUG'].

The score for a given position is the percentage calculated for the window ending at that position. The equation is as follows:

$$\mathbf{motif\ score_{position\ =\ i}}\ =\ \frac{\sum_j^n \#\ \mathbf{of\ bases\ overlaping\ motifs\ between\ position\ i\ -\ 50\ and\ i\ for\ exon\ j}}{\mathbf{50 \times n}} \times \mathbf{100}\%$$

Spliceosomal protein binding density

For the ENCODE eCLIP database on the binding of U2AF2 (3' splice sites) and PRPF8 (5' splice sites) at cryptic exons. We quarried U2AF2 binding peak in a 200 nt window, 150 nt upstream of 3' splice sites and 50 nt downstream of 3' splice sites. We identified robust binding of U2AF2 binding at annotated exons. By contrast, no U2AF2 binding peaks were identified at the cryptic exons. Analysis of PRFP8 also revealed no binding at the cryptic exons but robust binding at annotated exons.

Bar plot showing U2AF2 binding within a 200 nt window (150 nt upstream and 50 nt downstreatm) of the 3' splice site of cryptic exons and their immediate flanking exons in non-specific shRNA-expressing Control (Blue) and hnRNPM KD (Orange) cells.

RNA structure prediction

RNA structure prediction was done using the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) with default options. A script for generating RNA secondary structure images as well as the maximum absolute minimum free energy | MFE| values can be found at https://github.com/CC-Cheng-Splicing-lab-BCM/hnRNPM_CryEx_dsRNA. The MFE, expressed as negative kcal/mol, is a readout of RNA structure stability. In Unfold output, MFE values are presented as absolute values (|MFE|), which is converted as a positive value.

ISG score calculation

ISG gene signature was curated using MsigDB hallmark interferon alpha and beta gene sets (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp) and ssGSEA was conducted using ISG gene signature and its calculated score is represented as ISG score. The higher the ISG score the higher the interferon responses in cells. A script for calculating ISG scores can be found at https://github.com/CC-Cheng-Splicing-lab-BCM/hnRNPM_CryEx_dsRNA. hnRNPM-low and hnRNPM-high indicate first and fourth quantiles of normalized hnRNPM expression distribution in each TCGA cancer.

Immune infiltration score calculation

The infiltration levels of different immune cell populations were estimated by ssGSEA[71] in the R Bioconductor package Gene Set Variation Analysis (GSVA, v1.44.3) using default parameters. Gene sets from published xCell paper[75] were curated and used to derive ssGSEA scores. Calculated ssGSEA scores are used to represent immune infiltration scores. *P*-values were calculated between first and fourth quantiles of MED15 cryptic exon inclusion level distribution in Figure 6D and between high and low risk score tumors separated by mean in Figures 6E and S6D. After TPM normalization of the gene expression matrix for each TCGA cancer, feed each TPM normalized matrix to TIMER2.0 website (http://timer.cistrome.org) and retrieve immune infiltration estimation with TIMER[94] method for further immune cell infiltration comparisons.

Clinical and molecular data of TCGA study

Gene expression data was downloaded from the GDC portal (https://portal.gdc.cancer.gov). After filtering out low abundance genes that have less than 5 reads across half of the samples in all TCGA cancer types, we applied $\log_2$-transformed counts per million (CPM) normalization and the trimmed mean of M-values (TMM) adjustment on count matrix for each cancer. We then defined 1st quartile of hnRNPM expressed tumors as the hnRNPM-low tumors, and 4th quartile of hnRNPM expressed tumors as the hnRNPM-high tumors.

Multivariate survival model and risk score calculation

Four representative dsRNAs containing cryptic exons from LRP11, MED15, RBM34, and TRAPPC10 genes were chosen for multivariate survival analyses. Clinical outcome endpoints data were retrieved from TCGA Clinical Data Resource (CDR).[87] By sliced BAM files of cryptic splicing regions in these genes from TCGA, we calculated PSI values for these cryptic exons for each tumor across cancer types: BRCA-Basal, DLBC, GBM, LUSC, PRAD, and UVM. We then fit a multiple Cox regression model with the above-mentioned four cryptic exon PSIs and hnRNPM expression levels. Risk scores were extracted from the model and assigned to each tumor. Specifically, we used ‘‘survival’’ R package to perform the above analysis. We used coxph function to build the multi-variate model and the predict function with ‘‘risk’’ type parameter to calculate the risk score. Subsequently, all samples from each cancer were divided into high- or low-risk groups using risk score means as the cutoffs. A generalized script can be found at https://github.com/CC-Cheng-Splicing-lab-BCM/hnRNPM_CryEx_dsRNA.

QUANTIFICATION AND STATISTICAL ANALYSIS

All data are presented as mean ± standard error of the mean,[60] unless otherwise indicated. Value of n and other details are specified in figure legends. Statistical analyses were performed using either R or GraphPad Prism 10.1.1. Correlation was assessed using Pearson correlation. Statistical significance tests included Fisher’s exact tests and hypergeometric tests. p-value < 0.05 was considered statistically significant. $p < 0.05$(*), $p < 0.01$ (**), $p < 0.001$ (***), $p< 0.0001$ (****) where indicated.