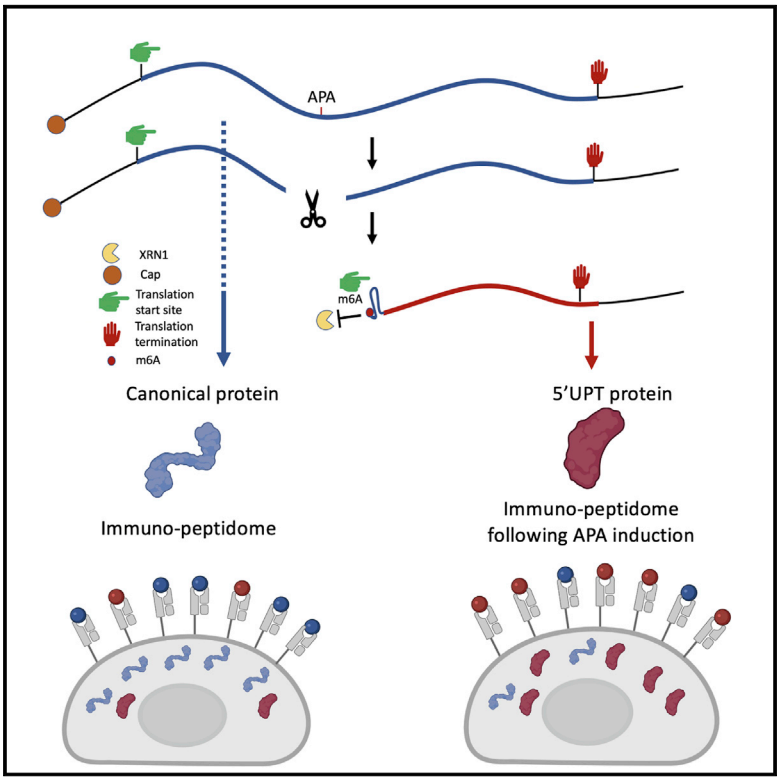
Molecular Cell

Article

Alternative cleavage and polyadenylation generates downstream uncapped RNA isoforms with translation potential

Graphical abstract



Highlights

- The human transcriptome contains thousands of uncapped poly(A) RNA transcripts (5′ UPTs)
- Alternative cleavage and polyadenylation (APA) is a major driver of 5′ UPTs biogenesis
- 5′ ends of 5′ UPTs are characterized by structured RNA and m6A modification
- 5′ UPTs CAP-independent translation alters the proteome and immunopeptidome repertoire

Authors

Yuval Malka, Ferhat Alkan, Shinyeong Ju, ..., Ran Elkon, Cheolju Lee, Reuven Agami

Correspondence

yuvalmalka79@gmail.com (Y.M.), r.agami@nki.nl (R.A.)

In brief

Malka et al. discovered, using various transcriptome and proteome techniques, a hidden layer of uncapped mRNA isoforms, which results from endonuclease cleavage and polyadenylation. These truncated variants have translation capacity, which globally contributes to the proteome and immunopeptidome diversity and explicitly regulates the translation of Bcl2 in a CAP-independent manner.

Malka et al., 2022, Molecular Cell *82*, 3840–3855
October 20, 2022
https://doi.org/10.1016/j.molcel.2022.09.036

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

Alternative cleavage and polyadenylation generates downstream uncapped RNA isoforms with translation potential

Yuval Malka,[1,*] Ferhat Alkan,[1] Shinyeong Ju,[2] Pierre-Rene Körner,[1] Abhijeet Pataskar,[1] Eldad Shulman,[3] Fabricio Loayza-Puch,[1] Julien Champagne,[1] Casper Wenzel,[1] William James Faller,[1] Ran Elkon,[3] Cheolju Lee,[2,4] and Reuven Agami[1,5,6,*]

[1]Division of Oncogenomics, Oncode Institute, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, Amsterdam, the Netherlands
[2]Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea
[3]Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv, 69978, Israel
[4]Division of Bio-Medical Science & Technology, KIST School, Korea University of Science and Technology, Seoul 02792, Republic of Korea
[5]Erasmus MC, Rotterdam University, Rotterdam, the Netherlands
[6]Lead contact
*Correspondence: yuvalmalka79@gmail.com (Y.M.), r.agami@nki.nl (R.A.)
https://doi.org/10.1016/j.molcel.2022.09.036

SUMMARY

The use of alternative promoters, splicing, and cleavage and polyadenylation (APA) generates mRNA isoforms that expand the diversity and complexity of the transcriptome. Here, we uncovered thousands of previously undescribed 5′ uncapped and polyadenylated transcripts (5′ UPTs). We show that these transcripts resist exonucleases due to a highly structured RNA and N6-methyladenosine modification at their 5′ termini. 5′ UPTs appear downstream of APA sites within their host genes and are induced upon APA activation. Strong enrichment in polysomal RNA fractions indicates 5′ UPT translational potential. Indeed, APA promotes downstream translation initiation, non-canonical protein output, and consistent changes to peptide presentation at the cell surface. Lastly, we demonstrate the biological importance of 5′ UPTs using Bcl2, a prominent anti-apoptotic gene whose entire coding sequence is a 5′ UPT generated from 5′ UTR-embedded APA sites. Thus, APA is not only accountable for terminating transcripts, but also for generating downstream uncapped RNAs with translation potential and biological impact.

INTRODUCTION

For most cellular mRNAs, the first step of mRNA translation involves the recognition of the 5′ CAP, a 7-methylguanosine (m7G) modification generated by the eukaryotic initiation factor 4E (eIF4E). 5′ CAP-bound eIF4F recruits the small (40S) ribosomal subunit associated with various translation initiation factors, enabling the efficient translation of eukaryotic mRNAs (Siddiqui and Sonenberg, 2015). Under various biological conditions, the cells can use an alternative mechanism for protein production in a 5′ CAP-independent manner (Shatsky et al., 2018). One such translation pathway is mediated by N^6-methyladenosine (m6A). Under heat shock stress, increased levels of YTHDF2, a YTH m6A RNA-binding protein, stimulate its translocation to the nucleus where it binds m6A in the 5′ UTR of a subset of stress-induced mRNAs and stimulates their CAP-independent translation (Zhou et al., 2015). 5′ CAP-independent, m6A-mediated translation is achieved by direct binding of the eukaryotic initiation factor 3 (eIF3), which is sufficient to recruit the 43S complex to initiate translation in the absence of the CAP-binding factor eIF4E (Meyer et al., 2015). Another well-characterized machinery for 5′ CAP-independent translation is the presence of highly structured internal ribosome entry sites (IRESs) (Jang et al., 1988; Pelletier and Sonenberg, 1988). These RNA structural elements have been found and validated mainly in viruses. However, the presence of IRES complex structures is rarely found in eukaryotic mRNAs undergoing 5′ CAP-independent mRNA translation (Stoneley and Willis, 2004).

Several transcriptomic and proteomic studies from recent years have pinpointed the expression of thousands of previously unannotated RNAs containing potential protein coding sequences (Ingolia et al., 2011, 2014; Kim et al., 2014). Many of these unannotated RNAs can potentially drive non-canonical protein production from embedded open reading frames (ORFs) and, by that, add proteins to the canonical proteome (Ruiz Cuevas et al., 2021). Interestingly, m6A modification preferentially induces mRNA translation at the first acceptable start codon, mainly at 5′ UTRs of full-length mRNAs, but fails to promote protein production from internal ORF sites (Meyer et al., 2015) even though ~35% of m6A modifications are located

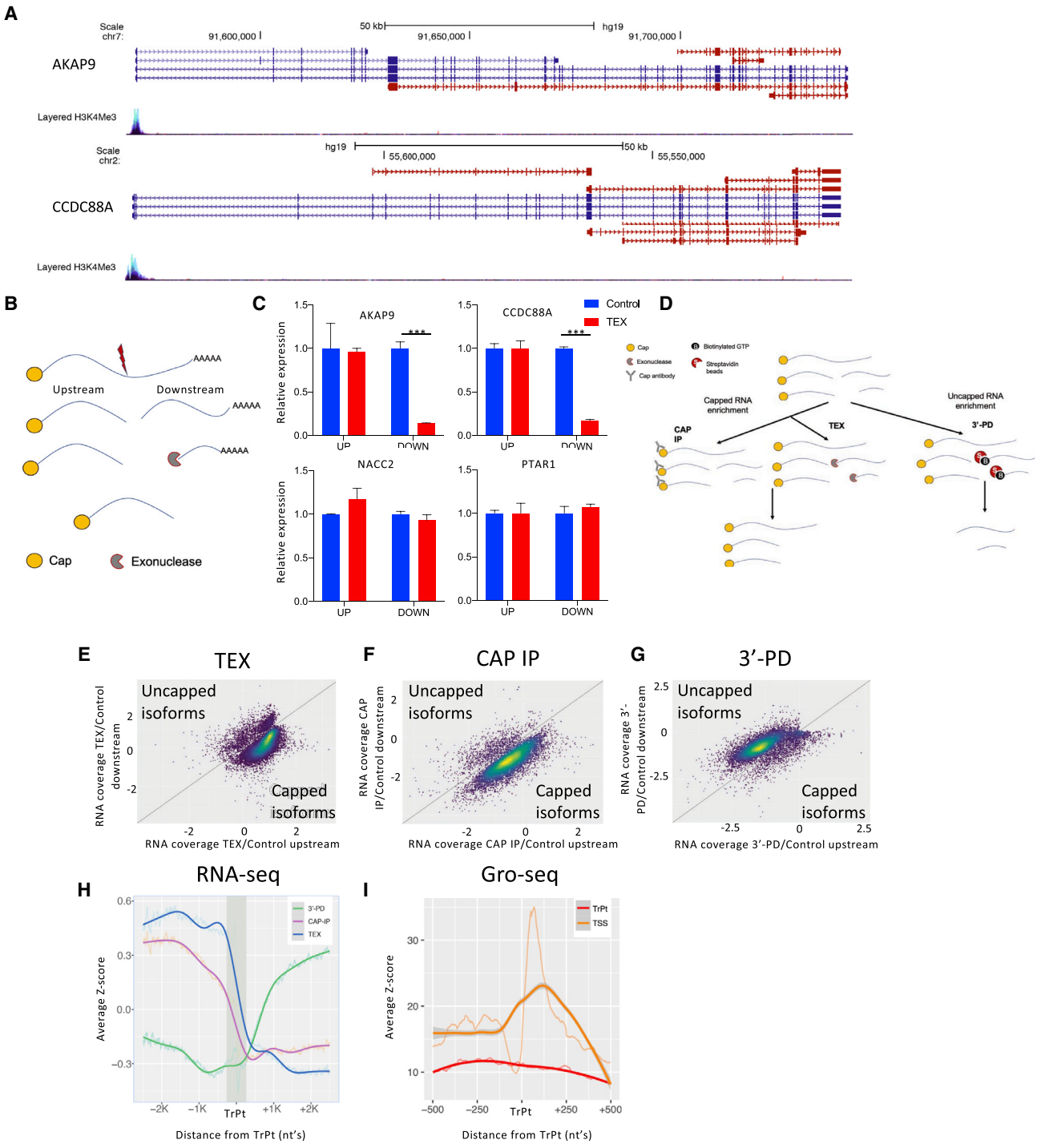


Figure 1. Vast expression of uncapped polyadenylated transcript isoforms

(A) AKAP9 and CCDC88A show multiple truncated variants (in red) with no histone methylation marks.

(B) A model for post-transcriptional processing of mRNA. mRNA is cleaved to generate a capped upstream and uncapped downstream isoform that is sensitive to TEX terminator 5′ phosphate-dependent exonuclease treatment.

(C) qRT-PCR analysis of upstream and downstream regions of AKAP9 and CCDC88A demonstrates high sensitivity to TEX treatment at the downstream region. ***$p < 0.001$ (two-tailed Student's t test).

(D) Poly(A)-selected RNA from U2OS cells was enriched for 5′ capped upstream isoforms using TEX treatment (TEX) or anti-CAP immunoprecipitation (CAP IP); uncapped downstream RNA isoforms were enriched by streptavidin bead pulldown of *in vitro* biotinylated-7-methylguanylate capped RNA (3′-PD).

(legend continued on next page)

within coding sequence regions (CDSs). Hence, while upstream ORFs are recognized and translated by scanning ribosomes before they reach the canonical ORFs, it is unclear how internal translation initiation occurs and is widely used in eukaryotes. Therefore, the initiating events of many translation incidents in eukaryotes is poorly understood, pinpointing a hidden mechanism of gene regulation.

In this study, we used various molecular and genomic methods to demonstrate the expression of thousands of mRNA isoforms with uncapped 5′ termini. These RNA fragments appear downstream of alternative cleavage and polyadenylation (APA) sites embedded in their coding and intron sequences and appear to resist exonuclease activity due to double-stranded RNA structures and m6A RNA modifications at their 5′ end. The biological relevance of these transcripts is manifested by their translation potential and their impact on the proteome and immunopeptidome.

RESULTS

Vast expression of uncapped polyadenylated transcript isoforms

The Encyclopedia of DNA Elements (ENCODE) project annotated widely prevalent truncated mRNA variants (Djebali et al., 2012). Some internal mRNA isoforms are generated by alternative promotors and can have a crucial biological function, such as the cell-cycle arrest induced by the expression of an alternative isoform of MDM2 following p53 activation by oncogenes and DNA damage (Barak et al., 1994; Pamudurti et al., 2017). Additional mechanisms of isoform generation are mRNA splicing, which results in the exclusion or inclusion of alternative exons and the expression of circular RNAs (circRNAs) by back-splicing (Pamudurti et al., 2017), and APA, which is responsible for the production of short mRNA isoforms (Lee et al., 2018; Singh et al., 2018). These too were shown to markedly contribute to cellular functionality (Elkon et al., 2013; Kristensen et al., 2019). Interestingly, none of these mechanisms can explain the vast expression of truncated polyadenylated mRNA isoforms. Figure 1A indicates two such examples, CCDC88A and AKAP9. The indicated isoforms (marked red) lack an adjacent annotated promoter and cannot be explained by alternative splicing. The presence of a poly(A) tail excludes the possibility of circRNAs.

Because the indicated isoforms are not adjacent to known promotors, their biogenesis may occur due to canonical transcript truncation. If true, the N7-methylguanosine (m7G) (CAP) added to the 5′ ends of the first exon of newly transcribed RNAs will be lost in these isoforms. Hence, we tested the presence of a CAP at the 5′ termini of mRNAs by treating RNA extracts with TEX, a terminator 5′ phosphate-dependent exonuclease that degrades uncapped RNA molecules (Figure 1B) (Malka et al., 2017). We first evaluated the TEX-induced changes in the presence of upstream and downstream regions of two genes with multiple downstream mRNA variants (CCDC88A and AKAP9) compared with control genes (NACC2 and PTAR1) that do not present multiple downstream variants (Figures 1A and S1A). Figure 1C shows that TEX treatment dramatically reduced the level of the downstream region in both test cases, while the upstream regions remained, as expected, mostly intact. In contrast, the levels of the upstream and downstream regions of the control genes remained mostly unaltered. We confirmed this result by RNA sequencing (RNA-seq) (Figure S1B). Altogether, we present here indications that some genes express 5′ uncapped polyadenylated transcript isoforms, which we named here 5′ UPTs.

Next, we assessed the abundance of 5′ UPTs. We combined three high-throughput techniques to distinguish between 5′ UPTs and the canonical capped transcripts (Figure 1D). We purified polyadenylated RNAs from U2OS cells and either enriched for capped transcripts using immunoprecipitations (IPs) with an anti-CAP antibody (CAP-IP), treated with TEX to degrade 5′ UPTs, or enriched for 5′ UPTs using an *in vitro* capping system with biotinylated guanosine triphosphate followed by pulldown of the newly capped fragments using streptavidin beads (3′-PD). We subjected the treated samples to RNA-seq analysis (e.g., Figure S1C). We inferred that prominent expression of 5′ UPTs should reduce read coverage along transcripts following CAP-IP and TEX treatments while causing an increase in the 3′-PD analysis when compared with untreated RNA samples. We used a computational algorithm that scans the overall normalized read coverage along transcripts to identify transition points to higher/lower RNA levels within transcripts (Gaussian Hidden Markov Model [HMM]) (Malka et al., 2017), followed by the Viterbi algorithm to distinguish the most probable transition points in read abundance for each transcript, and lastly the Kolmogorov-Smirnov test to determine statistical significance. We assumed that the identified significant transition points (TrPts) marked start sites of 5′ UPTs. To validate the reproducibility of the predicted TrPts, we analyzed each TEX-treated sample independently and found a strong overlap between all three samples (Figure S1D). Overall, TEX treatment identified 7,665 genes with a statistically significant TrPts in their coding sequence, CAP-IP 7,270 genes, and 3′-PD 7,780 genes (Figures 1E–1G and Tables 1-S3). This resulted in 3,192 common genes showing a significant coordinated change in reading coverage in all three treatments (41%–44% of the genes in either group) (Figure S1E). In addition, all three treatments showed either a sharp RNA decrease (TEX [57%], CAP-IP [39%]) or

(E–G) Computational analysis using a Hidden Markov Model (HMM) was applied to identify the most probable cleavage point (TrPt) for each transcript in the CDS region for TEX (overall of 7,665 cleaved genes, FDR < 0.01); CAP IP (7,270 genes); and 3′-PD (7,780 genes). Shown are density plots comparing the relative read coverage before and after treatment for upstream (x axis) and downstream (y axis) for TEX (C), CAP IP (D), and 3′-PD (E).

(H) For TEX, CAP IP, and 3′-PD treatments, we compared the local RNA-seq read coverage around TrPt sites with untreated control RNA-seq and plotted the average read coverage deviation (*Z* score of logFC values) at each position around the TrPt predictions. TEX, CAP IP, and 3′-PD treatments present a sharp deviation around the TrPt sites, including elevated read coverage for 5′ UPTs following 3′-PD and weaker coverage following TEX treatment and CAP IP (fitted lines are shown with original values in faded color).

(I) The average frequency of RNA polymerase II around canonical TSS for genes with predicted CDS TrPts in MCF7 cell line. In red, CDS TrPts; in yellow, canonical TSS (fitted lines are shown with original values in faded color).

increase (3′-PD) around TEX TrPt, pinpointing a common defined location of the predicted 5′ UPT start site (Figure 1H).

We further tested the stability, independence, and conservation of 5′ UPTs. To examine whether 5′ UPTs are stable and independent RNA molecules and not intermediate degradation fragments, we performed RNA-seq experiments in cells treated with α-amanitin (an RNA polymerase II- and III-specific inhibitor) for 6 h. HMM analysis after TEX treatment identified 4,792 genes with TrPts, indicating that 5′ UPTs are present long after transcription was inhibited (Figure S2A and Table S4). We also validated the existence of 5′ UPTs using full-length sequencing of native RNA strands by long-read Oxford Nanopore Technologies, a technology that highlights truncated mRNA isoforms that utilize an alternative first exon (Soneson et al., 2019). HMM analysis of this data indicated a significant reduction in the presence of truncated variants following TEX treatment (2,886 genes with TrPts) (Figure S2B and Table S5). Two examples are shown in Figure S2C. Lastly, we examined 5′ UPTs across diverse human cell lines. We extracted polyadenylated RNA from HeLa and HEK293 cells, mock or treated with TEX, and subjected them to deep sequencing and bioinformatics analysis as above. We identified 6,881 genes in HeLa and 7,518 in HEK293 cells containing TrPts (Figures S2D–E and Tables S6–S7, respectively). Interestingly, 78%–80% and 41%–46% of the identified TrPts were common to two or all cell lines, respectively (Figure S2F). To test the possibility that 5′ UPTs are transcription products derived from alternative promotors, we first searched for an epigenetic signature of H3K4Me3 in the proximity of predicted TrPts. We used TrPt sites of 5,525 genes derived from an HMM analysis following TEX treatment in MCF7 cells (Table S8). While the expected H3K4Me3 signal strongly appeared in the surrounding of canonical transcription sites (+/−500 nt), TrPts did not present any enrichment (Figure S2G). Second, we used global run-on sequencing (GRO-seq) of MCF7 cells and searched for accumulation of nascent transcription around TrPt sites that would indicate transcription start sites. While active promoters showed the expected accumulation in nascent RNA, TrPts did not (Figure 1I). Instead, nascent RNA abundance exhibited a stable and continuous distribution upstream and downstream to TrPts, indicating that 5′ UPTs are indeed processed RNA products cleaved from canonical transcripts.

Altogether, our results reveal a hidden layer of promotor-independent truncated uncapped RNA transcripts widely expressed in thousands of genes.

5′ UPTs are marked with a secondary RNA structure and m6A at their 5′ end

CAP modification of mRNAs at their 5′ termini is partly created to protect them from RNA exonucleases such as XRN1/2 (Furuichi et al., 1977). Hence, uncapped mRNAs are considered unstable. To test whether the observed 5′ UPTs are sensitive to XRN exonucleases, we treated Poly(A)-selected RNA extracts *in vitro* with purified XRN1 and compared its effect on RNA coverage with that of TEX-treated samples. While XRN1 and TEX had a comparable impact on the naturally uncapped mitochondrial RNAs (Figure 2A), XRN1 did not affect the coverage of nuclear-encoded genes (Figures 2B–C and Figure S3A). This implies the presence of a protective mechanism for the 5′ ends of 5′ UPTs.

Naturally uncapped RNAs, such as rRNAs, tRNAs, and some viruses, utilize various mechanisms to protect themselves from exonuclease decay, mostly based on RNA secondary structure (Akiyama et al., 2016; Chapman et al., 2014; Chernyakov et al., 2008). We therefore interrogated whether 5′ UPTs possess such properties that may protect them from 5′ end nuclease activity. First, we interrogated RNA structure *in vivo* and *in vitro* using datasets of click-selective 2′-hydroxyl acylation and profiling experiments (icSHAPE) (Lu et al., 2016; Spitale et al., 2015). icSHAPE captures RNA secondary structures at a transcriptome-wide level by measuring nucleotide flexibility for all four bases (Figure S3B). Interestingly, both *in vitro* and *in vivo* datasets revealed a sharp transition from single-stranded RNA to lower flexibility structured double-stranded RNA at the TrPt sites (Figures 2D and 2E), indicating structured RNA at their 5′ termini of many 5′ UPTs.

RNA secondary structures are determined by nucleotide composition and RNA modification (Liu et al., 2018; Roost et al., 2015; Zubradt et al., 2017). We therefore examined the frequency of various nucleotides surrounding TrPts and observed adenosine as the most frequent nucleotide at their 5′ ends (Figure S3C). As m6A is the most abundant RNA modification in eukaryotes, we searched for association with TrPt expression. Intriguingly, we observed enrichment of the canonical m6A motif (GGAC) around TrPt sites (Figure 2F). To further explore this point, we compared TrPt with known m6A sites using RMBase v2.0 (Xuan et al., 2018). While m6A topology often exhibits enrichment near stop codons (Dominissini et al., 2012; Meyer et al., 2012), we found a strong overlap between m6A sites and TrPts and 5′ UPT isoforms within CDSs (~77% out of 6,881 genes within 100 nt) (Figure 2G and Figure S3D).

The HMM analysis used above to identify TrPt sites is a statistical model that approximates the actual cleavage of mRNAs. Moreover, RNA-seq is conducted following a reverse transcription reaction, which may generate a bias in product prevalence due to RNA secondary structure and modified nucleotides. To circumvent these drawbacks and to study the nature of the 5′ terminus of 5′ UPTs at a single nucleotide resolution, we performed 5′ end sequencing (5′ seq) of uncapped RNAs (Figure 2H) (Pelechano et al., 2016). Indeed, 5′ end sites of uncapped RNAs globally correlated with TrPts, with adenosine being the most frequent first nucleotide (Figures S3E and S3F). Comparing the first adenosine from the 5′ seq reads with the RMBase v2.0 m6A database confirmed their m6A annotation (Figures 2I and S3G). m6A sites can functionally impact mRNA structures and stability (Roost et al., 2015). However, $ALKBH^{KD}$ did not globally affect 5′ UPT expression (Figure S3H, HMM analysis for control and $ALKBH^{KD}$, Tables 9–S10, respectively), suggesting a minor contribution to 5′ UPT stability. Altogether, we conclude that a structured RNA at the 5′ end of 5′ UPTs potentially explains their relative stability.

Translation potential of 5′ UPTs

Ribosome profiling (RP) experiments from recent studies have uncovered a surprisingly large variety of internal translation initiation sites and truncated ORFs in addition to the canonical ones

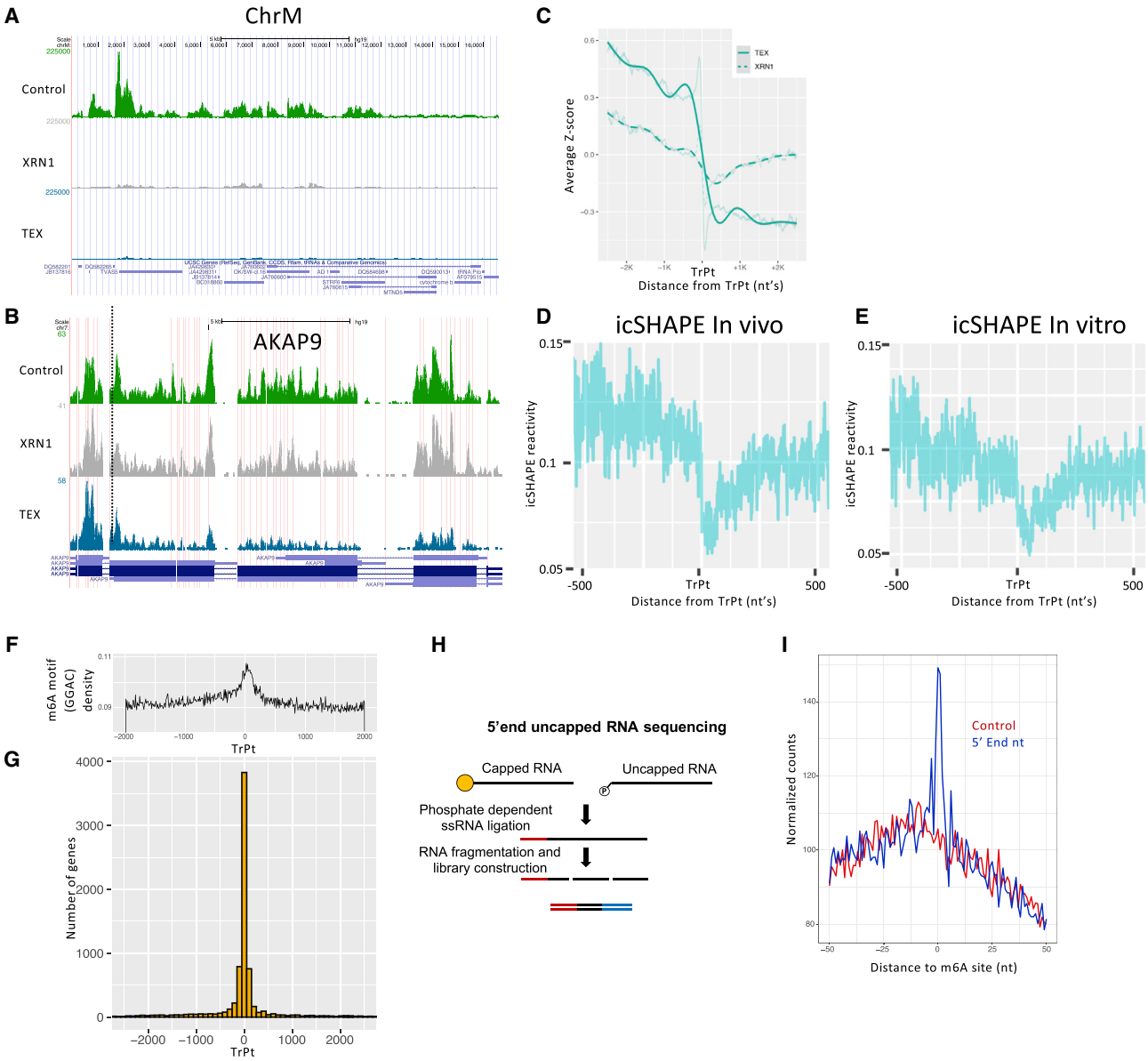


Figure 2. 5′ ends of 5′ UPTs are structured and enriched with m6A modification

(A and B) RNA-seq of mitochondrial chromosome and AKAP9 in control and after TEX/XRN1 treatments.

(C) Differential local read coverage analysis (*Z* score) surrounding the TrPt sites in RNA extracts treated with TEX/XRN1 vs. control (fitted lines are shown with original values in faded color).

(D and E) Analysis of icSHAPE data (*in vivo* and *in vitro*) present sharp transition from single-stranded RNA (higher average reactivity score) to structured double-stranded RNA downstream to TrPt sites.

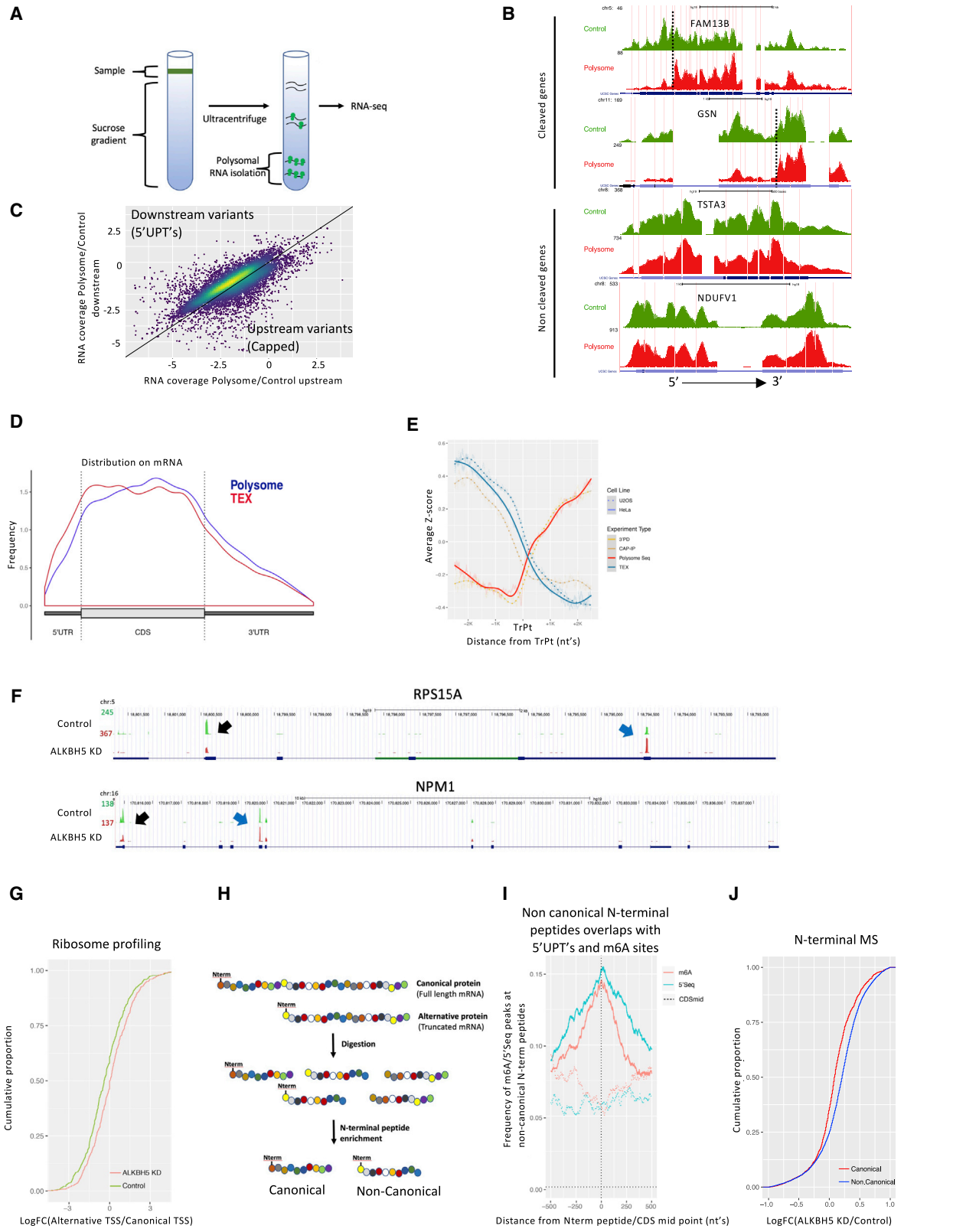
(F) Density plot of m6A-conserved motif (GGAC) shows high enrichment across TrPt sites. Distance from predicted TrPt in nts.

(G) Histogram of shortest distances (nts) from TrPt sites of 6,882 genes to known m6A sites (RMBase v2.0).

(H) A scheme depicting the sequencing strategy used to capture 5′ uncapped RNA. RNA linker (red) is directly ligated to 5′ phosphate uncapped RNA followed by RNA fragmentation, 3′ linker (blue), and PCR amplification.

(I) Overlapping m6A sites (RMBase v2.0) with 5′ end sequencing reads. Control positions were randomly selected from non-m6A adenosine positions from the same exons of known m6A sites.

(Chen et al., 2020; Ingolia et al., 2011; Kearse and Wilusz, 2017). However, this approach relies on the isolation of ribosome-protected fragments and thus fails to provide the full mRNA sequence context in which these non-canonical translation initiation sites occur. Alternatively, polysome fractionation, where RNAs loaded with ribosomes migrate slower in a sucrose gradient, is independent of RNAse treatment and can provide such knowledge (Figure 3A). Thus, to test whether 5′ UPTs can



(legend on next page)

serve as templates for mRNA translation, we performed polysome fractionation coupled to polyadenylated RNA-seq. As above, we compared read distribution along transcripts with total polyadenylated RNA using HMM analysis. Overall, we identified polysome-bound transcripts of 6,576 genes (Table S11) with statistically significant TrPts in their CDS. Interestingly, the vast majority of the identified genes showed increased RNA read coverage downstream of TrPts (Figures 3B and 3C). As control, we selected groups of genes with or without TrPt in their CDS and analyzed RNA read distribution compared with total polyadenylated RNA. Figures S4A and S4B show a slight increase in the read coverage along the CDS only in the group of genes with TrPts. The slight effect is expected as the data is not TrPt centralized.

Next, we crossed TrPts extracted from the TEX treatment and the polysome profiling procedure. We found 61%–64% overlapping genes, corresponding to 4,178 genes ($p < 9.428 \times 10^{-279}$) (Figure S4C). While TEX treatment enriched for capped mRNAs, polysome fractionation presented a mirror effect, indicating enrichment of 5′ UPTs in the heavy polysomal fractions (Figure 3D). We then analyzed the average read coverage around TrPts for the various treatments. While TEX and CAP-IP treatments exhibited a decrease in RNA coverage up to the TrPts and then stability downstream, polysomal RNA and 3′-PD data showed a sharp increase downstream of the TrPts (Figure 3E). Additionally, over half of TEX TrPts overlapped with polysomal RNA TrPts within a 350-nt window (Figure S4D). These results suggest a link between 5′ UPTs and protein production.

The presence of 5′ UPTs at heavy polysome fractions implies a CAP-independent translation mechanism. Recent studies have shown that mRNAs containing m6A at their 5′ end can recruit the 43S ribosomal subunit complex to initiate translation without the CAP-binding factor eIF4E (Coots et al., 2017; Meyer et al., 2015). Because our results demonstrated high enrichment of m6A at the 5′ end of 5′ UPTs, we asked whether this may initiate translation in a CAP-independent manner. To identify potential translation-start sites (TSSs) within 5′ UPTs, we used ribosome profiling (Ribo-seq) of cells treated with the initiation-specific translation inhibitor harringtonine. Evaluating internal alternative TSS usage revealed a significantly higher appearance in ALKBH5^{KD} than in control cells (Figures 3F and 3G). This observation is in line with a role of m6A in initiating mRNA translation from 5′ UPTs.

To further substantiate the translation potential of 5′ UPTs, we performed N-terminomics (N-pep), a method based on the enrichment of endogenous N-terminal peptides (Figure 3H). We selected the non-canonical N-terminal peptides (>50 amino acids downstream to the canonical TSS) and asked whether they are linked to the distribution of 5′ UPTs and m6A along the CDS. Figure 3I shows a strong overlap between uncapped 5′ end sequence and m6A sites (RNA immunoprecipitation sequencing [RIP-seq] HeLa cells) (Lin et al., 2019) and non-canonical N-terminal peptides. In addition, both *in vitro* and *in vivo* icSHAPE analyses demonstrated a sharp transition to lower flexibility structured double-stranded RNA at the N-terminal site of these peptides (Figure S4E), similar to TrPt sites shown previously (Figures 2D and 2E). We further analyzed changes in the distribution of N-terminal peptides in ALKBH5^{KD} cells. We confirmed the magnitude of ALKBH5^{KD} at the protein level (Figure S4F) and identified 10,481 N-terminal peptides (1,571 canonical and 8,910 non-canonical, Table S12). In line with our Ribo-seq data (Figure 3G), ALKBH5^{KD} stimulated the expression of non-canonical N-terminal peptides compared with canonical ones (Figure 3J). To exclude the possibility that the elevated non-canonical N-terminal peptides expression resulted from higher expression of 5′ UPTs in ALKBH5^{KD} cells, we compared their intensities with RNA levels, as determined by 5′ end sequencing. Figure S4G shows no significant difference between control and ALKBH5^{KD} reads at the non-canonical N-terminal peptides (Meyer et al., 2015). Thus, 5′ UPTs are engaged with initiating ribosomes and play a significant role in proteome diversity in an m6A-mediated, CAP-independent manner.

APA is a major driver of 5′ UPTs production

Next, we set to explore the most influential biogenesis pathway leading to the production of 5′ UPTs. Based on our observations, we postulated that 5′ UPTs are being generated through endonuclease activity. Previous work in mammalian cells produced widespread evidence for endonuclease cleavage events, but further investigation failed to connect them to Drosha, Dicer, and RNA-induced silencing complex (RISC) (Karginov et al.,

Figure 3. 5′ UPTs can be translated

(A) A scheme depicting the strategy for polysomal RNA isolation.

(B) RNA-seq of FAM13B and GSN shows enrichment in read coverage downstream of the dashed line in the polysomal RNA compared with the total polyA + RNA (green). TSTA3 and NDUFV1 representing genes with no such alterations.

(C) HMM analysis was applied for polysomal RNA vs. control total RNA data to identify CDS TrPts for each transcript (6,577 genes). Shown are density plots for upstream (x axis) and downstream (y axis) regions relative to the identified TrPts.

(D) Global RNA distribution of 4,178 genes with shared TrPts identified by TEX treatment and polysomal RNA analysis.

(E) *Z* score of TEX/CAP IP/3′-PD vs. control in U2OS cells, and TEX and polysomal RNA in HeLa cells around all statistically significant predicted TrPts (fitted lines are shown with original values in faded color).

(F) Ribo-seq of NPM1 and RPS15A uncovers TSS after harringtonine treatment in control (green) and ALKBH5^{KD} cells (red). Black arrows, canonical TSSs; blue arrows, alternative TSSs.

(G) Cumulative frequency distributions of alternative/canonical TSS expression in control and ALKBH5^{KD} cells (813 transcripts, two-sided Kolmogorov-Smirnov test [$p < 0.0001$]).

(H) Model of N-terminal peptides enrichment. Internal peptides were depleted, and N-terminal peptides were enriched and subjected to LC-MS/MS analysis.

(I) Peak calling analysis of m6A-RIP-seq data (red) (Lin et al., 2019) and 5′ end sequencing (blue) shows enrichment around internal N-terminal peptides, while coding-sequence midpoints from the same genes (dashed line) do not. To exclude canonical TSSs, we selected the non-canonical N-terminal peptides with a minimum distance of 50 amino acids downstream to canonical TSS.

(J) Cumulative frequency distributions of the log2 fold changes (ALKBH5^{KD}/control) for canonical N-terminal peptides and non-canonical (two-sided Kolmogorov-Smirnov test; p value $< 2.2 \times 10^{-16}$).

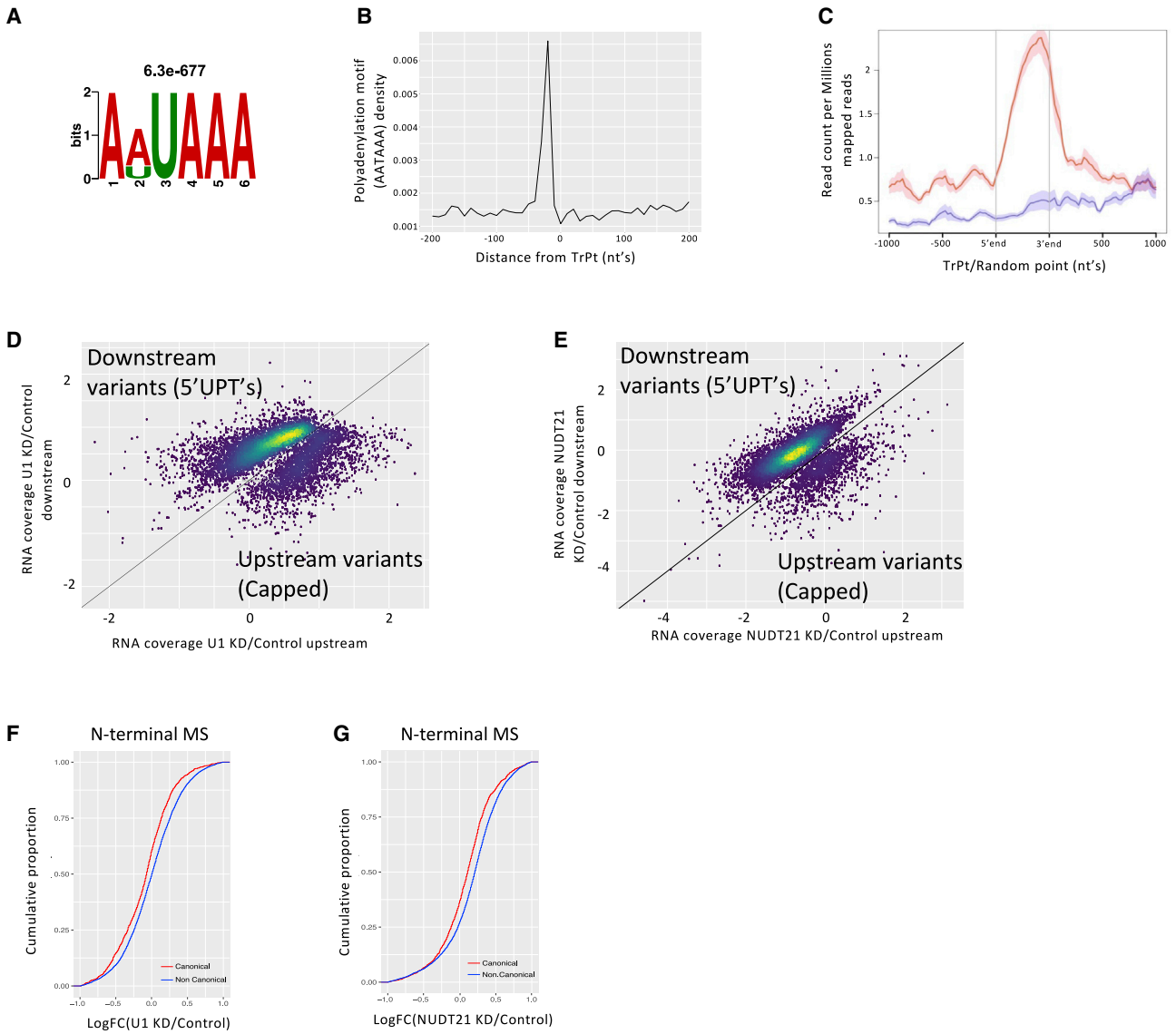


Figure 4. Generation of 5' UPTs by APA

(A) The regions +/−50 nt of TEX-treated TrPts were analyzed for enriched motifs using motif discovery algorithm DREME.

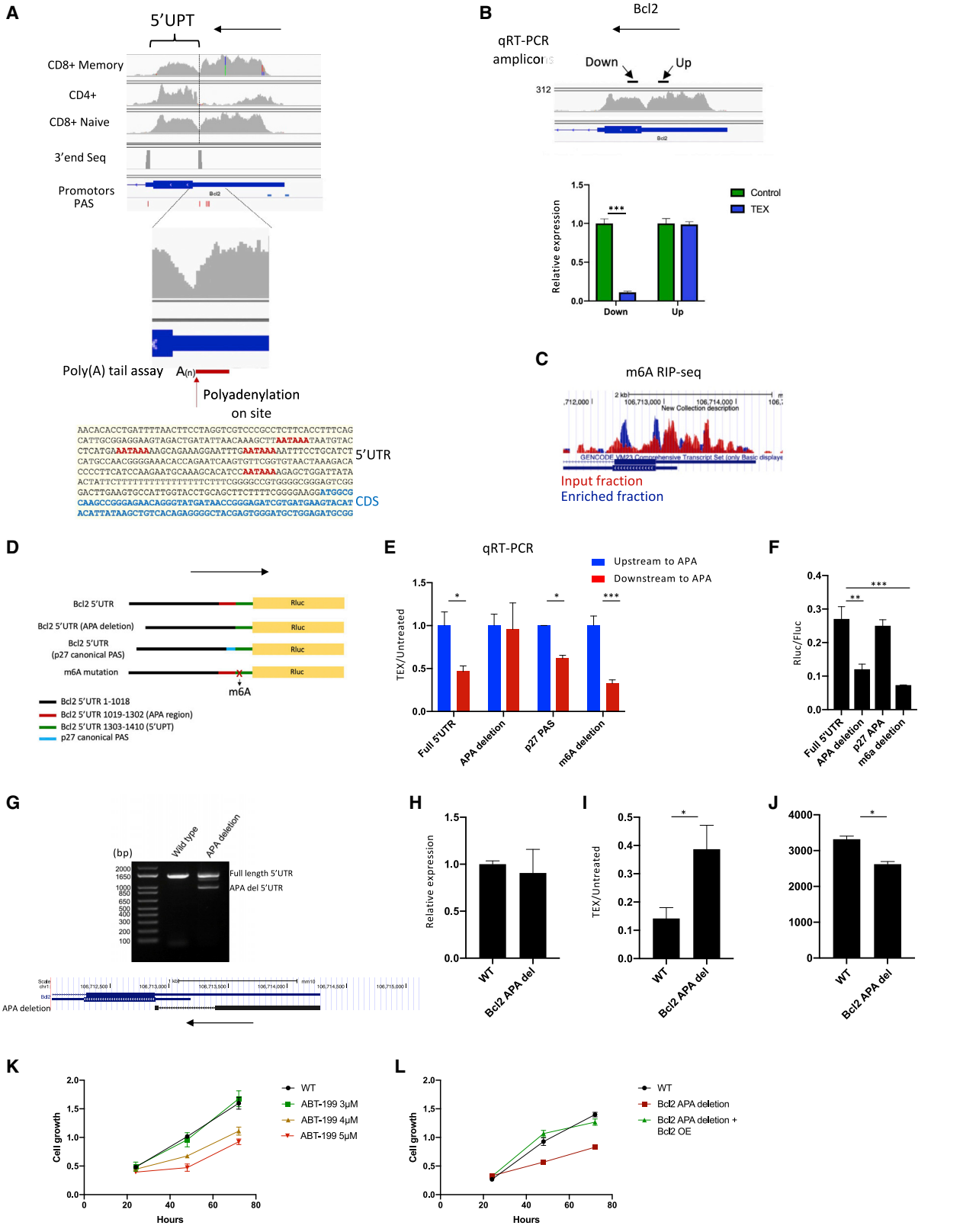
(B) Motif density plot of canonical polyadenylation signal AATAAA around TEX TrPts.

(C) Distribution of HeLa 3′ RNA-seq and random points surrounding TEX TrPts present accumulation of polyadenylation sites around TrPts.

(D and E) HMM analysis of RNA-seq data of NUDT21^{KD} (D) and U1^{KD} (E) vs. corresponding controls shows a high accumulation of downstream truncated isoforms following both knockdowns.

(F and G) Cumulative frequency distributions of the log2 fold changes (NUDT21^{KD}/U1^{KD} vs. corresponding control) for canonical N-terminal peptides and non-canonical N-terminal peptides (two-sided Kolmogorov-Smirnov test; NUDT21 p value < 2.2×10^{-16}, U1 p value < 5.529×10^{-12}).

2010). In addition, transcriptomic capped analysis of gene expression (CAGE) revealed widespread intra-exonoic capping that did not arise from conventional transcription initiation (Mercer et al., 2010, 2011). Interestingly, a consensus motif search analysis around TrPt sites identified A(A/T)TAAA (6.3×10^{-677}), a conserved polyadenylation signal (PAS) motif (Figure 4A), as the top hit. The PAS motif was enriched upstream of TrPts in CDSs, indicating that APA is likely a prominent upstream factor in the generation of 5′ UPTs (Figure 4B). To directly address this possibility, we performed 3′ end sequencing (3′ seq) of mRNAs that captures polyadenylation sites of mRNAs. Bioinformatics analysis of TrPt sites (11,052 genes with CDSs or 3′ UTR TrPts) showed a 29.1% overlap with active polyadenylation sites within +/−100-bp window, further supporting an overall correlation of 3′ ends with the identified TrPts (Figure 4C and Figure S5A), implicating the endo-nucleolytic cleavage activity at APA sites as a major principal event generating 5′ UPTs downstream of TrPts.



(legend on next page)

To experimentally evaluate the impact of APA usage on 5′ UPT biogenesis, we performed a series of knockdowns of the most prominent positive and negative APA regulators. Loss of either PCF11 or INTS11 (an APA termination factor and the endonuclease factor of integrator I, respectively) was recently shown to inhibit APA and promote 3′ UTR lengthening (Dasilva et al., 2021; Wang et al., 2019). We therefore knocked down PCF11 and INTS11 using small interfering RNA (siRNA) in HeLa cells (Figure S5B) and performed HMM and *Z* score analyses for each condition in TEX control and treated samples (Tables S13–S15 for control, PCF11^{KD}, and INTS11^{KD}, respectively, and Figure S5C–E). Figure S5D shows that PCF11^{KD} or INTS11^{KD} substantially reduced 5′ UPT expression. In parallel, we investigated the causal role of two of the most prominent nuclear APA factors, U1 and NUDT21, in the generation and translatability of 5′ UPTs. We used antisense morpholino oligos and siRNAs to knock down U1 and NUDT21 (Figure S5G) in HeLa cells (Oh et al., 2020). RNA-seq expression analysis (Tables S16–S17, respectively) revealed vast expression of truncated downstream-mRNA variants in both U1^{KD} and NUDT21^{KD} cells (Figures 4D and 4E). To link these mRNA variants with 5′ UPTs, we treated RNA extraction from each condition with TEX and analyzed by RNA-seq. We found increased TEX sensitivity by both U1^{KD} and NUDT21^{KD} (4,917 and 5,407 genes with TrPts in control morpholino and U1^{KD}, respectively [Tables S18–S19], 5,930 and 6,537 and genes for control siRNA and NUDT21^{KD}, respectively [Tables S20–S21]), supporting APA usage as a significant driver for 5′ UPTs biogenesis (Figure S5H–K). Comparative analysis of these data further substantiated this notion and showed higher sensitivity to TEX treatment in U1^{KD} and NUDT21^{KD} compared with their corresponding controls (Figure S5L–M).

Encouraged by the link between U1, NUDT21, and 5′ UPTs at the RNA level, we extended the analysis to N-terminal proteomics on cells with U1^{KD} and NUDT21^{KD} to evaluate their impact on protein production. We confirmed the magnitude of NUDT21 depletion (Figure S5N) at the protein level and found 9,141 and 11,264 (Tables S22–S23) N-term peptides in U1^{KD} and NUDT21^{KD}, respectively, that are shared with their specific control sample. We used this data to determine changes in the abundance ratio of non-canonical N-terminal peptides to canonical ones. Figures 4F and 4G show that both NUDT21^{KD} and U1^{KD} induced a global shift in the abundance of canonical N-terminal peptides to internal non-canonical compared with their respective control treatments. Altogether, the results presented here pinpoint APA usage as a major player in the biogenesis of 5′ UPTs and in the expression of truncated internal N-terminal proteomics.

Bcl2 is a 5′ UPT

Next, we wished to demonstrate the link between APA, m6A and 5′ UPTs, and the contribution of 5′ UPTs to gene expression using one prominent gene. The B cell lymphoma 2 (Bcl2) is a major apoptosis regulator involved in various diseases, including cancer (Youle and Strasser, 2008). Close inspection of Bcl2 mRNA reveals that its long 5′ UTR (1,410 bp) uniquely contains a cluster of several conserved and functional PASs (AATAAA) upstream to the start codon (Figure 5A). In line with this, analysis of several RNA-seq datasets from mouse T cells uncovered a noticeable cleavage point at the polyadenylation site and an unbalanced expression of the RNA levels from both sides of the cleavage region (Figure 5A). We therefore investigated Bcl2 RNA production downstream of the APA sites. We extracted RNA from mouse T cells and treated it with TEX. Indeed, while the expression of the 5′ UTR of Bcl2 was indifferent to TEX treatment, its coding sequence was vastly degraded (Figure 5B). These results indicate the unique situation that the entire Bcl2 protein expression originates from a 5′ UPT isoform. Interestingly, single-nucleotide-resolution mapping of m6A coupled with Ribo-seq showed that m6A promotes the translation of BCL2 in the human MOLM-13 cell line (Vu et al., 2017). We therefore examined the occurrence of m6A modification of Bcl2 mRNA using existing data of m6A RIP-seq from wild-type CD4^{+} T cells (Li et al., 2017). This identified an m6a site downstream to the APA sites and upstream to Bcl2-5′ UPT (Figure 5C). It was previously reported that CRISPR-Cas9-mediated disruption of Mettl3—a major

Figure 5. Bcl2 is a 5′ UPT controlled by APA sites at its 5′ UTR

(A) Different RNA-seq data from mouse T cells (CD8 memory, Sequence Read Archive: SRR1688717) (Russ, 2014); CD4, SRR1583932 (Wan, 2015); and CD8 naive, SRR1688691 (Russ, 2014) demonstrate a gap in read coverage (dashed line) and uneven RNA levels surrounding the gap site (middle sample) that is remote from promotor region (blue marks). In addition, 3′ seq (various tissues, SRP033205) (Nam, 2014) demonstrate active polyadenylation at the cleavage site downstream to a cluster of polyadenylation sequence motifs (AATAAA, red marks). Lower image (magnified); Poly(A) tail assay in mouse T cells shows an active polyadenylation site.

(B) qRT-PCR amplification upstream (Up) and downstream (Down) to the APA sites before and after TEX treatment (two-tailed Student's t test; ***$p < 0.001$).

(C) m6A RIP-seq data of wild-type CD4^{+} T cells (Li et al., 2017) (input SRR5735707, m6A-IP SRR5735709) shows m6A peak at the 5′ end of Bcl2-5′ UPT.

(D) pSTAR vectors (Slobodin et al., 2017) with either WT mouse Bcl2 5′ UTR (upstream to Renilla), deletion of PAS cluster region, substitution of the PAS cluster region with p27 canonical PAS, or mutation at m6A site were used to generate stable polyclonal Flp-In-293 cells.

(E) qRT-PCR analysis of the full-length Bcl2 5′ UTR, Bcl2 5′ UTR with the p27-PAS substitution, and Bcl2 5′ UTR with a mutation at the m6A site before and after TEX treatment (two-tailed Student's t test; *$p < 0.05$, ***$p < 0.001$).

(F) The relative Rluc/Fluc ratio (average ± SD) was determined in Flp-In-293 cells stably expressing the pSTAR vectors depicted in (H) (two-tailed Student's t test; **$p < 0.01$, ***$p < 0.001$).

(G) Genomic PCR analysis on the full length 5′ UTR of Bcl2 shows monoallelic deletion of the APA-containing region.

(H and I) qRT-PCR analysis of Bcl2 following 5′ UTR APA deletion shows no alteration in mRNA levels (H) but presents less sensitivity to TEX treatment (I) (two-tailed Student's t test; *$p < 0.05$).

(J) Median fluorescence intensity (MFI) of Bcl2 protein levels. Moderate reduction of protein levels results from the monoallelic deletion of APA site (data represent 3 independent experiments, two-tailed Student's t test; *$p < 0.05$).

(K and L) Cell growth was measured using quantification of crystal violet staining following treatment with different doses of ABT-199 (K) or different MEF clones (L) 24, 48, and 72 h after seeding (data represent 3 independent experiments).

m6A writer—reduces m6A levels and significantly decreases BCL2 protein levels (Vu et al., 2017). RNA-seq data of Mettl3^{KO} mice indicated lower RNA levels of Bcl2-5′ UPT compared with its 5′ UTR counterpart (Figures S6A–B), suggesting changes in RNA stability of either the capped 5′ UTR or Bcl2 5′ UPT, as seen in Figure 5A.

To further explore whether APA is required for Bcl2-5′ UPT expression, we cloned the entire murine 5′ UTR of Bcl2 upstream of the renilla luciferase gene (Rluc) in a Psi-Check2 dual-luciferase reporter vector that contains a *Flp*-InTM recombination system (pSTAR) (Slobodin et al., 2017) and stably inserted it into a fixed locus within *Flp-In*-HEK293 cells. As controls, we either deleted the APA sites or replaced them with the canonical polyadenylation signal of p27 (CDKN1B). In addition, we also mutated the single m6A site (RMBase v2.0—GGAC to CTCT) located at the 5′ UTR of the 5′ UPT to evaluate m6A contribution to translation (Figure 5D). 3′ end PCR analysis confirmed both the cleavage and polyadenylation downstream of the Bcl2-APA region, the lack of termination in its absence (APA deletion), and the restoration of 3′ end termination by p27 PASs (Figure S6C). TEX treatment of RNA extracts from these cells confirmed the expression of 5′ UPT-Rluc in the full Bcl2-5′ UTR and in the p27-PAS and the m6A deletion-containing vector (Figure 5E). In contrast, APA deletion diminished the expression of 5′ UPT-Rluc (Figure 5E). Interestingly, a potent luciferase expression was detected in cells transfected with the vector containing the full-length Bcl2-5′ UTR, while loss of APA and loss of m6A regions compromised luciferase expression (Figure 5F). Restoration of 3′ end processing by p27-PAS restored luciferase expression (Figure 5F). Altogether, these analyses pinpoint APA as an initiating event in the production of Bcl2-5′ UPT and downstream protein expression.

To substantiate APA function in generating BCL2, we used a dual CRISPR-Cas9 system to endogenously deplete the Bcl2-APA region (Vidigal and Ventura, 2015). We selected a clone and PCR-validated the deletion (Figure 5G). Figures 5H–5J show that while Bcl2-APA deletion did not significantly affect total mRNA levels of Bcl2, TEX treatment confirmed compromised expression of Bcl2-5′ UPTs, and flow cytometry analysis indicated reduced Bcl2 protein production.

Bcl2 is important in promoting cell survival and inhibiting pro-apoptotic pathways whose overexpression in cancer cells enables apoptosis evasion and increases malignancies (Niu et al., 2014). A specific Bcl2 inhibitor, ABT-199 (venetoclax), in combination therapy has provided an effective therapeutic strategy for several types of cancer (Pan et al., 2014). To examine the biological significance of 5′ UPTs, we tested the effect of Bcl2 expression in mouse embryonic fibroblasts. In line with a previous study, ABT-199 inhibited cell growth dose dependently (Figure 5K). Interestingly, a comparable reduction in cell growth was observed in Bcl2-APA-deleted cells, in correlation with a reduced expression of Bcl2 protein (Figure 5J). Ectopic expression of Bcl2 restored cell proliferation to levels similar to wild-type cells, indicating the key role of APA and 5′ UPT in maintaining Bcl2 expression in these cells (Figure 5L).

Since 5′ UPT expression is associated with strong RNA structure at the 5′ end (Figures 2D and 2E), we wished to determine the RNA structure at the 5′ end of Bcl2 5′ UPT. In Figure S6E, we show that the 5′ end of Bcl2 5′ UPT is a highly structured region that potentially can provide protection from endonuclease digestion (Lorenz et al., 2011). In addition, the sequence of the first 10 nt of this 5′ UPT predicted higher folding energy when compared with shuffled sequences that have the same mono-, di-, and tri-nucleotide frequencies (1,000 shuffles for each) (Figure S6F, top panel). Moreover, the first 10-nt region seems to have the highest opening energy when compared with other 10-nt regions on the input RNA sequence (Figure S6F, middle panel). A similar observation was made when these assays were repeated with different RNA lengths (80, 100, and 164 nts). Moreover, single nucleotide polymorphisms (SNPs) appearing in this region are unlikely to disrupt its RNA structure (Figure S6F, bottom panel). Thus, a strong secondary structure is suggested to support Bcl2-5' UPT expression.

5′ UPTs alter the immunopeptidome landscape

Our investigation of Bcl2 and the N-terminal mass spectrometry data indicated a significant contribution of 5′ UPTs to proteome diversity. We therefore addressed whether 5′ UPTs also can impact immunological properties of cells leading to diverse antigen presentation on human leukocyte antigen (HLA) molecules and whether this can potentially elicit immunoreactivity. As a first step, we assessed whether 5′ UPT translation products could principally be processed and present as immune peptides. We used the model peptide SIINFEKL from chicken ovalbumin that binds to mouse H-2K^b major histocompatibility complex (MHC) class I molecule (Dersh et al., 2019). We engineered MDA-MB-231 cells to express H-2K^b and a GFP reporter vector with a SIINFEKL sequence embedded in its 3′ UTR (Figure 6A). Upstream to the SIINFEKL sequence, we embedded either the Bcl2-APA region or a random sequence (Figure 6A). As expected, mRNA analysis with TEX treatment confirmed APA-induced 5′ UPT expression of SIINFEKL mRNA (Figure 6B). Moreover, flow cytometry analysis with an antibody-recognizing SIINFEKL/H-2K^b complex at the cell surface indicated APA-stimulated SIINFEKL presentation (Figure 6C). We then assessed whether this increased expression of SIINFEKL is biologically meaningful using co-culture of these cells with T cells derived from OT-1 mice designed to recognize SIINFEKL/H-2K^b protein complex. Indeed, the level of APA-induced SIINFEKL expression elicited a more potent immune response and cell-killing than control (Figure 6D).

Finally, we studied the global effect of APA induction on the HLA peptidome repertoire. For this purpose, we performed immunopeptidomics in NUDT21^{KD} cells. Overall, we detected 2.8–3.8-folds more HLA-I-associated peptides in NUDT21^{KD} compared with control cells, with 576 peptides shared in both (Figure 6E and Table S24). We predicted that if the increase in immunopeptides by NUDT21^{KD} is mediated by 5′ UPTs, then this should lead to a relative accumulation of C-terminal products. Indeed, Figures 6F and 6G show the accumulation of unique immune peptides identified in NUDT21^{KD} cells in the C-terminal part of proteins, while control cells show N-terminal accumulation. Thus, polypeptides derived from APA-dependent 5′ UPT mRNAs may be a source for neo-peptides and potentially elicit immune responses (Figure 6F).

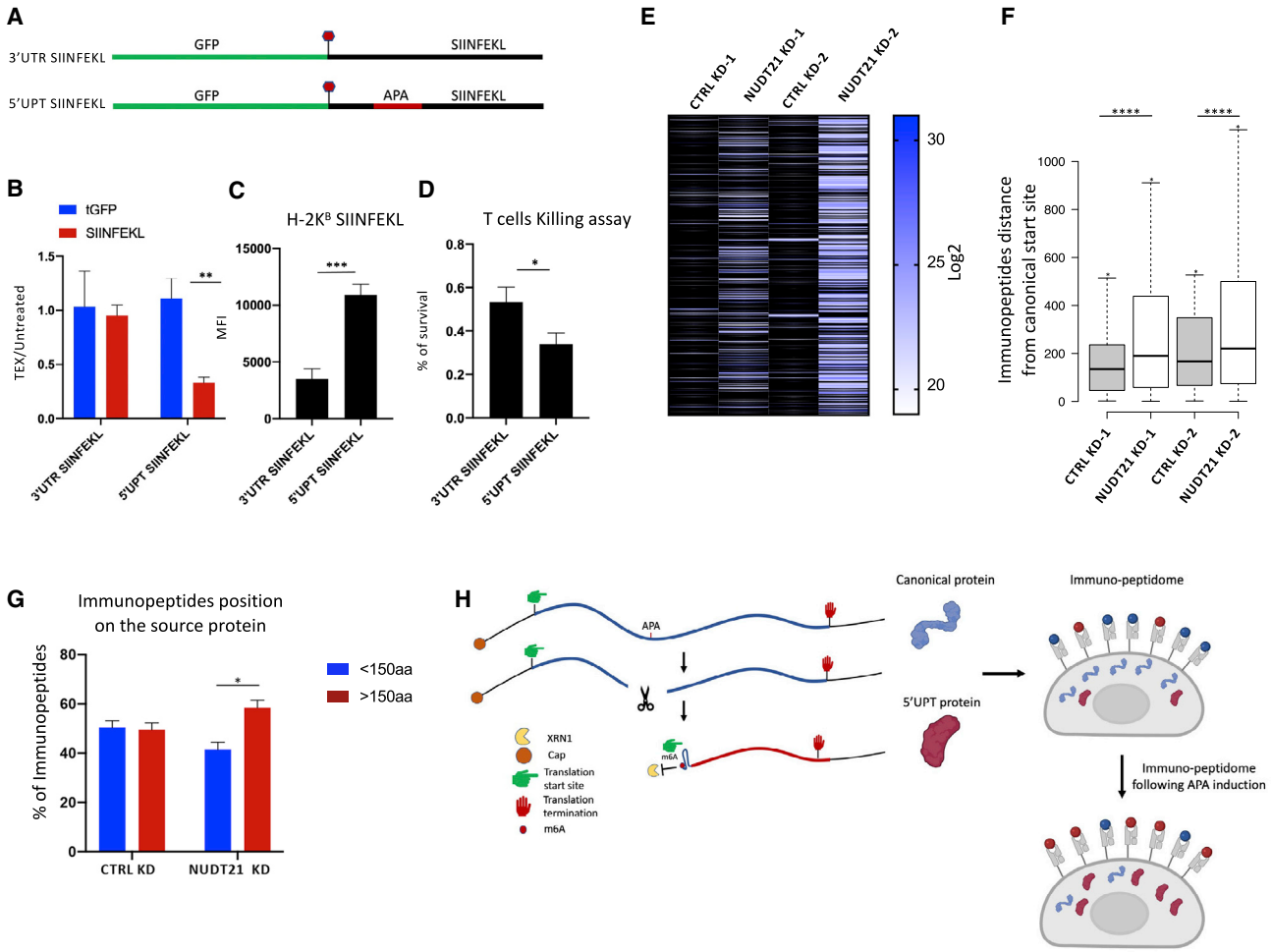


Figure 6. 5′ UPTs affect the immuno-peptide landscape of cells

(A) DNA sequence coding for SIINFEKL amino acid sequence was cloned downstream to tGFP ORF (in green) with an in-between insertion of 5′ UTR Bcl2 APA cluster (391 nts in red) or random sequence.

(B) qRT-PCR analysis following TEX treatment of amplicons corresponded to tGFP or SIINFEKL.

(C) Median fluorescence intensity (MFI) of H-2Kb-bound SIINFEKL peptides in MDA-MB-231 cells expressing H-2Kb in combination with tGFP-APA-SIINFEKL or tGFP-random-SIINFEKL reporters (data represent 3 independent experiments, two-tailed Student's t test).

(D) Cytotoxic T cells killing assay of SIINFEKL presenting cells incubated with OT-1 cells (2:1 ratio) for 24 h. Killing efficiency was determine by quantification of crystal violet staining.

(E) Heatmap showing HLA class I immunopeptides identified in either control siRNA or NUDT21 siRNA MDA-MB-231 transfected cells.

(F) Immunopeptides that are unique in each condition where selected. Boxplots depicting the distance of these peptides from the canonical start site of the source protein in the control vs. NUDT21^{KD} (two-tailed Wilcoxon test; ****$p < 0.0001$).

(G) We select the unique peptides into N/C-terminal groups (<150/>150 amino acids from the canonical start site of the source protein). Data are mean ± SD of either 2 control or 2 NUDT21^{KD} experiments (two-tailed Student's t test; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

(H) A schematic model depicting how APA induces the expression of downstream transcript isoforms. These isoforms are stabilized by RNA structure and can be used as templates for an uncapped non-canonical protein synthesis in a m6A-dependent manner. Following APA induction, global accumulation and translation of 5′ UPT alter HLA class I immunopeptidome repertoire.

DISCUSSION

Using diverse molecular and genomic methods, we show that thousands of mRNA variants, named here 5′ UPTs, are expressed as part of the human transcriptome. These 5′ UPTs are autonomous isoforms that are polyadenylated at their 3′ end, uncapped at their 5′ termini, and are the truncation products of their corresponding canonical transcripts. Moreover, we provide evidence for the contribution of 5′ UPTs to the human proteome and exploit their action in cell survival and immunology.

Generation of 5′ UPTs by APA

APA is a well-studied phenomenon whose greatest biological importance is considered to be the generation of shorter

upstream mRNA isoforms with altered function than the corresponding canonical transcripts (Gruber and Zavolan, 2019; Lee et al., 2018). The identification of mRNA isoforms downstream of APA suggests that the impact of APA on the transcriptome is underestimated. Here, we provide one example of Bcl2 whose expression depends both on APA sites embedded in its 5′ UTR and on the stability of a downstream 5′ UPT by m6A and RNA structure. This provides strong experimental evidence for the causal relationship between APA and the generation of downstream 5′ UPT isoforms.

Interestingly, recent studies described a strong link between APA and m6A modification. For example, m6A-containing transcripts are subjected to a higher APA usage (Molinie et al., 2016; Yue et al., 2018). In addition, the m6A reader YTHDC1 is associated with CPSF6, a 3′ end polyadenylation factor, and loss of YTHDC1 leads to extensive alternative polyadenylation in oocytes (Kasowitz et al., 2018). Thus, APA usage is linked to m6A modification and supports 5′ UPTs biogenesis.

5′ UPTs and the production of non-canonical proteins

Our results indicate the contribution of 5′ UPTs to the human proteome. Evidence for the translatability of these isoforms comes from their presence in polysome fractions, mapping non-canonical translation initiation start sites, and using N-terminal proteomics. We also connect this non-canonical protein production to APA and m6A. Interestingly, it has been shown that m6A residues within 5′ UTR sequences can be used as ribosome engagement sites to promote CAP-independent translation of mRNAs (Coots et al., 2017; Meyer et al., 2015). However, the efficiency of m6A-dependent translation is higher when the modification is located close to the 5′ end of the mRNA and lower when internally located (Meyer et al., 2015). Thus, the identification of enriched adenosine and m6A at the 5′ end of 5′ UPTs suggests that APA places internal m6A at the 5′ edge, resulting not only in increased mRNA stability, but also potential improvement in ribosome recruitment to stimulate non-canonical protein production.

Bcl2 expression and function

Our study pinpoints Bcl2 as a gene whose expression depends on APA sites embedded within its 5′ UTR, which generate 5′ UPTs that include its entire coding sequence. The 5′ UTR of Bcl2 is long and relatively highly conserved with multiple upstream open reading frames (uORFs) that are likely to disturb Bcl2 protein production. The generation of 5′ UPTs downstream of the APA sites is likely important for disconnecting the repressive long 5′ UTR from the coding region of the gene, making its translation m6A dependent. Indeed, recent data obtained by single-nucleotide-resolution mapping of m6A coupled with Ribo-seq indicated that m6A promotes the translation of Bcl2 in the human acute myeloid leukemia MOLM-13 cell line (Vu et al., 2017). This uncapped nature of Bcl2 expression implies a previously undescribed translation regulation mechanism. Interestingly, it has long been known that many cellular states (e.g., apoptosis, mitosis, cellular stress response) globally suppress CAP-dependent translation through various mechanisms (Holcik and Sonenberg, 2005; Spriggs et al., 2010). Also, exposure to various cellular stress agents drives widespread redistribution of m6A (Engel et al., 2018). It is therefore possible that the 5′ UPT-dependent expression of Bcl2 is required to escape inhibitory mechanisms of canonical protein production, promoting its vast expression during stress periods.

Contribution of 5′ UPTs to the immunopeptidome landscape

The immunopeptidome comprises the peptide repertoire presented by HLA class I and II molecules on the surface of cells. In cancer, the presentation of abnormal peptides in the malignant cells is critical for their recognition and clearance to suppress tumor progression. We demonstrate here that global induction of APA increases the immunoproteome repertoire, leading to a relative increase in peptides derived from the carboxyl terminus of genes (Figure 6H). This surprising result links the generation of 5′ UPTs by APA with peptide presentation and may indicate that many 5′ UPT-derived polypeptides are unstable and rapidly processed to generate peptides that are presented at the surface of cells. In light of the recurrent mutations in U1 appearing in multiple cancers (Shuai et al., 2019) and the reduced expression of NUDT21 in glioblastoma (Masamha et al., 2014), our results may open up new possibilities of cancer immunotherapy. Overall, our study reinforces a different perspective on the transcriptome and proteome and suggests that 5′ UPTs can contribute to various biological aspects of cell behavior.

Limitations of the study

In this study, we used HMM analysis to distinguish between capped and uncapped RNAs of the same genes in bulk RNA-seq data. This was achieved by enriching capped RNA or uncapped RNA in comparison with the total RNA from the same sample. Nevertheless, an accurate calculation of full-length transcripts vs. 5′ UPTs abundance may be challenging as HMM analysis detects one prominent cleavage site per transcript, while some genes likely contain multiple cleavage sites. Improvement of HMM analysis to multiple transition points is expected to improve accuracy in 5′ UPTs characterization.

Another limitation in our study is the link between m6A to 5′ UPT translation. While we present a correlative analysis between global m6A levels and translation potential of 5′ UTP following ALKBH5^{KD}, direct evidence is still needed to confirm the accountability of local m6A modification at the 5′ end of 5′ UPTs. Additional m6A-RIP-seq experiments are likely to establish a stronger link between 5′ end m6A and 5′ UPT translation.

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 SUBJECT DETAILS
  - Cell culture
  - OT-I T cells isolation and culture

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2022.09.036.

ACKNOWLEDGMENTS

R.A. is supported by the Dutch Cancer Society (KWF projects 13647, 11574) and the European Research Council (EEG-CEC/EU 832844). A.P. is supported by a long-term EMBO fellowship grant (EMBO ALTF 796-2018). S.J. and C.L. are supported by the National Research Foundation of Korea (grant number 2020R1A2C2003685). We thank Eran Rosenthal and Tommy Kaplan for technical assistance with the HMM model, Ittai Ben-Porath for comments on the manuscript, and Ron Kerkhoven and the Netherlands Cancer Institute (NKI-AVL) Genomics Core Facility in sequencing. We thank F. Van Gemert and H. te Riele for providing Bcl2 overexpression retrovirus.

AUTHOR CONTRIBUTIONS

Y.M. conceived the project, designed and performed experiments, analyzed data, and wrote the manuscript; R.A. conceived the project, wrote the manuscript, and supervised the project; P.-R.K., A.P., F.A., E.S., W.J.F., and R.E. analyzed data; S.J. and C.L. performed mass spectrometry; F.L.-P. performed ribosome profiling; J.C. performed T cell killing assay; C.W. performed experiments. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 5, 2022
Revised: June 13, 2022
Accepted: September 26, 2022
Published: October 20, 2022

REFERENCES

Akiyama, B.M., Eiler, D., and Kieft, J.S. (2016). Structured RNAs that evade or confound exonucleases: function follows form. Curr. Opin. Struct. Biol. *36*, 40–47. https://doi.org/10.1016/j.sbi.2015.12.006.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

Barak, Y., Gottlieb, E., Juven-Gershon, T., and Oren, M. (1994). Regulation of mdm2 expression by p53: alternative promoters produce transcripts with nonidentical translation potential. Genes Dev. *8*, 1739–1749. https://doi.org/10.1101/gad.8.15.1739.

Chapman, E.G., Moon, S.L., Wilusz, J., and Kieft, J.S. (2014). RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. Elife *3*, e01892. https://doi.org/10.7554/eLife.01892.

Chen, J., Brunner, A.D., Cogan, J.Z., Nunez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., and Weissman, J.S. (2020). Pervasive functional translation of noncanonical human open reading frames. Science *367*, 1140–1146. https://doi.org/10.1126/science.aay0262.

Chernyakov, I., Whipple, J.M., Kotelawala, L., Grayhack, E.J., and Phizicky, E.M. (2008). Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. Genes Dev. *22*, 1369–1380. https://doi.org/10.1101/gad.1654308.

Coots, R.A., Liu, X.M., Mao, Y., Dong, L., Zhou, J., Wan, J., Zhang, X., and Qian, S.B. (2017). m(6)A Facilitates eIF4F-Independent mRNA Translation. Mol Cell *68*, 504–514.e7. e507. https://doi.org/10.1016/j.molcel.2017.10.002.

Dasilva, L.F., Blumenthal, E., Beckedorff, F., Cingaram, P.R., Gomes Dos Santos, H., Edupuganti, R.R., Zhang, A., Dokaneheifard, S., Aoi, Y., Yue, J., et al. (2021). Integrator enforces the fidelity of transcriptional termination at protein-coding genes. Sci. Adv. *7*, eabe3393. https://doi.org/10.1126/sciadv.abe3393.

Dersh, D., Yewdell, J.W., and Wei, J. (2019). A SIINFEKL-Based System to Measure MHC Class I Antigen Presentation Efficiency and Kinetics. Methods Mol. Biol. 109–122. 1988. https://doi.org/10.1007/978-1-4939-9450-2_9.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108. https://doi.org/10.1038/nature11233.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206. https://doi.org/10.1038/nature11112.

Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. *14*, 496–506. https://doi.org/10.1038/nrg3482.

Engel, M., Eggert, C., Kaplick, P.M., Eder, M., Roh, S., Tietze, L., Namendorf, C., Arloth, J., Weber, P., Rex-Haffner, M., et al. (2018). The Role of m(6)A/m-RNA Methylation in Stress Response Regulation. Neuron *99*, 389–403.e9. e389. https://doi.org/10.1016/j.neuron.2018.07.009.

Furuichi, Y., LaFiandra, A., and Shatkin, A.J. (1977). 5'-Terminal structure and mRNA stability. Nature *266*, 235–239. https://doi.org/10.1038/266235a0.

Gruber, A.J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. Nat. Rev. Genet. *20*, 599–614. https://doi.org/10.1038/s41576-019-0145-z.

Guo, W., Li, Y., Liang, W., Wong, S., Apovian, C., Kirkland, J.L., and Corkey, B.E. (2012). Beta-mecaptoethanol suppresses inflammation and induces

adipogenic differentiation in 3T3-F442A murine preadipocytes. PLoS One *7*, e40958. https://doi.org/10.1371/journal.pone.0040958.

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.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat. Rev. Mol. Cell Biol. *6*, 318–327. https://doi.org/10.1038/nrm1618.

Ingolia, N., Brar, G., Stern-Ginossar, N., Harris, M., Talhouarne, G., Jackson, S., Wills, M., and Weissman, J. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep. *8*, 1365–1379. https://doi.org/10.1016/j.celrep.2014.07.045.

Ingolia, N., Lareau, L., and Weissman, J. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802. https://doi.org/10.1016/j.cell.2011.10.002.

Jang, S.K., Krausslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C., and Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J. Virol. *62*, 2636–2643. https://doi.org/10.1128/JVI.62.8.2636-2643.1988.

Karginov, F.V., Cheloufi, S., Chong, M.M., Stark, A., Smith, A.D., and Hannon, G.J. (2010). Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. Mol Cell *38*, 781–788. https://doi.org/10.1016/j.molcel.2010.06.001.

Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS Genet. *14*, e1007412. https://doi.org/10.1371/journal.pgen.1007412.

Kearse, M.G., and Wilusz, J.E. (2017). Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. *31*, 1717–1731. https://doi.org/10.1101/gad.305250.117.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36. https://doi.org/10.1186/gb-2013-14-4-r36.

Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. Nature *509*, 575–581. https://doi.org/10.1038/nature13302.

Kristensen, L.S., Andersen, M.S., Stagsted, L.V.W., Ebbesen, K.K., Hansen, T.B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. Nat. Rev. Genet. *20*, 675–691. https://doi.org/10.1038/s41576-019-0158-7.

Kulp, D., Haussler, D., Reese, M.G., and Eeckman, F.H. (1996). A generalized hidden Markov model for the recognition of human genes in DNA. Proc. Int. Conf. Intell. Syst. Mol. Biol. *4*, 134–142.

Lee, S.H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C.S., and Mayr, C. (2018). Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. Nature *561*, 127–131. https://doi.org/10.1038/s41586-018-0465-8.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, H.B., Tong, J., Zhu, S., Batista, P.J., Duffy, E.E., Zhao, J., Bailis, W., Cao, G., Kroehling, L., Chen, Y., et al. (2017). m(6)A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. Nature *548*, 338–342. https://doi.org/10.1038/nature23450.

Li, P., Zhou, X., Xu, K., Zhang, Q.C., et al. (2021). RASP: an atlas of transcriptome-wide RNA secondary structure probing data. Nucleic Acids Res *49*, D183–D191. https://doi.org/10.1093/nar/gkaa880.

Lin, X., Chai, G., Wu, Y., Li, J., Chen, F., Liu, J., Luo, G., Tauler, J., Du, J., Lin, S., et al. (2019). RNA m(6)A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. Nat. Commun. *10*, 2065. https://doi.org/10.1038/s41467-019-09865-9.

Liu, B., Merriman, D.K., Choi, S.H., Schumacher, M.A., Plangger, R., Kreutz, C., Horner, S.M., Meyer, K.D., and Al-Hashimi, H.M. (2018). A potentially abundant junctional RNA motif stabilized by m6A and Mg2+. Nat. Commun. *9*, 2761. https://doi.org/10.1038/s41467-018-05243-z.

Loayza-Puch, F., Drost, J., Rooijers, K., Lopes, R., Elkon, R., and Agami, R. (2013). p53 induces transcriptional and translational programs to suppress cell proliferation and growth. Genome Biol. *14*, R32. https://doi.org/10.1186/gb-2013-14-4-r32.

Lorenz, R., Bernhart, S.H., Honer 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.

Lu, Z., Zhang, Q., Lee, B., Flynn, R., Smith, M., Robinson, J., Davidovich, C., Gooding, A., Goodrich, K., Mattick, J., et al. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell *165*, 1267–1279. https://doi.org/10.1016/j.cell.2016.04.028.

Malka, Y., Steiman-Shimony, A., Rosenthal, E., Argaman, L., Cohen-Daniel, L., Arbib, E., Margalit, H., Kaplan, T., and Berger, M. (2017). Post-transcriptional 3 -UTR cleavage of mRNA transcripts generates thousands of stable uncapped autonomous RNA fragments. Nat. Commun. *8*, 2029. https://doi.org/10.1038/s41467-017-02099-7.

Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.B., Li, W., and Wagner, E.J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature *510*, 412–416. https://doi.org/10.1038/nature13261.

Mercer, T.R., Dinger, M.E., Bracken, C.P., Kolle, G., Szubert, J.M., Korbie, D.J., Askarian-Amiri, M.E., Gardiner, B.B., Goodall, G.J., Grimmond, S.M., and Mattick, J.S. (2010). Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. Genome Res. *20*, 1639–1650. https://doi.org/10.1101/gr.112128.110.

Mercer, T.R., Wilhelm, D., Dinger, M.E., Solda, G., Korbie, D.J., Glazov, E.A., Truong, V., Schwenke, M., Simons, C., Matthaei, K.I., et al. (2011). Expression of distinct RNAs from 3' untranslated regions. Nucleic Acids Res. *39*, 2393–2403. https://doi.org/10.1093/nar/gkq1158.

Meyer, K., Patil, D., Zhou, J., Zinoviev, A., Skabkin, M., Elemento, O., Pestova, T., Qian, S.B., and Jaffrey, S. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. Cell *163*, 999–1010. https://doi.org/10.1016/j.cell.2015.10.012.

Meyer, K., Saletore, Y., Zumbo, P., Elemento, O., Mason, C., and Jaffrey, S. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635–1646. https://doi.org/10.1016/j.cell.2012.05.003.

Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.X., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P., et al. (2016). m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. Nat. Methods *13*, 692–698. https://doi.org/10.1038/nmeth.3898.

Mukherjee, S., and Mitra, S. (2005). Hidden Markov Models, grammars, and biology: a tutorial. J Bioinform Comput Biol *03*, 491–526. https://doi.org/10.1142/s0219720005001077.

Nam, Jin-Wu, et al. (2014). Global analyses of the effect of different cellular contexts on microRNA targeting. Molecular Cell *53* (6), 1031–1043. https://doi.org/10.1016/j.molcel.2014.02.013.

Niu, X., Wang, G., Wang, Y., Caldwell, J.T., Edwards, H., Xie, C., Taub, J.W., Li, C., Lin, H., and Ge, Y. (2014). Acute myeloid leukemia cells harboring MLL fusion genes or with the acute promyelocytic leukemia phenotype are sensitive to the Bcl-2-selective inhibitor ABT-199. Leukemia *28*, 1557–1560. https://doi.org/10.1038/leu.2014.72.

Oh, J.M., Venters, C.C., Di, C., Pinto, A.M., Wan, L., Younis, I., Cai, Z., Arai, C., So, B.R., Duan, J., and Dreyfuss, G. (2020). U1 snRNP regulates cancer cell migration and invasion in vitro. Nat. Commun. *11*, 1. https://doi.org/10.1038/s41467-019-13993-7.

Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. Mol Cell *66*, 9–21.e7. e27. https://doi.org/10.1016/j.molcel.2017.02.021.

Pan, R., Hogdal, L.J., Benito, J.M., Bucci, D., Han, L., Borthakur, G., Cortes, J., DeAngelo, D.J., Debose, L., Mu, H., et al. (2014). Selective BCL-2 inhibition by ABT-199 causes on-target cell death in acute myeloid leukemia. Cancer Discov. *4*, 362–375. https://doi.org/10.1158/2159-8290.CD-13-0609.

Pedersen, B.S., Yang, I.V., and De, S. (2013). CruzDB: software for annotation of genomic intervals with UCSC genome-browser database. Bioinformatics *29*, 3003–3006. https://doi.org/10.1093/bioinformatics/btt534.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2016). Genome-wide quantification of 5'-phosphorylated mRNA degradation intermediates for analysis of ribosome dynamics. Nat. Protoc. *11*, 359–376. https://doi.org/10.1038/nprot.2016.026.

Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature *334*, 320–325. https://doi.org/10.1038/334320a0.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Ren, Y., Sun, C., Sun, Y., Tan, H., Wu, Y., Cui, B., and Wu, Z. (2009). PPAR gamma protects cardiomyocytes against oxidative stress and apoptosis via Bcl-2 upregulation. Vascul Pharmacol *51*, 169–174. https://doi.org/10.1016/j.vph.2009.06.004.

Roost, C., Lynch, S.R., Batista, P.J., Qu, K., Chang, H.Y., and Kool, E.T. (2015). Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J. Am. Chem. Soc. *137*, 2107–2115. https://doi.org/10.1021/ja513080v.

Ruiz Cuevas, M.V., Hardy, M.P., Holly, J., Bonneil, E., Durette, C., Courcelles, M., Lanoix, J., Cote, C., Staudt, L.M., Lemieux, S., et al. (2021). Most non-canonical proteins uniquely populate the proteome or immunopeptidome. Cell Rep. *34*, 108815. https://doi.org/10.1016/j.celrep.2021.108815.

Russ, Brendan E., et al. (2014). Distinct Epigenetic Signatures Delineate Transcriptional Programs during Virus-Specific CD8+ T Cell Differentiation. Immunity *41* (5), 853–865. https://doi.org/10.1016/j.immuni.2014.11.001.

Shatsky, I.N., Terenin, I.M., Smirnova, V.V., and Andreev, D.E. (2018). Cap-Independent Translation: What's in a Name? Trends Biochem. Sci. *43*, 882–895. https://doi.org/10.1016/j.tibs.2018.04.011.

Shuai, S., Suzuki, H., Diaz-Navarro, A., Nadeu, F., Kumar, S.A., Gutierrez-Fernandez, A., Delgado, J., Pinyol, M., Lopez-Otin, C., Puente, X.S., et al. (2019). The U1 spliceosomal RNA is recurrently mutated in multiple cancers. Nature *574*, 712–716. https://doi.org/10.1038/s41586-019-1651-z.

Siddiqui, N., and Sonenberg, N. (2015). Signalling to eIF4E in cancer. Biochem. Soc. Trans. *43*, 763–772. https://doi.org/10.1042/BST20150126.

Singh, I., Lee, S.H., Sperling, A.S., Samur, M.K., Tai, Y.T., Fulciniti, M., Munshi, N.C., Mayr, C., and Leslie, C.S. (2018). Widespread intronic polyadenylation diversifies immune cell transcriptomes. Nat. Commun. *9*, 1716. https://doi.org/10.1038/s41467-018-04112-z.

Slobodin, B., Han, R., Calderone, V., Vrielink, J.A.O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell *169*, 326–337.e12. e312. https://doi.org/10.1016/j.cell.2017.03.031.

Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M.D., and Hussain, S. (2019). A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. Nat. Commun. *10*, 3359. https://doi.org/10.1038/s41467-019-11272-z.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., and Chang, H.Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486–490. https://doi.org/10.1038/nature14263.

Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Translational regulation of gene expression during conditions of cell stress. Mol Cell *40*, 228–237. https://doi.org/10.1016/j.molcel.2010.09.028.

Stoneley, M., and Willis, A.E. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. Oncogene *23*, 3200–3207. https://doi.org/10.1038/sj.onc.1207551.

Vidigal, J.A., and Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat. Commun. *6*, 8083. https://doi.org/10.1038/ncomms9083.

Vu, L.P., Pickering, B.F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., MacKay, M., et al. (2017). The N(6)-methyladenosine (m(6)A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. Nat Med *23*, 1369–1376. https://doi.org/10.1038/nm.4416.

Wan, Chi-Keung, et al. (2015). IL-21-mediated non-canonical pathway for IL-1β production in conventional dendritic cells. Nature Communications7988. https://doi.org/10.1038/ncomms8988.

Wang, R., Zheng, D., Wei, L., Ding, Q., and Tian, B. (2019). Regulation of Intronic Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes. Cell Rep. *26*, 2766–2778.e6. e2766. https://doi.org/10.1016/j.celrep.2019.02.049.

Xuan, J.J., Sun, W.J., Lin, P.H., Zhou, K.R., Liu, S., Zheng, L.L., Qu, L.H., and Yang, J.H. (2018). RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. Nucleic Acids Res. *46*, D327–D334. https://doi.org/10.1093/nar/gkx934.

Youle, R.J., and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. Nat. Rev. Mol. Cell Biol. *9*, 47–59. https://doi.org/10.1038/nrm2308.

Yue, Y., Liu, J., Cui, X., Cao, J., Luo, G., Zhang, Z., Cheng, T., Gao, M., Shu, X., Ma, H., et al. (2018). VIRMA mediates preferential m(6)A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. Cell Discov *4*, 10. https://doi.org/10.1038/s41421-018-0019-0.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M., and Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. Nat. Methods *12*, 835–837. https://doi.org/10.1038/nmeth.3478.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594. https://doi.org/10.1038/nature15377.

Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S., and Rouskin, S. (2017). DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat. Methods *14*, 75–82. https://doi.org/10.1038/nmeth.4057.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-Cap antibody (1:100)	Merck	CAT # MABE419 Clone H-20; RRID:AB_2687977
mouse IgG1 anti-Flag	Sigma	CAT # A2220; RRID:AB_10063035
FITC anti-Bcl-2	Biolegend	CAT # 633503; RRID:AB_2028392
APC anti-mouse H-2Kb bound to SIINFEKL clone 25-D1.16	Biolegend	CAT # 141606; RRID:AB_11219595
Bacterial and virus strains		
DH5-alpha	Thermo Fisher Scientific	CAT # 18,265,017
Chemicals, peptides, and recombinant proteins		
RNA fragmentation reagent	Ambion	CAT # AM8740
α-amanitin (10 μg/mL)	Santa Cruz	CAT # SC-202440
Dharmafect I	Dharmacon	CAT # T-2001-03
Protein G Agarose beads	Sigma	CAT #E3403
Murine RNAse inhibitor	NEB	CAT #M0314S
proteinase K	NEB	CAT #P8107S
Biotin-11-GTP	Perkin-Elmer	CAT # NEL545001EA
Dynabeads MyOne Streptavidin C1	Invitrogen	CAT # 65,001
Terminator 5′-Phosphate-Dependent Exonuclease	Epicenter	CAT # TER51020
Trizol	Ambion	CAT # 15,596-018
XRN1	NEB	CAT #M0338S
Harringtonine	Santa Cruz	CAT # sc-204771A
CleanNGS	Gcbiotech	CAT # CNGS-0050
T4 DNA ligase	NEB	CAT #M0202T
T4 RNA Ligase 1	NEB	CAT #M0204S
Phusion DNA Polymerase	Thermo	CAT #F530L
NEBNext End Repair	NEB	CAT #E6050S
Cycloheximide	Sigma	CAT #C7698-5G
Gibson Assembly	NEB	CAT #E5510S
ABT-199	Abcam	CAT # ab217298
Critical commercial assays		
SensiFAST SYBR	Bioline	CAT # BIO-98020
mRNA Magnetic Isolation Module	NEB	CAT #E7490S
Vaccinia Capping System	NEB	CAT #M2080S
MICRO-Elute Kit	Norgen	CAT # 61,000
Oligotex	Qiagen	CAT # 70,042
NEBNext Ultra Directional RNA Library Prep Kit	NEB	CAT #E7420S
SMARTer® Stranded RNA-Seq Kit	Takara	CAT # 634,839
Illumina Truseq Stranded mRNA Library Prep kit	Illumina	CAT # 20,020,594
3′-mRNA-Seq Library Prep Kit	Lexogen	CAT # 015UG009V0211
Nanopore direct RNA sequencing	Oxford Nanopore	CAT # ONT SQK-RNA002
SuperScript III	Thermo	CAT # 18,080-051
USB Poly(A) tail-length assay kit	Affymetrix	CAT # 76,455 1KT
Deposited data		
RNA-seq data	Mouse T-cells	RR1688717
RNA-seq data	Mouse T-cells	Sequence Read Archive: SRR1583932

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNA-seq data	Mouse T-cells	Sequence Read Archive: SRR1688691
RNA-seq data	Mouse 3'seq	Sequence Read Archive: SRP033205
RNA-seq data	Mouse Mettl3 KD	GEO Series: GSE100048
RNA-seq data	Mouse m6A-RIP	GEO Series: GSE100048
RNA-seq data	U2OS	GEO Series: GSE84068
RNA-seq data	HEK293	GEO Series: GSE84068
RNA-seq data	M6A RIP-seq	GEO Series: GSE112795
RNA-seq data	Gro-seq	GEO Series: GSE53499
RNA-seq data	This study	GEO Series: GSE149204
Experimental models: Cell lines		
U2OS	ATCC	ATCC HTB-96
HEK293	ATCC	ATCC CRL-1573
HeLa	ATCC	ATCC CCL-2
HEK293/FRT	Thermo	R75007
Recombinant DNA		
Psi-Check2/FRT-Bcl2-5UTR	This study	N/A
Psi-Check2/FRT-Bcl2-5UTR-APA_MUT	This study	N/A
Psi-Check2/FRT-Bcl2-5UTR-P27_APA	This study	N/A
Psi-Check2/FRT-Bcl2-5UTR-m6A_MUT	This study	N/A
pOG44	Thermo	CAT #V600520
Software and algorithms		
TopHat v.2.0.13	PMID: 23,618,408 (Kim et al., 2013)	http://ccb.jhu.edu/software/tophat
Samtools	PMID: 19,505,943 (Li et al., 2009)	http://samtools.sourceforge.net
Bedtools	PMID: 20,110,278 (Quinlan et al., 2010)	http://code.google.com/p/bedtools
HMM	PMID: 29,229,900 (Malka et al., 2017)	https://github.com/eranroz/polyA
Htseq	PMID: 25,260,700 (Anders et al., 2015)	https://htseq.readthedocs.io/en/master/
RASP	PMID: 33068412 (Li et al., 2021)	http://rasp.zhanglab.net
CruzDB	PMID: 24037212 (Pedersen et al., 2013)	https://github.com/brentp/cruzdb

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Reuven Agami (r.agami@nki.nl).

Materials availability

All unique materials and reagents generated in this study are available from the lead contact with a completed material transfer agreement.

Data and code availability

- Genomic sequencing data have been deposited at the Sequence Read Archive (SRA) and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. SRA: SRR1688717
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

HeLa, U2OS, HEK293 and MEF cell lines were grown in Dulbecco's modified Eagle's medium, supplemented with 10% fetal calf serum, 100 units/ml penicillin, and 100 μg/mL streptomycin at 37°C. Cell lines were regularly tested for Mycoplasma contamination.

Cell lines were authenticated by expression analysis based on RNA-seq. To inhibit RNA polymerase II transcription activity, U2OS cells were incubated with α-amanitin (10 μg/mL; Santa Cruz SC-202440) for 6h at 37°C.

OT-I T cells isolation and culture

OT-I T cells were isolated using DynabeadsTM UntouchedTM Mouse CD8 Cells Kit (Invitrogen) according to manufacturer's protocol. T cells were initially maintained in RPMI 1640 Medium (RPMI 1640, Gibco) containing 10% human serum (One Lamda), 100U/mL of penicillin, 100 μg/ml of streptomycin, 100U/mL IL-2 (Proleukin, Novartis), 10 ng/mL IL-7 (ImmunoTools) and 10 ng/mL IL-15 (ImmunoTools).

METHOD DETAILS

Real-time PCR

1 μg of total RNA was reverse transcribed using Tetro cDNA synthesis kit following the manufacturer's instructions. Real-time PCR was performed using the SensiFAST SYBR real-time PCR kit (Bioline). Data were normalized to Humane endogenous control (GAPDH) and analyzed using the ΔΔCt model unless otherwise indicated.

The following primer sequences were used; PTAR1:
UP:
F: TTCAGGAGGAACCCACACAT.
R: CTTGACACACCAGCTCTCCA.
DOWN:
F: GCGGCATATTTTCTACCTTCA.
R: CTTCAGGCGTTTGGTTTCCT.
NACC2:
UP:
F: CGTACCACAACGAGGAGGAC.
R: GCTCAGGCTTCTCTTGCACT.
Down:
F: GGTCCTGAACGCTGTGAAAT.
R: GGCAGCATGGACTTGATCTT.
AKAP9:
Up:
F: TTCCTTTCTATCCCCAACCA.
R: CCATCCGACTGAGCTTTTCT.
Down:
F: CCGCCTACAAGAACTTGAACA.
R: GCAAATTTCCCAAATAGCACA.
CCDC88A:
Up:
F: AGTTCATGACCAGCCCTTTG.
R: ATTTGGAGCATGACCTGGTT.
Down:
F: TTCCCCTGGAAGTGAAGTTG.
R: CTGGCCAGACCACAGCTAAC.
GAPDH:
F: ACAACTTTGGTATCGTGGAAGG.
R: GCCATCACGCCACAGTTTC
mBcl2 up (Figure 4G):
F: ACTTTCCATGGACGCGTTT.
R: CGGGACGACCTAGGAAGTTA
mBcl2 down (Figure 4G):
F: ATAACCGGGAGATCGTGATG.
R: CAGGCTGGAAGGAGAAGATG
mBcl2 5′UTR (Figure 4I):
F: GGGGACTTCGTAGCAGTCAT.
R: CTTTCCGAAACGGGAAAAA.
Rluc (Figure 4I):
F: ATGGCTTCCAAGGTGTACGA.
R: GTAGCTGGAGGCAGCGTTAC.

3′end PCR (Fig. S7C):
F: CAATCTGGAAACCCTCCTGA.
R:TTTTTTTTTTTTTTTTTTGcagtAGATCGGAAGAGCGCTACACGACGCTCTTCCG ATCTactgCTTTTTTVN.
U1:
F: CTTACCTGGCAGGGGAGATA.
R: CAGTCCCCCACTACCACAAA (Used for revers transcription and qPCR)
NUDT21:
F: TTCGGCAACAAGTACATCCA.
R: AAATCTGGCTGCAACAGAGC.
PCF11.
F: GAAGAGGAGTCGGAAGGGAG.
R: TCCTCGGCTAGAATGGTCAG.
INTS11:
F: CCTGATCATGAAGGACAGCA.
R: TCTTGAGGTGGCTGTAGACG
sgRNA:
Bcl2:
TTTAAACGTGTAACTTGTAG.
GTCGGGACTTGAAGTGCCAT.

siRNA and morpholino knock down

siRNAs against NUDT21 and ALKBH5 were purchased from Life Technologies (NUDT21 siRNA catalog number 4392420, ID s21770 and s21772; ALKBH5 siRNA catalog number 4392420, ID s29686 and s29687; PCF11 siRNA catalog number 4392420 ID s28363 and s28364; INTS11 siRNA catalog number 4392420 ID s29893 and s29895; Negative control siRNA catalog number 4390844). Cells were transfected for using Dharmafect I reagent (Dharmacon) following the manufacturer's instructions (2 siRNAs per target gene) for 72h.

U1 and control morpholino oligonucleotide sequences were 5'-GGTATCTCCCCTGCCAGGTA AGTAT-3', and 5'-TGATAAGAA CAGATACTACACTTGA-3', respectively. Oligonucleotides were transfected into 15cm dish for 24h (12 millions of cells −190 pmol). All experiments were performed in two biological replicates.

Lentiviral production and transduction

For lentivirus production, 4 × 10^6 HEK293T cells were seeded per 100-mm dish, one day before transfection. For each transfection, 10 μg of the pCDH reporter, 5 μg of pMDL RRE, 3.5 μg pVSV-G AND 2.5 μg of pRSV-REV plasmids were mixed in 500 μL serum-free DMEM. Next, 500 μL of serum-free DMEM containing 63 μL of a 1 mg/mL PEI solution was added. The entire mix was vortexed and left for 15 min at room temperature, after which it was added to the HEK293T cells to be transfected. The next day, the medium was replaced by RPMI. The lentivirus-containing supernatants were collected 48 and 72 h after transfection, and snap-frozen in liquid nitrogen. Target cells were transduced on two consecutive days by supplementation of the lentiviral supernatant with 8 μg/mL polybrene (Sigma). One day after the last transduction, transduced cells were selected by the addition of 2 μg/mL puromycin to the medium.

5′- CAP RNA immunoprecipitation

5′- CAP RNA immunoprecipitation was performed as described previously (Malka et al., 2017). Briefly, Precleared RNA samples were then subjected to 5′-CAP RNA immunoprecipitation using 5 μL (1:100 dilution) anti-Cap antibody (Merck anti-m3G-cap, m7G-cap antibody, clone H-20) and incubated O/N in 4°C. Then, RNA immune complexes were incubated with 40 μL of Protein G Agarose beads for 4 h in RT following four washing steps using wash buffer (50 mM Tris, 150 mM NaCl, 0.1% Triton, and 1 μL murine RNAse inhibitor). Beads were then resuspended with 100 μL elution buffer (wash buffer containing 3 μL proteinase K NEB-P8107S and 0.1% SDS) and incubated for 15 min at 65°C. The elution step was repeated one more time. Finally, 800 μL QIAzol was added to the 200 μL eluted RNA.

3′-RNA capping and pulldown

3′-RNA capping and pulldown was performed as described previously (Malka et al., 2017). Briefly, uncapped poly(A) selected was capped using biotinylated GTP with the Vaccinia Capping System according to the manufacturer's instructions (NEB-M2080S—15 μL DEPC water, 2 μL 10× Capping Buffer, 1 μL Vaccinia Capping Enzyme, and 1 μL Biotin-11-GTP (Perkin-Elmer; NEL545001EA)). Next, RNA with biotinylated GTP was pulled down using Dynabeads MyOne Streptavidin C1 (Invitrogen; 65,001) according to the manufacturer's instructions.

Terminator phosphate-dependent TEX treatment

For RNA-seq analysis DNase I-treated, poly(A) selected RNA was treated with Terminator 5′-Phosphate-Dependent Exonuclease (Epicentre; TER51020) according to the manufacturer's instructions followed enzyme deactivation.

XRN1 treatment

For RNA-seq analysis DNase I-treated, poly(A) selected RNA was treated with XRN1 (NEB M0338S) according to the manufacturer's instructions followed enzyme deactivation.

Illumina RNA-sequencing

RNA was extracted from U2OS and HEK293 cells using QIAzol Reagent (Qiagen; 79,306) and Trizol reagent for HeLa cells (15,596-018, Ambion life technologies) according to the manufactures protocol followed by DNase I treatment. PolyA selected RNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB; E7490) for U2OS and HEK293 and Oligotex kit (QIAGEN) for HeLa cells and further processed with the NEBNext Ultra Directional RNA Library Prep Kit (NEB; E7420S) for U2OS and HEK293 or SMARTer Stranded RNA-Seq Kit (Takara; 634,839) and illumina Truseq Stranded mRNA Library Prep kit. for HeLa cells or 3′–3′-mRNA-Seq Library Prep Kit (lexogen; 015UG009V0211).

Nanopore RNA-sequencing

Poly(A) selected RNA (500 ng) of control or TEX treated samples were prepared for nanopore direct RNA sequencing generally following the ONT SQK-RNA002 kit protocol, including the optional reverse transcription step recommended by ONT. RNA sequencing on the MinION was performe using ONT R9 Flow Cells. The experiment was performed in two biological replicates.

5′end sequencing

20 μg of total RNA was polyA selected and then was treated with E.coli purified AlkB for demethylation of the RNA as previously describe (Zheng et al., 2015). Next, we construction uncapped 5′-specific sequencing libraries as previously describe (Pelechano et al., 2016). The experiment was performed in two biological replicates.

Polysomal RNA isolation

Polysomal RNA isolation was performed as described previously (Slobodin et al., 2017). Briefly, Sucrose gradients for separation of polysomes were usually prepared by gentle sequential addition of 2.2mL of the different sucrose solutions (e.i., 47, 37, 27, 17 and 7% in Tris-HCl pH = 7.5 (f.c. 20mM), $MgCl_2$ (f.c. 10mM) and KCl (f.c. 100mM), supplemented with 2mM DTT, Ribosafe RNase inhibitor (Bioline, 1 μL/mL) and CHX (100 μg/mL) into a 12 mL tube (Beckman, 9/16 × 3 ½ in.) and left overnight at 4°C to achieve continuous gradient prior to the centrifugation. Cells were treated with CHX and harvest after washing with PBS with CHX and lysed. The lysates were centrifuged 1300xg for 10 min at 4°C and the supernatants were transferred into new tubes. From the cleared lysates, 500 μL were loaded on top of each gradient, mounted on SW41TI rotor and centrifuged at 36000rpm for 2 h at 4°C. Following the centrifugation, each gradient was split into 15 equal fractions of 760 μL. Fractions 9-13 were collected for RNA isolation using TRIsure (Bioline) according to the manufacturer's instructions, polyA selected followed by RNA library preparation as previously describe (Takara; 634,839). The experiment was performed in two biological replicates.

Ribosome profiling (Ribo-seq)

Libraries from cultured cells were prepared as described previously (Loayza-Puch et al., 2013), with an addition of harringtonine treatment for 5 min that added to the cell culture medium (final concentration of 2 μg/mL -Santa Cruz sc-204771A) prior to CHX treatment. The experiment was performed in two biological replicates.

3′-end RNA-seq analysis

Trimming and filtering of raw reads

NextSeq basecall files were converted to FASTQ files using bcl2fastq (v.2.17.1.14). Reads were screened and preprocessed similarly to the described above, an additional step of removing polyA sequences from the 3′ ends of reads was performed with cutadapt, using a 75-mer oligo-A sequence as the "adapter" and a minimal overlap of 2.

Protein extraction and quantification

Frozen HeLa cells were resuspended in lysis buffer (0.2 M EPPS pH 8.0 (Merck, Germany), 6 M guanidine (Merck), 10 mM TCEP (Thermo Fisher Scientific, MA, USA), 40 mM 2-chloroacetamide (Merck), 1x Halt protease inhibitor (Thermo Fisher Scientific), 1 mM EDTA pH 8.0). After 10-min incubation on 95°C, cell pellets lysed by ultrasonication probe by VCX 750 Ultrasonic processor (Sonics, CT, USA). Cell lysates were centrifuged at 18,000 xg, 4°C for 10 min, only supernatants were transferred to another new 5mL centrifuge tube for further analysis. To precipitate the proteins, we added 8-fold of the supernatant volume of freezer cold (−20°C) acetone and 1-fold of freezer cold methanol and incubated the mixture in −80°C for 2 h. After washed twice with freezer cold methanol, precipitated protein resuspended in 0.2 M EPPS pH 8.0, 6M guanidine. Finally, protein concentration was determined by Quick Start Bradford protein assay (Bio-Rad Laboratories, CA, USA).

N-terminal protein enrichment

100 μg of protein was diluted with 0.2 M EPPS pH 8.0, 6M guanidine down to 4 mg/mL, then transferred to 5 mm NMR tube cap (Merck) and acetylated their α- and ε-amines using 200 mM D_6-acetic anhydride (Merck) and 200 mM pyridine (Merck) for 1 h.

To quench unreacted anhydride, added 0.2 M EPPS pH 8.0 for 9 volume of reaction mixture to pre-equilibrated 30 mg/1 cc HLB solid phase extraction (SPE) column (Waters) and then connected NMR tube cap with reaction mixture into inlet of SPE column and plugged its outlet with luer plug (Waters). The SPE column reactor incubated at 25°C for 1 h while end-over-end mixing. Proteins were digested with MS-grade trypsin (Promega) at 37°C for 4 h while shaking (1:50 trypsin:protein by mass). Newly synthesized α-amines by trypsin were depleted by addition of dried N-hydroxysuccinimide agarose bead (Thermo Fisher Scientific) and incubated at 25°C for 2 h while end-over-end mixing. The enriched peptides were washed by 1 mL of 50 mM EPPS pH 8, 150 mM sodium chloride then 5 mL of 0.1% formic acid in water MS-grade (Thermo Fisher Scientific). Finally, the enriched peptides were eluted from the SPE column with 1 mL of 0.1% formic acid in 50% (v/v) acetonitrile/water MS-grade (Thermo Fisher Scientific). The eluted peptides concentrated by evaporation in Speed Vac (Thermo Fisher Scientific). The peptides were reconstituted with 0.1% formic acid in 2% (v/v) acetonitrile/water MS-grade for mass spectrometry analysis. All experiments were performed in three biological replicates.

DNA cloning and luciferase analysis

Psi-Check2 vector with cloned FRT sequence and hygromycin-resistance gene served as the basic vector for the insertion of Bcl2 5′UTR at NheI site of the vector. Stable line expressing the constructs was established using the Flp-In system. Briefly, sub-confluent HEK-293 cells, containing a single pre-engineered flp recombinase target site (FRT) within the genome, were co-transfected with an insert shuttle vector Psi-Check2/FRT (carrying the Bcl2 5′UTR variants) together with a plasmid expressing the Flp-recombinase enzyme (pOG44). The plasmid ratio of Psi-Check2/FRT to pOG44 was 1:3, and the co-transfections were facilitated using Lipofectamine 2000. After 48 h of transfection, the cells were trypsinized and plated into four 150-mm dishes with fresh DMEM with selection medium containing 100 μg/ml hygromycin. Stable, hygromycin-resistant cells were selected for at least 3 weeks with media changes every third day. After 3 weeks of selection, small clusters of cell clones were isolated, allowed to expand. 72h before luciferase analysis and RNA extraction, 1mM beta mercaptoethanol was added to the full medium (Guo et al., 2012; Ren et al., 2009). Luciferase assay was measured using Dual-Luciferase Reporter assay kit (Promega) according to manufacture instructions. expression of Rluc was used to normalize the Fluc expression.

Poly(A) tail assay

Determination of poly(A) tail was performed with USB Poly(A) tail-length assay kit (Affymetrix) following the manufactor's protocol. In short, *in-vitro* transcribed RNA was G/I tailed, and reverse-transcribed with specific adapter primer. The poly G/I tailed cDNA was then amplified with a Bcl2 specific primer (GGAAAACACACCTGATTTTAACTTCCTAGGTCGTCC) and a provided universal PCR reverse primer. Finally, PCR products were resolved on a 2% agarose gel with ethidium bromide.

SIINFEKL-based peptide display

A DNA sequence coding for the amino acids LEQLESIINFEKL was cloned downstream of the tGFP-stop sequence in the pCDH-tGFP reporter constructs with APA/random sequence in between using Gibson assembly. Resulting plasmids were sequence-verified. These clones were infected into MDA-MB-231 H-2Kb expressing cells as describe above. For the detection of presented SIINFEKL peptides, cells were washed with PBS and detached using PBS-EDTA. Next, cells were washed in PBS-BSA (0.1%) and incubated with APC anti-mouse H-2Kb bound to SIINFEKL antibody (Biolegend, clone 25-D1.16, 141,606; 1:200) for 30 min. Next, the cells were washed three times and analyzed on a Attune NxT (Thermo). The data were analyzed using FlowJo V10 software (FlowJo).

T cell killing assay and clonogenic assay

200,000 MDA-MB-231 cells expressing H2-Kb and tGFP-APA-SIINFEKL and tGFP-random- SIINFEKL were seeded the day before co-culture in 24 well plate. Then, 100,000 OT-1 cells were added to the co-culture. After 1-day cells were refreshed, rinsed with PBS and fixed for 30 min at RT with 4% formaldehyde. Then cells were stained using Crystal Violet (0.1%) for 1 h at RT in the dark. Then wells were thoroughly washed with water and dried overnight. To quantify killing efficiency, wells were unstained using a 10% acetic acid solution and the wells were measured at 590nM using a TECAN.

Immunopeptidomics

MDA-MB-231 in 80% confluent where transfected with Dharmafect1to knock down NUDT21 or negative control siRNA as describe above (five 15cm dishes for replica, n = 2). 72h after transfection, the cells were treated with IFNγ to induce immunopeptide presentation and the cells from five dishes were scraped into 2mL/dish of cold lysis buffer (20mM Tris, pH 8.0, 100mM NaCl, 6mM MgCl2, 1mM EDTA, 60mM Octyl β-*d*-glucopyranoside, 0.2mM Iodoacetamide, 1.5% Triton X-100, 50xC0mplete Protease Inhibitor Tablet-EDTA free and PMSF) obtaining a total of ∼12mL lysate. This lysate was split into 2mL Eppendorf tubes, and incubated on ice for 15 min with 1ul of Benzonase (Sigma, E1014-25KU) to degrade nucleic acid. The lysates were then centrifuged at 4,000 rpm for 22 min at 4°C and the supernatants were transferred to another set Eppendorf tubes containing a mixture of 50μL pre-washed beads (Millipore Sigma, GE17- 0886-01) and 10μg of an HLA class I antibody (W6/32). The immune complexes were captured on the beads by incubating on a rotor at 4°C for 3h. The beads were washed to remove non-specifically bound material. In total, nine washing steps were performed; one wash with 1mL of cold lysis wash buffer (20mM Tris, pH 8.0, 100mM NaCl, 6mM MgCl2, 1mM EDTA, 60mM Octyl β-*d*-glucopyranoside, 0.2mM Iodoacetamide, 1.5% Triton X-100), four washes with 1mL of cold complete wash buffer

(20mM Tris, pH 8.0, 100mM NaCl, 1mM EDTA, 60mM Octyl β-*d*-glucopyranoside, 0.2mM Iodoacetamide), and four washes with 20mM Tris pH 8.0 buffer. Dry beads were stored at −80°C until mass spectrometry analysis was performed.

Immunopeptidomics mass spec

Dried in a vacuum concentrator. HLA-peptide elution and desalting to purify HLA-peptides, we followed HLA-peptide isolation method from Abelin et al. (10.1016/j.immuni.2017.02.007) We built in-house C18 stage tips using Empore C18 stationary phase (3M). To load and process liquid chromatography on stage tip, we used benchtop centrifuge at 1,000 x g mostly, if not, increased centrifugal force up to 10,000 x g. Stage tips were equilibrated with 2 x 100 uL methanol, 2 x 50 uL of 50% MeCN with 0.1% formic acid, and 2 x 100 uL of 1% formic acid. Samples were acidified by adding 6% MeCN with 10% formic acid with an equal volume of samples. Then, we transferred HLA-peptide including beads into stage tips. To elute HLA-peptide from HLA capturing bead, 50 uL of 1% formic acid and 2 x 50 uL of 10% acetic acid were added onto stage tips. For these steps, we collected every output during the steps, merged, and 150 uL of washed-out eluent was loaded onto stage tips, once. Again 50 uL of 1% formic acid was applied, collected, and reapplied once. Stage tips washed with 2 x 100 uL of 1% formic acid, and eluted with 20 uL of each 20% MeCN with 0.1% formic acid, 40% MeCN with 0.1% formic acid, 60% MeCN with 0.1% formic acid with increasing order of organic solvent percentage. Finally, eluents were combined and dried in a vacuum concentrator. The experiment was performed in two biological replicates.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mapping and peak calling

FASTQ files were aligned to the human hg19 genome using either STAR (STAR_2.5.0a) or TopHat (v.2.0.13) (Kim et al., 2013). Mouse FASTQ files were aligned with the same tools to the mm10 genome. TopHat aligned reads were aligned allowing up to 3 mismatches per read, a maximum gap of 5 bases, and a total edit distance of 8. Peaks were called using mapped reads from all samples, using findPeaks (Homer package, v.3.12) (Heinz et al., 2010) with "-region -size 200 -minDist 250 -strand" parameters. Peaks found within transcripts or up to 5 kb downstream of promoters were included in a GTF file and processed by the DE-Seq package. Exon-aligned reads were counted using HTSEQ (0.5.4p3).

For the Ribosome profiling analysis, reads in FASTQ format were trimmed for adapters using cutadapt, and were aligned to human hg19 trancriptome (gencode v19) using bowtie. Samtools (Li et al., 2009) and bedtools (Quinlan and Hall, 2010) were used for file format conversion (SAM to BAM and BAM to BED). Finally, CDS alignment was retained using PERL script which was covered into tabular format using custom PERL script, generating read count data for every one of the hundred windows covering the length of transcript, generated using BEDTOOLS. Peaks were identified with a custom R script using the findPeaks function from the package PRACMA. The fraction of reads for every peak identified in a transcript were identified by percentile normalization.

For 5′end seq samples in fastq format were aligned to the mouse transcriptome using gencode vM24 and the mouse genome version GRCm37 using TopHat v2.1.1 (with parameters -p24 –solexa-quals –no-coverage-search –no-novel-juncs –no-novel-indels).

m6A peakcalling was performed using MACS2 callpeak version 2.2.7.1 with standard parameters on IP versus Control files (Zhang et al., 2008).

Segmentation to upstream/downstream using transcript read coverage

Performed as described previously (Malka et al., 2017). Briefly, analysis was performed separately for each transcript/treatment (TEX, 3′-pulldown, CAP IP, XRN1, Polysomal RNA). First, the average read coverage for both treatment and control were calculated for overlapping 50 bp windows (in 20 bp offsets) along the coding regions and 3′-UTR. Coverage values were then transformed to log scale.

Next, we trained a linear regression model to correct for different transformations of the results in control and various treatments. For this, we considered all windows that overlap coding regions from all transcripts, and fitted a linear regression model $Y = aX + b$ for each treatment, where Y is the log-transformed set of coverages along the transcript for every treatment, and X is the log-transformed coverages for control. This allowed us to overcome experiment-specific bias and infer the offset and ratio between mean read coverage (per window) in every treated condition and control. We then applied the linear model to yield the expected read coverage for each transcript along the transcripts window, and deviations from the regression model were calculated.

Next, we trained an HMM (Kulp et al., 1996; Mukherjee and Mitra, 2005) with two internal states corresponding to upstream and downstream regions, allowing a single transition from the upstream to the downstream state. Each state was characterized with a Gaussian emission of continuous variables (with shared variance). The parameters of the model were optimized using the Baum–Welch (Expectation–Maximization) algorithm, using maximum likelihood estimations. Finally, the most probable upstream/downstream cleavage point was identified for every transcript/treatment as the maximum likelihood transition point between the upstream and downstream states, using the Viterbi algorithm.

Differential local read coverage analysis

To compare the local read coverages of TEX-treatment, XRN1-treatment, 5′end-sequencing, 3′end-sequencing, polysome-sequencing and m6A RIP-seq experiments to their control RNAseq (untreated) counterparts in a nucleotide-level resolution,

differential local read coverage tracks (*Z* score transformed values of logFC values at each position) were generated using a previously published custom Python script (https://github.com/eranroz/polyA/blob/master/zscore_track.py) (Malka et al., 2017). For plotting purposes, resulting genomic tracks were parsed with another custom Python script to obtain the exonic Z-scores with transcript-level coordinates. Later, data was centered to TrPt coordinates and at each transcript-level distance from predicted TrPt sites, positional Z-scores were averaged over all transcripts and position-specific averages were plotted together with a smoothed line using the ggplot2 library in R.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis

The dried peptide samples were reconstituted in 0.1% formic acid and injected from a cooled (10°C) autosampler into pre-packed EASY-spray column (50 cm × 75 μm, packed in-house) on an Eksigent nanoLC-ultra 1D plus UPLC system at a flow rate of 250 nL/min. Before use, the column was equilibrated with 96% solvent A (0.1% formic acid in water) and 4% solvent B (0.1% formic acid in acetonitrile). The peptides were eluted with a non-linear gradient from 4% to 32% solvent B over 150 min (Solvent B-Gradient Time; 4%-10min, 6.5%-25min, 13%-70min, 19%-115min, 22%-130min, 26%-145min, 32%-160min) and 32%–80% solvent B over 10 min followed by 40 min of organic wash and aqueous re-equilibration. The total run took 280 min. The column temperature was kept at 40°C by using an EASY-spray ionization source (Thermo Fisher Scientific). The UPLC system was coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific) operated in data-dependent acquisition (DDA) mode. Survey full-scan MS spectra (300–1600 *m/z*) were acquired with a resolution of 70,000. Source ionization parameters were as follows: spray voltage, 1.9 kV; capillary temperature, 275°C; S-lens RF level, 50.0. Up to 12 most intense ions in each full MS scan were fragmented and analyzed. MS/MS parameters were as follow: resolution, 35,000; automatic gain control target, 5×10^4; isolation window, 1.2 *m/z*; maximum injection time, 120 ms; fixed first mass, 120.0 *m/z*; normalized collision energy, 27; charge state acceptance, 1-5; dynamic exclusion duration, 60 s; intensity threshold, 2.5×10^4.

Database searching

Peptide identification from mass spectra performed by Proteome Discoverer (PD) version 2.2.0.388 (Thermo Fisher Scientific). All datasets were searched with a Sequest HT nodes which implemented in PD. We used default label-free quantitation (LFQ) workflow for Q Exactive with label-free quantitation for processing step and consensus step. For post-processing, we used Percolator node to maintain peptide level false discovery rate (FDR) lower than 0.01. The length of identified peptides set minimum 6 residues. Notable search parameters in Sequest HT node were: enzyme name, Trypsin (Semi); fragment mass tolerance, 0.05 Da; dynamic modifications for peptide N-terminus were Acetyl (+42.011 Da), Acetyl:2H(3) (+45.029 Da) and Glu- > pyro-Glu for N-terminal Glu residue (−18.011 Da); static modifications were Carbamidomethyl for any Cys residue (+57.021 Da) and Acetyl:2H(3) for any Lys residue (+45.029 Da) We used three databases to identify non-canonical translation peptides. To identify canonical peptides we searched the mass spectra with UniProt human reference database (release June 2019), to identify non-canonical peptides, we used OpenProt (ref: 10.1093/nar/gky936) database release 1.3 for human AltProts and Isoforms and to identify possible protein contaminants, we used cRAP (http://www.thegpm.org/crap) database released 2015 version only specifically categorized as dust/contact proteins. For LFQ quantitation we used Minora algorithm to quantitate regardless of PSM evidence across samples which similar algorithm with "match between run" setup of MaxLFQ (ref: 10.1074/mcp.M113.031591). we performed chromatographic retention time alignment with maximum 5 min shift and mass tolerance of 10 ppm. Peptide abundance were defined by peak area.

Database search and filtration

The RAW MS data files were analyzed by MaxQuant (1.6.10.43). Files were searched against the full canonical human proteome (Uniprot database, UniProt, 2019). The maximum allowed precursor mass tolerance was 20 ppm. N-terminal acetylation and methionine oxidation were set as variable modifications. A peptide spectrum match FDR of 0.05 was used, and no protein FDR was set. Enzyme specificity was set as 'unspecific', the 'match between runs' option was set with default settings and LFQ was set to a 'minimum ratio count' of 1. The obtained peptides were filtered for contaminants and reverse peptides.